



Fusing modalities by multiplexed graph neural networks for outcome prediction from medical data and beyond

Niharika S. D'Souza^{a,*}, Hongzhi Wang^a, Andrea Giovannini^b, Antonio Foncubierto-Rodriguez^b, Kristen L. Beck^a, Orest Boyko^c, Tanveer F. Syeda-Mahmood^a

^a IBM Research Almaden, San Jose, CA, USA

^b IBM Research Zurich, Switzerland

^c Department of Radiology, VA Southern Nevada Healthcare System, NV, USA

ARTICLE INFO

Keywords:

Multimodal fusion
Graph neural networks
Multiplex graphs

ABSTRACT

With the emergence of multimodal electronic health records, the evidence for diseases, events, or findings may be present across multiple modalities ranging from clinical to imaging and genomic data. Developing effective patient-tailored therapeutic guidance and outcome prediction will require fusing evidence across these modalities. Developing general-purpose frameworks capable of modeling fine-grained and multi-faceted complex interactions, both within and across modalities is an important open problem in multimodal fusion. Generalized multimodal fusion is extremely challenging as evidence for outcomes may not be uniform across all modalities, not all modality features may be relevant, or not all modalities may be present for all patients, due to which simple methods of early, late, or intermediate fusion may be inadequate. In this paper, we present a novel approach that uses the machinery of multiplexed graphs for fusion. This allows for modalities to be represented through their targeted encodings. We model their relationship between explicitly via multiplexed graphs derived from salient features in a combined latent space. We then derive a new graph neural network for multiplex graphs for task-informed reasoning. We compare our framework against several state-of-the-art approaches for multi-graph reasoning and multimodal fusion. As a sanity check on the neural network design, we evaluate the multiplexed GNN on two popular benchmark datasets, namely the AIFB and the MUTAG dataset against several state-of-the-art multi-relational GNNs for reasoning. Second, we evaluate our multiplexed framework against several state-of-the-art multimodal fusion frameworks on two large clinical datasets for two separate applications. The first is the NIH-TB portals dataset for treatment outcome prediction in Tuberculosis, and the second is the ABIDE dataset for Autism Spectrum Disorder classification. Through rigorous experimental evaluation, we demonstrate that the multiplexed GNN provides robust performance improvements in all of these diverse applications.

1. Introduction

The past decade has seen tremendous advances in targeted measurements being generated through multiple data modalities. In turn, this has provided novel perspectives and improved the understanding for many of the world's most challenging problems. The curation of large multimodal electronic health records has made it possible to capture information about a patient through multiple data-rich modalities. In turn, this allows us to obtain a holistic view of a patient's condition. For example, in complex diseases such as cancer (Subramanian et al., 2020), tuberculosis (Muñoz-Sellart et al., 2010) or autism spectrum disorder (Dsouza et al., 2021; D'Souza et al., 2021b; Li et al., 2020), evidence for a diagnosis or treatment outcome may be present in multiple modalities such as clinical, genomic, molecular, pathological

and radiological imaging. Studies have revealed that predicting disease specific targets, eg, diagnoses/treatment outcomes, is often a function of multiple patient-specific factors. For these applications, systematic collection and careful processing of multimodal data covering diverse, and often unstructured information about the patient has become essential (Muñoz-Sellart et al., 2010). At the same time, it is unclear what information is best captured in each modality and how best to combine them.

In such cases, multimodal fusion is needed because evidence for an entity, such as an event, or a disease may be present in more than one modality and no single modality is sufficient to reach strong conclusions. At the same time, determining how best to fuse the information is

* Corresponding author.

E-mail address: Niharika.Dsouza@ibm.com (N.S. D'Souza).

challenging since some modalities capture complementary, while others indicate contradictory information. In fact, modality features may be mutually exclusive, mutually reinforcing, or mutually correlated. A single modality may even be confirmatory in some cases, making other modalities redundant. Finally, not all modalities may be present for a sample, and those that are present could be spurious and error-prone. Despite the importance of the problem, there is currently no principled domain-agnostic approach to fuse modalities that is universally applicable.

Existing attempts to fuse modalities for outcome prediction can be divided into at least three approaches, namely, feature vector-based, statistical or graph-based techniques. Vector-based approaches perform early, intermediate, or late fusion (Subramanian et al., 2020; Baltrušaitis et al., 2018; Wang et al., 2021). In early fusion, features from different modalities are concatenated and then fed into a predictive model. Intermediate fusion methods compute modality-specific projections of features, that are folded into predictive models. Finally, late fusion aggregates the results of modality-specific predictors post-hoc rather than fusing the modality features themselves. Unfortunately, due to the underlying assumptions made by these methods, they may be inadequate for characterizing the broader range of possible relationships among modality features and their relevance for prediction.

Statistical approaches such as Canonical Correlation Analysis (CCA) are designed to identify highly-correlated features across and within modalities as modeled by a set of canonical variates of the data. Thus, CCA and its deep variants (Yang et al., 2019) can model feature dependence, either in the native representation or in a latent space (Subramanian et al., 2020). In the past, these representations have been useful for studying a variety of problems such as the identification of genotype-phenotype associations (Subramanian et al., 2018) in cancer, or for survival analysis in breast cancer (Subramanian et al., 2021) or recurrence of lung cancer (Subramanian et al., 2020). Nevertheless, these methods have limitations to the extent to which they can model possible relationships between modalities and are not guaranteed to learn discriminative patterns in their projections (Wang et al., 2019) natively.

Graph Neural Networks (GNN) are deep learning architectures designed for data structured as a graph. They distill complex connectivity information (i.e. dependence across data-streams) to guide a downstream inference task (Scarselli et al., 2008). For fusion, graph-based approaches have garnered significant interest for various clinical applications such as brain connectomics (D'Souza et al., 2021; D'Souza et al., 2020; Sebenius et al., 2021; D'Souza et al., 2021a). Here, the graph connectivity naturally arises as a consequence of the data geometry and acquisition (D'Souza and Venkataraman, 2023; D'Souza et al., 2019; Nandakumar et al., 2018, 2020). Alternatively, in the case of applications such as Cosmo et al. (2020), Zheng et al. (2022) involving unstructured imaging and non-imaging data, the fusion problem is less straightforward and the graph needs to be defined implicitly from the data. Thus, a latent graph learning from modality features is performed, either by concatenation (Cosmo et al., 2020) or by learnable weighted averaging (Zheng et al., 2022). Unfortunately, such collapsed representations (i.e. allowing only one type of edge connectivity/feature-relation) may be too constraining, and may miss subtleties of cross-modality and intra-modality dependence. To address this, the approach of Hou et al. (2023) keeps the modality identities separate through individualized graph convolutional networks, which are then mixed together using hyper-graph convolutions and hyper-edge cross-modal mixing. However, the modality and feature semantics may not be preserved through these mixed transformations. An alternative, yet close parallel to graph based approaches is the recent development of transformer representations (Parmar et al., 2018) which construct of a weighted fully connected graph between multimodal features/feature vectors. Here, all possible edge-weights inferred in a data-driven fashion via the attentional mechanism (Hou et al., 2022). While widely successful, especially in the image and text domain, such approaches often suffer from overfitting in applications where limited training data is available due to insufficient regularization.

1.1. Our contribution

In contrast, our multiplexed framework allows for the modalities to retain their individuality while still participating in exploring explicit relationships between the modality features. Specifically, our multiplexed graph explicitly models relationships within and across modality features via a two stage approach. First, we use an unsupervised learning (autoencoders) on unstructured raw modality features to convert them into a multiplex graph. This graph models feature interactions across and within modalities simultaneously. We then design, for the first time, a novel graph neural network for reasoning from this construction via structured message passing. To this end, we adopt the walk formulation native to multiplex graphs to implicitly leverage multi-hop feature dependence and relate these patterns to the outcome of interest.

As a sanity check, we evaluate our multiplexed GNN as a standalone module. We compare against existing graph neural frameworks designed for multi-relational graphs by performing experiments on two public benchmark graph datasets. Our model demonstrates robust performance improvements against state-of-the-art GNNs, suggesting enhanced representational capabilities. To demonstrate the efficacy of the multiplex formulation for multimodal fusion, we experiment on two large clinical databases with multi-omics data. We perform extensive experiments on the NIH-TB portals dataset for tuberculosis outcome prediction and the ABIDE dataset for autism spectrum disorder (ASD) classification. We demonstrate that by relaxing the fusing constraints through the multiplexed formulation, our method outperforms several state-of-the-art methods.

A preliminary version of our work appeared in D'Souza et al. (2022), which we extend in this work. In addition to more detailed presentation of the multiplexed formulation, we perform a more comprehensive evaluation to demonstrate the generality of the approach. Firstly, we perform a dedicated set of comparisons on the inference model (multiplexed GNN) on two separate benchmark graph datasets, AIFB and MUTAG. Here, we compare the multiplexed GNN against several state-of-the-art approaches for multi-relational reasoning. Second, we extend the comparisons presented on the outcome prediction for Tuberculosis in D'Souza et al. (2022) to include comparisons against recent fusion frameworks including representation learning based, graph based, and transformer frameworks. Given the challenge with using modest sized datasets for evaluation, we examine different dataset splits than those in our conference submission. We observe consistent improvements in terms of performance comparison against baselines, indicating the robustness of our approach. Third, we evaluate our fusion framework on a second clinical application, i.e. case/control discrimination on the ABIDE dataset, which is a large Autism study. Finally, we also perform a qualitative evaluation of the patterns mined by representation learning step for both datasets and examine the differential contribution of individual modalities for the prediction.

Taken together, these experiments highlight that the multiplexed formulation is a promising first step in the development of flexible yet powerful data-driven solutions for general multimodal fusion problems.

2. Materials and methods

To address the fusion problem, we would like to explicitly model various facets of cross-modal interactions. To this end, we propose to utilize the representation learning theory of multiplexed graphs and neural networks designed on such graphs.

2.1. Related works

In this section, we provide a short introduction to graph neural networks and multiplex graphs as explored in literature.

Graph Neural Networks: Graph Neural Networks (GNN) are deep learning frameworks designed for graphs that can distill connectivity

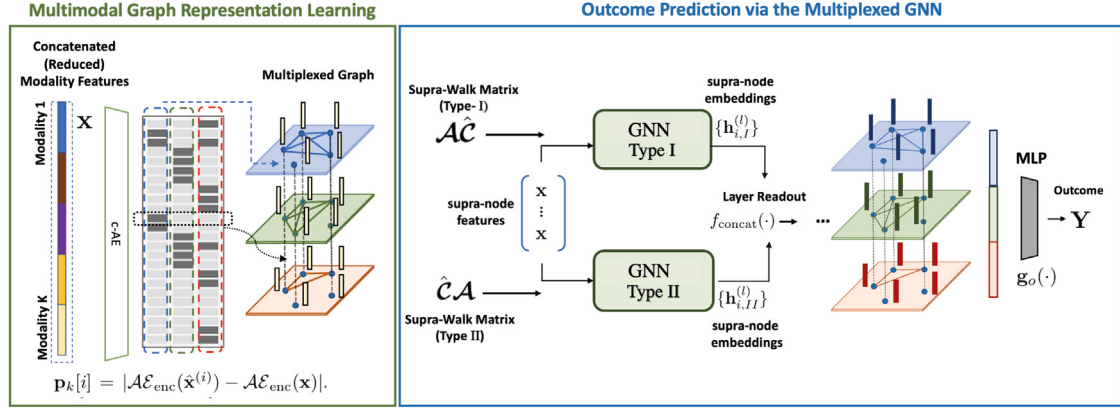


Fig. 1. Multiplexed Graph Based Fusion for Outcome Prediction. **Green Box:** A feature vector is formed by concatenating individual modality features. (The modality features may first be reduced through domain specific autoencoders (d-AE) for dimensionality reduction). The concatenated feature vector is projected into a common latent space using a concept auto-encoder (c-AE). Salient activations in the latent space are used to form the planes of the multiplexed graph. **Blue Box:** The multiplexed GNN uses message passing walks to combine latent concepts for inference. At the input layer, the node embeddings are copied over from the concatenated features $\mathbf{x}$ across the planes. During the message passing, the filtered embeddings maintain their plane-specific and node-specific identity. The output MLP aggregates these to map to the outcome $\mathbf{Y}$. We construct one multiplex graph per subject.

and attribute information to guide an inference task (Scarselli et al., 2008). Intuitively, they generalize the notion of convolutional operations in convolutional neural networks (CNN) (LeCun et al., 1995) designed on images to graph structured data. Due to their substantial expressive power and versatility, GNNs have become a popular tool across diverse applications in natural, physical and social sciences (Zhou et al., 2020)

Formally, a graph can be defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a vertex set $\mathcal{V}$ with number of nodes as $|\mathcal{V}| = P$. The set $\mathcal{E} = \{(i, j)\} \in \mathcal{V} \times \mathcal{V}$ denotes the edges linking pairs of nodes i and j . These relationships are captured as an adjacency matrix $\mathbf{A} \in \mathbb{R}^{P \times P}$. In the simplest case, the elements of this matrix are binary $\mathbf{A}[i, j] = 1$ if $(i, j) \in \mathcal{E}$ and zero otherwise. More generally, $\mathbf{A}[i, j] \in [0, 1]$, indicating the strength of connectivity between nodes i and j . A typical GNN schema comprises of two components: (1) a message passing scheme for propagating information across the graph and (2) task-specific supervision to guide the representation learning. Conventionally, each node i of the input graph $\mathcal{G}$ has a fixed input feature descriptor $\mathbf{x}_i \in \mathbb{R}^{D \times 1}$ associated with it. The message passing scheme ascribes a set mathematical operations occurring at each layer $l \in \{1, \dots, L\}$ of the GNN. Let $\mathbf{h}_i^{(l)} \in \mathbb{R}^{D' \times 1}$ be node feature for node i at layer (l) . GNNs infer the representations at subsequent layers $(l+1)$ by aggregating representations $(\{\mathbf{h}_j^{(l)}\})$ of nodes j that is connected to i . At the input layer (l) , we have:

$$\mathbf{h}_i^{(l+1)} = \phi(\{\mathbf{h}_j^{(l)}\}, \mathbf{A}; \theta^{(l)}) \quad \text{where } j : (i, j) \in \mathcal{E} \quad (1)$$

where $\phi(\cdot) : \mathbb{R}^{D'} \rightarrow \mathbb{R}^{D^{(l+1)}}$ is an aggregation function, and $\theta^{(l)}$ denote learnable parameters for layer l . $\mathbf{h}_i^{(0)} = \mathbf{x}_i$ at the input. From here, the node embeddings are used to estimate the outputs of the GNN via a mapping f_o : $\hat{\mathbf{Y}} = f_o(\{\mathbf{h}_i^{(L)}\})$. Depending on the task, the targets $\mathbf{Y}$ provide either graph, edge, or node level supervision during training. The parameters of the GNN are then estimated by optimizing a loss function $\mathcal{L}(\mathbf{Y}, \hat{\mathbf{Y}})$ via back-propagation for gradient estimation.

Conventional GNN architectures (Veličković et al., 2017; Duvenaud et al., 2015; Defferrard et al., 2016; Duvenaud et al., 2015; Fey, 2019; Kipf and Welling, 2016; Xu et al., 2018) propose different variants of message passing schemes with varying levels of expressive power. However, by design, these frameworks can handle graphs with only a single edge type (i.e. monoplex graphs) and are not suitable for multi-relational representations, such as those seen in multiplexed graphs

Multiplexed Graphs and their Applications: A multiplexed graph (Cozzo et al., 2018) is a type of multi-graph in which the nodes are grouped into multiple planes, each representing an individual edge-type. The information captured within a plane is multiplexed to other planes through vertical connections as shown in Fig. 1.

In the real world, data-entities may interact in multiple ways, giving rise to multiple different types of relationships between them (i.e. $K > 1$). The multiplexed graph has been known in literature and used for various modeling purposes (Kivelä et al., 2014; De Domenico et al., 2014; Ferriani et al., 2013; Maggioni et al., 2013). For example, simple graph theoretic properties (Musmeci et al., 2017; Tudisco et al., 2018) or tensor decompositions have been used to study financial data and other dynamical processes (Bartasaghi et al., 2022) or for link prediction (Matsuno and Murata, 2018). Random walks on the multiplex have been used for navigation planning (Guo et al., 2016). Diffusion dynamics on the multiplex have been adopted for applications in physical science (Gomez et al., 2013). However, its potential to address the fusion problem has been relatively unexplored in the literature and is the new perspective we bring.

2.2. Multiplexed graphs for multimodal fusion

The multiplexed graph is suitable for general multimodal fusion as it allows for different types of relationships between modality features to be captured through conceptual “planes”. Mathematically, we define a multiplexed graph as: $\mathcal{G}_{\text{Mplex}} = (\mathcal{V}_{\text{Mplex}}, \mathcal{E}_{\text{Mplex}})$, where $|\mathcal{V}_{\text{Mplex}}| = |\mathcal{V}| \times K$ and $\mathcal{E}_{\text{Mplex}} = \{(i, j) \in \mathcal{V}_{\text{Mplex}} \times \mathcal{V}_{\text{Mplex}}\}$. There are K distinct types of edges which can link two given nodes. Analogous to ordinary graphs, we have k adjacency matrices $\mathbf{A}_{(k)} \in \mathbb{R}^{P \times P}$, where $P = |\mathcal{V}|$, each summarizing the connectivity information given by the edge-type k . The elements of these matrices are binary $\mathbf{A}_{(k)}[m, n] = 1$ if there is an edge of type k between nodes $m, n \in \mathcal{V}$.

The nodes $\mathcal{V}_{\text{Mplex}}$ of the multiplex graph are produced by creating copies of nodes across the planes. We henceforth refer to these as *supra-nodes*. Within each plane, we connect supra-nodes to each other via a plane-specific adjacency matrix $\mathbf{A}_{(k)}$. These intra-planar connections allow us to traverse across the multi-graph according to individual relational edge-types. The information captured within a plane is multiplexed to other planes through vertical connections, thus connecting each supra-node with its own copy in other planes. These connections allow us to traverse across the planes and exploit cross-relational dependencies in tandem with in-plane traversal.

2.3. Multimodal graph representation learning

The construction in the Green Box in Fig. 1 is used to produce the multiplexed graph from the individual modality features. First, domain specific autoencoders (d-AE, trained on an MSE reconstruction loss) may be used to convert each modality into compact feature spaces.

This effectively reduces the dimensionality of the input representation, i.e. reduces the within modality redundancy.

To capture the cross-modal feature dependence, we first concatenate the reduced modality features and bring them to a common low dimensional subspace via a concept autoencoder (c-AE). The c-AE is trained to reconstruct the concatenated features using the MSE metric. From here, each latent dimension of the autoencoder represents an abstract aspect of the multimodal fusion problem. Specifically, features projected to be salient in the same latent dimension are likely to form meaningful joint patterns in relation to a specific task. The salient features form a “conceptual” plane of the multiplexed graph. The $|\mathcal{V}'_{\text{Mplex}}|$ “supra-nodes” of the multiplexed graph are produced by creating copies of features (i.e. nodes) across the planes. The edges between nodes in each plane represent features whose projections in the respective latent dimensions were salient (we provide implementation details for specific datasets in Section 3). Thus, each plane is endowed with its own topology. This acts as a proxy for the correlation between features with respect to the corresponding latent dimension. The connectivity in each plane is denoted by individual planar adjacency matrices.

Let the concatenated reduced modality features be denoted by the vector $\mathbf{x} \in \mathcal{R}^{P \times 1}$. To form the multiplexed graph planes, the c-AE projects $\mathbf{x}$ to a ‘conceptual’ latent space of dimension where $K \ll P$. We use the projections in the latent space to explore the dependence between the features. We infer within plane connectivity along each latent dimension by perturbing the features and recording those pairs of features giving rise to the largest incremental responses. Let $\mathcal{AE}_{\text{enc}}(\cdot) : \mathcal{R}^P \rightarrow \mathcal{R}^K$ be the c-AE mapping to the concept space. Let $\hat{\mathbf{x}}^{(i)}$ denote the perturbation of the input by setting $\hat{\mathbf{x}}^{(i)}[j] = \mathbf{x}[j] \forall j \neq i$ and 0 for $j = i$. Then for concept axis k , the perturbations are $\mathbf{p}_k[i] = |\mathcal{AE}_{\text{enc}}(\hat{\mathbf{x}}^{(i)}) - \mathcal{AE}_{\text{enc}}(\mathbf{x})|$. Thresholding $\mathbf{p}_k \in \mathcal{R}^{P \times 1}$ selects feature nodes with the strongest responses along concept k . To encourage sparsity, we retain the top one percent of salient patterns. We connect all pairs of such feature nodes with edge-type k via a fully connected (complete) subgraph between nodes thus selected (Fig. 1). Across the K concepts, we expect that different sets of features are prominent. The input features $\mathbf{x}_i$ are one dimensional node embeddings for node i .

Overall, the graph construction procedure implicitly models the interactions between the various modality features in a principled fashion. Accordingly, we connect supra-nodes within a plane to each other via the intra-planar adjacency matrix $\mathbf{A}_{(k)}$. These matrices allow us to traverse the multi-graph according to the edge-type k . Each supra-node in a given plane i is also connected with its own copy along other planes j via vertical connections. This connectivity is denoted by pairwise inter-planar adjacency matrices $\hat{\mathbf{C}}_{(i,j)}$, allowing for cross-planar traversal.

2.4. Multiplexed GNN for outcome prediction

We develop a novel graph neural network for outcome prediction from the multiplexed graph (Blue Box in Fig. 1). Our main motivation is to utilize the native formulation of the multiplex in the design of layerwise GNN reasoning steps.

A typical Graph Neural Network schema consists of two key components: (1) a message passing scheme for propagating information across the graph and (2) task-specific supervision. Recall that for GNNs designed for ordinary graphs (Kipf and Welling, 2016), the adjacency matrix $\mathbf{A}$ and its matrix powers allows us to keep track of graph neighborhoods (at arbitrary l hop distance) during message passing. In turn, cascading l GNN layers is equivalent to pooling information at each node i from its l -hop neighbors ($\mathcal{N}_i^{(l)}$) that can be reached by a walk starting at i . Using this guiding principle, we design, for the first time, a Multiplexed GNN that can emulate this behavior within its representation learning.

We first define two key quantities associated with the multiplex graph $\mathcal{G}_{\text{Mplex}}$ to design our message passing for inference. The *intra-planar adjacency matrix* $\mathbf{A} \in \mathcal{R}^{PK \times PK}$, and the *inter-planar transition control matrix* $\hat{\mathbf{C}} \in \mathcal{R}^{PK \times PK}$ (Cozzo et al., 2018) are denoted by:

$$\mathbf{A} = \bigoplus_k \mathbf{A}_{(k)} \quad ; \quad \hat{\mathbf{C}} = [\mathbf{1}_K \mathbf{1}_K^T] \otimes \mathbf{I}_P \quad (2)$$

where $\bigoplus$ denotes the Kronecker direct sum operation, $\otimes$ is the Kronecker product, $\mathbf{1}_K$ is the K vector of all ones, and $\mathbf{I}_P$ is the identity matrix of size $P \times P$.

By construction, $\mathbf{A}$ is block-diagonal and captures transitions across supra-nodes within each plane. Analogously, $\hat{\mathbf{C}}$ has identity matrices along off-diagonal blocks, implicitly restricting cross-planar transitions to be between supra-nodes corresponding to the same node (i.e. $s_i \leftrightarrow i \implies s_i = P(k-1) + i$ for some $k \in \{1, \dots, K\}$).

Note that supra-nodes across planes can be reached by using a combination of within and across-planar transitions. This provides comparable representational properties at a reduced complexity of $\mathcal{O}(PK)$ inter-planar edges instead of $\mathcal{O}(P^2K)$. Together, $\mathbf{A}$ and $\hat{\mathbf{C}}$ allow us to define multi-hop transitions on the multiplexed graph in a convenient factorized form.

2.4.1. Multiplex walks for message passing

A walk on the supra-nodes of the multiplexed graph is defined according to the following transition rules (Cozzo et al., 2018). A walk on $\mathcal{G}_{\text{Mplex}}$ combines within and across planar transitions to reach a supra-node $j \in \mathcal{V}'_{\text{Mplex}}$ from a given supra-node $i \in \mathcal{V}'_{\text{Mplex}}$. This gives rise to two types of steps: (1) A single intra-planar step or (2) A step that includes both an intra-planar step and an inter-planar step moving from one plane to another (this *can* be before or after the occurrence of an intra-planar step).

To recreate these transitions exhaustively, we have two supra-walk matrices. Specifically, $\mathbf{A}\hat{\mathbf{C}}$ encodes transitions where *after* an intra-planar step, the walk is *allowed* to continue in the same plane or transition to a different plane (Type I). Similarly, via $\hat{\mathbf{C}}\mathbf{A}$, the multiplexed walk *may* continue in the same plane or transition to a different plane *before* an intra-planar step (Type II). Starting at a given node i in a plane k_1 , Type I steps ($\mathbf{A}\hat{\mathbf{C}}$) indicate that the walk first transitions to a node j in the same plane k_1 through the intra-planar adjacency $\mathbf{A}$ followed by the transition control matrix $\hat{\mathbf{C}}$, allowing for transitions to the node j but in a different plane k_2 . Analogously, applying a Type II ($\hat{\mathbf{C}}\mathbf{A}$) step starting at a given node i in a plane k_1 may first transition across planes (by applying $\hat{\mathbf{C}}$, staying at node i but its copy in plane k_2) and then move to a different node j in plane k_2 (i.e. applying $\mathbf{A}$).

To provide an intuition for this mechanism, we take the toy multiplex graph in Fig. 2 with two planes and three nodes. For a walk starting at node i in plane k_1 , we show examples of nodes that can be reached using a multiplex walk with relevant portions of the factorization highlighted. Notice that the factorized forms defining the walk, i.e. $\mathbf{A}\hat{\mathbf{C}}$ and $\hat{\mathbf{C}}\mathbf{A}$ allow starting from a given supra-node to either reach a different node in the same plane (Fig. 2(a)), transition across planes but continue to remain at the corresponding node (Fig. 2(c)), and transition across planes and reach a different node in the new plane (Fig. 2(b) or (d)). Note that keeping the two sets of steps (Type I and II) distinct allows us to distinguish between cases (b) and (d) while accounting for both in the design of the GNN message passing described below.

Let $\mathbf{h}_{s_i}^l \in \mathcal{R}^{D \times 1}$ denote the (supra)-node representation for (supra)-node s_i associated with node i . We compute the filtered representation at layer $(l+1)$ this via the following operations:

$$\mathbf{h}_{s_i, I}^{(l+1)} = \phi_I(\{\mathbf{h}_{s_j}^{(l)}, s_j : [\mathbf{A}\hat{\mathbf{C}}][s_i, s_j] = 1\}) \quad (3)$$

$$\mathbf{h}_{s_i, II}^{(l+1)} = \phi_{II}(\{\mathbf{h}_{s_j}^{(l)}, s_j : [\hat{\mathbf{C}}\mathbf{A}][s_i, s_j] = 1\}) \quad (4)$$

$$\mathbf{h}_{s_i}^{(l+1)} = f_{\text{concat}}(\mathbf{h}_{s_i, I}^{(l+1)}, \mathbf{h}_{s_i, II}^{(l+1)}) \quad ; \quad \mathbf{g}_o(\{\mathbf{h}_{s_i}^{(L)}\}) = \hat{\mathbf{y}} \quad (5)$$

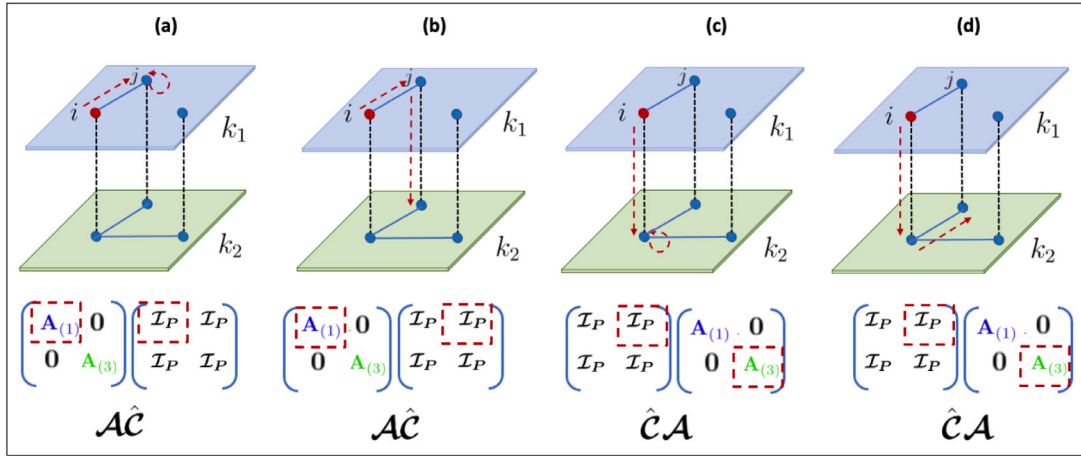


Fig. 2. Example of possible transitions within the multiplex graph with node (i) in plane (k_1) is the starting point (red node). Sub-figures ($a - d$) correspond to different types of end-points that can be reached using the portion of the factorization highlighted in red below the corresponding examples. Taken together, the factorized supra-matrix forms allow coverage across all nodes of the multiplex across planes, while maintaining the distinctiveness of the order in which within- and across-planar transitions are enumerated.

Here, $f_{\text{concat}}(\cdot)$ concatenates the Type I and Type II representations. At the input layer, we have $\mathbf{d}^{(0)} = \mathbf{x} \otimes \mathbf{1}_K$, where $\mathbf{x} \in \mathcal{R}^{|V| \times 1}$ are the node inputs (concatenated modality features). $\phi_{I/II}$ are graph neural network transformations, such as a Graph Isomorphism Network (Xu et al., 2018), Graph Convolutional Network (Kipf and Welling, 2016) layers etc. Eqs. (3)–(5) performs message passing according to the neighborhood relationships given by the supra-walk matrices. For each subject, we can thus construct one multiplexed graph with $\mathbf{Y}$ being the vector patient-specific prediction target. Thus, $\mathbf{g}_o(\cdot)$ is a graph readout network that maps to the one-hot encoded outcome. It is a two-layered fully connected network with Leaky ReLU activation. Finally, the learnable parameters of the Multiplex GNN can be estimated via standard backpropagation.

2.4.2. Implementation details

We chose the model architecture and hyperparameters (learning rate, weight decay, number of epochs, number of planes K etc.) for our framework and the comparison baselines using grid-search and performance on the validation set as a guide for generalization. We utilize the ADAMw optimizer (Loshchilov and Hutter, 2017) due to its empirical stability and robustness. All models were implemented using the Deep Graph Library (v=0.6.2) in PyTorch (v=0.10.1) and trained on an 2.3 GHz 8-Core Intel i9 machine with 64 GB RAM.

As mentioned earlier, we perform two sets of evaluations. The first experiment is designed to validate the Multiplex GNN message passing alone (i.e. semi-supervised node classification on benchmark graph datasets). The second set of experiments form the main clinical evaluation on the multimodal fusion problem. We utilize Graph Convolutional layers (GCN) (Kipf and Welling, 2016) for message passing (i.e. $\{\phi_I(\cdot), \phi_{II}(\cdot)\}$ from Eqs. (3)–(5) for the semi-supervised node classification task and Graph Isomorphism Network (GIN) layers (Xu et al., 2018) for the multimodal fusion task, in each case with $L = 2$ layers. Our non-linearity of choice was LeakyReLU (negative slope= 0.01) due to its empirical robustness during training. In the semi-supervised setting, the node based readout $\mathbf{g}_o(\cdot)$ computes a linear combination of the supra-embeddings, followed by a softmax to estimate the (node-specific) logits at the output. An analogous procedure was performed at the readout layers for the baseline GNNs. For the multimodal fusion task, we perform graph based classification task with one graph constructed per subject. Therefore, for all methods, we attach a graph readout function $\mathbf{g}_o(\cdot)$ after the GNN layers. For each case, $\mathbf{g}_o(\cdot)$ this was a two-layered MLP (with hidden layer widths as $\{100, 20\}$), with LeakyReLU (negative slope= 0.01) activation. Applying a softmax to the graph outputs gives the logits for classification. We used the Cross Entropy loss for training all models in every experiment.

Software platform- multimodal fusion toolkit (MMMT). We introduce the multimodal-model-toolkit (MMMT, pronounced mammut): a platform for accelerating research and development designed to handle multimodal data. The code is available in open source (Multimodal-Models-Toolkit). This platform automates the entire analysis pipeline for multimodal data. MMMT has been designed to have a modular structure and integrates the following main steps: (1) Data Loading (2) Data Representation: Unimodal/Multimodal (3) Training (4) Inference (5) Evaluation. Along with the modular structure, MMMT enables users to launch multiple model computations by configuring a readable YAML file and by calling a single starting script. The multiplexed framework and several baselines models are available as a part of this current release. We envision that this effort would accelerate the rapid testing and benchmarking of multimodal modeling frameworks in the future.

3. Experiments and results

We divide the evaluation of our contributions into two parts. In the first section, we are interested in examining the efficacy of the Multiplexed GNN as a tool for reasoning from general multi-graphs. Here, the primary focus is on *verifying the validity of the GNN message passing design* in relation to state-of-the-art graph neural networks designed to operate on multi-graphs. We accordingly run experiments on two benchmark graph datasets, where the multi-graph is predefined.

On the other hand, the main focus of the second set of experiments is to evaluate the utility of the Multiplexed formulation (Multiplex Graph Representation Learning + Inference) for *general purpose multimodal fusion of medical data*. In these experiments, we compare the pipeline in Fig. 1 against existing models developed to address the fusion problem.

Evaluation metrics. Most prediction tasks considered in this section have a multi-class classification setting. We evaluate the performance using AU-ROC (Area Under the Receiver Operating Curve), which measures the one-vs-all tradeoff between the True Positive Rate (TPR) and False Positive Rate (FPR) considering the class of interest as the conventional “positive class”. We also report weighted average AU-ROC as an overall summary statistic. We also report standard deviation values on the AU-ROCs. Note that for comparisons where only two target classes are present, we report only one combined AU-ROC measure of performance.

For all experiments and models, we rely on 10 randomly generated train/test/validation splits for each model. We evaluate statistical differences between the baselines and our method via the DeLong test (DeLong et al., 1988) computed on the distribution of the class-wise AU-ROCs.

3.1. Evaluating the multiplex GNN design

As a sanity check on the design of our novel GNN framework, we would like to verify its applicability for reasoning from pre-constructed multi-graphs. To this end, we use two public benchmark datasets, commonly adopted in literature.

3.1.1. Datasets

To test the ability of the Multiplexed GNN for reasoning from multi-graphs, we evaluate on two benchmark datasets for semi-supervised node classification, namely, the AIFB and the MUTAG datasets. Note that the classification problem being studied here is performed in a semi-supervised transductive setting, with supervision provided by the labeled training examples.

AIFB Dataset: The AIFB dataset described the staff, research group, and publications of the AIFB research institute at the University of Karlsruhe (Bloehdorn and Sure, 2007). It contains 8285 entities, nearly 29k edges, and 45 different relationships or edge types. We define a multiplex graph on these entities. The dataset contains labels for 176 members of a research group with 4 different research subgroups. The goal is to predict the affiliation of a researcher to one of the subgroups. Since we have only 176 labeled examples, we utilized 122/18/36 (Train/Validation/Test) splits during evaluation. We instantiated 8 dimensional learnable features for each node which are updated within the overall optimization scheme.

MUTAG dataset: MUTAG is a graph dataset introduced in Debnath et al. (1991) in which the goal is to predict the mutagenicity of a collection of nitroaromatic compounds. Thus there are two classes, is-mutagenic or not. In the node classification setting, there are 27163 nodes with 23 relationship types. The dataset has 340 labeled nodes, with each node belonging to one of two classes. We utilize ten random 238/34/68 Train/Val/Test splits. We instantiated 8 dimensional learnable features for each node which are updated within the overall optimization scheme.

3.1.2. Baseline comparisons

We compare against several state-of-the-art deep networks designed to operate on multiplexed graphs.

Relational GCN: The relational GCN (Schlichtkrull et al., 2018) is a GNN designed to handle multi-relational edges between a given set of nodes. Essentially, a separate GCN layer is used for each relation-type (i.e. plane). The distinct embeddings thus obtained are aggregated post-hoc at each layer by averaging them. Note that this implicitly assumes independence across different edge-types, and may ignore vital information arising from the complementarity in multi-edge relationships.

Multi-Dimensional GCN: The multi-dimensional GCN (Ma et al., 2019) extends the R-GCN by adding a learnable cross-modal graph attention to the node embeddings. Nevertheless, this framework does not directly model higher order paths in the multi-graph arising from mixing and matching across and within relational neighborhoods.

Multi-Layered GNN: The multi-layered GNN (Grassia et al., 2021) is a recent framework designed to model the multi-planar graph in Fig. 1. It separates the inter $\mathcal{A}$ and intra-planar $\hat{\mathcal{C}}$ adjacency matrices into individualized message passing layers and concatenates the embeddings after. Both branches use Graph Attention Networks (Veličković et al., 2017). Note that this procedure introduces an artificial separation between the cross-planar and intra-planar edges at each operation.

Multi-Behavioral GNN: This framework was designed for prediction from multiplexed graphs in recommendation systems arising from various user-item behavioral preferences. It first collapses the multiplexed graph into a single “quotient” graph i.e. average of the intra-planar adjacency matrix across planes. Node embeddings are pre-filtered using

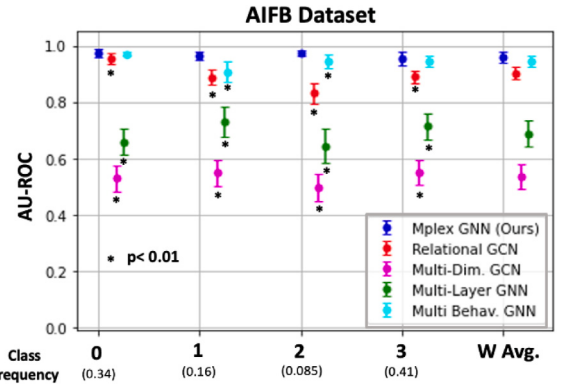


Fig. 3. Performance the on the semi-supervised node classification on the AIFB dataset as measured by per-class and weighted average AU-ROC. We display mean performance along with standard errors. Higher values indicate improved performance. Comparisons marked with * reach a statistical significance threshold of $p < 0.01$ according to the De-Long test, as measured for individual class AU-ROCs. The Multiplex GNN outperforms all of the baselines for multi-class classification compared to the baseline multi-graph neural networks.

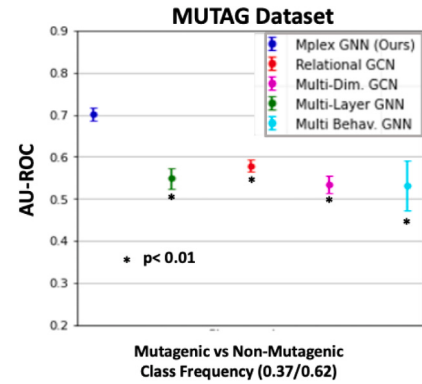


Fig. 4. Performance the on the semi-supervised node classification on the MUTAG dataset as measured by the overall AU-ROC. We display mean performance along with standard errors. Higher values indicate improved performance. Comparisons marked with * reach a statistical significance threshold of $p < 0.01$ according to the De-Long test. The Multiplex GNN outperforms all of the baselines for classification compared to the baseline multi-graph neural networks.

a GCN (Kipf and Welling, 2016) based on quotient connectivity. Next, the filtered embeddings are input to a second GCN which uses the combined supra-adjacency matrix $\hat{\mathcal{A}} = \mathcal{A} + \hat{\mathcal{C}}$. In contrast, our framework tries to exploit the native formalism of the multiplex walks instead of collapsing the graph adjacency matrices during message passing.

3.1.3. Classification results

Figs. 3 and 4 illustrate the per-class AU-ROC and weighted average AU-ROC distribution of our framework (MplexGNN) against the multi-GNN baselines for the AIFB and MUTAG datasets respectively. Along with each errorbar plots, we indicate the frequency of the corresponding class within the dataset. We observe that the Multiplexed GNN provides the best per class and overall AU-ROC for all of the comparisons. The standard errors for our model are also considerably smaller than the baseline predictions as well. The class imbalance makes the classification task challenging, particularly for the minority classes. This partially contributes to the poor performance of the baselines on these classes, where the averaging or separation of connectivity information across planes may miss key patterns relevant to the task at hand. In contrast, our framework is a more efficient alternative for reliably uncovering stable multi-relational patterns from graph data that are relevant for prediction. We conjecture that the performance

Table 1
Dataset description of the NIH-TB dataset.

Modality	CT	Genomic	Demographic	Clinical	Regimen	Continuous
Native Dimensionality	2048	4081	29	1726	233	8
Rank	250	300	24	183	112	8
Reduced Dimensionality	128	64	8	128	64	4

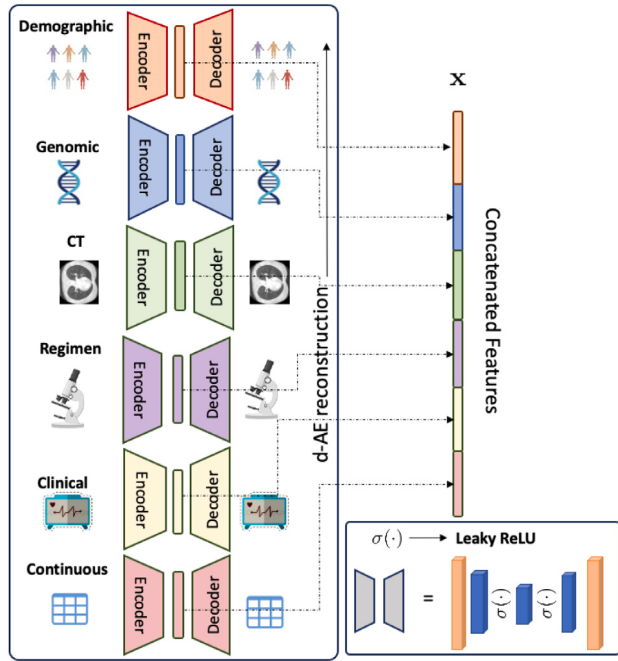


Fig. 5. For the NIH-TB dataset, domain specific autoencoders (d-AE) are used to convert individual modality features to compact feature spaces and reduce the within-modality redundancy. These are trained on an MSE reconstruction loss. The architecture of the encoder-decoder pairs is shown in the box at the bottom right corner and the dimensionality of the reduced features is indicated in Table 1. Then, the reduced features are concatenated and fed to the model in Fig. 1 to convert it into a multiplexed graph representation.

gain is due to the walk-based message passing design, which is more naturalistic and thus suitable for reasoning from multiplexed graphs.

3.2. Evaluating outcome prediction from multimodal medical data

To demonstrate the clinical utility of the multiplexed formulation for multimodal fusion, we employ two multi-omics datasets.

3.2.1. Datasets

NIH-TB Dataset: We consider the Tuberculosis Data Exploration Portal (Gabrielian et al., 2019) consisting of 3051 patients with five different treatment outcomes (Died, Still on treatment, Completed, Cured, or Failure) with the class frequencies as: 0.21/0.11/0.50/0.10/0.08 and five modalities. For each subject, we have features available from demographic, clinical, regimen and genomic recordings with chest CTs available for 1015 of them. We have a total of 4081 genomic, 29 demographic, 1726 clinical, 233 regimen features that are categorical, and 2048 imaging and 8 miscellaneous continuous features. Information that may directly be related to treatment outcomes, eg. drug resistance type, were removed from the clinical and regimen features. For genomic data, 81 single nucleotide polymorphisms (SNPs) from the causative organisms *Mycobacterium tuberculosis* (Mtb) known to be related to drug resistance were used. For 275 of the subjects, we also assemble the raw genome sequence from NCBI Sequence Read Archive. This provides a more fine-grained description of the biological sequences of

Table 2
Dataset description of the ABIDE dataset.

Modal.	Anatomical	Functional	Phenotypic	Connectomics
Nat. Dim.	6	10	48	256

the causative pathogen (Seabolt et al., 2019). Briefly, we performed a *de novo* assembly process on each Mtb genome to yield protein and gene sequences. We utilized InterProScan (Jones et al., 2014) to further process the protein sequences and extract the functional domains, i.e. sub-sequences located within the protein's amino acid chain responsible for the enzymatic bioactivity of a protein. This provides a total of 4000 functional genomic features. Finally, for the imaging modality, the lung was segmented via multi-atlas segmentation (Wang and Yushkevich, 2013) followed by a pre-trained DenseNet (Huang et al., 2017) to extract a 1024-dimensional feature vector for each axial slice intersecting the lung. The mean and maximum of each feature were then assembled to give a total of 2048 features. Missing features are imputed from the training cohort using mean imputation for all runs. Table 1 provides a description of the dataset. We indicate the d-AE bottleneck inferred from the validation dataset along with the rank of the data matrix for each modality. We thus have a total of $P = 128 + 64 + 8 + 128 + 64 + 4 = 396$ features and $K = 32$ latent dimensions in the c-AE. Fig. 5 provides a pictorial overview of this processing pipeline.

ABIDE Dataset: The Autism Brain Imaging Data Exchange I (ABIDE) (Di Martino et al., 2014) is a collection of functional magnetic resonance imaging (rs-fMRI) scans and corresponding phenotypic data from 24 different collection sites for subjects with Autism Spectrum Disorder (ASD) against normal controls (NC). Analogous to Zheng et al. (2022), we select 871 subjects for Autism classification with 468 NC and 403 ASD subjects, i.e. which are roughly class balanced. The average subject ages are 16.84 ± 7.23 for the NC cohort (90 males / 378 females) and 17.07 ± 7.95 for the ASD group (288 males / 115 females). We follow the data curation steps outlined in Zheng et al. (2022). Functional MRI data is processed according to the Configurable Pipeline for the Analysis of Connectomes (C-PAC) (Craddock et al., 2013). Functional connectomes are computed using the Pearson's correlation coefficient between regional time-series from the fMRI data. Thus, four modalities are available in total, including 48 phenotypic measures, 6 automated anatomical assessment metrics, 10 automated functional assessment metrics, and 256 fMRI connectomics measures. For this dataset, we use the concatenated features ($P = 256 + 10 + 6 + 48 = 320$) directly as input to the c-AE ($K = 32$), as the individual features were not found to have much redundancy (low rank structure) (see Table 2).

3.2.2. Baseline comparisons

We compare our multiplexed framework against three types of fusion models, namely: (a) No fusion or single modality predictors (b) traditional fusion approaches (c) ablation studies on the model components and design, and (d) graph based fusion models developed for medical applications.

No Fusion: We first run baseline predictive models on the individual modality features without fusing them as a benchmark. We use a two

layered multi-layered perception (MLP) with hidden layer widths as 400 and 20 and LeakyReLU activation (neg. slope=0.01).

Early Fusion: Individual modality features are first concatenated and then fed through a neural network with the same architecture as described above.

Intermediate Fusion: In this comparison, we perform intermediate fusion after the d-AE projection by using the concatenated feature x as input to a two layered MLP (hidden width: 150 and 20, LeakyReLU activation). Since the ABIDE features do not use the d-AE reduction, we feed the hidden representation from the c-AE to this model. Both comparisons evaluate the benefit of using graph based fusion via the c-AE latent encoder.

Kronecker Product Fusion: An alternative intermediate fusion strategy is to combine individual modality-specific representations via the Kronecker Fusion model in [Chen et al. \(2020\)](#) developed for outcome prediction. It creates an intermediate representation by creating an M th-order tensor by taking a Kronecker product of the modality features (M being the number of modalities). The intermediate representation is then fed into a classification Multi-Layered Perceptron (MLP) that maps to the classification targets. We fix the modality-specific projection networks to be MLP with two hidden layers (hidden dimensions 32 and output dimension 10), leaky ReLU activations (negative slope=0.01). The classification network is an MLP with two hidden layers of dimensionality 32 and ReLU activations. Training hyperparameters are set according to [Chen et al. \(2020\)](#)

Late Fusion: We utilize the late fusion framework of [Wang et al. \(2021\)](#) to combine the predictions from the modalities trained individually in the No Fusion baseline. This framework leverages the uncertainty in the individual classifiers to improve the robustness of outcome prediction. We used the hyperparameters in [Wang et al. \(2021\)](#).

Metric Learning based Fusion: A recent alternative to our multiplexed representation using autoencoders (i.e. d-AE and c-AE projections) is the metric-learning late fusion framework of [Cheerla and Gevaert \(2019\)](#). This approach first projects each of the modality representations to a space of fixed dimensionality, followed by modality-specific discriminative networks (MLPs) trained on classification (cross-entropy loss) against the prediction targets. For fusion, they introduce a 'similarity loss' to encourage all modality-specific logits for a given patient (considered pairwise across batches) to be similar to each other, but dissimilar across patients. The network is trained jointly on the two losses. For our application, we design the projection and classification networks to be two layered MLPs (hidden dimensionality=64, LeakyReLU activation). Hyperparameters for training are set according to [Cheerla and Gevaert \(2019\)](#).

rGCN on a Multiplexed Graph: This baseline utilizes the multigraph representation learning (Blue Box of [Fig. 1](#)), but replaces the Multiplex GNN feature extraction with the rGCN framework of [Schlichtkrull et al. \(2018\)](#). Since the width, depth and graph readout is the same as with the Multiplex GNN, this helps evaluate the expressive power of the walk based message passing in [Eq. \(5\)](#).

rGCN w/o Latent Encoder: For this comparison, we utilize the reduced features after the d-AE, but instead create a multi-layered graph with the individual modalities in different planes. Within each plane, nodes are fully connected to each other after which a two layered rGCN ([Schlichtkrull et al., 2018](#)) model is trained. Effectively, *within modality* feature dependence may still be captured in the planes, but the concept space is not used to infer the *cross-modal* interactions.

GCN on monoplex feature graph: This baseline also incorporates a graph based representation, but does not include the use of latent concepts to model within and cross-modal feature correlations. Essentially,

we construct a fully connected graph on x instead of using the (multi-) conceptual c-AE space and train a two layered Graph Convolutional Network ([Kipf and Welling, 2016](#)) for outcome prediction.

Transformers for fusion: Lack of clear semantic correspondences across features from different modalities is a challenging aspect of multimodal fusion. In this baseline, we explore the use a transformer architecture ([Hou et al., 2022](#)) to define the relationships between all pairs of multimodal features. We design a two layered transformer encoder followed by a two layered MLP (hidden layer width = 256, ReLU activation) to map the concatenated filtered representation to the class label for each subject. Since this baseline admits all possible modality feature interactions, it effectively constructs a weighted fully connected multimodal graph between modality features. This baseline could be considered as an ablation that directly evaluates the benefit of restricting interactions between features via the multiplex graph representation learning.

Latent Graph Learning: This baseline was developed on multimodal data ([Cosmo et al., 2020](#)) and introduces latent patient graph learning from the modality features via a graph-attention (GAT-like ([Veličković et al., 2017](#))) formulation. However, unlike our model, this baseline concatenates the modality features upfront and constructs a single-relational (patient-patient) graph that is learned as a part of the training. Hyperparameters are set according to [Cosmo et al. \(2020\)](#).

Multimodal Graph Learning: This baseline ([Zheng et al., 2022](#)) also performs latent graph learning on a single-relational patient-patient graph similar to the previous model. Instead of concatenating the modality features to learn the graph, this model performs an attention based aggregation across modality features to construct the node features. Hyperparameters are set according to [Zheng et al. \(2022\)](#). Note that this and the previous baselines perform node-based classification for fusion instead of graph-level prediction.

Hybrid Graph Convolutional Network (HGCN): The HGCN model is a very recent graph based fusion baseline developed by [Hou et al. \(2023\)](#). This method consists of identical graph convolutional layers (GCN) to process the individual modality features and then convert them into a modality specific hypergraph connecting all features of each modality to a hypernode densely. The obtained modality representations are combined using a hypergraph convolutional layer (HCN) that implements a hyperedge mixing module to provide a unified mixed modality representation. Finally, the mixed representation is added to modality specific embedding to obtain a joint representation for training a classifier.

3.2.3. Outcome prediction performance

[Figs. 6–7](#) and [Figs. 8–9](#) illustrate the outcome prediction results for the NIH-TB and ABIDE datasets.

We notice that all fusion frameworks (ours, as well as several baselines) largely improve the performance over the single modality outcome classifiers. This observation highlights the need for fusing multi-modal data for clinical applications. Our framework outperforms the traditional fusion baselines (Early, Intermediate, and Uncertainty based Late Fusion, Kronecker Product Fusion and Metric Learning), for a majority of the comparisons for both datasets. This suggests that the implicit assumptions made by these frameworks may be too restrictive for modeling nuanced relationships between modality features.

The rGCN-multigraph baseline is an ablation that replaces the Multiplexed GNN with the framework in [Schlichtkrull et al. \(2018\)](#) (which also performed second to best among the comparisons in [Section 3.1.2](#)). We utilize the same multiplexed representation as learned by the c-AE autoencoder latent space. The performance gains within this ablation suggest that the Multiplexed GNN is better suited for task-specific reasoning from multigraphs. The added representational power is likely due to the design of the message passing ([Eq. \(5\)](#)). Along similar lines, rGCN w/o latent encoder and the GCN on the monoplex feature graph

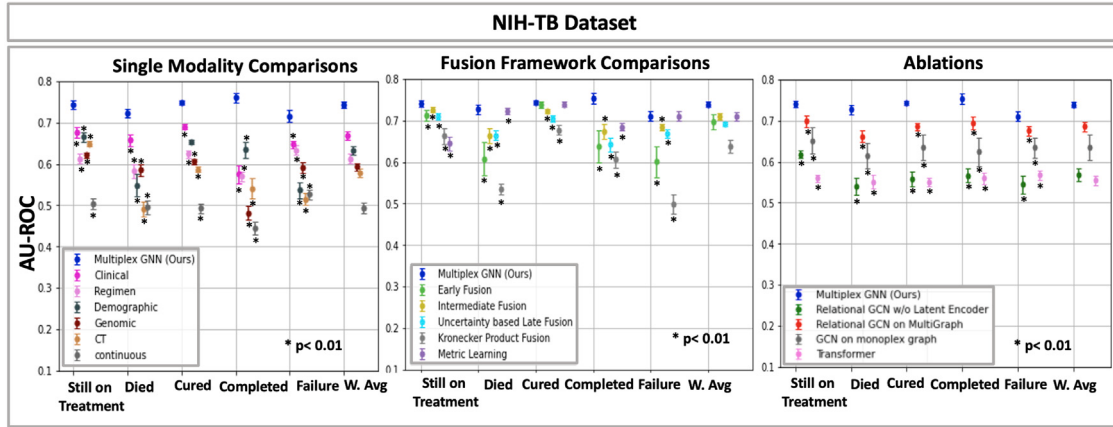


Fig. 6. Outcome prediction performance on the NIH-TB dataset as measured by per-class and weighted average AU-ROC. We display mean performance along with standard errors. Higher values indicate improved performance. Comparisons marked with * reach a statistical significance threshold of $p < 0.01$ according to the De-Long test, as measured for individual class AU-ROCs. The figure illustrates comparisons against (Left): Single Modality Predictors (Middle): Traditional (early/intermediate/late) Fusion, Metric Learning, Kronecker Product Fusion (Right): Ablations of the framework. We note that the multiplexed GNN provides improved performance for multi-class classification on all comparisons, with a majority of them achieving the statistical threshold.

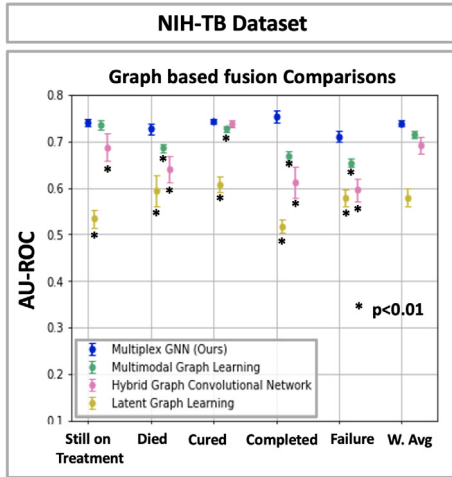


Fig. 7. Outcome prediction performance on the NIH-TB dataset as measured by per-class and weighted average AU-ROC. Higher values indicate improved performance. Comparisons marked with * reach a statistical significance threshold of $p < 0.01$ according to the De-Long test, as measured for individual class AU-ROCs. We display mean performance along with standard errors. This figure illustrates comparisons against recent graph based fusion frameworks for medical data. We note that the multiplexed GNN provides improved performance for multi-class classification on all comparisons, achieving the statistical threshold in all cases.

baseline comparisons are generic graph based approaches, which we adapt for fusion. They allow us to examine the benefit of using the salient activation patterns from the c-AE latent concept space to infer the multi-graph representation. Specifically, the former separates the modality features into fully connected multi-planar graph. The latter constructs a fully connected graph with a single edge-type on the concatenated modality features. Our framework provides large gains over these baselines. This highlights the efficacy of our graph construction. We surmise that the salient learned conceptual patterns are more successful at uncovering cross modal interactions between features that are explanative of patient outcomes.

We observe that the multiplex GNN approach outperforms the transformer baseline on both datasets. We posit that this difference in performance arises as a direct consequence of the transformer architectural design which seeks to estimate *all possible pairwise feature interactions* in a data-driven fashion. This is often extremely challenging

for applications in the healthcare domain owing to the high data dimensionality coupled with limited dataset sizes. In contrast, the multiplexed graph construction pre-selects for connectivity patterns deemed to be most salient, implicitly regularizing the feature graph representation. At the same time, the multiplex GNN walk factorizations (i.e. $\mathcal{A}\hat{\mathcal{C}}$ and $\hat{\mathcal{C}}\mathcal{A}$) are deliberately designed to combine within and across planar connection patterns, allowing nodes across planes to seamlessly communicate with each other during message passing. We believe that these two design choices contribute to the improved representational power over transformers for this application.

Finally, our framework outperforms the latent graph learning multimodal fusion models, and the hybrid graph convolutional network for the NIH-TB dataset (Fig. 7). It is a close second for ABIDE (Fig. 9) in terms of mean performance, but provides smaller standard errors. As opposed to the collapsed patient-patient graphs constructed in Zheng et al. (2022), Cosmo et al. (2020), our approach allows for subject-specific modeling of feature interactions in the construction of the multigraph. The Hybrid GCN (Hou et al., 2023) constructs patient-specific graphs and processes multimodal features in a modality specific (using graph and hyper-graph convolutions) as well as mixed fashion (via a hyper-edge mixing network). However, the feature and modality semantics may not be preserved through the intermediate mixing transformations. In contrast, the multiplex GNN design ensures that the filtered supra-node embeddings maintain their planar and nodal (i.e. feature and modality) identities throughout, allowing for fine-grained yet patient-specific reasoning. We believe that this relaxation of constraints allows us to generalize well to challenging multi-outcome prediction tasks such as the NIH-TB dataset.

Overall, these observations highlight key representational aspects of our framework, and demonstrate the efficacy for two different outcome prediction tasks. Given the clinical relevance, a promising direction for exploration would be to extend frameworks for explainability in GNNs (for example, via subgraph exploration (Yuan et al., 2021)) to Multiplex GNNs to automatically highlight patterns relevant to downstream prediction.

4. Discussion

4.1. Examining the multimodal graph representation

After having extensively evaluated the prediction performance for various applications, we would like to better understand the multiplexed graph representation learned as a part of the multimodal fusion. Recall that the planes of our multiplex graphs refer to directions in

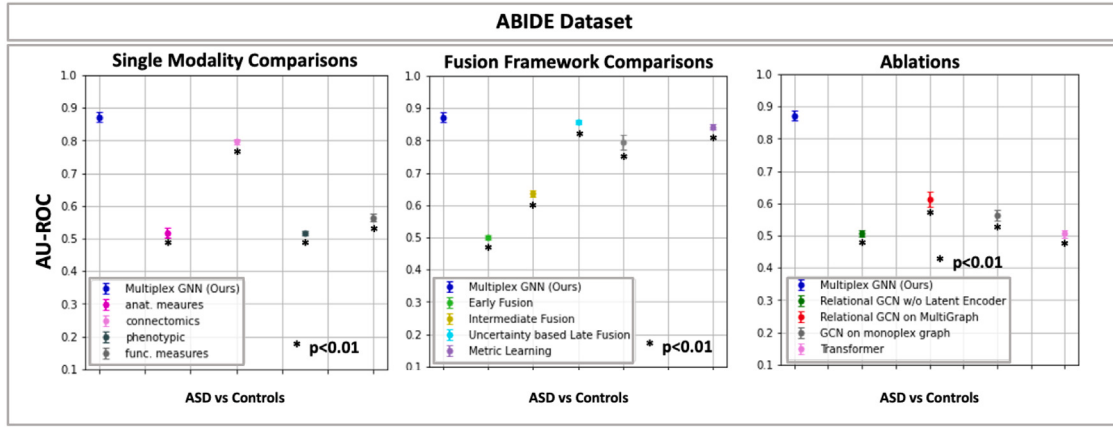


Fig. 8. Classification performance on the ABIDE dataset as measured by the AU-ROC metric. We display mean performance along with standard errors. Higher values indicate improved performance. Comparisons marked with * reach a statistical significance threshold of $p < 0.01$ according to the De-Long test, as measured for individual class AU-ROCs. The figure illustrates comparisons against (Left): Single Modality Predictors (Middle): Traditional (early/intermediate/late) Fusion Frameworks, Metric Learning, Kronecker Product Fusion, (Right): Ablations of the framework. We note that the multiplexed GNN provides improved performance for classification on all comparisons, achieving the statistical threshold in all cases.

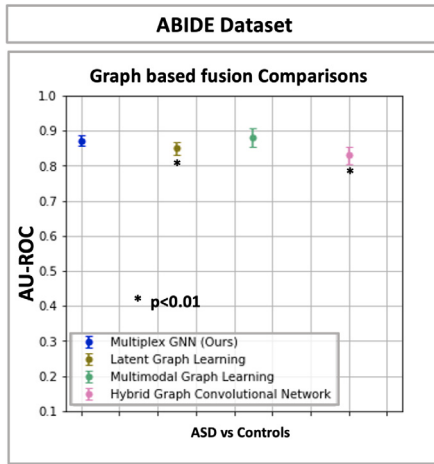


Fig. 9. Classification performance on the ABIDE dataset as measured by the AU-ROC metric. Comparisons marked with * reach a statistical significance threshold of $p < 0.01$ according to the De-Long test, as measured for individual class AU-ROCs. We display mean performance along with standard errors. This figure illustrates comparisons against recent graph based fusion frameworks for medical data. The Multiplexed GNN is a close second on this task, but results in smaller standard errors.

the concept autoencoder (c-AE) latent space. Each direction allows us to uncover different types of dependence between the multimodal features. For a given subject, the multimodal features deemed to be salient with respect to each latent direction induce a unique connectivity pattern per plane. The multiplexed GNN then leverages this unique topology to robustly uncover complementary discriminative patterns from the modality features.

Since this connectivity is different for each patient, we aim to visualize how these connection patterns look on average across the dataset via a heatmap display. This is used to determine which connections are more frequently retained, and thus likely more important for outcome prediction. Specifically, we display the feature connectivity heatmap obtained from three of $K = 32$ planes from one of the runs. Fig. 10 for the NIH-TB dataset and Fig. 11 for the ABIDE dataset, ordered from left to right from most to least sparse. Each connection is weighted by the relative frequency of occurrence across patients, i.e. relative population-level density. Brighter values imply that the pair of features were connected with each other in the multiplex graph of more subjects. Since the connection patterns are symmetric by construction, we

display only the lower half of this matrix. For simplicity, we group features from each modality into contiguous blocks.¹

We notice that for both datasets, the c-AE picks out a large number of cross-modal interactions (highlighted in the red box), in tandem with intra-modality feature connections (block diagonal connections of the matrix, highlighted by the green box). While sparse overall, a number of connections in each matrix have a considerable frequency of occurrence, indicating that they are mined fairly consistently (about 20–60 percent of subjects for the ABIDE dataset and 10–40 percent for the NIH-TB dataset). Comparing across datasets, we do see differences in the distribution of the connection strengths in the heatmap, with those for the ABIDE dataset having relatively higher consensus values across subjects for the strongest connections. More interestingly, each plane (i.e. latent direction in the concept space) induces a unique graph topology, as seen by comparing the connectivity patterns across the example plots from a given dataset. This is in line with our desire to automatically uncover complementary modes of interaction (via the c-AE latent space) from multimodal data. The Multiplexed GNN utilizes the complementarity in the representations in a principled fashion during message passing for reasoning

Finally, we would like to examine the contribution that individual data modalities have for the fusion task for outcome prediction. To this end, we run a second ablation study that assesses the differential modality contribution. In this experiment, we remove features corresponding to a given modality, one modality at a time; then we re-estimate the multiplex graph (by retraining the c-AE) and re-train the Multiplexed GNN for prediction. We use the same train-test-validation folds as in the main experiments. Next, we quantify the change in the weighted AU-ROC obtained by excluding one modality at a time. Table 3 denotes the results for the NIH-TB dataset and Table 4 for the ABIDE dataset. First, we observe that removal of any given modality lowers the overall performance for both applications. This suggests that none of the modalities are likely redundant in relation to the prediction task. Additionally, removing Clinical features for the NIH-TB dataset and Connectomics features for ABIDE dataset lead to the largest performance drop. This indicates that the feature connectivity patterns associated with these modalities are considered informative for the respective prediction tasks.

¹ Note: For the NIH-TB dataset, the features are the outputs of the d-AE dimensionality reduction and are not the raw input features.

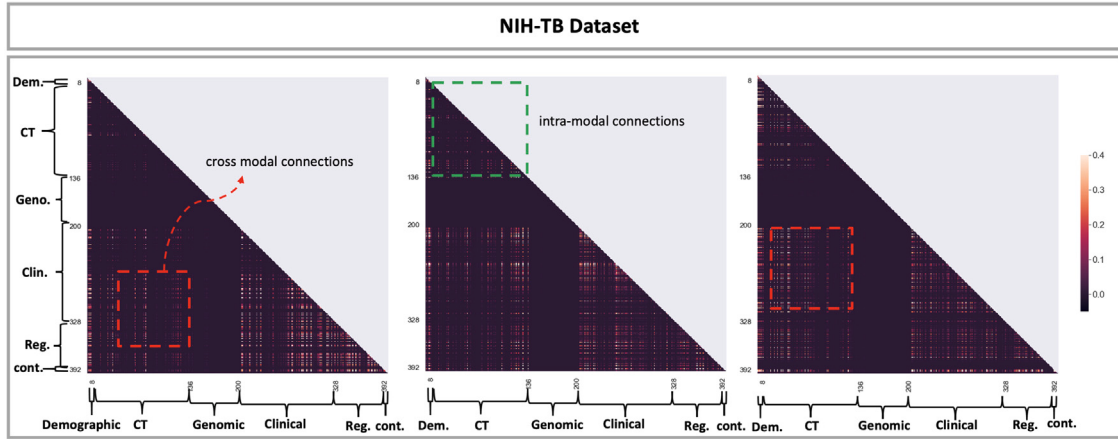


Fig. 10. Average within-planar connectivity topology for three of the thirty-two multiplex planes for the NIH-TB dataset. We display the relative frequency of occurrence of the connection across subjects. Brighter values indicate stronger consensus across subjects. Reduced modality features have been grouped together into contiguous blocks in the matrix. Each plane captures a different set of interaction patterns (i.e. unique connectivity topology). While overall sparse, we observe that several within-modality (block diagonal, examples highlighted in green) and cross modal (off diagonal, examples highlighted in red) are consistently observed across individuals.

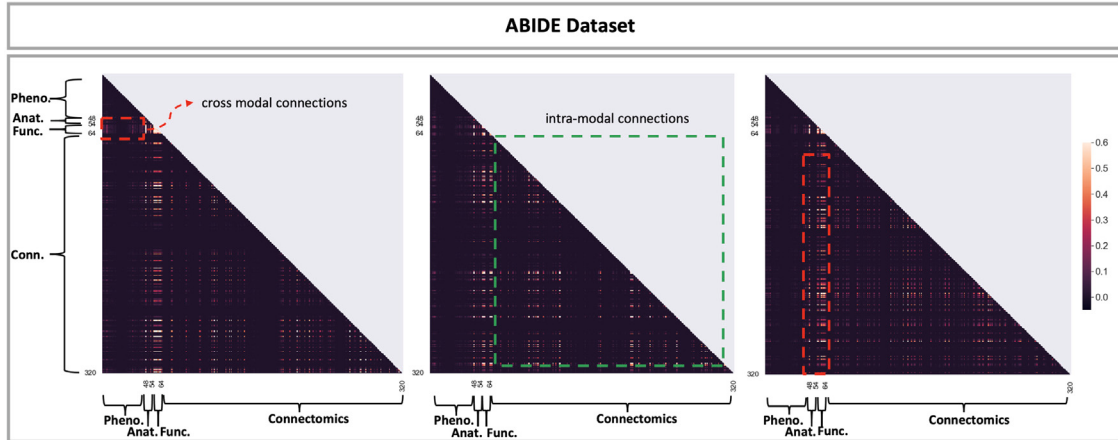


Fig. 11. Average within-planar connectivity topology for three of the thirty-two multiplex planes for the ABIDE dataset. We display the relative frequency of occurrence of the connection across subjects. Brighter values indicate stronger consensus across subjects. Reduced modality features have been grouped together into contiguous blocks in the matrix. Each plane captures a different set of interaction patterns (i.e. unique connectivity topology). While overall sparse, we observe that several within-modality (block diagonal, examples highlighted in green) and cross modal (off diagonal, examples highlighted in red) are consistently observed across individuals.

Table 3
Modality feature ablation on the NIH dataset.

Removed modality	W. avg. AU-ROC	∇
Reference	0.74 ± 0.008	—
CT	0.71 ± 0.012	−0.03
Genomic	0.71 ± 0.006	−0.03
Clinical	0.70 ± 0.005	−0.04
Regimen	0.71 ± 0.008	−0.03
Demographic	0.71 ± 0.008	−0.03
Continuous	0.73 ± 0.009	−0.01

Table 4
Modality feature ablation on the ABIDE dataset.

Removed modality	AU-ROC	∇
Reference	0.87 ± 0.014	—
Anatomical	0.62 ± 0.052	−0.25
Functional	0.67 ± 0.046	−0.19
Phenotypic	0.74 ± 0.045	−0.13
Connectomics	0.54 ± 0.009	−0.33

4.2. Advantages and application potential

We have developed a novel graph neural framework grounded in the native formalism of multiplex graphs to address problems of generalized multimodal fusion in medical data. Going one step beyond simple statistical measures, the multiplexed graph construction allows us to uncover nuanced non-linear notions of dependence between modality features via the latent space of autoencoder representations. The Multiplexed GNN layers allow the node features to retain their individuality in terms of the plane (relationship type) in the filtered representation. This admits more explainable intermediate representations in comparison to the baselines, i.e. provides us with the ability to explicitly reason at the granularity of both the nodal and planar representations. Conversely, the GNN baselines in Section 3.1.2 and graph based/traditional fusion baselines in Section 3.2.2 collapse this information, either in the multimodal representation or in the inference step. We conjecture that added flexibility contributes to the improved generalization power in both sets of experiments.

As such, this model makes very mild assumptions about the nature of the multimodal data. Moreover, we have demonstrated the utility of the Multiplexed GNN in both the transductive (semi-supervised) and inductive (fully supervised) inference setting. The general principles and machinery developed in this work would likely be useful to a wide variety of applications beyond the medical realm.

4.3. Limitations and future work

As mentioned previously, lack of clear semantic correspondences between multimodal features is a challenging aspect of fusion applications. The development of multiplexed graph neural networks for this is an important preliminary step towards developing flexible machinery to explore the dependence across features and modalities in a principled fashion. At the same time, the current approach suffers from some limitations. Specifically, our model implementation separates the modality specific feature extraction from the multiplexed graph construction and inference steps. End-to-end optimization of the two is challenging given the non-differentiability of our multiplexed graph construction step. As a result, the current implementation relies on converting the imaging and non-imaging information to 1D feature signals which are then fused within the multiplexed GNN. It is important to note that such feature extraction is not universal, is modality-dependent, and often introduces heterogeneous characteristics specific to the application under consideration. In this light, an important future direction of exploration is the integration of modality-specific feature extraction modules end-to-end with the inference framework to allow them to inform each other during training. In turn, this would allow for more targeted task-specific representation learning, which is currently not addressed sufficiently within this work.

In problems of multimodal fusion, data acquisition is a fairly contrived and expensive process. In many applications, especially in medical domain, modalities may often be only partially observed, missing in totality, or noisy in acquisition. Simple methods such as mean based imputation may be inadequate as they miss key patterns in the data. In an aim to mitigate such practical challenges, another active line of exploration is to extend the framework to handle missing or ambiguous data in the multiplexed representation by leveraging statistical and graph theoretic tools in conjunction with message passing.

By definition, the multiplexed formulation assumes that the planes share a common set of nodes which are connected by vertical connections. An active direction exploration is to relax this assumption into multi-layered graph representations (D'Souza et al., 2023) for generalized fusion frameworks. Here, each plane may be allowed to have an arbitrary number of nodes and inter-planar cross linkages across planes may also be present. While the transformer model is one instance where dense connectivity patterns are permitted, such models do not always work well in practice when limited training data is available due to poor regularization. Fundamentally, this aspect would require a rethinking of notions of dependence (connectivity) for constructing such graphs, as well as extending the message passing scheme to handle the more general case.

5. Conclusion

We have introduced a novel framework based on multiplexed graphs that can combine diverse multimodal data for outcome prediction. We have explored the potential of this framework in the medical domain using two clinical multi-omics datasets. Our Multi-modal Graph Representation Learning transforms individual modality features into abstract concept spaces, which allows us to tease apart complex cross modal dependencies between features. We have also developed a novel multiplexed graph neural network that can systematically track information flow within the multi-graph via message passing walks. Our GNN formulation provides the necessary flexibility to mine rich representations from multimodal data. Overall, this provides for improved outcome prediction performance against several state-of-the-art baselines. Finally, our framework makes very few assumptions and could be easily applied to a variety of fusion problems in general AI domains.

Declaration of competing interest

The authors do not have any additional conflicts of interest to declare.

Data availability

The datasets used are public and the code has been made available (link in paper).

References

- Baltrušaitis, T., Ahuja, C., Morency, L.-P., 2018. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (2), 423–443.
- Barteseaghi, P., Clemente, G.P., Grassi, R., 2022. A tensor-based unified approach for clustering coefficients in financial multiplex networks. *Inform. Sci.* 601, 268–286.
- Bloehdorn, S., Sure, Y., 2007. Kernel methods for mining instance data in ontologies. In: *The Semantic Web*. Springer, pp. 58–71.
- Cheerla, A., Gevaert, O., 2019. Deep learning with multimodal representation for pancreatic cancer prognosis prediction. *Bioinformatics* 35 (14), i446–i454.
- Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F., 2020. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans. Med. Imaging* 41 (4), 757–770.
- Cosmo, L., Kazi, A., Ahmadi, S.-A., Navab, N., Bronstein, M., 2020. Latent-graph learning for disease prediction. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II* 23. Springer, pp. 643–653.
- Cozzo, E., de Arruda, G.F., Rodrigues, F.A., Moreno, Y., 2018. *Multiplex Networks*. Springer International Publishing, Cham, <http://dx.doi.org/10.1007/978-3-319-92255-3>, URL <http://link.springer.com/10.1007/978-3-319-92255-3>.
- Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S.S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., et al., 2013. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). *Front. Neuroinform.* 42, 10–3389.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S., Arenas, A., 2014. Mathematical formulation of multilayer networks. *Phys. Rev. X* (ISSN: 21603308) 3, 041022. <http://dx.doi.org/10.1103/PHYSREVX.3.041022>, URL <https://journals.aps.org/prx/abstract/10.1103/PhysRevX.3.041022>.
- Debnath, A.K., Lopez de Compadre, R.L., Debnath, G., Shusterman, A.J., Hansch, C., 1991. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *J. Med. Chem.* 34 (2), 786–797.
- Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* 29, 3844–3852.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 837–845.
- Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19 (6), 659–667.
- D'Souza, N.S., Nebel, M.B., Crocetti, D., Robinson, J., Mostofsky, S., Venkataraman, A., 2021. M-gcn: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations. In: *Med. Imag. Deep Learn.*. PMLR, pp. 119–130.
- D'Souza, N.S., Nebel, M.B., Crocetti, D., Robinson, J., Mostofsky, S., Venkataraman, A., 2021a. A matrix autoencoder framework to align the functional and structural connectivity manifolds as guided by behavioral phenotypes. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII* 24. Springer, pp. 625–636.
- D'Souza, N.S., Nebel, M.B., Crocetti, D., Robinson, J., Wymbs, N., Mostofsky, S.H., Venkataraman, A., 2021b. Deep sr-DDL: Deep structurally regularized dynamic dictionary learning to integrate multimodal and dynamic functional connectomics data for multidimensional clinical characterizations. *NeuroImage* 241, 118388.
- D'Souza, N.S., Nebel, M.B., Crocetti, D., Wymbs, N., Robinson, J., Mostofsky, S., Venkataraman, A., 2020. A deep-generative hybrid model to integrate multimodal and dynamic connectivity for predicting spectrum-level deficits in autism. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII* 23. Springer, pp. 437–447.
- D'Souza, N.S., Nebel, M.B., Wymbs, N., Mostofsky, S., Venkataraman, A., 2019. A coupled manifold optimization framework to jointly model the functional connectomics and behavioral data spaces. In: *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings* 26. Springer, pp. 605–616.
- D'Souza, N.S., Venkataraman, A., 2023. mSPD-NN: A geometrically aware neural framework for biomarker discovery from functional connectomics manifolds. In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 53–65.

- D'Souza, N.S., Wang, H., Giovannini, A., Foncubierta-Rodriguez, A., Beck, K.L., Boyko, O., Syeda-Mahmood, T., 2022. Fusing modalities by multiplexed graph neural networks for outcome prediction in tuberculosis. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference*, Singapore, September 18–22, 2022, Proceedings, Part VII. Springer, pp. 287–297.
- D'Souza, N.S., Wang, H., Giovannini, A., Foncubierta-Rodriguez, A., Beck, K.L., Boyko, O., Syeda-Mahmood, T., 2023. MaxCorrMGNN: A multi-graph neural network framework for generalized multimodal fusion of medical data for outcome prediction. In: *Workshop on Machine Learning for Multimodal Healthcare Data*. Springer, pp. 141–154.
- Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P., 2015. Convolutional networks on graphs for learning molecular fingerprints. *arXiv preprint arXiv:1509.09292*.
- Ferriani, S., Fonti, F., Corrado, R., 2013. The social and economic bases of network multiplexity: Exploring the emergence of multiplex ties. *Strateg. Organ.* (ISSN: 14761270) 11, 7–34. <http://dx.doi.org/10.1177/1476127012461576>.
- Fey, M., 2019. Just jump: Dynamic neighborhood aggregation in graph neural networks. *arXiv preprint arXiv:1904.04849*.
- Gabrielian, A., Engle, E., Harris, M., Wollenberg, K., Juarez-Espinosa, O., Glogowski, A., Long, A., Patti, L., Hurt, D.E., Rosenthal, A., Tartakovsky, M., 2019. TB DEPOT (data exploration portal): A multi-domain tuberculosis data analysis resource. In: Gao, F. (Ed.), *PLoS One* (ISSN: 1932-6203) 14 (5), e0217410. <http://dx.doi.org/10.1371/journal.pone.0217410>.
- Gomez, S., Diaz-Guilera, A., Gomez-Gardenes, J., Perez-Vicente, C.J., Moreno, Y., Arenas, A., 2013. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.* 110 (2), 028701.
- Grassia, M., De Domenico, M., Mangioni, G., 2021. mGNN: Generalizing the graph neural networks to the multilayer case. *arXiv preprint arXiv:2109.10119*.
- Guo, Q., Cozzo, E., Zheng, Z., Moreno, Y., 2016. Levy random walks on multiplex networks. *Sci. Rep.* 6 (1), 1–11.
- Hou, W., Huang, H., Peng, Q., Yu, R., Yu, L., Wang, L., 2022. Spatial-hierarchical graph neural network with dynamic structure learning for histological image classification. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Springer Nature Switzerland, Cham, pp. 181–191.
- Hou, W., Lin, C., Yu, L., Qin, J., Yu, R., Wang, L., 2023. Hybrid graph convolutional network with online masked autoencoder for robust multimodal cancer survival prediction. *IEEE Trans. Med. Imaging*.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4700–4708.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* (Oxford, England) (ISSN: 1367-4811) 30 (9), 1236–1240. <http://dx.doi.org/10.1093/bioinformatics/btu031>.
- Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kivelä, M., Arenas, A., Barthélemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A., 2014. Multilayer networks. *J. Complex Netw.* (ISSN: 2051-1310) 2, 203–271. <http://dx.doi.org/10.1093/COMNET/CNU016>, URL <https://academic.oup.com/commnet/article/2/3/203/2841130>.
- LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. In: *The handbook of brain theory and neural networks*, Vol. 3361, No. 10. p. 1995.
- Li, X., Zhou, Y., Dvornek, N.C., Zhang, M., Zhuang, J., Ventola, P., Duncan, J.S., 2020. Pooling regularized graph neural network for fmri biomarker analysis. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference*, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23. Springer, pp. 625–635.
- Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Ma, Y., Wang, S., Aggarwal, C.C., Yin, D., Tang, J., 2019. Multi-dimensional graph convolutional networks. In: *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, pp. 657–665.
- Maggioni, M.A., Breschi, S., Panzarasa, P., 2013. Multiplexity, growth mechanisms and structural variety in scientific collaboration networks. (ISSN: 13662716) pp. 185–194. <http://dx.doi.org/10.1080/13662716.2013.791124>, URL <https://www.tandfonline.com/doi/abs/10.1080/13662716.2013.791124>.
- Matsuno, R., Murata, T., 2018. MELL: effective embedding method for multiplex networks. In: *Companion Proceedings of the the Web Conference 2018*. pp. 1261–1268.
- Muñoz-Sellart, M., Cuevas, L., Tumato, M., Merid, Y., Yassin, M., 2010. Factors associated with poor tuberculosis treatment outcome in the southern region of ethiopia. *Int. J. Tuberculosis Lung Dis.* 14 (8), 973–979.
- Musmeci, N., Nicosia, V., Aste, T., Di Matteo, T., Latora, V., 2017. The multiplex dependency structure of financial markets. *Complexity* 2017.
- Nandakumar, N., D'Souza, N.S., Craley, J., Manzoor, K., Pillai, J.J., Gujar, S.K., Sair, H.I., Venkataraman, A., 2018. Defining patient specific functional parcellations in lesional cohorts via Markov random fields. In: *Connectomics in NeuroImaging: Second International Workshop, CNI 2018, Held in Conjunction with MICCAI 2018*, Granada, Spain, September 20, 2018, Proceedings 2. Springer, pp. 88–98.
- Nandakumar, N., D'souza, N.S., Manzoor, K., Pillai, J.J., Gujar, S.K., Sair, H.I., Venkataraman, A., 2020. A multi-task deep learning framework to localize the eloquent cortex in brain tumor patients using dynamic functional connectivity. In: *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology: Third International Workshop, MLCN 2020, and Second International Workshop, RNO-AI 2020, Held in Conjunction with MICCAI 2020*, Lima, Peru, October 4–8, 2020, Proceedings 3. Springer, pp. 34–44.
- Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A., Tran, D., 2018. Image transformer. In: *International Conference on Machine Learning*. PMLR, pp. 4055–4064.
- Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2008. The graph neural network model. *IEEE Trans. Neural Netw.* 20 (1), 61–80.
- Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M., 2018. Modeling relational data with graph convolutional networks. In: *European Semantic Web Conference*. Springer, pp. 593–607.
- Seabolt, E.E., Nayar, G., Krishnareddy, H., Agarwal, A., Beck, K.L., Terrizzano, I., Kandogan, E., Roth, M., Mukherjee, V., Kaufman, J.H., 2019. OMXWare, A Cloud-Based Platform for Studying Microbial Life at Scale. *arXiv:1911.02095*.
- Sebenius, I., Campbell, A., Morgan, S.E., Bullmore, E.T., Liò, P., 2021. Multimodal graph coarsening for interpretable, MRI-based brain graph neural network. In: *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing. MLSP, IEEE*, pp. 1–6.
- Subramanian, V., Chidester, B., Ma, J., Do, M.N., 2018. Correlating cellular features with gene expression using CCA. In: *2018 IEEE 15th International Symposium on Biomedical Imaging. ISBI 2018, IEEE*, pp. 805–808.
- Subramanian, V., Do, M.N., Syeda-Mahmood, T., 2020. Multimodal fusion of imaging and genomics for lung cancer recurrence prediction. In: *2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, IEEE*, pp. 804–808.
- Subramanian, V., Syeda-Mahmood, T., Do, M.N., 2021. Multimodal fusion using sparse CCA for breast cancer survival prediction. In: *2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE*, pp. 1429–1432.
- Tudisco, F., Arrigo, F., Gautier, A., 2018. Node and layer eigenvector centralities for multiplex networks. *SIAM J. Appl. Math.* 78 (2), 853–876.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Wang, H., Subramanian, V., Syeda-Mahmood, T., 2021. Modeling uncertainty in multimodal fusion for lung cancer survival analysis. In: *2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE*, pp. 1169–1172.
- Wang, L., Wu, J., Huang, S.-L., Zheng, L., Xu, X., Zhang, L., Huang, J., 2019. An efficient approach to informative feature extraction from multimodal data. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 01. pp. 5281–5288.
- Wang, H., Yushkevich, P., 2013. Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation. *Front. Neuroinf.* 7, 27.
- Xu, K., Hu, W., Leskovec, J., Jegelka, S., 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.
- Yang, X., Liu, W., Liu, W., Tao, D., 2019. A survey on canonical correlation analysis. *IEEE Trans. Knowl. Data Eng.* 33 (6), 2349–2368.
- Yuan, H., Yu, H., Wang, J., Li, K., Ji, S., 2021. On explainability of graph neural networks via subgraph explorations. In: *International Conference on Machine Learning*. PMLR, pp. 12241–12252.
- Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., Yang, Y., Zhao, Y., 2022. Multi-modal graph learning for disease prediction. *IEEE Trans. Med. Imaging* 41 (9), 2207–2216.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: A review of methods and applications. *AI Open* 1, 57–81.