Anomaly Detection in Human Brain via Inductive Learning on Temporal Multiplex Networks

Ali Behrouz ALIBEZ@CS.UBC.CA
University of British Columbia
Vancouver, BC, Canada

Margo Seltzer MSELTZER@CS.UBC.CA
University of British Columbia
Vancouver, BC, Canada

Abstract

The human brain is at the center of complex neurobiological systems, and understanding its structural and functional mechanisms remains an intriguing goal for neuroscience research. While magnetic resonance imaging (MRI) is one of the most widespread and important sources of neurological data, it poses daunting analysis challenges. Due to recent advances in graph theory and machine learning on graphs, representing the connections of the human brain as a network has become one of the most pervasive analytical paradigms. However, most existing graph machine learning-based methods suffer from a subset of four critical limitations: They are ① designed for one type of data (e.g., fMRI or sMRI) and one individual subject, limiting their ability to use complementary information provided by different images. ② designed in supervised or transductive settings, limiting their generalizability to unseen patterns. ③ designed for classifying brain networks, limiting their ability to reveal underlying patterns that might cause the symptoms of a disease or disorder. ④ frequently unable to scale to large numbers of samples. To address the first limitation, we suggest using multiplex networks–networks with different types of connections– to model the network of different data samples. We present ADMire, an inductive and unsupervised anomaly detection method for multiplex brain networks that can detect anomalous patterns in the brains of people living with a disease or disorder. It uses two different casual multiplex walks, *inter-view* and *intra-view*, to automatically extract and learn temporal network motifs. It then uses an anonymization strategy to hide node identity, keeping the model inductive. We then propose a novel negative sample generator strategy for multiplex networks that lets our model learn anomalous patterns in an unsupervised manner. Our experiments on Parkinson's Disease, Attention Deficit Hyperactivity Disorder, and Autism Spectrum Disorder show the efficiency and effectiveness of our approach in detecting anomalous brain activity in people living with these diseases or disorders.

1. Introduction

Over the recent years, the field of neuroscience and brain imaging research has undergone a significant shift in focus from region-specific analyses to network models (Bassett and Sporns, 2017; Park and Friston, 2013; Mišić and Sporns, 2016), largely due to the rapid development of modern neuroimaging technology. Network models of the brain represent regions of interest (ROIs) as nodes and calculate pairwise similarities between regions to form edges (Bassett and Sporns, 2017; De Domenico, 2017), usually derived from functional

Magnetic Resonance Imaging (fMRI) or structural Magnetic Resonance Imaging (sMRI). These models have demonstrated their effectiveness in enhancing our understanding of brain diseases and disorders (Chatterjee et al., 2021; Finn et al., 2015; Preti et al., 2017). As a result, empirical data on brain networks has substantially increased in size and complexity, leading to a strong demand for appropriate tools and methods to model and analyze this data (Preti et al., 2017).

In recent years, there has been significant interest in machine learning methods for analyzing graph-structured data in various domains, such as drug discovery (Xiong et al., 2019), brain network classification (Behrouz and Hashemi, 2022; Abrate and Bonchi, 2021), and protein networks (Gao et al., 2023). While several studies have demonstrated the effectiveness of machine learning on graphs for analyzing human brain networks, most focus only on graph or node classification tasks (Kan et al., 2022c; Hashemi et al., 2023; Zhu et al., 2022a; Cui et al., 2022b). These tasks involve detecting diseases (Zhu et al., 2022a), predicting biological features (Kan et al., 2022c), and identifying functional systems within the brain (Behrouz and Hashemi, 2022). However, detecting abnormal brain activity in people with neurological disorders is a crucial step in understanding the causal mechanisms of symptoms and facilitating early detection and development of medical treatments. Moreover, most consider a single brain network (from a single type of neuroimage and a single subject), which can be noisy or incomplete (Agrawal et al., 2020; De Domenico, 2017; Zhang et al., 2020). To address this limitation, De Domenico (2017) suggests using static multiplex networks. Multiplex networks are graphs where nodes can be connected by different types of edges (Kivelä et al., 2014; Hashemi et al., 2022). Edge types can be the brain network of different subjects or different neuroimaging modalities (see §3.2).

Limitation of Previous Methods. Although anomaly detection in graphs is a well-studied problem and several methods have been designed, brain networks have three unique traits that make directly applying existing graph anomaly detection models impractical: ① Noisy data: a single neuroimaging data sample can be extremely noisy and inaccurate, which hinders the identifications of biological insights into the structure of brain networks. Existing general anomaly detection methods can only use a single brain image, making them sensitive to noise, or are required to aggregate different neuroimages as a pre-processing step, missing complex brain activity in each brain image. ② Multimodal neuroimaging: while several studies discussed the importance of using different neuroimage types (e.g., fMRI, sMRI, etc.) in brain network analysis as different modalities provide complementary information to each other (Zhang et al., 2018c; Zhu et al., 2022b), existing works are limited to a single type of neuroimages and are unable to incorporate the information about different modalities. ③ Time alignment: exisitng methods assume that the timestamps are the same across different graphs. However, while modeling neuroimage data as *temporal* brain networks, the timestamps might be shifted and be not aligned across brain images of different subjects.

Previous studies on anomaly detection in brain networks not only suffer from ① and ② mentioned above but also suffer from two more limitations: ⓘ These studies assume pre-defined anomaly patterns or man-made features. Such approaches do not easily generalize to the brain activity of different individuals. Moreover, in a real-world scenario, brain activity might be more complex in nature, and it is nearly impossible to detect anomalies with high accuracy using pre-defined patterns/roles. ⓘⓘ These methods are designed for static brain networks, missing temporal properties and the dynamics of brain activity over time. In

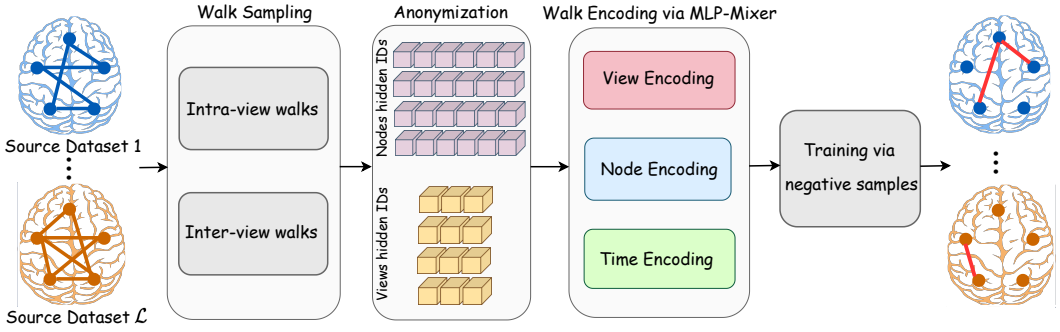


Figure 1: **Schematic of the ADMire model**. ADMire consists of four main stages called (1) Walk Sampling, (2) Anonymization, (3) Walk Encoding, and (4) Training via generating negative samples.

addition to the above limitations, both groups suffer from an important drawback. All these methods are designed in the transductive setting, which limits their generalizability to unseen nodes or patterns. Due to the complex and potentially different brain activity in different subjects, the costly process of obtaining neuroimaging data, and also the lack of labeled data for different diseases/disorders, it is a must for effective brain anomaly detection methods to learn from data in an inductive and unsupervised manner, which enables generalizing to unseen nodes or patterns.

Generalizable Insights about Machine Learning in the Context of Healthcare

To mitigate the above limitations, we introduce ADMire (**A**nomaly **D**etection in **M**ulti**i**plex B**r**ain N**e**tworks). ADMire uses two different and novel temporal walks, inter-view and intra-view walks, to capture the causal relationships between brain activities across different views and within a single view, respectively, over time. Next, it uses an anonymization method based on the correlation between network motifs to hide the identity of nodes and views, keeping the model inductive during training. To learn the structural and temporal properties of the network, ADMire uses a novel encoding process. It encodes the information about each walk by mixing the encoding of the sequence of nodes that appears in the walk via an MLP-Mixer (Tolstikhin et al., 2021). To overcome noise in the data and/or to take advantage of complementary information provided by different neuroimage modalities, ADMire uses a new attention mechanism to incorporate the node encodings obtained from different views. To mitigate the time alignment issue, we use a non-periodic time encoding module that encodes each timestamp. Finally, we design a new negative sample generator algorithm that lets the model be trained in an unsupervised manner. In our experimental evaluation, we first use synthetic datasets to show the superior performance of ADMire over baselines, the importance of its critical components, and the importance of multiplex modeling. Second, we use real-world datasets to show how ADMire can be used to detect abnormal brain activities in a control group with brain disease or disorder.

In summary, this work highlights the significance of ① inductive and unsupervised machine learning-based anomaly detection in understanding the anomalous activities of the human brain that might cause a brain disease or disorder, and ② modeling neuroimaging

data (e.g., fMRI and sMRI) as *multiplex* brain networks to overcome noise in data or to use complementary information provided by multimodal brain networks.

2. Related Work

To situate our research in a broader context, we briefly review related work. For additional discussion see Appendix B.

Temporal Graph Learning. Learning from temporal networks has been a widely studied topic in the literature (Longa et al., 2023). The first major group uses Graph Neural Networks (GNNs) to learn the node encoding and Recurrent Neural Networks (RNNs) to keep these encodings updated over time (Seo et al., 2018; Zhao et al., 2019; Peng et al., 2020; Wang et al., 2020b; You et al., 2022; Hashemi et al., 2023). Recently, more sophisticated learning methods for temporal graphs have been designed based on temporal random walks (Wang et al., 2021; Jin et al., 2022; Behrouz et al., 2023), line graphs (Chanpuriya et al., 2023), GraphMixer Cong et al. (2023), neighborhood representation (Luo and Li, 2022), and subgraph sketching (Chamberlain et al., 2023). However, all these methods differ from our approach as they are designed for monoplex temporal graphs and cannot easily be extended to multiplex networks.

Multiplex Graph Learning. Several methods have been proposed to learn node embeddings on multiplex networks by integrating information from individual relation types (Cen et al., 2019; Pio-Lopez et al., 2021; Yan et al., 2021; Chang et al., 2015; Xie et al., 2021; Wang et al., 2020a; Jing et al., 2021; Park et al., 2020). Other work proposed graph convolutional networks for multiplex graphs (Behrouz and Hashemi, 2022; Cheng et al., 2021; Zhang et al., 2018a). Zhang et al. (2018b) proposed a method that uses a latent space to integrate the information across multiple views. All these methods are designed for static multiplex networks and in the transductive setting, while brain networks are temporal by nature and require inductive learning due to the cost and small size of neuroimaging datasets.

Feature Learning and Anomaly Detection in Brain Networks. In recent years, several studies focused on analyzing brain networks to understand and distinguish healthy and diseased human brains (Jie et al., 2016; Chen et al., 2011; Wee et al., 2011). Recently, due to the success of GNNs in analyzing graph-structured data, deep models have been proposed to predict brain diseases by learning the brain networks structure (Kan et al., 2021; Cui et al., 2021; Kan et al., 2022a; Zhu et al., 2022a; Cui et al., 2022b). All these methods are designed for graph or node classification and cannot easily be extended to edge-anomaly detection. In addition, several anomaly detection methods have been proposed to find anomalous regions, or subgraphs in the brain, which can cause a disease (Chatterjee et al., 2021; Zhang et al., 2016; Liu et al., 2020). All these methods are designed for node or subgraph anomaly detection tasks in *single* brain networks and cannot easily be extended to the edge-anomaly detection task in *multiplex* brain networks. Also, these methods are not learning-based and consider pre-defined patterns/rules for anomalies, limiting their ability to generalize to complex brain activity.

Anomaly Detection in Multiplex Networks. Several non-machine learning methods for anomaly detection in static multiplex networks have been proposed based on eigenvector centrality (Mittal and Bhatia, 2018), clique/near-clique structures(Bindu et al., 2017),

multi-normality (Bansal and Sharma, 2020), node centrality (Maulana and Atzmueller, 2020), and persistence summary (Ofori-Boateng et al., 2021). These models are not able to learn from data and are limited to pre-defined rules/patterns. To address this issue, recently, learning-based methods have been proposed. AnomMAN (Chen et al., 2022) uses an auto-encoder module and a GCN-based decoder to detect node anomalies in static multiplex networks. All of these approaches are limited to static multiplex networks and are designed to detect topological anomalous subgraphs, nodes, or events, and cannot identify anomalous edges. The only exception is a Gnn-based anomaly detection method in multiplex networks. AnoMulY (Behrouz and Seltzer, 2022). However, AnoMulY is designed for the transductive setting and also cannot scale to multiplex brain networks with a large number of views (see § 5.1). For more discussion see Appendix B.

3. Methods

3.1. Preliminaries

We first precisely define temporal multiplex networks. Next, we motivate the use of multiplex networks in modeling brain activity, and finally, we formalize the problem of edge anomaly detection in temporal multiplex networks.

Definition 1 (Temporal Multiplex Networks):
A temporal multiplex network $\mathcal{G} = \{G_r\}_{r=1}^{\mathcal{L}} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$, can be represented as a sequence of connections with different types that arrive over time, i.e., $\mathcal{E} = \{(e_1, t_1), (e_2, t_2), \dots\}$, where $\mathcal{V}$ is the set of nodes, $\mathcal{L}$ is the set of relation types, $\{e_1, e_2, \dots\} \subseteq \mathcal{V} \times \mathcal{V} \times \mathcal{L}$, and $\mathcal{X} \in \mathbb{R}^{|\mathcal{V}| \times f}$ is a matrix that encodes node attribute information for nodes in $\mathcal{V}$. Given a relation type r, we use $G_r = (\mathcal{V}, \mathcal{E}_r, \mathcal{X})$ to denote the corresponding graph of the relation type r (a.k.a r-th view of the graph), and we denote the set of vertices in the neighborhood of $u \in \mathcal{V}$ in relation r as $\mathcal{N}_r(u)$. Given time t, we use $\mathcal{E}_r^t(u) = \{(e, t') \in \mathcal{E}_r | u \in e \text{ and } t' < t\}$ to represent the set of connections attached to a node u in relation type r before certain time t.

Our goal is to detect anomalous incoming edges. Specifically, given the current time t_{now}, for each edge $e = (u, v, r, t_{\text{now}}) \in \mathcal{E}$, we produce an anomaly score $\varphi(e)$.

3.2. Motivations

Given a neuroimaging dataset, we answer the following three questions:
Q1: How does multiplex modeling improve upon modeling a single network? While modeling the human brain as a network has gained much attention in the neuroscience community in recent years (Lynn and Bassett, 2019; Liu et al., 2017), most studies have focused on a single type of simple brain networks (Liu et al., 2020; Chatterjee et al., 2021). However, recent research on brain network analysis suggests that different modalities of brain networks provide complementary information (Zhang et al., 2018c; Zhu et al., 2022b). The fusion of multiple modalities can lead to consistent improvements in brain analysis. Additionally, several studies suggest that brain networks generated from an individual can be noisy and incomplete (Lanciano et al., 2020; Behrouz and Seltzer, 2022; Behrouz et al., 2022). Consequently, researchers are exploring the possibility of studying the human brain without necessarily discarding or aggregating the vast amount of data available (De Domenico, 2017).

Multiplex brain networks, where each view represents a type of neuroimaging data or the neuroimage of an individual, are accurate models that can capture the complementary information provided by different neuroimaging data and increase the model's robustness to noise and incompleteness in individual neuroimages.

Q2: How can neuroimaging data be modeled as (temporal) multiplex networks? We focus on three ways to model neuroimaging data as multiplex networks.

① **Activity in different frequency bands**: in the context of fMRI images, previous works utilize filtering procedures to extract signals within a particular frequency range, typically between 0.01 and 0.1 Hz (De Domenico, 2017; De Vico Fallani et al., 2014). However, the selection of the frequency band carries significant implications for the functional representation of the brain. In fact, existing methods do not distinguish the contributions coming from different frequency bands, overlooking the contributions of different ranges, and instead concentrate on a single frequency range. To this end, De Domenico et al. (2016) shows that brain signals in a range between 0.01 and 0.25 Hz, in steps of 0.02 Hz provide unique information and should be neither aggregated nor neglected. We suggest using multiplex brain networks, where each view represents the correlations graph of signals in a specific range.

② **Multimodal brain networks**: Recently, several studies discussed the importance of using different neuroimage types (e.g., fMRI, sMRI, Diffusion Tensor Imaging (DTI), etc.) in brain network analysis as different modalities of brain networks provide complementary information (Zhang et al., 2018c; Zhu et al., 2022b). In this case, a multiplex brain network is a multimodal brain network, where each view represents the obtained brain network from a specific type of neuroimage (e.g., fMRI or sMRI).

③ **Different subjects**: Previous studies discuss the challenges of the existence of noise in a brain network generated from an individual (Lanciano et al., 2020; Zhang et al., 2020). Most existing methods aggregate (e.g., averaging) the data from different individuals to mitigate the noise in the dataset (Chatterjee et al., 2021; Abrate and Bonchi, 2021). However, this aggregation discards complex patterns in each individual's brain activities, causing missing information. Moreover, in brain network disease/disorder analysis, it is known that individuals having the same disease or disorder share similar patterns (Kan et al., 2022b; Cui et al., 2022c), which means that disorder/disease-specific anomalous activities require consideration of the brain networks of different subjects. In this case, in a multiplex brain network, each view represents the brain network of an individual.

Q3: How can a method automatically learn the type of the data modeling? When modeling neuroimage data as multiplex networks, there are two main advantages: ① different views can provide complementary information and help to learn brain activities in a more effective and robust manner, and ② there might be causal effects between different views, and capturing them can improve performance. However, one of the main challenges in designing machine learning models on multiplex brain networks is to learn the type of data modeling. In multimodal brain networks, the activities observed in one type of neuroimage (e.g., fMRI) at a previous timestamp may be correlated with and contribute to another activity in a different type of neuroimage (e.g., sMRI) at the current timestamp. However, when each view represents the brain of a different subject, the brain activities between different individuals are not causally correlated, although different views may offer complementary information about the brain activities of people with the same disease or disorder. Accordingly, based

on different multiplex modeling of neuroimages, there is a spectrum of causal relationships between different views of the network. We automatically learn whether there is a causal effect between activities in different neuroimages and incorporate the complementary information provided by these images. To this end, we design two different temporal multiplex walks to capture the causal effect and the dynamics in different views over time. Next, we use a learnable neural layer to combine the information provided by these two walks by automatically learning their importance from the data.

3.3. Anonymous Multiplex Temporal Walk

We first define two variants of multiplex temporal walk and then design an anonymization process to hide nodes' identity in walks, keeping the model inductive in the training phase. The main intuition of our approach is to use multiplex temporal walks as a proxy of temporal motifs in multiplex networks and extract the causality of the existence of an edge in a specific type of connection over time. That is, inspired by Wang et al. (2021), a multiplex temporal walk starts from a connection of interest and backtracks several adjacent edges over time to encode the underlying causality of network dynamics. However, as discussed in §3.2, the main challenge to extracting the causality of the existence of an edge in multiplex networks is that we have two types of network motifs that depend on how we model the problem using multiplex networks. To this end, we design two multiplex temporal walks to extract the causality within a specific view and across different views.

Inter-view Temporal Walk. To capture the correlation between different views and extract the causality of an edge from different views of the network, in the inter-view temporal walk, we let the walker walk across views. Accordingly, an inter-view temporal walk W_{inter} on temporal multiplex networks can be represented as:

$$W_{\text{inter}} = ((u_0, r_0, t_0), (u_1, r_1, t_1), \ldots, (u_m, r_m, t_m)) \,;\;\; t_0 > t_1 > \cdots > t_m, \tag{1}$$

where $(u_i, u_{i+1}, r_{i+1}, t+1) \in \mathcal{E}$. That is, not only does the walker walk over time and capture the temporal causality of an edge, but also can walk over different views to capture the dependencies of connections in different views, taking advantage of complementary information provided by different types of relations (e.g., fMRI and sMRI). We let $W_{\text{inter}}(i)$ denote the i-th element of the temporal multiplex walk, (u_i, r_i, t_i). Also, we use $W_{\text{inter}}(i, 0), W_{\text{inter}}(i, 1)$, and $W_{\text{inter}}(i, 2)$ to refer to u_i, r_i, and t_i, respectively.

Intra-view Temporal Walk. While there is no causal relationship between different types of interactions (e.g., brain networks of different subjects), we limit our walks to a specific type of connection. Given a type of relation r, an intra-view temporal walk W^r_{intra} on a view of a temporal multiplex network can be represented as:

$$W^r_{\text{intra}} = ((u_0, r, t_0), (u_1, r, t_1), \ldots, (u_m, r, t_m)) \,;\;\; t_0 > t_1 > \cdots > t_m, \tag{2}$$

where $(u_i, u_{i+1}, r, t+1) \in \mathcal{E}$. However, different views still provide complementary information (see §3.2). To take advantage of this complementary information, in §3.4, we design an attention mechanism that incorporates the information of different types of connections.

How to sample temporal multiplex walk? As discussed in previous studies, newer connections in temporal networks are often more informative (Wang et al., 2021; Jin et al.,

2022). To this end, we use a biased sampling method with hyperparameter μ to control the importance of recent connections. Given the time of a previously sampled edge, t_0, we sample an adjacent edge at time t with probability proportional to $\exp(\mu(t-t_0))$. In multiplex networks, the correlation of different pairs of views can be different (Park et al., 2020; Behrouz and Seltzer, 2022) and for connections in a given view r, a subset of views might play more important roles in causality extraction. Accordingly, in inter-view temporal walks, we use a biased sampling method and sample link (u_1', u_2', r', t') after previously sampled link (u_1, u_2, r, t) with probability proportional to $\psi(r, r')$. In fact, $\psi(r, r')$ shows the importance of view r for view r'. In §3.4, we discuss how to calculate $\psi(r, r')$. See Appendix C for the pseudocode of this procedure.

Given a (potential) link (u, v, r, t), we use the above procedure to generate M inter-view and M' intra-view walks with m steps starting from each of nodes u and v. We use $\mathcal{S}_{\text{inter}}(u)$, $\mathcal{S}_{\text{intra}}(u)$, $\mathcal{S}_{\text{inter}}(v)$, and $\mathcal{S}_{\text{intra}}(v)$ to store started walks from u and v, respectively.

Remark 1 *Note that intra-view walks are not limited to the relation type of the given link. That is, although all steps in an intra-view walk must have the same type, this type can be different from relation type r. This is the key to using complementary information provided by different views, as we discussed later. We use $\mathcal{S}^{r'}_{intra}(u)$ and $\mathcal{S}^{r'}_{intra}(v)$ to store started walks from u and v within view r'.*

Anonymization Process. Recent studies argue that traditional anonymization methods (e.g., (Micali and Zhu, 2016)) suffer from several limitations and suggest using an anonymization process that can capture the correlation between different walks (Wang et al., 2021; Jin et al., 2022; Behrouz et al., 2023). Moreover, in multiplex networks, we need to hide the identity of both nodes and views (e.g., relation types) to keep the model inductive. Given a (potential) link (u, v, r, t), let $w_0 \in \{u, v\}$. To capture the correlation across different walks, which could be a key to reflecting the network dynamics, for a given node w that appears on at least one walk in $\mathcal{S}_{\text{inter}}(u) \cup \mathcal{S}_{\text{inter}}(v)$, we use a relative vector $\mathcal{C}(\mathcal{S}_{\text{inter}}(w_0), w) \in \mathbb{Z}^{m+1}$ that represents the number of times in $\mathcal{S}_{\text{inter}}(w_0)$ that node w appears at certain positions. That is,

$$\mathcal{C}_i\left(\mathcal{S}_{\text{inter}}(w_0), w\right) = \left|\{W_{\text{inter}} | W_{\text{inter}} \in \mathcal{S}_{\text{inter}}(w_0), w = W_{\text{inter}}(i, 0)\}\right|, \ \forall\, 0 \le i \le m. \quad (3)$$

Similarly, we define $\mathcal{C}(\mathcal{S}_{\text{intra}}(w_0), w)$ over intra-view temporal walks. Until now node identities were accessible, but we now remove node identities and use only these four vectors in the training phase to represent each node, thereby hiding their identity:

$$\textsc{Id}(w) = \{\mathcal{C}(\mathcal{S}_{\text{inter}}(u), w), \mathcal{C}(\mathcal{S}_{\text{inter}}(v), w), \mathcal{C}(\mathcal{S}_{\text{intra}}(u), w), \mathcal{C}(\mathcal{S}_{\text{intra}}(v), w)\}. \quad (4)$$

Given a set of walks (e.g., $\mathcal{S}_{\text{inter}}(w_0)$), we count the number of times we see a relation type at certain positions when we start from a specific relation type to capture the correlation of different views. To this end, for a given relation type r, we use a relative vector $\mathcal{C}^{\text{view}}(\mathcal{S}_{\text{inter}}(w_0), r) \in \mathbb{Z}^{m+1}$ that counts times in $\mathcal{S}_{\text{inter}}(w_0)$ that a relation with type r appears at certain positions:

$$\mathcal{C}^{\text{view}}_i\left(\mathcal{S}_{\text{inter}}(w_0), r\right) = \left|\{W_{\text{inter}} | W_{\text{inter}} \in \mathcal{S}_{\text{inter}}(w_0), r = W_{\text{inter}}(i, 1)\}\right|, \ \forall\, 0 \le i \le m. \quad (5)$$

Accordingly, we use $\textsc{Id}^{\text{view}}(r) = \{\mathcal{C}^{\text{view}}(\mathcal{S}_{\text{inter}}(u), r), \mathcal{C}^{\text{view}}(\mathcal{S}_{\text{inter}}(v), r)\}$ to hide the identity of view r. Note that, although intra-view walks are within a single view, we still need to hide the identity of the view and we use the same $\textsc{Id}^{\text{view}}(r)$ as above.

3.4. Neural Encoding

We next present our fast and simple, yet effective and generalizable, neural network to encode temporal multiplex walks so that we can extract structural and temporal information from the network. The intuition of this neural encoding is to use anonymous temporal multiplex walks to learn the structural and temporal properties as well as casual rules of the network in an inductive manner.

Previous walk-based studies see each walk as a sequence of nodes and use sequence encoders (e.g., RNN or Transformers) to encode each walk (Wang et al., 2021; Jin et al., 2022). However, these sequence encoders are time-consuming and can limit the generalizability of the encoding method in temporal graphs (Cong et al., 2023). Accordingly, we design a neural encoding that first uses a time-encoding function to learn to distinguish different timestamps and then uses an MLP-Mixer (Tolstikhin et al., 2021) to mix and encode the hidden identity of nodes in a walk.

Time Encoding. Existing methods in temporal graph learning (Cong et al., 2023; Wang et al., 2021) use random Fourier features (Kazemi et al., 2019) to encode time. However, this approach captures only periodicity in the data, while in brain activity patterns we also need to learn non-periodic patterns dependent on the progression of time (e.g., in task-based fMRI). To this end, we also add a learnable linear term to the feature representation of time encoding. That is, we encode a given time t as:

$$\mathscr{T}(t) = (\boldsymbol{\omega}_l t + \mathbf{b}_l) \,||\, \cos(t\boldsymbol{\omega}), \tag{6}$$

where $\boldsymbol{\omega}_l, \mathbf{b}_l \in \mathbb{R}$ and $\boldsymbol{\omega} \in \mathbb{R}^d$ are learnable parameters, and $||$ denotes concatenation.

Node Encoding. We now define a node encoding function $\zeta(.)$ that encodes each node w based on $\textsc{Id}(w)$. However, since the concept and task of intra-view and inter-view walks are different, we first break the $\zeta(.)$ function over these walks, called $\zeta_{\text{intra}}(.)$ and $\zeta_{\text{inter}}(.)$, respectively, and then interpolate between them by a learnable parameter λ to obtain $\zeta(.)$.

For each node w that appears on at least one walk in $\mathcal{S}_{\text{inter}}(u) \cup \mathcal{S}_{\text{inter}}(v)$, we use *one* simple MLP to encode the w's hidden identities:

$$\zeta_{\text{inter}}(w) = \texttt{MLP}\left(\mathcal{C}(\mathcal{S}_{\text{inter}}(u), w)\right) + \texttt{MLP}\left(\mathcal{C}(\mathcal{S}_{\text{inter}}(v), w)\right). \tag{7}$$

While inter-view walks by walking across different views naturally capture the causal relationship and correlation between different types of connections, intra-view walks have a different role and capture causality within one type of connection. However, to take advantage of complementary information in multiplex networks, we need to aggregate the information provided by inter-view walks in different views. However, the importance of views might be different (e.g., one disease might be more correlated with functional connectivity than structural connectivity). We design an attention mechanism that learns the importance of each view for other views. Existing attention mechanisms (Behrouz and Seltzer, 2022; Park et al., 2020) for *general* multiplex networks assume that the importance

of each view for each node is different, while in our experimental evaluations on multiplex *brain* networks, we usually observe that the importance of one view for different nodes is almost the same (see Appendix D for a detailed discussion). Given two arbitrary views r_1, r_2, let $\eta(r_1)$ and $\eta(r_2)$ be the learned encoding of r_1 and r_2. The importance of r_2 for r_1, $\psi(r_1, r_2)$, is defined as:

$$\psi(r_1, r_2) = \frac{\exp\left(\sigma\left(\boldsymbol{a}^T.[\mathbf{W}^{\text{att}}\,\eta(r_1)\;||\;\mathbf{W}^{\text{att}}\,\eta(r_2)]\right)\right)}{\sum_{r'\in\mathcal{L}}\exp\left(\sigma\left(\boldsymbol{a}^T.[\mathbf{W}^{\text{att}}\,\eta(r_1)\;||\;\mathbf{W}^{\text{att}}\,\eta(r')]\right)\right)}, \tag{8}$$

where $\boldsymbol{a}$ and $\mathbf{W}^{\text{att}}$ are learnable parameters and $\sigma(.)$ is an activation function (e.g., ReLU). Given a relation type $r' \in \mathcal{L}$, we define view-based node encoding $\zeta^{r'}_{\text{intra}}(w)$ as:

$$\zeta^{r'}_{\text{intra}}(w) = \texttt{MLP}\left(\mathcal{C}(\mathcal{S}^{r'}_{\text{intra}}(u), w)\right) + \texttt{MLP}\left(\mathcal{C}(\mathcal{S}^{r'}_{\text{intra}}(v), w)\right). \tag{9}$$

Next, we aggregate these node embeddings to incorporate information from different views and obtain $\zeta_{\text{intra}}(w)$:

$$\zeta_{\text{intra}}(w) = \sum_{r'\in\mathcal{L}} \psi(r, r')\zeta^{r'}_{\text{intra}}(w). \tag{10}$$

Now, we use a learnable parameter λ to automatically learn the importance of each $\zeta_{\text{intra}}(w)$ and $\zeta_{\text{inter}}(w)$ based on the data. This formulation lets our model learn to interpolate between Equation 7 and Equation 10, which enables it to be flexible in each way the neuroimaging data is modeled (§3.2). Therefore, $\zeta(w)$ is defined as:

$$\zeta(w) = \zeta_{\text{intra}}(w) + \lambda \times \zeta_{\text{inter}}(w). \tag{11}$$

When there is no causal relation between different views (e.g., when views are brain networks of different subjects), our model is expected to set $\lambda \approx 0$ (see §5.1).

View Encoding. For each node $r \in \mathcal{L}$, we use *one* simple MLP to encode the r's hidden identities:

$$\eta(r) = \texttt{MLP}\left(\mathcal{C}^{\text{view}}(\mathcal{S}_{\text{inter}}(u), r)\right) + \texttt{MLP}\left(\mathcal{C}^{\text{view}}(\mathcal{S}_{\text{inter}}(v), r)\right). \tag{12}$$

Walk Encoding. Given a walk $\hat{W} \in \{W_{\text{inter}}, W_{\text{intra}}\}$, we use node encoding function $\zeta(.) : \mathbb{Z}^{(m+1)\times 4} \to \mathbb{R}^{k_1}$ to encode hidden node identities and $\eta(.) : \mathbb{Z}^{(m+1)\times 2} \to \mathbb{R}^{k_2}$ to encode hidden view identities. We then concatenate their outputs with the embedding of the node's corresponding timestamp. Finally, we use an MLP-Mixer (Tolstikhin et al., 2021) to mix these encodings to obtain the walk encoding:

$$\textsc{Enc}(\hat{W}) = \textsc{Mean}\left(\mathbf{H}_{\text{token}} + \mathbf{W}^{(2)}\sigma\left(\texttt{LayerNorm}\left(\mathbf{H}_{\text{token}}\right)\mathbf{W}^{(1)}\right)\right), \tag{13}$$

where the i-th row of $\mathbf{H}_{\text{token}}$ is defined as:

$$\mathbf{H}^i_{\text{token}} = \left[\zeta\left(\textsc{Id}\left(\hat{W}(i,0)\right)\right)\;||\;\eta\left(\textsc{Id}^{\text{view}}\left(\hat{W}(i,1)\right)\right)\;||\;\mathscr{T}(t_i)\right]. \tag{14}$$

In the above equations, $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$, $\mathbf{W}^{(1)}_{\text{token}}$, and $\mathbf{W}^{(2)}_{\text{token}}$ are learnable parameters, `LayerNorm` is layer normalization (Ba et al., 2016) and $\sigma(.)$ is a nonlinear function (e.g., Gaussian error linear units, GeLU (Hendrycks and Gimpel, 2020)).

Anomaly Score. To assign an anomaly score to a given link $e = (u, v, r, t) \in \mathcal{E}$, we first sample temporal multiplex walks and then encode each walk $W \in \mathcal{S}_{\text{inter}}(u) \cup \mathcal{S}_{\text{inter}}(v) \cup \mathcal{S}_{\text{intra}}(u) \cup \mathcal{S}_{\text{intra}}(v)$ as described above. Next, we use an $\textsc{Agg}(.)$ function (e.g., mean-pooling) to aggregate walks' encodings and encode link e. Finally, we use a 2-layer perceptron to make the anomaly score:

$$\varphi(e) = \texttt{MLP}\left(\frac{1}{M + M'} \sum_{\hat{W}} \textsc{Enc}(\hat{W})\right), \tag{15}$$

where M and M' are the numbers of inter-view and intra-view walks.

3.5. ADMIRE Framework

We next explain how we use the view-aware edge encoding method to detect anomalous interactions. Figure 1 illustrates the ADMire framework.

Negative Sample Generator. We generate negative samples to train ADMire in an unsupervised manner. Previous anomaly detection methods mostly use (simple or biased) random negative samples (Zheng et al., 2019; Behrouz and Seltzer, 2022), which limit their generalizability to real anomalous patterns (Poursafaei et al., 2022). Moreover, these methods are designed for simple networks and cannot generalize to anomalous patterns in multiplex networks (see §5.1). Inspired by Poursafaei et al. (2022), we design a novel negative sampling method for temporal *multiplex* networks.

Let $\mathcal{E}_{\text{train}}$ and $\mathcal{E}_{\text{t}}$ be the set of edges in the training set and in timestamp t, respectively. For each edge in the training set $e = (u, v, r, t) \in \mathcal{E}$, we generate three types of negative samples: ① Inter-view negative samples: We use these negative samples so our model learns to detect connections that are anomalous across different views. We randomly generate a negative connection with relation type r with probability inversely proportional to the number of views in which this connection appears. The intuition is that if two nodes are already connected with several types of connections, a connection of yet another type is unlikely to be an anomalous connection. ② Intra-view negative samples: Here, we follow previous negative sampling generation methods (Zheng et al., 2019; Behrouz and Seltzer, 2022) and randomly change one endpoint of a connection to another node and keep the type of connection unchanged. ③ Historical negative samples: we generate negative edges from the set of edges that have been observed during previous timestamps but are absent in the current timestamp. That is, we randomly sample an edge $e \in \mathcal{E}_{\text{train}} \cap \bar{\mathcal{E}}_{\text{t}}$.

Training and Loss Function. Let $\mathcal{E}_{\text{train}}$ be the set of edges in the training set and $\mathcal{E}_{\text{neg}}$ be the set of generated negative samples. For each link $e \in \mathcal{E}_{\text{train}} \cup \mathcal{E}_{\text{neg}}$ we generate temporal multiplex walks to find view-aware edge encoding of e. Next, we use the margin-based pairwise loss (Bordes et al., 2013) to train the model. To avoid overfitting, we also use an $L2$-regularization loss, $\mathscr{L}_r^{reg}$, which is the summation of the $L2$ norm of all trainable parameters. This produces the loss function:

$$\mathscr{L} = \sum_{(u,v,r,t) \in \mathcal{E}_{\text{train}}} \sum_{(u',v',r,t) \in \mathcal{E}_{\text{neg}}} \max\left\{0, \gamma + \varphi(u, v, r, t) - \varphi(u', v', r, t)\right\} + \lambda \mathscr{L}^{reg}, \tag{16}$$

where $0 \leq \gamma \leq 1$ is the margin between normal and negative sampled edges.

4. Data and Experimental Settings

4.1. Data and Preprocessing

We evaluate ADMireusing three real-world datasets, PD (Day et al., 2019), ADHD (Brown et al., 2012), and ASD (Craddock et al., 2013), as well as three synthetic datasets. Each of the datasets represents one type of multiplex brain network modeling proposed in §3.2.

PD Dataset. Attention dysfunction is a common symptom of Parkinson's disease (PD) and has a significant impact on quality of life. This dataset (Day et al., 2019) uses the Attention Network Test (ANT) (Fan et al., 2005) and is designed to study three aspects of attention: alerting (maintaining an alert state), executive control (resolving conflict), and orienting. It consists of structural and functional MRI images of participants with and without PD, with six repetitions of the ANT task (Fan et al., 2005). It contains data for 25 subjects (7 female, age = 66.1 ± 10.0 yrs, years since disease onset = 8.4 ± 4.8) in the PD group and 21 subjects (12 female, age = 62.1 ± 9.9 yrs) in the healthy control group. We model the data using a temporal multiplex brain network with two views, 114 ROIs, and six timestamps (fMRI during each task). The first view represents the brain network obtained from fMRI, while the second view represents that generated from T1-weighted structural MRI.

ADHD Dataset. This dataset consists of resting fMRI data taken from USC Multimodal Connectivity Database (USCD) (Brown et al., 2012). The dataset contains data for 50 subjects (27 female, age = 9.84 ± 3.57 yrs) in the PD group and 50 subjects (25 female, age = 12.74 ± 4.1 yrs) in the typically developed (TD) control group. This dataset is preprocessed [1]. We model this data using a temporal multiplex brain network with 50 views, 190 ROIs, and 10 timestamps; each view represents the brain network of an individual.

ASD Dataset. This dataset consists of resting fMRI data taken from the Autism Brain Imaging Data Exchange (ABIDE) (Craddock et al., 2013); it contains data for 45 subjects (23 female, age = 23.1 ± 8.1 yrs) in the ASD group and 45 subjects (22 female, age = 25.4 ± 8.9 yrs) in the typically developed control group. We have followed the five pre-processing strategies denoted as DPARSF, followed by Band-Pass Filtering with different filters in a range between 0.01 and 0.25 Hz, in steps of 0.02 Hz. This range and steps are previously motivated by (De Domenico et al., 2016). We model this data using a temporal multiplex brain network with 12 views, 116 ROIs, and 10 timestamps; the i-th view represents the brain network obtained by filtering the fMRI values in the range $[0.01 + (i-1) \times 0.02, 0.01 + i \times 0.02]$ Hz.

Synthetic Datasets. We use synthetic datasets to show ① the effectiveness of ADMire in detecting anomalous connections compare to baselines, ② the importance of each element in our framework (ablation study), and ③ the advantage of modeling brain images as multiplex networks compared to modeling them as monoplex networks. Since the ground truth label for anomaly detection (specifically in brain networks) is difficult to obtain (Akoglu et al., 2015) (ground truth is unknown in many real neuroimaging data), we follow the methodology used in existing studies (Akoglu et al., 2015; Zheng et al., 2019; Yu et al., 2018b; Behrouz and Seltzer, 2022) and synthetically inject anomalous edges into our brain networks of people in the control group (healthy or TD) from our datasets. Accordingly, the nature of our

1. https://ccraddock.github.io/cluster_roi/atlases.html

synthetic datasets is real brain networks; however, synthetically anomalous connections are added to mitigate the lack of labeled data.

Pre-pocessing. Unless stated otherwise, for preprocessing and constructing brain networks from original fMRI and DTI data, we use the FSL toolbox and BrainGB (Cui et al., 2022a). Each edge in the fMRI brain networks shows that the statistical correlation between its endpoint is more than 80-th percentile of the distribution of correlation values.

4.2. Evaluation Approach/Study Design

The goal of our experiments on synthetic data is to validate our claims and to compare the performance of ADMire to different baselines. We follow previous studies (Behrouz and Seltzer, 2022; Zheng et al., 2019) and use the area under the ROC curve (AUC) as the metric of comparison. We use AnoMulY (Behrouz and Seltzer, 2022), GOutlier (Aggarwal et al., 2011), NetWalk (Yu et al., 2018b), AddGraph (Zheng et al., 2019), ML-GCN (Behrouz and Hashemi, 2022), and MNE (Zhang et al., 2018b) as baselines in the transductive setting. Since there is no prior work on inductive learning in multiplex networks, we compare our model with inductive network embedding in monoplex networks, CAW-N (Wang et al., 2021), TGAT (da Xu et al., 2020), and EvolveGCN (Pareja et al., 2020). For the detailed explanation of baselines see Appendix F. Next, we use real-world datasets to study abnormal connections in the brain of people living with PD, ADHD, or ASD. In each real-world experiment, we use the neuroimages of people in the control group (either healthy or TD) to train our model. Once our model is trained, we test it by using the neuroimages of people in the condition group (living with PD, ADHD, or ASD). In the inductive setting, we follow previous works (Wang et al., 2021; Jin et al., 2022) and randomly hide 10% of nodes in the training phase. The code, datasets, and supplements are available[2].

5. Results

5.1. Results on Synthetic Experiments

Effectiveness Evaluation. We first compare the effectiveness of ADMire with baselines, including state-of-the-art edge anomaly detection in multiplex networks, in detecting synthetic anomalous connections. Table 1 reports the AUC of methods on different datasets. ADMire outperforms all baselines by a significant margin (min = 6.42% and max = 18.04% performance improvement over the best baseline) in the transductive setting. There are four reasons for ADMire's superior performance: ① ADMire outperforms monoplex methods as it is a multiplex method and is able to learn from different data sources, image modalities, or frequency bands by using inter-view walks and attention mechanism over intra-view walks that incorporates complementary information from different views. ② It outperforms multiplex methods as it can capture the complex underlying rules of brain activities through its two causal walks over time. Also, its anonymization process hides the identity of nodes and views and captures the correlation between walks, increasing its ability to generalize better to unseen patterns in the test data. Previous methods use GRU cells to update node embedding over time, limiting their ability to generalize to unseen patterns in the test

2. Due to the double-blind reviewing policy, we will release the repository after reviewing phase.

Table 1: Performance comparison (AUC).

	Methods	PD		ADHD		ASD	
	Anomaly %	1%	5 %	1%	5 %	1%	5 %
Transductive	**Monoplex Methods**						
	GOutlier	$61.42_{\pm 1.04}$	$59.98_{\pm 2.21}$	$65.37_{\pm 0.93}$	$64.70_{\pm 2.09}$	$60.85_{\pm 0.97}$	$59.13_{\pm 1.86}$
	NetWalk	$69.71_{\pm 1.99}$	$69.02_{\pm 2.83}$	$70.29_{\pm 2.15}$	$69.86_{\pm 2.58}$	$69.07_{\pm 2.20}$	$68.52_{\pm 2.55}$
	AddGraph	$71.94_{\pm 1.64}$	$70.33_{\pm 3.01}$	$71.89_{\pm 1.48}$	$70.11_{\pm 2.06}$	$71.30_{\pm 1.38}$	$70.96_{\pm 2.02}$
	Multiplex Methods						
	MNE	$70.39_{\pm 1.22}$	$70.54_{\pm 1.30}$	$73.78_{\pm 2.14}$	$72.31_{\pm 2.36}$	$70.19_{\pm 1.62}$	$69.94_{\pm 1.98}$
	ML-GCN	$68.50_{\pm 1.46}$	$68.33_{\pm 1.73}$	-*	-*	$69.56_{\pm 2.25}$	$69.35_{\pm 2.84}$
	AnoMulY	$78.07_{\pm 3.25}$	$79.85_{\pm 2.89}$	-*	-*	$77.14_{\pm 2.37}$	$77.08_{\pm 1.79}$
	ADMire	$\mathbf{85.09}_{\pm \mathbf{1.76}}$	$\mathbf{84.98}_{\pm \mathbf{2.12}}$	$\mathbf{88.67}_{\pm \mathbf{2.73}}$	$\mathbf{88.53}_{\pm \mathbf{1.76}}$	$\mathbf{91.06}_{\pm \mathbf{1.48}}$	$\mathbf{89.95}_{\pm \mathbf{2.51}}$
Inductive	EvolveGCN	$55.18_{\pm 3.10}$	$55.06_{\pm 2.31}$	$57.23_{\pm 2.81}$	$57.41_{\pm 2.54}$	$56.89_{\pm 2.48}$	$56.21_{\pm 2.34}$
	TGAT	$59.34_{\pm 2.76}$	$58.72_{\pm 3.39}$	$60.19_{\pm 2.86}$	$60.10_{\pm 3.15}$	$60.28_{\pm 3.58}$	$59.93_{\pm 3.50}$
	CAW-N	$75.85_{\pm 1.67}$	$75.90_{\pm 1.85}$	$71.64_{\pm 1.26}$	$71.02_{\pm 1.61}$	$71.31_{\pm 2.34}$	$71.96_{\pm 2.87}$
	ADMire	$\mathbf{84.72}_{\pm \mathbf{2.45}}$	$\mathbf{84.31}_{\pm \mathbf{2.27}}$	$\mathbf{88.03}_{\pm \mathbf{2.23}}$	$\mathbf{88.97}_{\pm \mathbf{1.84}}$	$\mathbf{90.49}_{\pm \mathbf{2.12}}$	$\mathbf{90.28}_{\pm \mathbf{2.45}}$

* Training time exceeds the threshold.

Table 2: Ablation study (AUC).

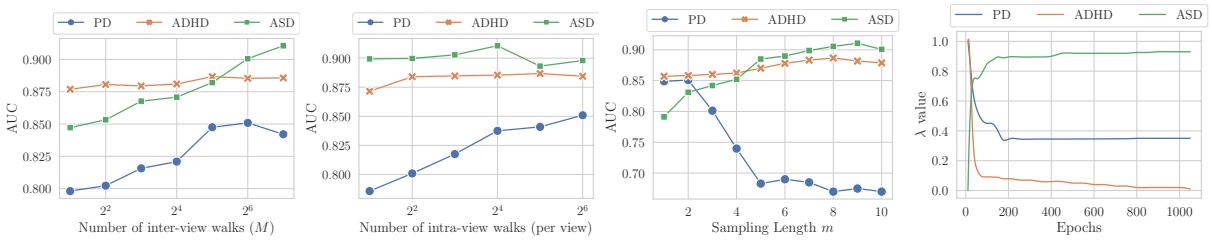
	Methods	PD	ADHD	ASD
1	ADMire	**85.09**	**88.67***	**91.06**
2	w/o inter-view	78.59	**88.73***	80.65
3	w/o intra-view	77.14	69.59	79.62
4	w/o λ ($\lambda = 1$)	80.42	80.36	89.30
5	w/o attention	84.79	86.14	86.57
6	w/o time encoding	84.16	82.78	85.92
7	w/o inter-view NS	84.77	84.28	83.46
8	w/o intra-view NS	79.91	78.75	81.09
9	w/o historical NS	84.68	84.16	84.31
10	w/ RNN	83.90	85.32	89.13
11	Monoplex-ADMire	76.52	72.07	74.15

* There is no causal relation between views.

data. ③ ADMire is a stream-based method and use a time encoding module to capture time information, while the baselines are snapshot-based and aggregate links into network snapshots, which remove some useful time information (Wang et al., 2021). ④ It is scalable with respect to the number of views and can be trained on many data sources, image modalities, or frequency bands. AnoMulY, the state-of-the-art, as well as ML-GCN use different Gnn modules for each view, making them infeasible for large networks with a large number of views (e.g., ADHD dataset with 50 subjects).

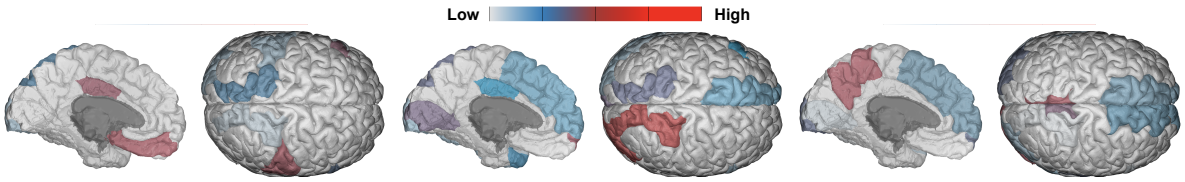
Finally, Table 1 reports the performance of ADMire and inductive baselines in the inductive setting. We attribute ADMire's superior performance (with min = 11.08% and max = 26.89% improvement over the best baseline) to two main reasons: ① ADMire is an end-to-end anomaly detection method with an exclusive design of generating negative samples and training process, while baselines are designed to learn the temporal and structural properties of the network. ② ADMire is a multiplex method, while baselines are monoplex methods.

Ablation Studies. We further conduct ablation studies to validate the effectiveness of critical components of ADMire. The results are summarized in Table 2. Rows 2 and 3 show the effectiveness of inter-view and intra-view walks. The only exception is removing the inter-view walks in the ADHD dataset. As we discussed in § 3.4, when there is no causal relation between views, inter-view walks are not informative and our model is expected to learn to ignore these walks (sets $\lambda = 0$). Accordingly, removing these walks from the ADMire cannot much change the performance on the ADHD dataset, when there is no causal relation between views. Rows 4 and 5 show the importance of the learnable parameter λ and attention mechanism to incorporate the information of different views. Rows 7, 8, and 9 show the importance of our new negative sample generator. When using RNN instead of MLP-Mixer in the walk encoding phase (row 10), we gain better performance due to its ability to learn the dependency of nodes' encoding in a walk. Finally, the last row shows the superior performance of multiplex ADMire over monoplex ADMire, when using only one brain network generated from an individual, image modality, or frequency band. These results show the importance of multiplex modeling of neuroimages. We further discuss the importance of multiplex modeling in Appendix G.1.



(a) Intra-view walks (b) Inter-view walks (c) Walk Length (d) Evolution of λ

Figure 2: The effect of hyperparameters on the performance (a-c), and λ evolution (d).



(a) Results on PD (b) Results on ADHD (c) Results on ASD

Figure 3: The distribution of anomalous edges in condition groups.

Parameter Sensitivity. We systematically analyze the effect of hyperparameters used in ADMire on the performance. Figure 2(a) shows that only a small number of intra-view walks are enough to achieve competitive performance. While more inter-walks improve the performance until some point, the performance gain is saturated after that. A similar pattern can be seen for increasing the number of intra-view walks (Figure 2(b)). Note that this figure reports the number of intra-view walks per view. Accordingly, it is expected to see more performance gain on datasets with a smaller number of views (e.g., PD). Figure 2(c) shows that ADMire might achieve the best performance at a certain walk length, while the exact value depends on the complexity of higher-order motifs that are required to learn underlying network dynamic law as well as the number of views. Note that since inter-view walks cross over different views, networks with a large number of views might need longer walks to learn the causal relation in different views. Finally, Figure 2(d) shows the evolution of λ in training. As expected, in datasets with no causal relationship between different views (e.g., ADHD), ADMire learns to set $\lambda \leq 0.1$ in a few numbers of epochs. For other datasets, it shows that ADMire converges very quickly to the best value of λ.

5.2. Results on Real-world Datasets

In this section, we report our findings from applying ADMire on real-world datasets. We train our model on the healthy control group and then test it on the condition group (living with PD, ADHD, ASD) to find anomalous brain activities of people in the condition group. Additional visualizations and results on real-world datasets can be found in Appendix G.2.

Parkinson's Disease. In this experiment, we focus on abnormal *brain structure* and *functional* activities of PD patients. Since the brain of each individual in each task might also have complex exclusive activities, we need to focus on common or more frequently

appeared anomalous connections between ROIs over different subjects to capture abnormal activities that might be correlated to PD. To this end, we show how anomalous connections found by ADMire are distributed in the brain of people living with PD. Figure 3(a) reports the average distribution of anomalous edges in the brain networks of people living with PD. Most anomalous edges found by ADMire have a vertex in either *Posterior Cingulate*, *Superior Parietal*, *Medial Orbitofrontal*, *Pars Opercularis*, or *Supramarginal Gyrus* ($\geq 95\%$ of all found anomalies). Next, we apply ADMire on the healthy control group to see whether these findings are exclusive to the PD group and to identify possible noise in the dataset. We observe that ADMire finds 94.2% less anomalous connections in the healthy control group, most of which have a node in either *Temporal Pole* or *Anterior Insula*.

Attention Deficit Hyperactivity Disorder. Following the previous part, we first focus on the abnormal brain functional activities of subjects in the ADHD group. Figure 3(b) shows the average distribution of anomalous edges in the brain networks of subjects in the condition ADHD group. Most abnormal connections found by ADMire have an endpoint in either *Frontal Pole*, *Right Lateral Occipital Cortex*, *Lingual Gyrus*, *Left Temporal Pole*, or *Right Superior Parietal Lobule* ($\geq 95\%$ of all found anomalies). Applying ADMire on the healthy control group, we observe that ADMire finds 89.6% less anomalous connections in the healthy control group, most of which have an endpoint in either *Planum Polare* or *Angular Gyrus*. Interestingly, these findings are consistent with previous studies on ADHD, using voxel-wise estimation of regional tissue volume changes (Wang et al., 2007), abnormality in DTI images (Lei et al., 2014), and Forman–Ricci curvature changes (Chatterjee et al., 2021), which shows the potential of ADMire in revealing abnormal connections that might be correlated to a brain disease or disorder.

Autism Spectrum Disorder. Figure 3(b) shows the average distribution of anomalous edges in the brain networks of subjects in the condition ASD group. Most abnormal connections found by ADMire have an endpoint in either *Right Superior Temporal Gyrus*, *Right Cerebellum Cortex*, *Right Precuneus*, *Frontal Pole*, *Left Lateral Occipital* ($\geq 95\%$ of all found anomalies). Applying ADMire on the healthy control group, ADMire finds 93.7% less anomalous connections in the healthy control group, most of which have an endpoint in either *Temporal Pole* or *Posterior Cingulate Cortex*. Although several works have studied ASD and found different abnormality patterns, there is still no known ASD biomarker (Müller and Linke, 2021). However, a part of our findings about the abnormal activity in the cerebellum cortex is consistent with previous studies (Rogers et al., 2013).

6. Discussion

In this paper, we discuss the importance of using multiple neuroimage data sources and suggest three approaches–different frequency bands, image modalities, and subjects– to model neuroimages as multiplex networks, taking advantage of complementary information provided by these multiple data sources. Next, we present ADMire, an end-to-end inductive unsupervised learning method on multiplex networks to detect abnormal brain activities that might cause a brain disease or disorder. ADMire uses inter-view (resp. intra-view) temporal walks to implicitly extract network motifs and causal relationships across different views (resp. within a view), and adopts novel anonymization based on the correlation between

network motifs to hide the identity of nodes and views. Next, it uses an MLP-Mixer is used to encode the sequence of nodes in a walk. In the training phase, we design a new negative sample generation to train the model in an unsupervised manner.

Our experimental results show the importance of using multiple neuroimage data and also the power of ADMire to effectively identify temporal and structural anomalies in the human brain. We evaluate the performance of ADMire in the inductive setting and show ADMire can generalize well to detect abnormal connections between unseen ROIs. This is especially important in brain analysis, due to the cost of neuroimage data and potential noise in them. Finally, we report our findings from applying ADMire on the brain networks of subjects with PD, ADHD, and ASD. While some findings are new, several found abnormal activities are compatible with previous studies that used costly, non-learning, and/or supervised approaches. New findings also show the potential of ADMire in detecting abnormal activities that are missed by previous studies and might cause a brain disease or disorder. We leave this question that whether the found anomalies are biomarkers of these diseases/disorders for future works.

Limitations This study has two main limitations: ① Simple negative samples: it is unreasonable to naively assume that the anomalous brain activities are equal to random negative samples as the brain activities of each individual are exclusive and complex, and this simple method might introduce bias into the negative samples. One important future direction is to address this limitation by using non-contrastive learning methods or generating more sophisticated negative samples. Also, another potential approach is to use a module to learn to generate hard negative samples (Du et al., 2023). ② Black-box method: Health-related domains are highly sensitive and require interpretable or explainable models. However, ADMire is a black-box approach and the lack of interpretability is one of its major limitations. Fortunately, the learning process of ADMire is based on extracting network motifs and pairing it with neural network interpretation technique (Montavon et al., 2018) is a potential future work to make it explainable and address its lack of transparency in prediction (Behrouz and Seltzer, 2023).

Acknowledgments

This research was enabled in part by support provided by WestGrid (https://www.westgrid.ca) and The Digital Research Alliance (https://alliancecan.ca/en). We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Carlo Abrate and Francesco Bonchi. Counterfactual graphs for explainable classification of brain networks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery; Data Mining*, KDD '21, page 2495–2504, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467154. URL https://doi.org/10.1145/3447548.3467154.

Charu C. Aggarwal, Yuchen Zhao, and Philip S. Yu. Outlier detection in graph streams. In *2011 IEEE 27th International Conference on Data Engineering*, pages 399–409, 2011. doi: 10.1109/ICDE.2011.5767885.

Uday Agrawal, Emery N. Brown, and Laura D. Lewis. Model-based physiological noise removal in fast fmri. *NeuroImage*, 205:116231, 2020. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2019.116231. URL https://www.sciencedirect.com/science/article/pii/S1053811919308225.

Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, May 2015. ISSN 1573-756X. doi: 10.1007/s10618-014-0365-y. URL https://doi.org/10.1007/s10618-014-0365-y.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization, 2016.

Monika Bansal and Dolly Sharma. Ranking and discovering anomalous neighborhoods in attributed multiplex networks. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, page 46–54, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450377386. doi: 10.1145/3371158.3371164. URL https://doi.org/10.1145/3371158.3371164.

Danielle S. Bassett and Olaf Sporns. Network neuroscience. *Nature Neuroscience*, 20(3): 353–364, Mar 2017. ISSN 1546-1726. doi: 10.1038/nn.4502. URL https://doi.org/10.1038/nn.4502.

Ali Behrouz and Farnoosh Hashemi. Cs-mlgcn: Multiplex graph convolutional networks for community search in multiplex networks. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, CIKM '22, page 3828–3832, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392365. doi: 10.1145/3511808.3557572. URL https://doi.org/10.1145/3511808.3557572.

Ali Behrouz and Margo Seltzer. Anomaly detection in multiplex dynamic networks: from blockchain security to brain disease prediction. In *NeurIPS 2022 Temporal Graph Learning Workshop*, 2022. URL https://openreview.net/forum?id=UDGZDfwmay.

Ali Behrouz and Margo Seltzer. ADMIRE++: Explainable anomaly detection in the human brain via inductive learning on temporal multiplex networks. In *ICML 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH)*, 2023. URL https://openreview.net/forum?id=t4H8acYudJ.

Ali Behrouz, Farnoosh Hashemi, and Laks V. S. Lakshmanan. Firmtruss community search in multilayer networks. *Proc. VLDB Endow.*, 16(3):505–518, nov 2022. ISSN 2150-8097. doi: 10.14778/3570690.3570700. URL https://doi.org/10.14778/3570690.3570700.

Ali Behrouz, Farnoosh Hashemi, Sadaf Sadeghian, and Margo Seltzer. Cat-walk: Inductive hypergraph learning via set walks, 2023.

P.V. Bindu, P. Santhi Thilagam, and Deepesh Ahuja. Discovering suspicious behavior in multilayer social networks. *Computers in Human Behavior*, 73:568–582, 2017. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2017.04.001. URL https://www.sciencedirect.com/science/article/pii/S0747563217302303.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

Jesse A Brown, Jeffrey D Rudie, Anita Bandrowski, John D Van Horn, and Susan Y Bookheimer. The UCLA multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. *Front Neuroinform*, 6:28, November 2012.

Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. Representation learning for attributed multiplex heterogeneous network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1358–1368, 2019.

Benjamin Paul Chamberlain, Sergey Shirobokov, Emanuele Rossi, Fabrizio Frasca, Thomas Markovich, Nils Yannick Hammerla, Michael M. Bronstein, and Max Hansmire. Graph neural networks for link prediction with subgraph sketching. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=m1oqEOAozQU.

Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. Heterogeneous network embedding via deep architectures. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 119–128, 2015.

Sudhanshu Chanpuriya, Ryan A. Rossi, Sungchul Kim, Tong Yu, Jane Hoffswell, Nedim Lipka, Shunan Guo, and Cameron N Musco. Direct embedding of temporal network edges via time-decayed line graphs. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Qamz7Q_Ta1k.

Tanima Chatterjee, Réka Albert, Stuti Thapliyal, Nazanin Azarhooshang, and Bhaskar DasGupta. Detecting network anomalies using forman–ricci curvature and a case study for human brain networks. *Scientific Reports*, 11(1):8121, Apr 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87587-z. URL https://doi.org/10.1038/s41598-021-87587-z.

Gang Chen, B Douglas Ward, Chunming Xie, Wenjun Li, Zhilin Wu, Jennifer L Jones, Malgorzata Franczak, Piero Antuono, and Shi-Jiang Li. Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional mr imaging. *Radiology*, 259(1):213, 2011.

Ling-Hao Chen, He Li, and Wenhao Yang. Anomman: Detect anomaly on multi-view attributed networks, 2022. URL https://arxiv.org/abs/2201.02822.

Jiafeng Cheng, Qianqian Wang, Zhiqiang Tao, Deyan Xie, and Quanxue Gao. Multi-view attribute graph convolution networks for clustering. In *Proceedings of the Twenty-Ninth*

International Conference on International Joint Conferences on Artificial Intelligence, pages 2973–2979, 2021.

Weilin Cong, Si Zhang, Jian Kang, Baichuan Yuan, Hao Wu, Xin Zhou, Hanghang Tong, and Mehrdad Mahdavi. Do we really need complicated model architectures for temporal networks? In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ayPPc0SyLv1.

Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7:27, 2013.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Brainnnexplainer: An interpretable graph neural network framework for brain network based disease analysis. *arXiv preprint arXiv:2107.05097*, 2021.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xuan Kan, Antonio Aodong Chen Gu, Joshua Lukemire, Liang Zhan, Lifang He, Ying Guo, and Carl Yang. BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks. *IEEE Transactions on Medical Imaging (TMI)*, 2022a.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 375–385. Springer, 2022b.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 375–385, Cham, 2022c. Springer Nature Switzerland. ISBN 978-3-031-16452-1.

da Xu, chuanwei ruan, evren korpeoglu, sushant kumar, and kannan achan. Inductive representation learning on temporal graphs. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rJeW1yHYwH.

Trevor K M Day, Tara M Madhyastha, Mary K Askren, Peter Boord, Thomas J Montine, and Thomas J Grabowski. Attention network test fMRI data for participants with parkinson's disease and healthy elderly. *F1000Res*, 8:780, June 2019.

Manlio De Domenico. Multilayer modeling and analysis of human brain networks. *GigaScience*, 6(5), 02 2017. ISSN 2047-217X. doi: 10.1093/gigascience/gix004. URL https://doi.org/10.1093/gigascience/gix004. gix004.

Manlio De Domenico, Shuntaro Sasai, and Alex Arenas. Mapping multiplex hubs in human functional brain networks. *Front Neurosci*, 10:326, July 2016.

Fabrizio De Vico Fallani, Jonas Richiardi, Mario Chavez, and Sophie Achard. Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philos Trans R Soc Lond B Biol Sci*, 369(1653), October 2014.

Boxin Du, Lihui Liu, Jiejun Xu, Fei Wang, and Hanghang Tong. Neural multi-network diffusion towards social recommendation. *arXiv preprint arXiv:2304.04994*, 2023.

Jin Fan, Bruce D McCandliss, John Fossella, Jonathan I Flombaum, and Michael I Posner. The activation of attentional networks. *Neuroimage*, 26(2):471–479, 2005.

Emily S. Finn, Xilin Shen, Dustin Scheinost, Monica D. Rosenberg, Jessica Huang, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671, Nov 2015. ISSN 1546-1726. doi: 10.1038/nn.4135. URL https://doi.org/10.1038/nn.4135.

Ziqi Gao, Chenran Jiang, Jiawen Zhang, Xiaosen Jiang, Lanqing Li, Peilin Zhao, Huanming Yang, Yong Huang, and Jia Li. Hierarchical graph learning for protein–protein interaction. *Nature Communications*, 14(1):1093, 2023.

Farnoosh Hashemi, Ali Behrouz, and Laks V.S. Lakshmanan. Firmcore decomposition of multilayer networks. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 1589–1600, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450390965. doi: 10.1145/3485447.3512205. URL https://doi.org/10.1145/3485447.3512205.

Farnoosh Hashemi, Ali Behrouz, and Milad Rezaei Hajidehi. Cs-tgn: Community search via temporal graph neural networks. In *Companion Proceedings of the Web Conference 2023*, WWW '23, New York, NY, USA, 2023. Association for Computing Machinery. doi: 10.1145/3543873.3587654. URL https://doi.org/10.1145/3543873.3587654.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2020.

Biao Jie, Mingxia Liu, Xi Jiang, and Daoqiang Zhang. Sub-network based kernels for brain network classification. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 622–629, 2016.

Ming Jin, Yuan-Fang Li, and Shirui Pan. Neural temporal walks: Motif-aware representation learning on continuous-time dynamic graphs. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=NqbktPUkZf7.

Baoyu Jing, Chanyoung Park, and Hanghang Tong. Hdmi: High-order deep multiplex infomax. In *Proceedings of the Web Conference 2021*, pages 2414–2424, 2021.

Xuan Kan, Hejie Cui, Ying Guo, and Carl Yang. Effective and interpretable fmri analysis via functional brain network generation. *arXiv preprint arXiv:2107.11247*, 2021.

Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. *arXiv preprint arXiv:2205.12465*, 2022a.

Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. FBNETGEN: Task-aware GNN-based fMRI analysis via functional brain network generation. In *Medical Imaging with Deep Learning*, 2022b. URL https://openreview.net/forum?id=oWFphg2IKon.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022c. URL https://openreview.net/forum?id=1cJ1cbA6NLN.

Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*, 2019.

Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 07 2014. ISSN 2051-1310. doi: 10.1093/comnet/cnu016.

Tommaso Lanciano, Francesco Bonchi, and Aristides Gionis. Explainable classification of brain networks via contrast subgraphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining*, KDD '20, page 3308–3318, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379984. doi: 10.1145/3394486.3403383. URL https://doi.org/10.1145/3394486.3403383.

Du Lei, Jun Ma, Xiaoxia Du, Guohua Shen, Xingming Jin, and Qiyong Gong. Microstructural abnormalities in the combined and inattentive subtypes of attention deficit hyperactivity disorder: a diffusion tensor imaging study. *Scientific reports*, 4(1):6875, 2014.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SJiHXGWAZ.

Jiaxin Liu, Wei Zhao, Ye Hong, Sheng Gao, Xi Huang, Yingjie Zhou, Alzheimer's Disease Neuroimaging Initiative, et al. Learning features of brain network for anomaly detection. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 900–905. IEEE, 2020.

Jin Liu, Min Li, Yi Pan, Wei Lan, Ruiqing Zheng, Fang-Xiang Wu, and Jianxin Wang. Complex brain network analysis and its applications to brain disorders: A survey. *Complexity*, 2017:8362741, Oct 2017. ISSN 1076-2787. doi: 10.1155/2017/8362741. URL https://doi.org/10.1155/2017/8362741.

Antonio Longa, Veronica Lachi, Gabriele Santin, Monica Bianchini, Bruno Lepri, Pietro Lio, Franco Scarselli, and Andrea Passerini. Graph neural networks for temporal graphs: State of the art, open challenges, and opportunities, 2023.

Yuhong Luo and Pan Li. Neighborhood-aware scalable temporal network representation learning. In *The First Learning on Graphs Conference*, 2022. URL https://openreview.net/forum?id=EPUtNe7a9ta.

Christopher W. Lynn and Danielle S. Bassett. The physics of brain network structure, function and control. *Nature Reviews Physics*, 1(5):318–332, May 2019. ISSN 2522-5820. doi: 10.1038/s42254-019-0040-8. URL https://doi.org/10.1038/s42254-019-0040-8.

Razvan Marinescu, Arman Eshaghi, Daniel Alexander, and Polina Golland. Brainpainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes. *arXiv preprint arXiv:1905.08627*, 2019.

Asep Maulana and Martin Atzmueller. Centrality-based anomaly detection on multi-layer networks using many-objective optimization. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, volume 1, pages 633–638, 2020. doi: 10.1109/CoDIT49905.2020.9263819.

Silvio Micali and Zeyuan Allen Zhu. Reconstructing markov processes from independent and anonymous experiments. *Discrete Applied Mathematics*, 200:108–122, 2016.

Ruchi Mittal and MPS Bhatia. Anomaly detection in multiplex networks. *Procedia Computer Science*, 125:609–616, 2018.

Bratislav Mišić and Olaf Sporns. From regions to connections and networks: new bridges between brain and behavior. *Current Opinion in Neurobiology*, 40:1–7, 2016. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2016.05.003. URL https://www.sciencedirect.com/science/article/pii/S095943881630054X. Systems neuroscience.

Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital signal processing*, 73:1–15, 2018.

Ralph-Axel Müller and Annika Linke. Functional connectivity in autism spectrum disorders: Challenges and perspectives. *Brain Network Dysfunction in Neuropsychiatric Illness: Methods, Applications, and Implications*, pages 239–272, 2021.

Dorcas Ofori-Boateng, I Segovia Dominguez, C Akcora, Murat Kantarcioglu, and Yulia R Gel. Topological anomaly detection in dynamic multilayer blockchain networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 788–804. Springer, 2021.

Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 5363–5370, 2020.

Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5371–5378, 2020.

Hae-Jeong Park and Karl Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342(6158):1238411, 2013. doi: 10.1126/science.1238411. URL https://www.science.org/doi/abs/10.1126/science.1238411.

Hao Peng, Hongfei Wang, Bowen Du, Md Zakirul Alam Bhuiyan, Hongyuan Ma, Jianwei Liu, Lihong Wang, Zeyu Yang, Linfeng Du, Senzhang Wang, et al. Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Information Sciences*, 521:277–290, 2020.

Léo Pio-Lopez, Alberto Valdeolivas, Laurent Tichit, Élisabeth Remy, and Anaïs Baudot. Multiverse: a multiplex and multiplex-heterogeneous network embedding approach. *Scientific Reports*, 11(1):1–20, 2021.

Farimah Poursafaei, Andy Huang, Kellin Pelrine, and Reihaneh Rabbany. Towards better evaluation for dynamic link prediction. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. URL https://openreview.net/forum?id=1GVpwr2Tfdg.

Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160:41–54, 2017.

Tiffany D Rogers, Eric McKimm, Price E Dickson, Dan Goldowitz, Charles D Blaha, and Guy Mittleman. Is autism a disease of the cerebellum? an integration of clinical and pre-clinical research. *Front Syst Neurosci*, 7:15, May 2013.

Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *International conference on neural information processing*, pages 362–373. Springer, 2018.

Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Peter Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. MLP-mixer: An all-MLP architecture for vision. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=EI2KOXKdnP.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=rklz9iAcKQ.

Jian-zhe Wang, Tian-zi Jiang, Qing-jiu Cao, and Yufeng Wang. Characterizing anatomic differences in boys with attention-deficit/hyperactivity disorder with the use of deformation-based morphometry. *American Journal of Neuroradiology*, 28(3):543–547, 2007.

Qifan Wang, Yi Fang, Anirudh Ravula, Ruining He, Bin Shen, Jingang Wang, Xiaojun Quan, and Dongfang Liu. Deep partial multiplex network embedding. In *Companion Proceedings of the Web Conference 2022*, WWW '22, page 1053–1062, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391306. doi: 10.1145/3487553.3524717. URL https://doi.org/10.1145/3487553.3524717.

Xiao Wang, Yuanfu Lu, Chuan Shi, Ruijia Wang, Peng Cui, and Shuai Mou. Dynamic heterogeneous information network embedding with meta-path based proximity. *IEEE Transactions on Knowledge and Data Engineering*, 2020a.

Xiaoyang Wang, Yao Ma, Yiqi Wang, Wei Jin, Xin Wang, Jiliang Tang, Caiyan Jia, and Jian Yu. Traffic flow prediction via spatial temporal graph neural network. In *Proceedings of The Web Conference 2020*, pages 1082–1092, 2020b.

Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, and Pan Li. Inductive representation learning in temporal networks via causal anonymous walks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=KYPz4YsCPj.

Chong-Yaw Wee, Pew-Thian Yap, Wenbin Li, Kevin Denny, Jeffrey N Browndyke, Guy G Potter, Kathleen A Welsh-Bohmer, Lihong Wang, and Dinggang Shen. Enriched white matter connectivity networks for accurate identification of mci patients. *Neuroimage*, 54 (3):1812–1822, 2011.

Yuanzhen Xie, Zijing Ou, Liang Chen, Yang Liu, Kun Xu, Carl Yang, and Zibin Zheng. Learning and updating node embedding on dynamic heterogeneous information network. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, page 184–192, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450382977. doi: 10.1145/3437963.3441745. URL https://doi.org/10.1145/3437963.3441745.

Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16):8749–8760, 2019.

Yuchen Yan, Si Zhang, and Hanghang Tong. Bright: A bridging algorithm for network alignment. In *Proceedings of the Web Conference 2021*, pages 3907–3917, 2021.

Jiaxuan You, Tianyu Du, and Jure Leskovec. Roland: Graph learning framework for dynamic graphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, page 2358–2366, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393850. doi: 10.1145/3534678.3539300. URL https://doi.org/10.1145/3534678.3539300.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *IJCAI*, pages 3634–3640, 2018a. URL https://doi.org/10.24963/ijcai.2018/505.

Wenchao Yu, Wei Cheng, Charu C. Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang. Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 2672–2681, New York, NY, USA, 2018b. Association for Computing Machinery. ISBN 9781450355520. doi: 10.1145/3219819.3220024. URL https://doi.org/10.1145/3219819.3220024.

Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. Generalized latent multi-view subspace clustering. *IEEE transactions on pattern analysis and machine intelligence*, 42(1):86–99, 2018a.

Hongming Zhang, Liwei Qiu, Lingling Yi, and Yangqiu Song. Scalable multiplex network embedding. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 3082–3088. International Joint Conferences on Artificial Intelligence Organization, 7 2018b. doi: 10.24963/ijcai.2018/428. URL https://doi.org/10.24963/ijcai.2018/428.

Jingyuan Zhang, Bokai Cao, Sihong Xie, Chun-Ta Lu, Philip S Yu, and Ann B Ragin. Identifying connectivity patterns for brain diseases via multi-side-view guided deep architectures. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 36–44. SIAM, 2016.

Ru-Yuan Zhang, Xue-Xin Wei, and Kendrick Kay. Understanding multivariate brain activity: Evaluating the effect of voxelwise noise correlations on population codes in functional magnetic resonance imaging. *PLOS Computational Biology*, 16(8):1–29, 08 2020. doi: 10.1371/journal.pcbi.1008153. URL https://doi.org/10.1371/journal.pcbi.1008153.

Xi Zhang, Lifang He, Kun Chen, Yuan Luo, Jiayu Zhou, and Fei Wang. Multi-View graph convolutional network and its applications on neuroimage analysis for parkinson's disease. *AMIA Annu Symp Proc*, 2018:1147–1156, December 2018c.

Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858, 2019.

Li Zheng, Zhenpeng Li, Jian Li, Zhao Li, and Jun Gao. Addgraph: Anomaly detection in dynamic graph using attention-based temporal gcn. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 4419–4425. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/614. URL https://doi.org/10.24963/ijcai.2019/614.

Yanqiao Zhu, Hejie Cui, Lifang He, Lichao Sun, and Carl Yang. Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 272–276. IEEE, 2022a.

Yanqiao Zhu, Hejie Cui, Lifang He, Lichao Sun, and Carl Yang. Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis, 2022b.

Appendix A. Reproducibility

The implementation of ADMire is available in https://github.com/ubc-systopia/ADMIRE.

Appendix B. Additional Related Work

To situate our research in a broader context, we briefly review research in ① temporal graph learning methods, ② multiplex graph learning, ③ feature learning in brain networks, ④ anomaly detection in brain networks, and ⑤ anomaly detection in multiplex networks.

Temporal Graph Learning. Learning from temporal networks has been a widely studied topic in the literature (Longa et al., 2023). The first group of methods uses a Graph Neural Network (GNN) as a feature encoder and then uses a sequence model on top of the GNN to capture temporal properties (Peng et al., 2020; Wang et al., 2020b; Yu et al., 2018a). The second group uses Recurrent Neural Networks (RNNs) with a GNN layer replacing the linear layer to learn from the temporal network (Li et al., 2018; Seo et al., 2018; Zhao et al., 2019; You et al., 2022). Recently, more conceptually complicated learning methods for temporal graphs have been designed based on temporal random walks (Wang et al., 2021; Jin et al., 2022), line graphs (Chanpuriya et al., 2023), neighborhood representation (Luo and Li, 2022), and subgraph sketching (Chamberlain et al., 2023). Cong et al. (2023) design a simple but effective temporal edge encoding method and show that self-attention mechanisms and RNNs are not essential for temporal graph learning. However, all these methods differ from our approach as they are designed for simple temporal graphs and cannot easily be extended to graphs with different types of edges (multiplex networks).

Multiplex Graph Learning. In the literature, multiplex networks (also known as multi-view, multilayer, or multi-dimensional networks) are graphs with a node type but multiple edge types (relations) (Kivelä et al., 2014). Several methods have been proposed to learn network embeddings on multiplex networks by integrating information from individual relation types (Cen et al., 2019; Pio-Lopez et al., 2021; Yan et al., 2021; Chang et al., 2015; Xie et al., 2021; Wang et al., 2020a). Other work proposed Graph Convolutional Networks (GCNs) methods for multiplex networks (Behrouz and Hashemi, 2022; Cheng et al., 2021; Zhang et al., 2018a). Inspired by Deep Graph Infomax (Veličković et al., 2019), Park et al. (2020) and Jing et al. (2021) proposed unsupervised approaches to learn node embeddings by maximizing the mutual information between local patches and the global representation of the entire graph. Zhang et al. (2018b) proposed a method that uses a latent space to integrate the information across multiple views. Recently, Wang et al. (2022) proposed DPMNE to learn from incomplete multiplex networks. All these methods are designed in the transductive setting for static multiplex networks, which is different from our formulation.

Feature Learning in Brain Networks. In recent years, several studies focused on analyzing brain networks to understand and distinguish healthy and diseased human brains (Jie et al., 2016; Chen et al., 2011; Wee et al., 2011). Recently, due to the success of GNNs in analyzing graph-structured data, deep models have been proposed to predict brain diseases by learning the graph structures of brain networks (Kan et al., 2021; Cui et al., 2021; Kan et al., 2022a; Zhu et al., 2022a; Cui et al., 2022b). All these methods are designed for the

graph or node classification and cannot easily be extended to the edge-anomaly detection task.

Anomaly Detection in Brain Networks. In addition to predicting disease in brain networks, understanding the cause of the disease is important. To this end, several anomaly detection methods have been proposed to find anomalous connections, regions, or subgraphs in the brain, which can cause a disease (Chatterjee et al., 2021; Zhang et al., 2016; Liu et al., 2020). All these methods are designed for node or subgraph anomaly detection tasks and cannot easily be extended to the edge-anomaly detection task.

Anomaly Detection in Multiplex Networks. The problem of anomaly detection in multiplex networks has recently attracted attention. Mittal and Bhatia (2018) use eigenvector centrality, page rank centrality, and degree centrality as handcrafted features for nodes to detect anomalies in static multiplex networks. Bindu et al. (2017) proposed a node anomaly detection algorithm in static multiplex networks that uses handcrafted features based on clique/near-clique and star/near-star structures. Bansal and Sharma (2020) defined a quality measure, Multi-Normality, which uses the structure and attributes together of each view to detect attribute coherence in neighborhoods between layers. Maulana and Atzmueller (2020) use centrality of all nodes in each view and apply many-objective optimization with full enumeration based on minimization to obtain Pareto Front. Then, they use Pareto Front as a basis for finding suspected anomaly nodes. Chen et al. (2022) proposed AnomMAN that uses an auto-encoder module and a GCN-based decoder to detect node anomalies in static multiplex networks. Although this model can learn from the data, it is limited to static networks, and it treats each view equally in the *Structure Reconstruction* step. Finally, Ofori-Boateng et al. (2021) developed a new persistence summary and used it to detect events in dynamic multiplex blockchain networks. All of these approaches are designed to detect topological anomalous subgraphs, nodes, or events, and cannot identify anomalous edges. Moreover, these methods, except AnoMAN (Chen et al., 2022), are based on pre-defined patterns/roles or handcrafted features, while real-world network anomalies have complex nature. Therefore, these models cannot be generalized to different domains, limiting their application.

The only exception and also the closest method to our approach is AnoMulY (Behrouz and Seltzer, 2022), a Gnn-based anomaly detection method in multiplex networks. However, this method suffers from four main limitations: ① Transductive learning: The AnoMulY framework is designed in a transductive setting and cannot be applied to unseen nodes/patterns. In contrast, ADMire anonymizes nodes in such a way to work in the inductive setting. ② Memory and scalability: The AnoMulY framework is snapshot-based. That is, it requires storing the entire snapshot of the temporal network at each timestamp, which consumes a great deal of memory. Moreover, since it uses different Gnn modules for each type of connection, it cannot be utilized for multiplex brain networks with a large number of views (e.g., in datasets with a large number of participants). However, ADMire is a streaming method, requiring only constant memory (see Appendix C). Moreover, our random walk encoder scales to brain multiplex networks with more than 100 views. ③ Lack of generalizability: The AnoMulY framework uses a simple negative sampling method by randomly changing one endpoint of a connection to learn anomalous interactions. While this negative sampling method is fast and lets the model be trained in

Algorithm 1: Temporal multiplex walk sampling procedure

input : The edge set $\mathcal{E}$, previously sampled node w_p in view r_p at time t, and hyperparameter μ
output : Next sampled connection (w_n, w_p, r_n, t_n)
for $e = (w_n, w_p, r_n, t_n) \in \mathcal{E}^{t_p}(w_p)$ **do**
  Sample $b \sim$ Uniform$(0, 1)$;
  if $b < q_{r_p}^{t_p}(w_p, e) \times \varphi(r_p, r_n)$ **then**
    return $e = (w_n, w_p, r_n, t_n)$;
  end
end
return EOA;

an unsupervised manner, these negative sample generator methods are too simple and can cause poor performance in more complicated datasets (Poursafaei et al., 2022). ADMire introduces a novel negative sampling method for multiplex networks and shows its efficacy in § 5.1. ④ Many hyperparameters: The AnoMulY framework has many hyperparameters that require tuning before the model achieves good performance. However, tuning these hyperparameters is difficult in real-world datasets, limiting its applications. In ADMire, there are only four hyperparameters, which can simply be tuned based on the dataset properties.

Appendix C. Efficient Sampling

The first step in our sampling is to compute the sampling probability of an incoming connection in relation type r. For an incoming edge $e = (u, v, r, t)$ we compute the probabilities $q_r^t(w)$ for $w \in \{u, v\}$ as follows:

$$q_r^t(w, e) = \frac{\exp(\mu t)}{\sum_{(w_0, t') \in \mathcal{N}_r^t(w)} \exp(\mu t')}, \tag{17}$$

where $\mathcal{N}_r^t(w)$ represents the set of w's neighbor in view r and before time t. This probability needs to be computed one time when arrives and does not need to be updated anymore. Also, for calculating the probability of sampling this connection after a connection from another relation type r', we simply multiply this probability by $\varphi(r, r')$.

Algorithm 1 shows the sampling procedure. Given a previously sampled connection in view r at time t, we sample the next connection in view r' at tme $t' < t$ with a probability proportional to $\exp(\mu(t' - t)) \times \varphi(r, r')$. It is not hard to show that Algorithm 1 sample the next connection with a probability proportional to $\exp(\mu(t_n - t_p)) \times \varphi(r_p, r_n)$. Inspired by Wang et al. (2021), in our experiments, we store most k recent connections with $k \propto \mathcal{O}\left(\frac{1}{\mu}\right)$. The intuition is that if we sort connections in $\mathcal{E}^t(w_p)$ by their timestamp $\{t_i\}_{i=1}^h$, and assume that $\exp(\mu(t_i - t))$ are i.i.d., the probability of sampling j-th connection is:

$$\mathbb{P}[\text{sampling } j\text{-th connection}] = \frac{\prod_{i=1}^{j} \exp(\mu(t_i - t)) \times \varphi(r_p, r_i)}{\sum_{i=1}^{h} \prod_{s=1}^{i} \exp(\mu(t_s - t)) \times \varphi(r_p, r_s)}.$$

It is not hard to see that this probability is very small when we increase the value of j. Accordingly, in practice, we only need to store a constant number of the most recent connections at each time.

Appendix D. Attention Mechanism: Motivation

As we discussed, in multiplex networks the importance of views might be different. For example, one disease might be more correlated with functional connectivity than structural connectivity, or a brain network of an individual can be noisy and we need to automatically ignore it in the training process. To this end, we design an attention mechanism that learns the importance of each view for other views. Existing attention mechanisms (Behrouz and Seltzer, 2022; Park et al., 2020) are designed for *general* multiplex networks, assuming that the importance of each view for each node is different. Although these mechanisms are more general, they are required to learn many parameters, limiting their scalability to large networks with a large number of views. Here, in our experimental evaluations on multiplex *brain* networks, we observe that the importance of one view for different nodes is almost the same. Given a view r and a node u, we use $\Omega(u, r)$ to show the importance of view r for node u. We use the attention mechanism proposed by Behrouz and Seltzer (2022) instead of our attention mechanism and train the model on PD, ADHD, and ASD datasets. While it requires $\approx 1.8\times$ training time, we observe that $\Omega(u, r) \approx \Omega(v, r)$ for any given view r and arbitrary nodes u and v. That is, given a view r, the maximum variance of $\Omega(u, r)$ for different nodes u is 0.02, 0.05, and 0.02 in PD, ADHD, and ASD datasets, respectively. Therefore, since in a multiplex brain network we might have a large number of views (e.g., a large number of subjects, a large number of image modalities, or a large number of frequency bands), we design a more efficient and scalable attention mechanism that learns the importance of each view for other views (independent of nodes). One can interpret this attention mechanism as a model that learns the correlation between each pair of views.

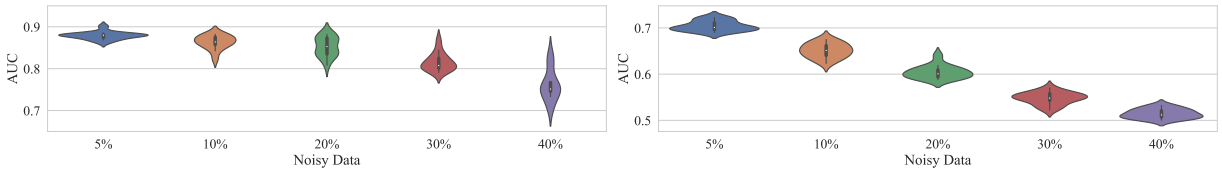
Appendix E. Experimental Setting Details

We tune hyper-parameters by cross-validation, and search the hyper-parameters over ① $\mu \in \{0.5, 1, 2, 4\} \times 10^{-5}$, ② Inter-view sampling number $M \in \{32, 64, 128, 256\}$, ③ Intra-view sampling number per view $M' \in \{8, 16, 32, 64\}$, ④ Walk length $m \in \{2, 4, 8, 12\}$. Also, in the training, we use a learning rate of 0.0001, hidden dimension 100 in MLP-Mixer, and batch size of 600.

To visualize the average distribution of anomalous connections, we use BrainPainter (Marinescu et al., 2019) with the Desikan-Killiany atlas.

Appendix F. Baselines

Since AnoMulY (Behrouz and Seltzer, 2022) is the only competitor method on edge anomaly detection in multiplex networks, we also compare ADMire with single-layer edge anomaly detection methods: GOutlier (Aggarwal et al., 2011) builds a generative model for edges in a node cluster. NetWalk (Yu et al., 2018b) uses a random walk to learn a



(*a*) Performance of ADMire (*b*) Performance of monoplex ADMire

Figure 4: The advantage of multiplex brain networks over monoplex brain networks.

unified embedding for each node and then dynamically clusters the nodes' embeddings. AddGraph (Zheng et al., 2019) is an end-to-end approach that uses an extended GCN in temporal networks. Finally, we compare with two multiplex network embedding baselines, ML-GCN (Behrouz and Hashemi, 2022) and MNE (Zhang et al., 2018b). We apply K-means clustering on their obtained node embeddings for anomaly detection (Yu et al., 2018b).

In the inductive setting, since there is no inductive learning (or anomaly detection) method on multiplex networks that we are aware of, we compare ADMire with inductive learning methods on monoplex networks. CAW-N (Wang et al., 2021) is an inductive method that uses causal anonymous walks to extract network motifs and a novel set-based anonymization process that keep model inductive by hiding the identity of nodes during the training phase. EvolveGCN (Pareja et al., 2020) uses a RNN to estimate the GCN parameters for the future snapshots. TGAT (da Xu et al., 2020) uses GAT (Veličković et al., 2018) to extract node representations where the nodes' neighbors are sampled from the history and then encodes temporal information via random Fourier features.

Appendix G. Additional Experimental Results

G.1. Noisy Brain Images

As we discussed in § 3.2, one of the main motivations for modeling neuroimaging datasets as multiplex networks is to make the model more robust against noise in each brain image. To validate it, in this experiment, we add Gaussian noise to a subset of brain images ($5\%, 10\%, 20\%, 30\%$ and 40%) in the ADHD dataset. We model the noisy dataset as a multiplex brain network and use it to train ADMire. Next, as a baseline, following previous methods (Lanciano et al., 2020; Zhang et al., 2020), we take the average of all brain images in the noisy dataset and use it to train the monoplex ADMire. Figure 4 reports the performance of ADMire and monoplex ADMire with varying the size of noisy samples. Not only ADMire achieves superior performance with a significant margin, but it also shows to be more robust against noise than the monoplex ADMire. This experiment shows the importance of multiplex modeling and also the effectiveness of the proposed attention mechanism that can learn to ignore noisy samples.

G.2. Additional Results on Real-world Datasets

In this section, we present additional visualization of results provided in § 5.1. Figure 5, Figure 6, and Figure 7 present the average distribution of anomalous edges in PD, ADHD, and ASD groups.

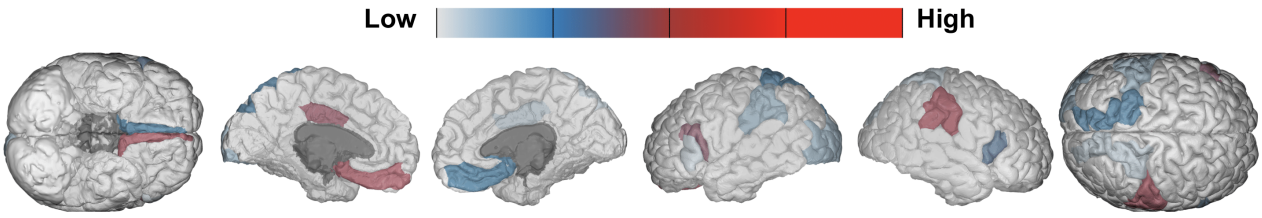


Figure 5: The distribution of anomalous edges in PD group.

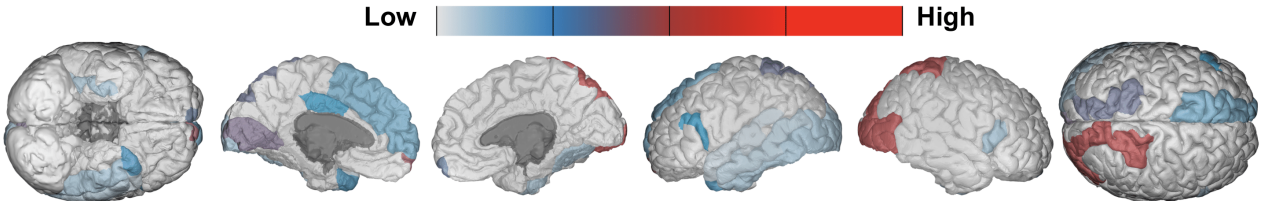


Figure 6: The distribution of anomalous edges in ADHD group.

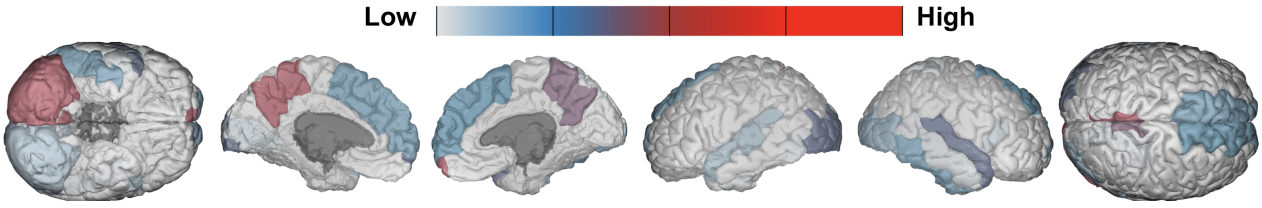


Figure 7: The distribution of anomalous edges in ASD group.