Multimodal deep learning on hypergraphs

Arya, D.

Publication date
2022
Document Version
Final published version

Multimodal Deep Learning on Hypergraphs

Devanshu Arya

Multimodal Deep Learning on Hypergraphs

Devanshu Arya

Multimodal Deep Learning on Hypergraphs

Devanshu Arya

This book was typeset by the author using LaTeX 2_ε.

Cover design: Kratika Singh

Multimodal Deep Learning on Hypergraphs

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het college voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 17 juni 2022, te 11.00 uur

door

Devanshu Arya

geboren te Patna, India

Promotiecommissie

Promotor:	Prof. dr. M. Worring	Universiteit van Amsterdam
	dr. S. Rudinac	Universiteit van Amsterdam
Overige leden:	prof. dr. A. Hanjalic	Technische Universiteit Delft
	dr. T.N. Kipf	Google Research
	prof. dr. C.G.M. Snoek	Universiteit van Amsterdam
	prof. dr. ing. Z.J.M.H. Geradts	Universiteit van Amsterdam
	prof. dr. E. Kanoulas	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

UNIVERSITEIT VAN AMSTERDAM

The work described in this thesis has been carried out at the MultiX group of the University of Amsterdam.

CONTENTS

1

INTRODUCTION

From early childhood to old age, humans gain and revise their understanding of the world by inferring knowledge through the analysis of *objects* and forming mental *relations* among those objects [28]. The basis for the object analysis are sensory inputs, whereas the relations are formed by contextualizing these inputs. We can make inference about an object by the raw sensory data associated with it, often referred to as its content, and relating it to pre-existing knowledge representations or other sensory inputs. Doing so, we create an understanding about a scene, an activity or an event. For example, a person who has never seen "a house boat" or "a white peacock" can easily imagine these combinations by simply interpreting the content of each word and forming relations between them. Thus, for making any inference, a structural understanding of the raw sensory data in combination with their relations is pivotal.

The very thought of *relating* objects makes us typically assume the relations would be *pairwise* and that is how we commonly represent relations in a machine, namely using *graphs*. A graph is a data structure describing a set of objects, represented as *nodes*, and their pairwise relationships, represented as *edges*. For example, a simple financial transaction between two individuals can be represented by an edge between two nodes representing those individuals. Graphs have been the most ubiquitous data structures for representing relations and using them to discover relevant information in a data collection. This is due to their capability to combine node-level information with the underlying inter-node relations. However, making any inference using only pairwise relations is often insufficient in real-world scenarios. Consider a simple visual scene of "a room consisting of a chair, a desk, a person and a picture of human anatomy". Accurately making even a simple inference about the person using pairwise relations between person-chair or person-desk in this scenario is highly unlikely. However, if we include all the objects and analyze this *group relation* of person-chair-desk and human anatomy picture, we can make a fair judgement about the person being a doctor and the room to be a clinic. These group relations termed as *higher-order* relations — involving more than two objects at a time — are crucial for humans to gain insights. Higher-order relations are commonly encountered in many domains, such as medical science (e.g., coexisting diseases/symptoms), pharmacology (e.g., reacting chemicals), bibliometrics (e.g., collaborating researchers), people analytics (e.g., a team) and social networks (e.g., groups of users and the posts among them). These relations capture a group of objects, where each group can exhibit different properties and the higher-order relations can dynamically change over time. Therefore, representing relations in real-world datasets as pairwise connections using graphs is sub-optimal in capturing the complex information. Using higher-order relations can enhance the representation capabilities of a data structure.

Just as humans exploit higher-order relations to make sense of the world, machines should also be capable of exploiting them for making better inferences. However, as mentioned above, modeling the higher-order relations with graphs leads to loss of information. The pairwise relations fail to represent all the higher-order relations among the objects and, do not correctly capture the collective flow of information. A collection of intersecting higher-order relations can be much better represented using *hypergraphs*. Hypergraphs are graph-like structures that allow edges (called "*hyperedges*" or "*hyperlinks*") spanning over more than two nodes. In a hypergraph there exist two types of relations, intra-group relations between nodes within a hyperedge and inter-group relations between nodes across hyperedges. In a quest to better understand, learn and infer such relations, in this thesis we propose novel approaches for hypergraph representation learning. In particular, we introduce a range of methods for structuring the representations and computations of deep neural network-based models on hypergraphs. Our hypergraph representations eventually allow for improved generalization in learning from multimodal data consisting of complex higher-order relations.

This thesis takes a broad view of representation learning on hypergraphs. We try to simultaneously learn about the content of an object stored as features on a node and higher-order relations among the objects represented by hyperedges. In particular, our focus is on the development of hypergraph learning frameworks that can capture the group relations on dynamically evolving real-world datasets. We seek to answer the following main research question:

How to learn higher-order relations using hypergraphs?

One of the earliest efforts in developing machine learning algorithms for hypergraphs came from Zhou et. al in 2007 [206]. They generalize the methodology of spectral clustering, which originally operates on undirected graphs, to hypergraphs and further develop algorithms for hypergraph embedding and classification. Recent advancements in the field of geometric deep learning [30] have presented formulations on graph structured data for the tasks of node classification [91], link prediction [200], or the classification of graphs [202]. Most of early methods do not generalize to the outlined problem of learning higher-order relations.

In this thesis, we argue for the introduction and design of deep learning models which can accurately learn higher-order relations within datasets represented using hypergraphs. Some of the major challenges in devising such a learning algorithm include extracting relational information from the complex hypergraph structure, combining content based information with the hypergraph structure, scalability to multiple modalities, adaptability to the dynamic nature of real-world datasets, and generalizability of the model across multiple data domains. We begin our research by examining the extent to which the structure of a hypergraph can capture higher-order relational information as compared to a graph. This leads to the first sub-question:

What information can be extracted from the hypergraph structure alone?

Hypergraphs have been shown to represent higher-order relations but their capability to capture and devise predictive modeling algorithms using these relations is yet to be

explored. In Chapter 2, we study the extent to which a hypergraph can be used to capture relational information among objects and how to utilize those relations to make useful predictions. We show that representing objects in a social network by mere pairwise relations tends to under exploit the underlying nuances, leading to substantial loss in information. In fact, much of the interesting information in social networks is captured by jointly observing the inter-group relations with intra-group relations. In this chapter, we design a framework that uses only the structure of a hypergraph to simultaneously capture these relations for performing several classification and recommendation tasks. This leads to the second research question:

What is the added value of content-based relations in deep learning on hypergraphs?

As outlined above, accurately learning the representation of an object requires modeling the content of an object and its relations. In Chapter 3, we define relations among objects by using the content comprising their intrinsic characteristics. These characteristics can be defined based on the available information of an object within a data collection. The underlying idea is to enhance a hypergraph model to learn heterogeneous relations among objects. For example, two social network users can be linked based on their friendship relations or based on their political views reflected in the content they interact with. We hypothesize that the flexibility of the hypergraph structure can facilitate designing deep learning models that can represent and learn such relations as well. Thus, we exploit content-based relations using hypergraphs to construct a learning framework. As a use case, this framework is used to make predictions on the communication behavior of users in an online discussion forum. However, these communication behaviors can evolve over time and thus can lead to structural changes in the hypergraph. Accommodating these changes into the model and making new inferences is of utmost importance in many applications. This leads to the third research question:

How to interactively add temporal structural changes to a hypergraph learning model?

One of the major challenges for learning on graphs and hypergraphs has been their rigidity in adapting to any structural change in terms of addition or removal of nodes and edges. These changes can be induced by temporal changes in the content or because of incremental insight in the collection by the user. In reality, relations between nodes change over time and, ideally, a deep learning model should be capable of adapting to the changes by learning these new sets of relations. To that end in Chapter 4 we present HYPER-MATRIX, a novel large-scale interactive hypergraph learning framework that enables temporal hypergraph exploration with a user relevance feedback model. This interactive learning approach is capable of incorporating user's responses into a pre-existing hypergraph learning model and providing relevant results in interactive time. Doing so, HYPER-MATRIX integrates user relevance feedback with a deep learning model, improving the quality of predictions over time. Such an approach enhances the learning framework to incorporate structural changes within a hypergraph based on external domain knowledge. However, such an approach possesses the problem of scalability to many modalities and to larger datasets due to the incremental matrix updation framework. This leads to the third research question:

Is it possible to scale deep learning on hypergraphs to datasets with many modalities?

While we were able to formulate a hypergraph learning framework, in order for multimedia analytics to be a true enabler of knowledge gain, we must address the problem of scalability to datasets with many modalities. These modalities can include, for example, image, video, audio, text (in the form of a comment, tag or post)or emoji. HYPERLEARN offers the possibility of deploying a hypergraph learning framework on highly multimodal data, alleviating the limitations on scaling the framework to many modalities. Moreover in a parallel computing setting, adding new modalities to the model requires only an additional computing unit keeping the computational time unchanged when such nodes are available, which brings representation learning to truly multimodal datasets. In Chapter 5, we demonstrate the feasibility of such a framework in experiments on multimedia datasets featuring higher-order relations. The relations and the number of objects in these datasets are static and do not change with time, making the entire learning framework transductive. In real world scenarios, however, there are temporal changes in a dataset leading to addition/removal of objects (nodes) as well as relations (edges). To perform inferences on unseen objects, an inductive learning framework is required. This leads to our final research question:

How can we develop deep learning models for dynamically evolving higher-order relations on hypergraph?

For long-term analysis of practially any real-world dataset, a learning framework should be adaptive to its dynamically evolving nature. This includes not only the structural changes induced by modifying the relations between objects, but also the addition of an increasing number of new objects over time. Thus, making inferences on these never encountered objects has become a priority in many applications. The final chapter of this thesis addresses the challenging problem of inductive learning on dynamic hypergraphs which aims at making predictions on unseen nodes and relations. In Chapter 6, we propose HYPERMSG comprising a message passing strategy to accurately and efficiently propagate information through the hyperedges, thereby learning the higher-order relations. HYPERMSG is scalable and generalizable to datasets from domains ranging from citation networks and social multimedia networks to brain activity networks in neuroscience.

The combination of the answers to the research questions yields a complete framework enabling the machine to interactively assist the user in multimedia analytics on even very large and heterogeneous collections.

LIST OF PUBLICATIONS

This thesis is based on the following publications:

- **Chapter 2** is based on "Exploiting relational information in social networks using geometric deep learning on hypergraphs", published in *International Conference*

on Multimedia Retrieval (ICMR), 2018 [12], by **Devanshu Arya** and Marcel Worring.

- **Chapter 3** is based on "Predicting Behavioural Patterns in Discussion Forums using Deep Learning on Hypergraphs", published in the conference *IEEE Content Based Multimedia Indexing (CBMI)*, 2019 [11], by **Devanshu Arya**, Stevan Rudinac and Marcel Worring.

- **Chapter 4** is based on "Visual Analytics for Temporal Hypergraph Model Exploration", published in *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2020 [54], by Maximilian T Fischer, **Devanshu Arya**, Dirk Streeb, Daniel Seebacher, Daniel A Keim, Marcel Worring.

- **Chapter 5** is based on "HyperLearn: a distributed approach for representation learning in datasets with many modalities", published in *ACM International Conference on Multimedia (ACMMM)*, 2019 [10], by **Devanshu Arya**, Stevan Rudinac and Marcel Worring.

- **Chapter 6** is based on "Adaptive Neural Message Passing for Inductive Learning on Hypergraphs", under review at *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2021 [8], by **Devanshu Arya**, Deepak K. Gupta, Stevan Rudinac, Marcel Worring.

For all chapters except Chapter 4, the ideas, text, figures, and experiments originate in majority from the first author. All other authors had important advisory roles, helped with running experiments and/or directly contributed in writing to a small number of individual sections of the above listed papers. For Chapter 4, the work was a collaboration with University of Konstantz where the author of this thesis was responsible for developing the interactive hypergraph learning framework. The visualization of this framework was carried out by University of Konstantz.

During his PhD, the author has further contributed to the following publications:

- "Fusing Structural and Functional MRIs using Graph Convolutional Networks for Autism Classification", published in *Medical Imaging with Deep Learning (MIDL)*, 2020 [9], by **Devanshu Arya**, Richard Olij, Deepak K Gupta, Ahmed El Gazzar, Guido van Wingen, Marcel Worring and Rajat Mani Thomas.

- "Livestock Monitoring with Transformer", published in *British Machine Vision Conference (BMVC)*, 2021 [167], by Bhavesh Tangirala, Ishan Bhandari, Daniel Laszlo, Deepak K Gupta, Rajat M Thomas and **Devanshu Arya**.

- "Rotation Equivariant Siamese Networks for Tracking", published in *IEEE CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021 [65], by Deepak K Gupta, **Devanshu Arya** and Efstratios Gavves.

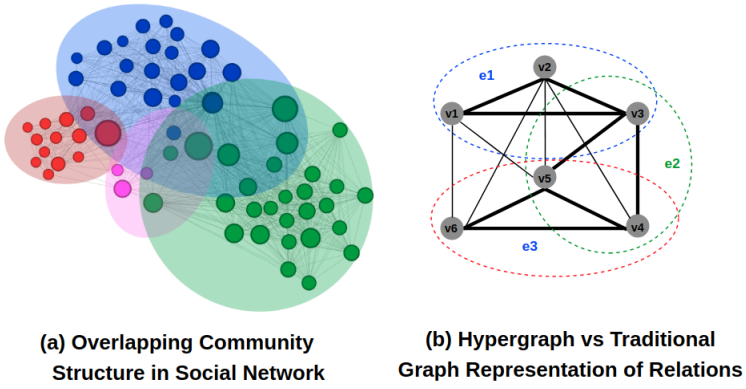
2

EXPLOITING RELATIONAL INFORMATION IN SOCIAL NETWORKS USING GEOMETRIC DEEP LEARNING ON HYPERGRAPHS

Online social networks are constituted by a diverse set of entities including users, images and posts which makes the task of predicting interdependencies between entities challenging. We need a model that transfers information from a given type of relations between entities to predict other types of relations, irrespective of the type of entity. In order to devise a generic framework, one needs to capture the relational information between entities without any entity dependent information. However, there are two challenges: (a) a social network has an intrinsic community structure. In these communities, some relations are much more complicated than pairwise relations, thus cannot be simply modeled by a graph; (b) there are different types of entities and relations in a social network, taking into account all of them makes it difficult to formulate a model. In this paper, we claim that representing social networks using hypergraphs improves the task of predicting missing information about an entity by capturing higher-order relations. We study the behavior of our method by performing experiments on CLEF dataset consisting of images from Flickr, an online photo sharing social network.

2.1 INTRODUCTION

The structure of an online social network contains an enormous amount of information within the intrinsic relationships among entities. Capturing implicit relations within these structure allows to perform tasks such as clustering, classification and link prediction. In order to extract these relational information sources, data representation plays a key role. In multimedia, the problem of learning one type of relations between entities to predict other types of relations has been a topic of significant interest. In particular, exploiting relations in online social networks, brings up the problem of generalization across different types of entities. The entities can be users in social communication networks (Facebook, Twitter), images/videos in media sharing networks (Flickr, Instagram), posts in discussion forums (Reddit, Quora) or resources in 'sharing economy' networks (Airbnb, Uber). There exists a multitude of relations within these social networks, even for a small dataset which reveals another hindrance to extract meaningful information. One of the key solutions for these problems is to design a model that can efficiently capture large amounts of relational information between entities, so one can perform tasks irrespective of any entity specific knowledge. Hence, there is a need for a representation which is scalable and a formulation which is neutral to all kinds of social networks.



(a) Overlapping Community Structure in Social Network

(b) Hypergraph vs Traditional Graph Representation of Relations

Figure 1: Figure (a) represents the overlapping community structure within a social network with the size of node proportional to its degree. As can be seen, the higher degree nodes forms a sub-unit to which other low degree nodes are densely connected. Figure (b) shows a comparison between the hypergraph and traditional graph representation. The higher-order relation between vertices cannot be captured using the pairwise edges. However, it can be easily captured using the hyperedges e1,e2 and e3.

Traditional graph-based representations of a social network leads to a loss in information, as it implicitly takes into account only pairwise connections between the entities. These pairwise relations fail to represent higher-order relations among the entities. Moreover, a simple binary representation of relations does not depict a collective flow of information. For instance, consider Twitter which has users, tweets, hashtags, lists etc.. Representing even a ternary relation of a user releasing a tweet containing multiple hashtags is infeasible using traditional graphs [100]. Or even in the simple case of a co-authorship network, one cannot know whether three or more authors that link together in the network were co-authors of the same paper or not [206] as seen in Fig.21(b). Thus to capture higher-order relations in social networks, traditional graph-based representation of the network proves to be insufficient.

Another characteristic of social networks is the presence of overlapping communities [113]. Most importantly a social network possesses the special property of being a scale-free network [123] [126]. Scale-free networks are a class of power-law networks where the nodes that have many connections (high-degree nodes) tend to be connected to other nodes with many connections, while they are surrounded by many small clusters of low-degree nodes. In other words, social networks contain a structure of communities, where smaller communities in the network are joined to larger communities by highly connected nodes that play the role of local hubs [140]. Graphically, these communities form subunits within the network which show relatively high levels of connection within them and a lower connectivity among them as seen in Fig.21(a). This implies that high-degree nodes in the core of a subunit are crucial for an efficient flow of information and to maintain strong connectivity in these networks. To efficiently capture the relational information within a social network one needs to exploit this dense overlapping community structure.

Various approaches have been proposed in the past to exploit relational information in social networks [119] [36] [131] [81] [115]. However, they do not fully capture the structural features shared within the overlapping community structure of the entities. In

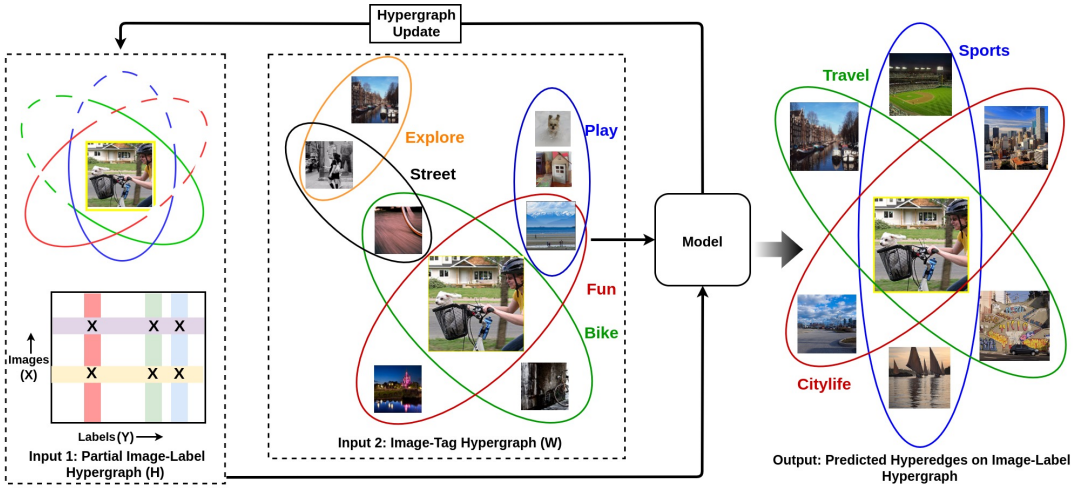


Figure 2: An example of the proposed method on the CLEF dataset of Flickr. The goal is to predict any type of metadata for images given two sets of input. Input 1 is the partial information about the images in one of the metadata spaces represented by an incomplete hypergraph (implying an incomplete incidence matrix). Input 2 is the complete information for the images in the other metadata space. Finally, the output is the set of predicted hyperedges between the images in the partial metadata space.

order to utilize this property, in this paper we represent social networks as hypergraphs where each entity is represented as a set of vertices and the edges represent the overlapping relations between them. A hypergraph [25] is a generalization of the simple graph in which the edges, called hyperedges, are arbitrary non-empty subsets of the vertex set and may therefore connect any number of vertices. The hyperedges form the key difference between a hypergraph and a traditional graph. The nodes are kept the same as in a simple graph but a hyperedge can connect even all the nodes at once as compared to a traditional graph where an edge is always a connection between 2 nodes. Especially, a set of multimodal entities in a social network can be viewed as a hypergraph whose vertices are the individuals and whose hyperedges are the communities. Finally, a hypergraph representation can be computationally advantageous as compared to the simple graph model since the incidence matrix of a hypergraph requires less storage space in depicting the same volume of information [185]. In this way, a hypergraph is a natural framework to capture the community structure as well as higher-order relations between nodes in the network.

In order to infer relations between entities using only its relational information requires not only a good representation of data, but also a robust powerful model to integrate them. In this work, we propose a methodology which can perform multiple tasks and can be generalized to all social networks. We develop a model to predict missing information (metadata) about an entity by learning relations between entities, without requiring any content-specific features of the entity. To build a multi-functional model for predicting missing metadata, we introduce a multi-graph convolutional neural network model for hypergraphs based on the recent works in deep learning on graphs, specifically graph convolutional networks.

Deep convolutional neural networks [94] have been proven to offer an efficient framework to extract deep meaningful statistical patterns in signals like image, speech or

video, in which there is a latent Euclidean structure. However, most of the definitions of convolution, utilize the properties of stationarity and locality which holds for Euclidean data spaces. Recent works [48] [91] on *geometric deep learning* aims to extend the framework of convolutional neural networks to data represented on graphs. The key idea in geometric deep learning is to devise a method for representation learning that can capture structural information within non-Euclidean domains, especially graphs. The applications of graph convolutional network ranges from describing shapes in different human poses [150], semi-supervised classification of authors in citation networks [91] and learning molecular fingerprints [43]. However, regular graph CNNs provide only a partial solution for learning dense information within a group of nodes. The two basic drawbacks of such models in a social network scenario have been due to the compact structure of data which makes the model unscalabale and the inability to share relational information across entities. Hence, we focus on developing a framework that can represent the scale-free properties of a social network and is generalizable to all social networks.

The points below highlight the contributions of this paper:

- We propose a generic framework which can transfer relational information from one type of relation to predict other types of relations between entities. Our approach is entity independent and captures higher-order relations by using hypergraph-based representation of a social network.
- We formulate a model for geometric deep learning on hypergraphs to perform tasks such as multi-label classification, link prediction and recommendation. Our results shows significant improvement as compared to previous graph-based methods.
- We further establish that a hypergraph-based representation of a social network is the most efficient way to build a model for learning the same volume of information in a network as compared to traditional pairwise simple or weighted graphs.

2.2 BACKGROUND

In this section, we introduce some background on the three concepts on which we base our methodology i.e. hypergraphs, matrix completion and geometric deep learning.

2.2.1 *Notation and Formulation of a Hypergraph*

A hypergraph G is formally represented as $H = (V, E)$, where V is a set of vertices and E is a set of hyperedges where each $e \in E$ is a subset of V. The *degree* of a hyperedge e, denoted as $\delta(e)$, is the number of vertices in e. In case of a simple graph, $\delta(e) = 2$ and hence they are known as "2-graph". The diagonal matrices containing the degrees of all vertices (v) and hyperedges (e) are denoted by D_v and D_e respectively. We say that there is a *hyperpath* between vertices v_1 and v_k when there is a sequence of distinct vertices and hyperedges $v_1, e_1, v_2, e_2, ..., e_{k-1}, v_k$ such that $\{v_i, v_{i+1}\} \subseteq e_i$ for $1 \leq i \leq k-1$. One of the key differences of hypergraphs as compared to pairwise simple graphs is its representation using an incidence matrix. It is represented by a $|V| \times |E|$ matrix H with

entries $h(v,e) = 1$ if $v \in e$ and 0 otherwise. A simple graph is commonly represented by a square $|V| \times |V|$ matrix A which is known as the adjacency matrix, such that its element a_{ij} is 1 when there is an edge from vertex i to vertex j and 0 when there is no edge. There are many advantages of the incidence matrix (H) over the adjacency matrix (A) to model relational data [185]. The three key advantages are: (i) the incidence matrix of the hypergraph requires less storage space in comparison with the graph adjacency matrix to represent the same volume of information; (ii) hypergraph incidence matrices require fewer operations for matrix-vector multiplication; and (iii) most importantly, the benefits of using the Laplacian of a hypergraph incidence matrix which will be discussed in detail in section 2.4.2.

2.2.2 *Matrix Completion*

Matrix completion is the task of finding the missing values of a partially observed $p \times q$ matrix M. That is, we only observe a sparse set E of observations $M_{i,j} : \forall (i,j) \in E$, with $|E| << pq$. The goal is to estimate the rest of the values $M_{i,j} \notin E$. A particularly popular model is to assume that the values lie in a smaller subspace, resulting in M being a low-rank matrix, which leads to solving a rank minimization problem. Let $\Xi(\bullet)$ be the projection operator selecting only those entries that lie in the set E and let R be the target matrix to be reconstructed using $M_{i,j} : \forall (i,j) \in E$. Then the rank minimization problem is given by:

The number of unknown variables in this formulation are in the order of $p \times q$ which makes it practically unscalable for large matrices. One of the solutions is to use a factorized representation of matrix R i.e. $R = XY^T$, where X, Y are $p \times r$ and $q \times r$, matrices respectively with $r << min(p,q)$, which formulates as eq.6.1 [154].

Low-rank further implies linear dependence of rows/columns of M which can be utilized to constraint the space of solutions to be smooth. In many scenarios, the rows/columns form communities which can further optimize computation by incorporating proximity information among rows/columns. Recent work on geometric matrix completion has shown the importance of these relations by using them as side information to the matrix completion problem [82] [125] [24] [138]. It assumes that there exists a graph $G_r = (V^X, E^X)$ whose adjacency matrix encodes the relationships between the p rows of X and a graph $G_c = (V^Y, E^Y)$ for q rows of Y. The geometric matrix completion can then be written as

$$\min_{X,Y} \frac{1}{2}\|X\|^2_{G_r} + \frac{1}{2}\|Y\|^2_{G_c} + \lambda_{X,Y}\|\Xi(M - XY^T)\|^2_F \tag{2.1}$$

where, $\|X\|^2_{G_r} = trace(X\Delta_r X^T)$ and $\|Y\|^2_{G_c} = trace(Y\Delta_c Y^T)$ are the graph Dirichlet semi-norm for rows and columns respectively. Δ_r and Δ_c are the row and column laplacian matrices.

2.2.3 *Geometric Deep Learning*

As defined by Bronstein et.al. [30]; "Geometric deep learning is an umbrella term for emerging techniques attempting to generalize (structured) deep neural models to

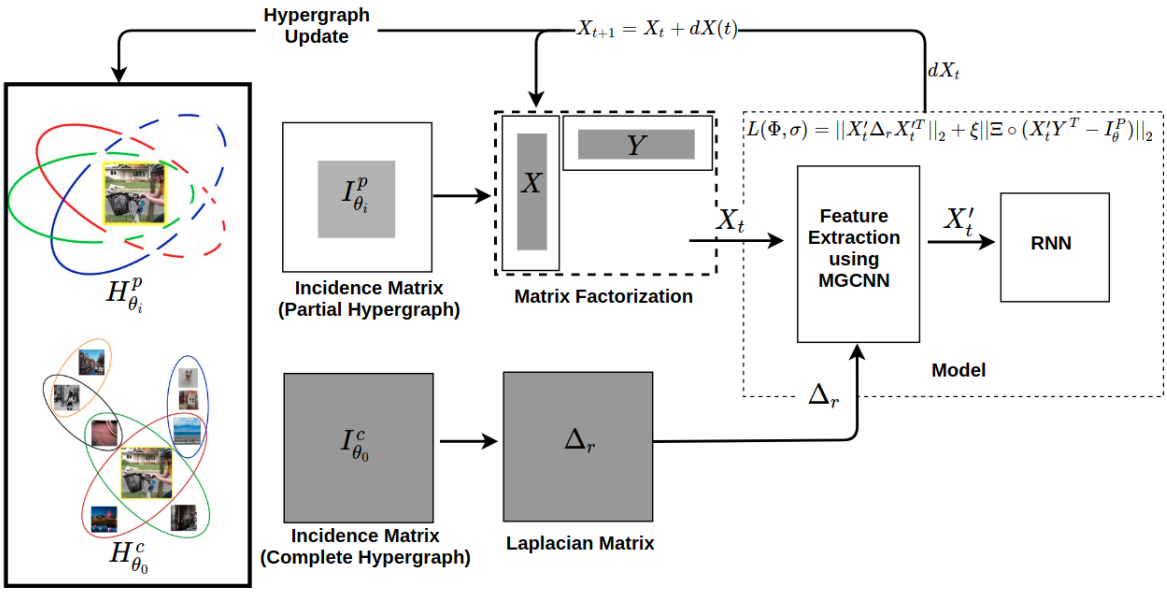


Figure 3: A block-diagram of the proposed model. The input to the model are the partial $H^p_{\theta_i}$ and complete $H^c_{\theta_0}$ hypergraphs corresponding to the two types of metadata θ_i and θ_0 respectively. $I^c_{\theta_0}$ and $I^p_{\theta_i}$ are the incidence matrices (the grey color depicts data used for training whereas white represents missing data). The model updates the hypergraph incrementally by updating the incidence matrix using geometric deep learning based model.

non-Euclidean domains such as graphs and manifolds". One of the early attempts on generalizing neural network to graphs are due to Scarselli et.al. in 2005 [63], who proposed a combination of recurrent neural networks and random walk models called Graph Neural Networks (GNN). The first formulation of convolution neural networks on graphs used the definition of convolutions from graph signal processing in the spectral domain [31].

In this work, we focus on applying convolution network on graphs in order to learn the intrinsic relations in social networks. A convolutional layer in the spectral domain is defined as

$$f_l^{out} = \xi\Big(\sum_{l'=1}^{p} \Phi_k \hat{G}_{l,l'} \Phi_K^T f_{l'}^{in}\Big) \tag{2.2}$$

where, $F_{in}^{n\times p} = (f_1^{in}...f_p^{in})$ and $F_{out}^{n\times q} = (f_1^{out}...f_p^{out})$ represent the p and $q-$dimensional input and output signals on the vertices of the graph. Φ_k is the $n \times k$ matrix of the eigenvectors from the spectral decomposition of the graph. $\hat{G}_{l,l'}$ are the learnable spectral filters and ξ is the ReLU non-linearity. Further advancement to this definition have been proposed in order to make a *Graph Convolution Network* (GCN) generic and scalable [71] [43] [91].

2.3 RELATED WORK

We review two categories of related work: studies on context/network based learning of relational information in social networks and applications of deep learning approaches on graphs.

2.3.1 *Learning Relational Information in Social Networks*

Several approaches have been introduced for learning relations within a social network. They can be grouped in five main categories based on their representation of social network data.

In the first one, [110] proposed one of the earliest approaches where they use network topology in which, they model a social network as a simple homogeneous graph where each node represents an entity and each link denotes social relationships. [179] proposed Link Prediction using Social Features ($LPSF$), based on features extracted from patterns of prominent interactions across the network for each entity pair. These features are very useful in identifying similar node pairs, even when they are far away. They propose a simple yet powerful model to capture relations between entities. However, these approaches are not made generic across all networks and lack scalability due to their dependence on pre-defined methods for feature extraction.

In the second type of approach [196] proposed to model social networks as pairwise heterogeneous graphs as opposed to homogeneous ones and apply a random walk algorithm to calculate link proximity. Online photo sharing networks have been of particular interest in learning relations due to the generation of large amounts of metadata. [119] proposed a graphical model that treats image classification as a problem of simultaneously predicting binary labels for a network of photos. They represent each image by a node while the edges are formed between images that have some common property. The first two approaches model social relations between entities as pairs and then apply a structural learning algorithm. These approaches can be scalable to large networks but they still fail to capture any higher-order relations. Therefore, they cannot make use of the community structure in social networks leading to loss in information. Moreover, their applicability to real-world networks is confined as they use parametric methods for modeling relations [81].

The third category of approaches represents data on an ego network, which consist of a focal node (”ego”) and the nodes to whom ego is directly connected to (”alters”). Egos and alters are tied to each other by social relations, in [45] and [99], the authors propose to learn social circles by representing the data in ego networks. Li et. al. [104] further study the problem of profiling user attributes in social networks by capturing the correlation between attributes and social connections in an ego network. These approaches however does not generalize for all types of social networks and to learn all kinds of relations between entities and metadata.

Another kind of approaches are based on hypergraph theory. Hypergraph-based models have been widely used in the multimedia domain for solving the problems of community detection [205] [111], multi-label classification [35] [160], tag-based social image searching [58], music recommendation [32] and link prediction in social

networks [100]. In this work, we use hypergraph-based approach to represent data, for precisely capturing the high-order relations, in order to build a generic framework for classification, recommendation and link prediction.

Several "non-graph" based approaches to exploit relational information across domains have also been a field of particular interest. Earlier works on multi-domain collaborative filtering includes interaction-associated information of users and items as side information for recommendation. Cross-domain collaborative filtering (CDCF) [73] has recently started to draw significant research attention. The basic concept of CDCF is to borrow rating knowledge for each user from some related auxiliary domains, whose rating matrices are relatively dense, to alleviate the rating sparsity problem in the sparse target domain. These approaches rely mostly on implicit domain correlations that are mined solely from user preference data, and and no explicit links are exploited. There are two major questions surrounding this approach [148]. First, what could be the common knowledge that can be transferred/shared between different domains, and, second, what could be the optimal way to transfer/share knowledge between different domains [133].

2.3.2 *Application of Deep Learning based Approaches on Graphs*

There has been a recent surge of interest to formulate deep learning methods on non-euclidean domain especially in graphs. The effectiveness of deep learning graph-based approaches ranges from computer graphics [29] to chemistry [48]. The spectral graph convolutional neural networks (*GCN*), originally proposed in [31] and extended in [43] have proven effective in classification of handwritten digits and news texts. [91] proposed a simplified *GCN* for semi-supervised classification of authors in a citation network. In the computer vision community, *GCN* has been extended by [118] to describe shapes in different human poses, [150] to demonstrate classification of point clouds and [124] for image and 3D shape analysis. In multimedia, [143] proposed an approach to categorize user posts for political extremism content based on their discussion topics. Deep learning on graphs for social networks is yet to be explored for their ability to uncover hidden relations between multimedia items. In this paper, we take a step further to devise a generic model for learning relations in social networks using geometric deep learning methods.

2.4 PROPOSED MODEL

In this work, we define a trainable graphical model that treats predicting metadata for an entity, as the unified problem of generating sets of hyperedges across entities. The basic hypothesis of the model is that entities related through one set of metadata carry imperative information which can be learnt to predict other relational properties between them. In this paper, we will use $H_{\Theta}^{p/c}$ (with $I_{\Theta}^{p/c}$ as its incidence matrix) to denote a partial(p) or a complete (c) hypergraph. The subscript Θ is the type of metadata used to construct the hyperedges i.e. $\Theta = t/l/g/u$ for tags (t), labels (l), groups (g) and users (u) respectively. The inputs to the model are: (a) a complete hypergraph of entities constructed using one type of metadata denoted by $H_{\Theta}^{c}$and (b) a partial hypergraph on the same sets of entities constructed from the required metadata denoted by H_{Θ}^{p}. The

training of the model has three phases: constructing the model by formulating it as a factorized matrix completion problem, relational feature extraction using geometric deep learning and finally updating the partial hypergraph by predicting hyperedges across entities. Fig.3 shows these three phases as a block diagram.

2.4.1 *Formulating Hyperedge Prediction as Matrix Completion*

The computational advantage of using a hypergraph for the above mentioned problems instead of a simple graph is the representation of its vertices and edges by an incidence matrix. As compared to traditional graphs where the incidence matrix has an additional constraint of only two non-zero values in each column ("2-graph" property), the incidence matrix of a hypergraph can have as many as all non-zero values in each column. Therefore, generating hyperedges in a hypergraph can be termed equivalent to the problem of filling missing entries in its corresponding incidence matrix. In this work, we represent the relation between entities using their metadata by hypergraphs. Each entity corresponds to a vertex and the edges depict all unique values of the corresponding metadata. The respective incidence matrix ($I_{\Theta}^{p/c}$) is of dimension $n \times \theta$, where n is the total number of entities and θ are the unique values corresponding to the metadata Θ. Hence, the problem of predicting hyperedges between entities in H_{Θ}^{p} reduces to completing the incidence matrix I_{Θ}^{p} with multiple missing entries corresponding to each column and at least one known entry in each row, where each row is an entity and the columns are values of the metadata.

$$\min_{X} \|X\|_{H_{\Theta}^{c}}^{2} + \lambda_X \|\Xi \circ (I_{\Theta}^{p} - XY^{T})\|_{F}^{2} \tag{2.3}$$

2.4.2 *Feature Extraction using Multi-Graph CNNs on Hypergraph*

The second phase of our model aims at jointly extracting features from H_{Θ}^{c} and H_{Θ}^{p}. In this way, we can transfer the relational information from the complete hypergraph H_{Θ}^{c} to predict the missing hyperedges in H_{Θ}^{p}. In this paper, we devise our solution based on recent work on multi-graph convolution (*MGCNN*) [125]. It uses the formulation for GCN using recurrent Chebyshev polynomials which simplifies eq.2.2 [43]. The motivation behind multi-graph convolution is that, a Fourier transform of a 2-dimensional signal can be simplified by formulating it as applying a one-dimensional Fourier transform to its rows and columns. In particular, multi-graph convolution proposes a method of matrix completion, given the rows and columns of a matrix possess relational information within themselves. In our framework, we extract features combining I_{Θ}^{p} and I_{Θ}^{c} by stacking multi-graph CNN layers given by

$$X_t' = \sum_{j=0}^{q} \Phi_j T_j(\Delta_r) X_t \tag{2.4}$$

where Φ_j are the learnable filter coefficients, $\Delta_r^{n \times n}$ is the row-hypergraph Laplacian and T_j is the representation of filters using Chebyshev polynomials. In this way a multi-graph CNN on $X_t^{n \times q}$ with a single channel produces a k dimensional output $X_t'^{n \times q \times k}$.

	Task1	**Task2**	**Task3**	**Task4**
θ_0	Tags (t)	Tags(t)	Tags (t)	Labels (l)
θ_i	Labels (l)	User (u)	Groups (g)	Tags (t)
$\|\|rel_{\theta_i}\|\|$	613,014	51,804	70,226,414	91,485,864
$\|\|rel(H^c_{\theta_0})\|\|$	45,766	45,766	45,766	55,396
$\|\|rel(G^c_{\theta_0})\|\|$	85,802	85,802	85,802	95,766

Table 1: Table showing the details about the 4 tasks. The goal is to predict relations given a partial set of rel_{θ_i} and complete set of relations represented on hypergraph ($rel(H^c_{\theta_0})$) or on simple and weighted-graph ($rel(G^c_{\theta_0})$).

The other advantage of the above formulation is the use of the Laplacian to encode information from data defined on hypergraph H^c_Θ. The Laplacian matrix of a hypergraph has been shown to be useful for learning higher-order relations [1] [160], spectral clustering of edges [206] and to measure the relatedness between two entities [32]. In this paper, we use the normalized hypergraph Laplacian matrix (Δ_r) [206] given by where D_v and D_e are the vertex and edge degree matrices of hypergraph H^c_Θ respectively, $\mathbb{I}$ is the identity matrix and I^{cT}_Θ is the transpose of incidence matrix I^c_Θ. The Laplacian will be used for incorporating the structure of the hypergraph H^c_Θ in eq.2.4. In this way, we extract relational features using combined information from the complete and the partial hypergraph.

2.4.3 *Incremental Updates of the Hypergraph*

The next step is to diffuse the features extracted by coupling the structures of the two hypergraphs (H^c_Θ and H^p_Θ). The partial hypergraph is updated incrementally as a consequence of the completion of its incidence matrix. We use a Recurrent Neural Network (RNN) [72] to predict small incremental changes (dX) to the matrix X [125]. One of the main advantages of using an RNN for predicting accurate small changes is its ability to store information for longer temporal steps. The model is finally trained by feeding the features extracted from multi-graph CNN (X'_t) to an RNN and perform training by using the minimization eq.2.3 in geometric matrix completion as the loss function.

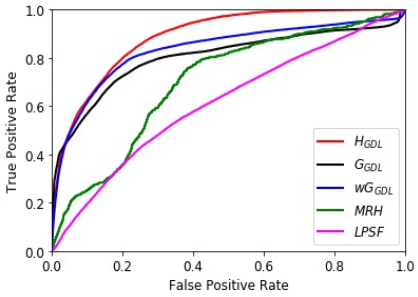
$$\mathscr{L}(\Phi,\sigma) = \|X'_{t,\sigma}\Delta_r {X'_{t,\sigma}}^T\|_2 + \lambda_X\|\Xi \circ (X'_{t,\sigma}Y^T - I^P_\theta)\|_2 \tag{2.5}$$

where $X't$ is the feature extracted by multi-graph convolution with Φ as the learning coefficient, σ denotes the parameter for RNN and the subscript t denotes the number of diffusion iterations.

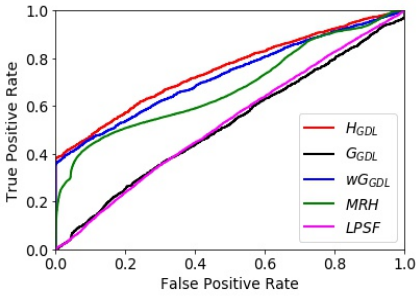
2.5 EXPERIMENTS

In this section, we perform extensive experiments to show the advantages of learning higher-order relations in a social network using geometric deep learning on hypergraphs as compared to other approaches. We design our experiments to investigate the following:

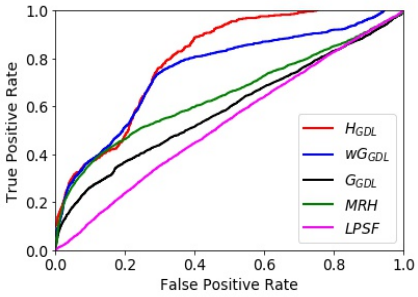
- Performance of the proposed generic framework to predict multiple types of relations between entities



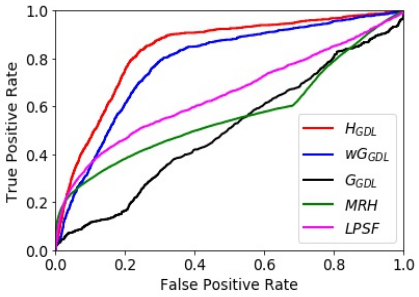
(a) Task1: Multi-Label Image Classification



(b) Task2: Image-User Link Prediction



(c) Task3: Group Recommendation



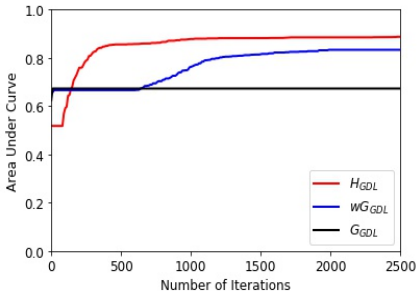
(d) Task4: Tag Recommendation

Figure 4: Experiment 1 - Receiver Operating Characteristics (ROC) curve showing the performance of the models on each of the 4 tasks. The hypergraph-based geometric deep learning model (H_{GDL}) has significant advantage as compared to other methods.

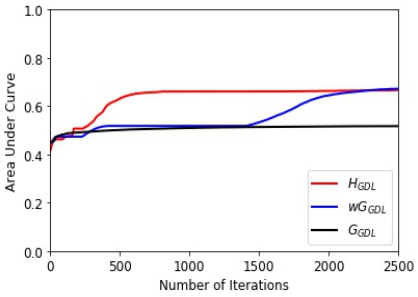
- Advantages of using geometric deep learning over existing simple graph as well as hypergraph-based learning
- Efficiency in representing relational information of a network using hypergraphs as compared to pairwise simple graph representation

To evaluate our model, we explore the online photo sharing social network Flickr, which generates a huge amount of metadata and hence relations for each image. Flickr has been particularly very popular in using social network metadata for image classification among other implications [81] [119]. The metadata, such as user-generated tags and community-curated groups in Flickr are used by people as a means to communicate with other people, and as a means to describe the image and its location. But not every image is annotated with all the information, hence using relational information can be highly informative in unveiling the missing information of every image.

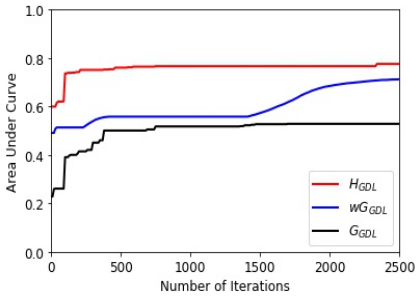
Data Setup For our experiments, we study the CLEF dataset [119] comprising of images from Flickr which has social network metadata and has labels provided by human annotators for each image. The dataset consists of 4,546 images with 99 labels (l), 21,192 tags (t), 10,575 groups (g) and 2,663 users (u). We show that our framework can be used as a generic multi-functional setup for generating information for an image by performing 4 types of tasks using our model: $Task1$: Multi-Label Image Classification, $Task2$: Image-User Link Prediction, $Task3$: Group Recommendation and $Task4$: Tag Recommendation. Given a set of known-metadata (θ_0) for each image, we first construct the complete hypergraph $H^c_{\theta_0}$. Our goal is to predict other sets of partially known metadata (θ_i) associated with the images.



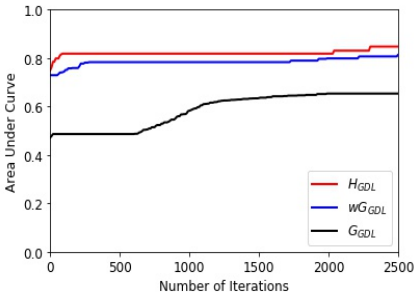
(a) Task1: Multi-Label Image Classification



(b) Task2: Image-User Link Prediction



(c) Task3: Group Recommendation



(d) Task4: Tag Recommendation

Figure 5: Experiment 2 - Figure showing the rate of learning with each iteration of the proposed model using hypergraph (H_{GDL}), weighted graph (wG) and simple graph (G). As can be seen, the hypergraph-based model converges faster for all the 4 tasks implying a better representation to learn relational information.

Training The total number of relations, $\|rel_{\theta_i}\|$ between the images and the target metadata (θ_i) is tabulated in Table2. As seen from the table, each image has multiple values of metadata in common with other images, resulting in a multitude of relations. We randomly sample 40% of these relations and keep them aside to use them as test set. The remaining relations are used to construct the partial hypergraph $H^p_{\theta_i}$ for training the model along with the complete hypergraph $H^c_{\theta_0}$.

Evaluation To show the efficiency in representing social network information with a hypergraph, we compare our result with the data represented by hypergraph (H), simple graph (G) and weighted graph (wG) using the same model. Simple graph (G) indicates a binary relation between entities with a value 1 if the two entities share at least one common value of the metadata. The weighted graph (wG) is constructed by assigning weights equal to the count of values of the metadata common between two entities. The hypergraph-based representation reduces the total number of relations between entities and the known metadata significantly by representing higher-order relations as community. This can be seen from Table2 where $\|rel(H^c_{\theta_0})\|$ and $\|rel(G^c_{\theta_0})\|$ denote the total number of relations in a hypergraph and graph based representation respectively.

We evaluate the performance of our geometric deep learning based model as compared to the previous hypergraph based algorithm (MRH) [32] [100] and a graph-based model trained on social network features ($LPSF$) [179] [180] for the same tasks. $LPSF$ as mentioned under the first approach in section 2.3.1, trains a neural network on popular features like Page Rank, Number of Common Neighbors, Preferential Attachment etc. extracted from a social network. We use the notation H_{GDL}, wG_{GDL} and G_{GDL} for our

geometric deep learning (*GDL*) model on hypergraph, weighted graph and regular graph respectively.

2.5.1 *Experiment 1: Learning Relational Information in a Social Network*

We start our experimental evaluation by showing the performance of our model, *MRH* and *LPSF* on the 4 tasks. To evaluate the performance of our model and show its advantages over other methods, we show the Receiver Operating Characteristic (ROC) curves for each tasks. The ROC curve depicts how well a model is able to predict the presence/absence of a relation among images with the corresponding metadata. Fig.8 shows the performance of the models on the 4 tasks. The Geometric Deep Learning based approach outperforms existing hypergraph-based *MRH* and graph-based *LPSH* methods in all the 4 tasks. This confirms, the significant advantage of using a hypergraph representation of the network as compared to simple and weighted graphs using the same model. Most importantly, this proves the advantage of learning relations using geometric deep learning techniques as compared to existing hypergraph-based model.

2.5.2 *Experiment 2: Measuring the efficiency of representing data from social network*

To explore the advantage of representing a social network using hypergraphs as compared to traditional graphs, we evaluate their efficiency in learning relational information. We compare the rate of convergence of our algorithm on the three graph frameworks mentioned above i.e. hypergraph (H), weighted graph (wG) and simple graph (G) on the 4 tasks. The faster the algorithm converges, the better the framework is in capturing the same volume of relational information. We plot the area under the ROC curve against the number of iterations used to update the matrix incrementally. As can be seen from Fig.5, the hypergraph-based representation converges faster than simple and weighted graphs for all the 4 tasks. This concludes the efficiency of a hypergraph in capturing information which makes it the best choice to represent data on a social network.

2.6 CONCLUSION

In this paper, a generic method to exploit relational information between entities in a social network for predicting missing information about an entity has been presented. In contrast with traditional graph representation, we model a social network using hypergraphs. We show the importance of using hypergraphs in order to capture all types of entities and either the pair wise or high-order relations among them to avoid loss of any information. Moreover, our approach is content independent i.e. it does not depend on any entity-specific information and hence can be generalized to all types of social networks. We formulate the learning problem as matrix completion on graphs and extend the methods on geometric deep learning to hypergraphs. We evaluate our model on 4 tasks: multi-label image classification, image-user link prediction, group and tag recommendation in a Flickr dataset. Experimental results show a significant advantage in representing social networks by hypergraphs and using deep learning based method for exploiting relational information within the network. We also prove the computational

effectiveness of representing the same volume of information from a social network on a hypergraph as compared to the traditional pairwise graphs.

3

PREDICTING BEHAVIOURAL PATTERNS IN DISCUSSION FORUMS USING DEEP LEARNING ON HYPERGRAPHS

Online discussion forums provide open workspace allowing users to share information, exchange ideas, address problems, and form groups. These forums feature multimodal posts and analyzing them requires a framework that can integrate heterogeneous information extracted from the posts, i.e. text, visual content and the information about user interactions with the online platform and each other. In this paper, we develop a generic framework that can be trained to identify communication behavior and patterns in relation to an entity of interest, be it user, image or text in internet forums. As the case study we use the analysis of violent online political extremism content, which has been a major challenge for domain experts. We demonstrate the generalizability and flexibility of our framework in predicting relational information between multimodal entities by conducting extensive experimentation around four practical use cases.

3.1 INTRODUCTION

A large amount of visual and textual content is posted daily in different social networking and content sharing platforms, where users can express their thoughts and share experiences. Pervasive nature of internet and social media has not only made it possible to communicate and demonstrate radical views and intentions, but also to connect to other persons with similar interests. Due to this high reachability and popularity of social media, people also use these platforms for planning events and mobilizing others for protests, public demonstrations, promoting violent extremist ideologies, and spreading racist opinions. The problem of automatic identification of such online radicalization and prediction of social unrest is of paramount importance for law enforcement agencies. It requires collection, fusion and analysis of 'weak signals' or 'digital traces' which are present on social media. Current analysis techniques focus mostly on hashing and filtering of known extremist multimedia items. However, aiding domain experts in rigorous large-scale empirical analysis requires novel multimodal tools designed to handle unstructured data from diverse information channels, especially internet discussion forums.

Discussion forums are a type of social multimedia network where people can meet, form groups, discuss common interests and exchange ideas. Through the use of discussion forums, it is also possible for members of the public, whether supporters or detractors of a group, to engage in debate. This may assist the terrorist group in adjusting their position and tactics and, potentially, increasing their levels of support and general appeal [39, 61, 183]. This is also noted in Europol's annual terrorism situation

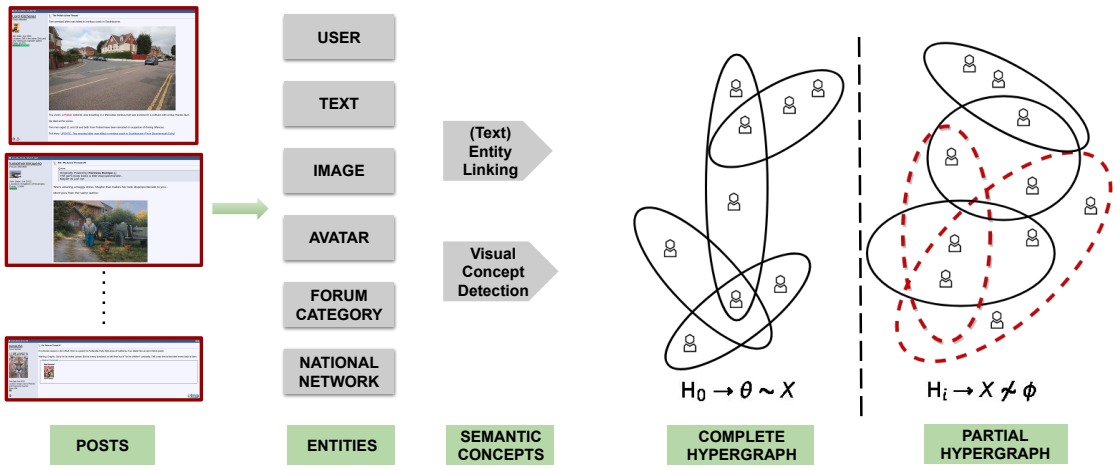


Figure 6: An example of the proposed pipeline of the framework. It represents a typical post with multimodal entities in Stormfront, a white nationalist, white supremacist and neo-Nazi Internet forum followed by the extraction of semantic concepts from the post and then construction of a hypergraph framework using the entities and concepts as proposed in [12]. H_0 and H_i represents complete and partial hypergraph. The goal is to predict missing information i.e. generate hyperedges (red-dotted line) on H_i.

and trend report for 2012, which warns that internet forums present effective means for addressing target audiences, and "recruiting" supporters with no off-line links to terrorist organizations [52]. By just analyzing the content after it has been shared in "extremosphere" of these forums can lead to a delay in detecting critical events, which can prove to be a massive loss in the future. Hence, an effective framework is needed for predicting future communication behaviour between users or communities within the (often implicit) social networks hosted by these forums. Major challenge in developing such framework is the presence of low quality content in contextual metadata and the large volume of information in internet forums.

In this paper, we construct a pipeline, as shown in figure 21, which can be used to identify communication behavior and patterns in violent online extremism forums. Our framework is built upon the methodology developed in [12] where the authors presented an approach to predicting links and groups between entities (which can be images, users, posts, groups etc.) within a social multimedia network such as Flickr. An entity can be any of the visible constituents of a post in social network. For example, in Instagram, entities consist of tags, image, video, user, location and caption. In this work, we extend the methodology proposed in [12] for the use with heterogeneous entities in discussion forums, which contain more unstructured information. It can enhance the Law Enforcement Agencies (LEAs) to exploit future interactions between entities within a network, for instances: (a) *mob formation* - which type of people or forums a particular user(s) might be interested in to interact with, (b) *deciphering hidden messages from an image* - what kinds of images can be associated with a post or (c) *content classification* - which type of category a post might belong to. These use cases requires analysis of posts at a particularly high semantic level. Hence, we focus on extracting semantic concepts, such as topics, personages, locations and gender from text using entity linking and visual concepts (such as TRECVID [152] and ImageNet [44] concepts) from images and videos.

The usage of such semantic concepts is important since the results are intended to be interpretable by the end users.

To facilitate such applications, we extract semantic (visual and text) concepts from the posts and then train a model based on relational information between entities in the social network. These relations can either be formed between entities of different modality or between entity and its metadata information. Uncovering hidden relations between multimedia items has long been a topic of research in multimedia information retrieval. Graph-based approaches have been prominent for their ability to represent and analyze such problems. Recent works [30] [48] [43] on geometric deep learning aim at formulating convolutional neural networks to data represented on graphs. The key idea in geometric deep learning is to devise a method for representation learning that can capture structural information within non-Euclidean domains, especially graphs. However, there are two major challenges for their wider adoption in learning multimodal relations in discussion forums. Firstly, graph-based approaches are hindered by the challenges related with associating and learning semantic concepts due to the presence of unstructured textual data and low quality images. In addition, inefficient representation of multimedia post can lead to loss of available information or capture it only partially.

The main contributions of the paper are:

- We present a framework, which combines semantic concepts and contextual relations between entities in a discussion forum to predict communication behavior and patterns in relation to various types of entities.
- We demonstrate the flexibility of such framework in tackling a variety of potential use cases arising during the analysis of violent political extremism forums. Multimedia analysis is further shown effective in aiding the domain experts involved in the qualitative analysis of these forums.
- Our experiments provide insights into the usefulness of the relations between individual modalities and semantic features, which can be exploited to unravel implicit information about diverse entities in a discussion forum.

The remainder of this paper is organized as follows. In Section 2 we provide an overview of related work. Then in Section 3 we introduce our approach and in sections 4 and 5 we present the experimental setup and results. Section 6 concludes the paper.

3.2 RELATED WORK

This section describes and discusses related work on the methods for learning patterns and behaviours of entities in a social network.

Understanding how users behave when they connect to social networking sites creates opportunities for richer studies of social interactions, better detection of irregular behavior and improved design of content distribution systems. Jin et al. [80] presented an elaborate survey on the importance of analysis and characterization of user behaviours in online social networks, highlighting the different perspectives that are shaping the ongoing work in the field. The need for analysis of user behaviors have now become even more interesting with the rise of social multimedia network. The presence of multimodal

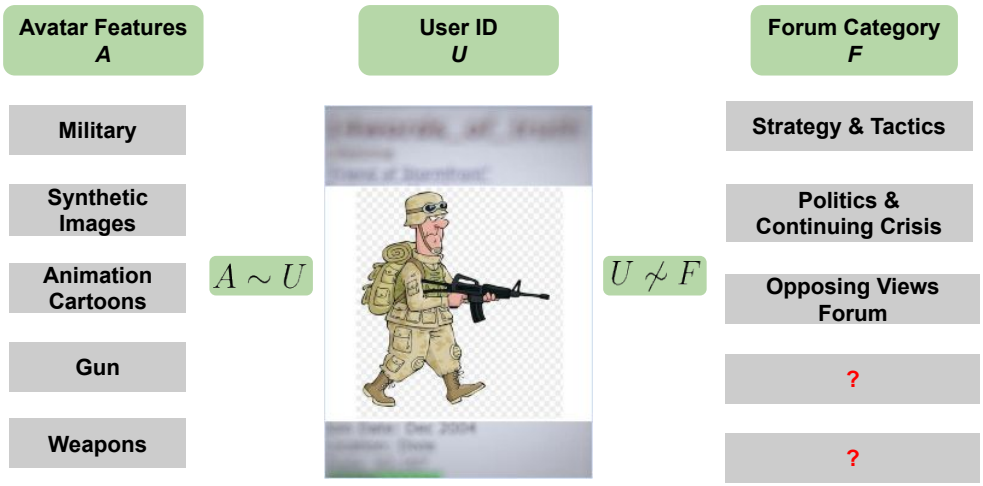


Figure 7: An example from Use Case 4, where the known information for a user is his/her User ID (U), Avatar Features (A) and partially known Forum Categories that he/she posted. The goal is to predict these unknown forum categories using the relations $A \sim U$ and $U \not\sim F$. Similar, examples can be drawn from the other use cases.

entities in social network, has shifted the primary focus of multimedia community to go beyond the structural analysis of network [179] [196] and towards the analysis of content at a higher semantic level [119].

Moreover, the advantages of combining semantic concepts with the network structure has made graph-based approaches very popular. The application of graph-based methods on social network varies from link prediction [45], discovering social circle in ego networks [99], music recommendation by combining social media information and music content [32] to categorizing violent online political extremism content [143]. Hypergraphs [25], in particular, were proven to be highly efficient in capturing relational information in multimodal social networks [12] [100] [58].

3.3 APPROACH

Given a multimodal post from an arbitrary discussion forum, our goal is to construct a framework that can encode semantic concepts about an entity in the form of relations, and then predict valuable information about them. Our framework can, in general, allow a user to perform either of two tasks: (1) predict implicit relations between multimodal entities or (2) extract additional semantic concept about entities. We combine these two tasks by using a common graph-based approach that can provide users with the flexibility to train a model according to their usage. Below we describe the three main facets of our formulation - representation, information flow and generalizability.

3.3.1 *Representation*

We represent information in a multimodal post using hypergraphs due to their numerous advantages over traditional graph based representation, one of them being efficient capture of higher-order relations [12] [206]. A hypergraph is a generalization of the

graph in which the edges, called hyperedges, are arbitrary non-empty subsets of the vertex set and may therefore connect any number of vertices. The nodes are kept the same as in a graph but a hyperedge can connect even all the nodes at once as compared to a traditional graph where an edge is always a connection between 2 nodes. In particular, a set of multimodal entities in a social network can be viewed as a hypergraph whose vertices are the individuals and whose hyperedges are the common properties between them [100].

The other advantage of using hypergraphs is the ease of modifying definitions of nodes and hyperedges. In this work, we will exploit this property in order to merge the two tasks mentioned in section 3.3. For both tasks, we construct a hypergraph in which its nodes represent the main entity (for which relations/information needs to be predicted) connected through hyperedges which can represent either entities from other modality or metadata information about the main entity. So, the problem reduces to that of generating hyperedges across the main entity which in-turn can represent (a) relations between multimodal entities or (b) relations between metadata and the main entity. In this way, we can devise a learning algorithm which can merge both tasks and is devoid of any loss in available information.

3.3.2 *Information Flow*

The next challenge is to construct a pipeline extracting entities and concepts from posts to then learn relational information. We aim at encoding information from the available set of relations for an entity to predict the unknown sets of relations. For extracting semantic information from a post, we employ entity linking for text, where the idea is to link the text to an external knowledge base such as Wikipedia [122, 143] and for visual concepts we extract 346 TRECVID semantic concepts [152]. Further, we use a robust model for extracting features and learn relational information [12] from the posts represented on hypergraphs. This model is based on geometric matrix completion solution initially proposed in [125]. We formulate the relation prediction task as a matrix completion problem, where rows and columns represent two separate entities. This matrix is derived from the incidence matrix of partial hypergraph, where the vertices forms the rows and the edges forms the columns. For example, the images posted in discussion forums will be represented on the rows of a matrix (and vertices of hypergraph) while the columns (corresponding edges in hypergraph) can be the forum categories. Thus, each entry of this matrix will have a binary value representing presence/absence of an image in a particular forum category. The aim is to complete this matrix using auxiliary information about images from the complete hypergraph. Hence, to extract combined relational features we use Multi-Graph Convolution Networks (MGCNN). To explain further, we give a brief background about low rank matrix completion and Multi-Graph Convolution Networks.

Low rank matrix completion involves recovering a matrix $M \in \mathbb{R}^{N_1 \times N_2}$ of rank $R << min(N_1, N_2)$ from a subset of its entries Ω. To concisely put, given partial observation of M over an index set $\Omega \subset (1, 2, .., N_1) \times (1, 2, .., N_2)$ the task is to select the matrix with the lowest rank. Let X denotes the matrix to recover and M_Ω is the set of the known entries. However, rank minimization is an NP-hard optimization problem and in many real world matrix completion problems, the entities defined on rows and/or

columns share many common attributes. These entities can thus be encoded using graphs by exploiting their proximity information. Incorporating proximity information forces the solution for matrix completion task to be smooth on these row and column graphs. For the row graph, entities defined in the rows forms the vertices and each row of the matrix can be thought of as signals defined on its vertices. In order to combine information from both the row and column graphs, [82] used the concept of Graph Fourier Transform on matrices. Taking into account that Fourier Transform operation is separable and symmetric, the two dimensional transforms can be computed as sequential row and column one-dimensional transforms. Hence the corresponding Fourier transform of matrix X is given by $\mathcal{F}(X) = \Phi_r^T X \Phi_c$, where Φ_r and Φ_c are the eigenvectors of row and column graphs with $\mathbb{L}_r$ and $\mathbb{L}_c$ as the corresponding Laplacian matrices respectively. Further, Monti et.al. [125] proposed Multi-Graph Convolutional Networks (MGCNN) that aims at extracting spatial features from the matrix. Given a matrix $X \in \mathbb{R}^{N_1 \times N_2}$, MGCNN is given by

$$\widetilde{X} = \sum_{j,j'=0}^{q} \theta_{j,j'} T_j(\mathbb{L}_r) X T_{j'}(\mathbb{L}_c) \tag{3.1}$$

where, $\Theta = \theta_{j,j'}$ is the $(q+1) \times (q+1)$ represents coefficient of filters and $T_j(.)$ denotes the Chebyshev polynomial of degree j. Using this equation as the convolutional layer of MGCNN, it produces q output channels ($N_1 \times N_2 \times q$) for matrix $X \in \mathbb{R}^{N_1 \times N_2}$ having a single input channel .

In this way, we extract features and then combine all the information channels in one framework using both the semantic concepts and their contextual relations with the entities.

3.3.3 *Generalizability*

The proposed framework should be generalizable to different use cases consisting of any form of relevant information. Let X and θ be the main entity and the known concept/entity respectively. From all the available information, we know all the relations of entities in X to that with θ, let this relation be represented by ($\theta \sim X$). We aim to predict all the relations of X with another type of concept/entity(ϕ) for whom we know partial relations ($X \nsim \phi$). We define ($\theta \sim X$) and ($X \nsim \phi$) on two different hypergraphs H_0 and H_i respectively and then use the learning model proposed in [12] to complete the partial information in H_i. We will represent the complete framework by ($\theta \sim X \nsim \phi$). Our generic approach representing relations on hypergraph and the learning model make the framework applicable in a variety of settings and use cases. Especially in case of discussion forums, adding any semantic concepts would not alter the pipeline of the proposed framework.

Table 2: Table showing the data setup of the 4 use cases. For training, the framework uses all the relations $\theta \sim X$ and 40% of $X \sim \phi$. The goal is to predict the rest of $X \sim \phi$ relations.

	Use Case 1	**Use Case 2**	**Use Case 3**	**Use Case 4**
$\boldsymbol{\theta}$	E: 3,319	F: 39	E: 3,319	A: 290
$\mathbf{X}$	P: 10,000	U: 10,252	P: 10,000	U: 10, 252
$\boldsymbol{\phi}$	U: 4,218	T: 258	F: 31	F: 39
$\boldsymbol{\theta \sim X}$	$E \sim P$: 36,640	$F \sim U$: 40,418	$E \sim P$: 36,640	$A \sim U$: 51,260
$\boldsymbol{X \sim \phi}$	$P \sim U$: 10,000	$U \sim T$: 29,321	$P \sim F$: 55,280	$U \sim F$: 40,418

3.4 EXPERIMENTAL SETUP

3.4.1 *Dataset*

We use Stormfront, a white nationalist, white supremacist and neo-Nazi Internet forum as the testbed for this study. The forum contains 40 high-level categories, indicating topics of discussion, ranging from "Politics and Continuing Crises", "Strategy and Tactics" and "Ideology and Philosophy" to the topics relevant to national chapters, e.g. "Stormfront en Francais" and "Stormfront en Espanoly Portugués". Typically, in Stormfront, a user posts a content (text, images, videos, links etc.) in one of the relevant forum categories. These users often have a small avatar image and/or somewhat larger profile picture. Recently, Rudinac et al. [143] deployed graph convolutional neural networks for classifying 2 million user posts from Stormfront. In this paper, we use the same data setup as mentioned in [143] for all the use cases. We include the following items of a post from the dataset for our experiments:

- Post ID (P): Unique ID given to each post
- User (U): User who posted it
- Avatar Features(A): Features extracted from display picture of user's avatar
- Forum Category (F): Type of category in which a post has been shared
- User Topics (T): Topic of interest for a user
- Semantic Entities (E): Relevant entities present in textual content of a post

3.4.2 *Use Cases*

To demonstrate the effectiveness of our framework we conduct a set of experiments organized around the following use cases:

Use Case 1: $(E \sim P \nsim U)$ In the first case, we aim to predict potential users who would be interested in interacting with a certain post. This can help in understanding and tracking communication patterns of certain users. In this case, users are the partial

information for the posts while the semantic entities associated with the post will be used for learning the relations between the posts.

Use Case 2: **($F \sim U \nsim T$)** [143] the authors conjecture that the user preferences are a good predictor of the post category. We take motivation from this result for the second use case. Often we require more information about a particular user, but due to very limited user activity, it is not possible to extract such information. We conjecture that the community formed by users posting in different forum categories can have sufficient clues which can be exploited to extract more information (user topics) about an arbitrary user.

Use Case 3: **($E \sim P \nsim F$)** It is often necessary to categorize posts based on their national network due to the formation of local mob or even event organizations. The extracted semantic entities might carry sufficient information for such classification, as the national chapters are characterised by a certain number of topics, more frequent use of national (i.e. non-English) language, and a closed group of users discussing the matters of regional relevance.

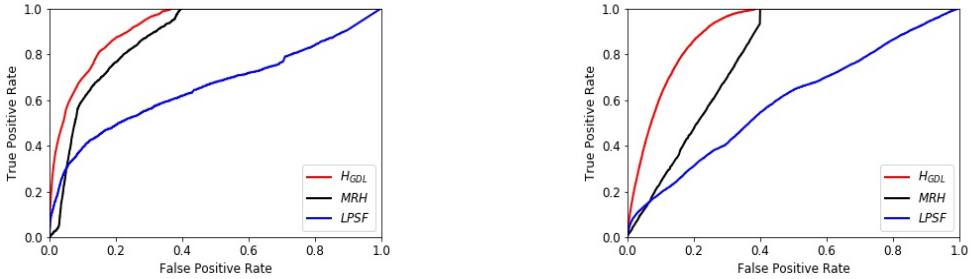
Use Case 4: **($A \sim U \nsim F$)** We aim at identifying the properties of avatars specific of a particular post category. This is to investigate which semantic concepts are more commonly appearing in a certain forum categories, for example "For Stormfront Ladies Only" forum as compared to the other forums on Stormfront. The particular use case came from the domain experts investigating the role and portrayal of women in right-wing extremist networks. Based on the avatars and profile pictures, we are trying to identify users likely to be associated with a particular, specialised discussion forum (e.g. "dating advice" or "religion"). Figure 7 shows an example of the user's features extracted from visual concepts and the partial information of his/her forum categories.

Table 3: Performance of our approach (H_{GDL} as compared with standard models MRH and LPS F

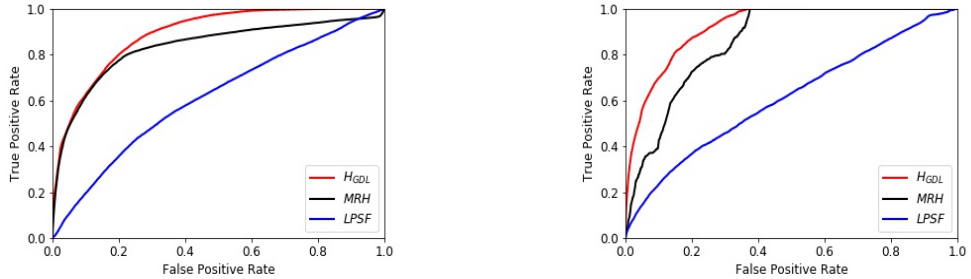
	Use Case 1		**Use Case 2**		**Use Case 3**		**Use Case 4**	
	AUC	EER'	AUC	EER'	AUC	EER'	AUC	EER'
H_{GDL}	86.4	78.5	88.7	80.3	80.6	72.5	89.2	83.0
MRH	79.1	69.9	77.8	70.2	78.4	70.6	84.6	76.5
LPSF	63.1	60.2	60.3	57.9	71.2	68.4	62.9	60.3

3.4.3 *Experiments*

We start our experimental evaluation by showing the performance of our framework on all the 4 use cases. The corresponding number of entities used in each use-cases and the total number of relations formed among them is shown in Table 2. As seen from the table, each entity has multiple values in common with other entities, resulting in a multitude of relations. For example, a user (U) can post in multiple forum categories about various issues. This will account for a large number of $F \sim U$ and $U \sim T$ relations. For our experimental setup, we randomly sample 40% of these relations and keep them aside to use as a test set. The remaining relations are used to construct the partial hypergraph H_i for training the model.



(a) Use Case 1 *(b) Use Case 2*



(c) Use Case 3 *(d) Use Case 4*

Figure 8: Receiver Operating Characteristics (ROC) curve showing the performance of the models on each of the 4 use cases. The hypergraph-based geometric deep learning model (H_{GDL}) has significant advantage as compared to other methods on all the 4 use cases.

3.5 RESULTS

We report the results of our framework and compare them to hypergraph based algorithm (*MRH*) [32, 100] and a graph-based model trained on social network features (*LPS F*) [179] for the same tasks. *LPS F* trains a neural network on popular features like Page Rank, Number of Common Neighbors, Preferential Attachment etc. extracted from a social network. To evaluate the performance of our model and show its advantages over other methods, we plot the Receiver Operating Characteristic (ROC) curves for each task. The ROC curve depicts how well a model is able to predict the presence/absence of any information in an entity. Figure 8 shows the performance of the models on the 4 use cases.

To further quantify the results, we calculate AUC(Area Under Curve of the ROC plot) and $EER' = 100\% - EER$ (Equal Error Rate) for all the three methodologies. EER corresponds to the point on the ROC curve that corresponds to an equal probability of miss-classifying a positive or negative sample. Both these numbers are very important indicators of a model's overall performance, the higher the AUC and EER' the higher the accuracy of the system. We show them in Table 3, where it can be seen that our model, in general, overperforms alternatives in predicting information of any type of entities.

3.6 CONCLUSION

In this paper, we constructed a framework which can be used to learn and predict relational information within a discussion forum. The generalizability of the framework provides the flexibility to the end users to formulate their own use cases irrespective of domain-specific constraints. As a test bed for our study we use the analysis of a realistic collection of data from Stormfront, a violent online political extremism forums. The experiments are conducted around research questions raised by the domain experts and demonstrate the effectiveness of our approach in providing implicit information about users of a forum. The results confirm the merit of our approach to geometric deep learning on hypergraphs and suggest that in case of multimodal data, the proposed framework can be used for designing a case study. Finally, on four example use cases we demonstrated that this technique may be a valuable asset to domain experts performing qualitative analysis of violent online political extremism.

4

VISUAL ANALYTICS FOR TEMPORAL HYPERGRAPH MODEL EXPLORATION

Many processes, from gene interaction in biology to computer networks to social media, can be modeled more precisely as temporal hypergraphs than by regular graphs. This is because hypergraphs generalize graphs by extending edges to connect any number of vertices, allowing complex relationships to be described more accurately and predict their behavior over time. However, the interactive exploration and seamless refinement of such hypergraph-based prediction models still pose a major challenge. We contribute HYPER-MATRIX, a novel visual analytics technique that addresses this challenge through a tight coupling between machine-learning and interactive visualizations. In particular, the technique incorporates a geometric deep learning model as a blueprint for problem-specific models while integrating visualizations for graph-based and category-based data with a novel combination of interactions for an effective user-driven exploration of hypergraph models. To eliminate demanding context switches and ensure scalability, our matrix-based visualization provides drill-down capabilities across multiple levels of semantic zoom, from an overview of model predictions down to the content. We facilitate a focused analysis of relevant connections and groups based on interactive user-steering for filtering and search tasks, a dynamically modifiable partition hierarchy, various matrix reordering techniques, and interactive model feedback. We evaluate our technique in a case study and through formative evaluation with law enforcement experts using real-world internet forum communication data. The results show that our approach surpasses existing solutions in terms of scalability and applicability, enables the incorporation of domain knowledge, and allows for fast search-space traversal. With the proposed technique, we pave the way for the visual analytics of temporal hypergraphs in a wide variety of domains.

4.1 INTRODUCTION

A significant volume of real-world data consists of entities and their relationships and can accordingly be modeled mathematically using graph-based approaches. Such approaches are widely applied in many domains, ranging from natural and social sciences to engineering and business. Examples include modeling biological and chemical processes like protein-protein interactions [137], relationships in computer [182] as well as human communication networks [132], or knowledge network exploration in business processes [70]. Whereas static graphs can represent the fixed relationships between entities, using an undirected or directed graph as a model, many of the examples presented above are more accurately described as processes with complex interrelations that may change or evolve.

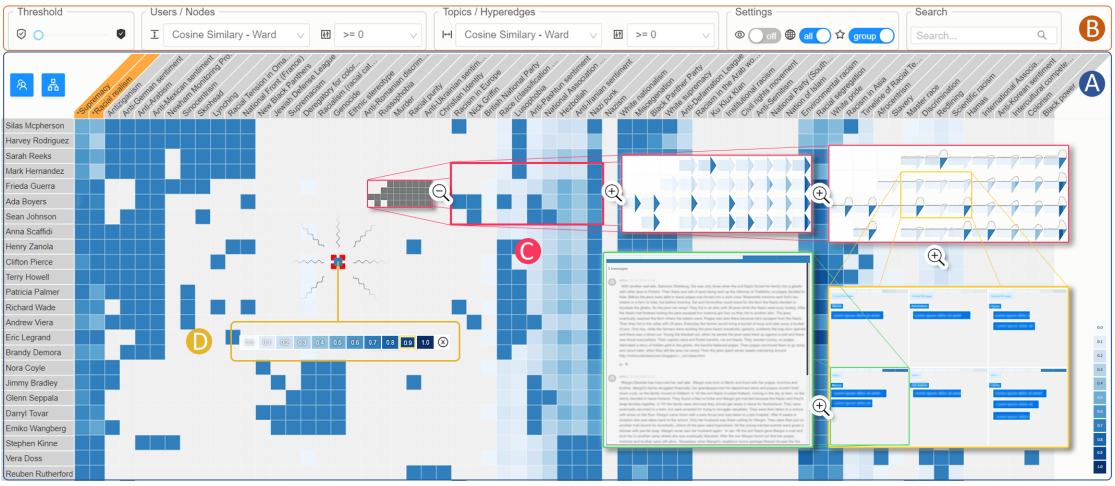


Figure 9: HYPER-MATRIX, *a novel approach to explore and refine temporal hypergraph models using visual analytics. The interactive multi-level matrix-based visualization* A *enables the inspection of the model, together with the upper interface* B *. The main area shows the second semantic zoom level applied to an obfuscated real-world dataset in criminal investigations, while the five insets* C *show the other drill-down levels for exploration. The technique allows to interactively* D *contribute domain knowledge, the resulting implications have ripple effects on the whole machine learning model, thereby refining it.*

Here, geometric deep learning methods together with interactive visualization can help to more accurately model, predict, and explore the model evolution. Considering, for example, conversations, a topic is a time-dependent grouping encompassing users, which cannot be described using a static graph. This evolution of relations should be modeled by dynamic networks. Compared to regular graphs, using edges or separate node types, such modeling often reflects the actual process more accurately. Dynamic networks are, however, more challenging to model and have traditionally been modeled as regular, undirected graphs, mainly due to computational and visualization limitations. In recent years, modeling has extended to dynamic networks [95], but some limitations remain.

Consequently, one can take a step further and use temporal hypergraphs. Hypergraphs generalize graphs by extending edges to connect any number of vertices, allowing complex relationships to be described more accurately [147] while reducing ambiguity and network inflation. Utilizing temporal hypergraph prediction models, however, introduces its own set of challenges.

First, as the model structure is more complex, it is relevant how the information is communicated to the analyst through visualization (cf. [69]) and how domain knowledge feedback is incorporated. Static hypergraphs can be considered as standard sets, with different visualizations available [4]. Temporal hypergraphs, meanwhile, add a time-dependent evolution, making it harder to convey the relevant information meaningfully.

Secondly, many traditional graph-based concepts cannot directly be applied to hypergraphs. Hyperedges, as arbitrary sized sets of connected nodes, add another order of complexity. In previous works [11, 13], we presented how geometric deep learning can be applied to hypergraphs and showed how this method could be leveraged to predict behavioral patterns in social media hypergraph models.

Consequently, the incorporation of machine learning techniques into an interactive model to more accurately predict changes in the hypergraph due to changes in the data introduces new problems. While deep learning avoids assiduous manual feature engineering and algorithm design, it reduces explainability and accountability of the results. Domain experts usually have some domain-specific intuition—a mental model and structure—about inherent and implicit relations and groupings not available in the data, enabling them to judge the plausibility of hypotheses and to steer the exploration. Yet, they face difficulties articulating their domain knowledge through machine learning into the predictions and tracing its influence. This holds especially for very complex models, like temporal hypergraphs. The knowledge formalization requires a very detailed a priori understanding of the problem by domain experts, which is not always available. For the same reason, it is challenging to capture the knowledge independently of the model without rapid, iterative feedback. Hence, the machine learning outcome often correlates strongly with the adequacy of the initial problem modeling and the quality of the training data, while domain expertise and domain knowledge are frequently not leveraged to their full potential.

To address these issues, we present HYPER-MATRIX, making the following contributions:

- A novel, interactive framework for temporal hypergraph exploration through the use of semantic zooming relying on a multi-level matrix-based approach and various exploration concepts.
- The extension of a geometric machine learning architecture [11, 13] with a relevance feedback model.
- A tight coupling between the visualization and the machine learning relevance feedback model for evaluation and seamless refinement, offering the integration of domain knowledge and making the corresponding model changes visually transparent.
- One case study describing an application of the technique to the law enforcement field.
- A formative evaluation with law enforcement experts using real-world communication data, demonstrating that our technique surpasses existing solutions, enabling the effective analysis of large amounts of information in a targeted way.

Our approach bridges the gap between visual exploration and separate model training, allowing domain experts to enhance the machine learning predictions with implicit domain knowledge in the same step as evaluating and exploring the temporal hypergraph model predictions.

4.2 RELATED WORK

This research is an entry into the interactive temporal hypergraph model exploration in the context of explainable support by machine learning. In the literature hypergraphs are studied from both a visualization as well as a machine learning perspective. In the following discussion, we adhere to the same distinction and relate our work to the visualization of temporal hypergraphs as well as their application in machine learning.

4.2.1 *Visualization of Hypergraphs*

We first shortly discuss the situation for (static) hypergraphs as well as dynamic graphs, before looking at temporal hypergraphs. Hypergraphs can be considered as a set of sets. The survey on set visualizations by Alsallakh et al. [4] shows that several visualizations are applicable to hypergraphs. Hypergraphs are often drawn as regular graph networks or bipartite networks. When making their dimensionality explicit, they can be drawn as subsets—like Venn diagrams or radial sets—or in node-link form [175], using colored hulls or other, specifically adapted approaches [120]. A third possibility is to use a matrix-based approach, which improves scalability [86]. Subsets and node-link diagrams suffer from limited scalability, quickly leading to occlusion and clutter. Bound in the number of visual attributes they can employ, these techniques typically reach their constraints in the order of one or two dozens of hyperedges [173]. Further, they are difficult to extend with a temporal component, having already used up most visual attributes.

In comparison to set based approaches, dynamic graphs change over time, leaving the choice [20] between employing animation or an additional timeline component. The former puts significant strain on the mental map when many connections change, while the latter is limited by the available screen space in the number of discrete timesteps it can show. The survey [20] also points out that node-link diagrams remain the most commonly used type of visualization. However, these approaches mostly lack the extendability to hypergraphs.

When studying temporal hypergraphs, the issues arising from the dimensionality and the temporal nature all build up. Indeed, there is almost no prior work on the visualization of temporal hypergraphs specifically. Two notable exceptions exist, which allow visualizing—but not modifying or refining—temporal hypergraphs: First, the recent works by Valdivia et al. [171–173]. Their visualization approach is also shown later in Figure 14c as part of the case study. Second, the previous work by Streeb et al. [157] introduces an in-line visualization of the temporal evolution. Valdivia et al. begin to tackle the research gap by proposing PAOHvis, thereby claiming to provide the "first [...] highly readable representation of dynamic hypergraphs". While this is a strong claim to make, the literature review showed a broad diversity between the approaches, but none—except the two mentioned above—is directly suitable for temporal hypergraph visualization, supporting this conclusion. Utilizing the previously discussed approaches as substitutes for a tailored visualization often does not adequately leverage the additional information available with temporal hypergraphs and does not address the tasks that come with hypergraph topology and evolution. For those, we refer to Section 4.2.3. Shortcomings in existing approaches include, for example, Streeb et al. providing only the prediction abstraction level in their visual interface (cf. Level 3 in Section 4.4.1). Similarly, this is true for Valdivia, although they support coloring by a group. This can lead to information overload, as filtering using thresholds is the only way to reduce the information. In contrast, usage of semantic zoom enables an exploration of the complete hypergraph (cf. Section 4.4.1) without the need to preliminary apply filters while enabling tailored visualizations showing detailed information when focusing on different abstraction levels. Prominent examples of matrix-based visualizations are the Zoomable Adjacency Matrix Explorer [50] that enables users to zoom and pan with interactive performance from an overview to the most detailed views and the visual analysis system

of Behrisch et al. [22]. It features a flexible semantic zoom to navigate through sets of matrices at different levels of detail. Further, both Streeb and Valdivia, only support sorting by weights and average (cf. size ordering in Section 4.4.2), compared to our default matrix-based sorting, improving cluster identification. Significantly, all existing approaches aim at analyzing a fixed hypergraph model. None focus on interactively working *with* the model and iteratively improving it (cf. Sections 4.3.2 and 4.4.3).

At last, while not strictly related to the research on temporal hypergraphs *per se*, we want to mention approaches that are, at least partly, similar to ours, and also conventional tools so far applied in practice. Here we concentrate on how hypergraph-like data is handled in the law enforcement field, relevant for the case study and the evaluation through domain experts (see also Sections 4.5 and 4.6). The visual analysis of communication data—but without any hypergraph visualization or a tunable model—is not novel and has been researched both from the analytical side [116] as well as the visualization side [186]. Also, the idea of semantic zooming for matrix-like visualizations has been described previously [174], however, in a different way and in the area of software management. Further, it was also described how an overlay magic lens [60] can be used instead of zooming, to keep the context and allow for faster search space traversal from locations far apart, which we partly employ for the partition hierarchy (Section 4.4.2). In practice, for the law enforcement field, we found that data which benefits from a hypergraph modeling, like communication patterns or process analysis, is prevalent, but not supported by any system. Gephi [17] is sometimes used, but analysts often prefer Pajek [18, 19], as it supports larger networks. The most popular tool is IBM i2 Analyst's Notebook's [78] graph component due to the prevalence and familiarity in this domain.

4.2.2 *Machine Learning for Hypergraph Models*

Learning with hypergraphs was introduced by Zhou et al. [206] to model high-order correlations for semi-supervised classification and clustering. It generalizes the efficient methodology of spectral clustering to hypergraphs by proposing a label propagation method to minimize the differences in labels of vertices sharing the same hyperedge. The correlation among hyperedges was further explored by Hwang et al. [77], assuming that highly correlated hyperedges have similar weights. More recent works [30] concentrate on parametric learning of weights using propagation of node features across hyperedges [53, 191].

Understanding communication patterns of users on social networking sites has created opportunities for richer studies of social interactions and better prediction of behavioral patterns. In multimedia, link prediction on hypergraphs has been a popular topic of research in social network analysis. This includes predicting metadata information such as tags and groups for entities in social networks, e.g., images from Flickr [13], music recommendation by exploiting network proximity information of users in Last.fm [32] and predicting higher-order links (such as tweets with a specific hashtag) in Twitter [100]. Besides, hypergraph learning models are being used in multimodal data analysis to integrate complementary information from multiple modalities effectively. Liu et al. [114] proposed a multi-hypergraph learning method to handle incomplete multimodal data for disease diagnosis in neuroimaging and Arya et al. [10] proposed a framework to

learn a compact representation for each modality in a multimodal hypergraph using a tensor-based representation. These works have shown the importance of hypergraph based learning for predicting implicit links within a network. However, none of these approaches pose an interactive learning formulation that can assimilate user feedback as an external source of information to either improve the predictive capability of a model or to even change the intrinsic properties such as learnable parameters of a model. In this work, we extend our previous work [11] on link prediction in communication networks capable of fine-tuning the trained model by incorporating external relevance feedbacks.

4.2.3 *Tasks for Evaluation of Temporal Hypergraph Models*

Tasks in temporal hypergraph analysis relate to dynamic networks and set comparisons. A task taxonomy of the former is provided in the survey by Beck et al. [20], and for the latter in the survey by Alsallakh et al. [5]. For temporal hypergraphs, in particular, the tasks sometimes substantially differ; for example, one being the analysis of changes of both connections and attributes over time. The proposed technique does not directly fit with any existing task taxonomy, positioning itself between disciplines [6]. For a discussion on existing taxonomies and their applicability to temporal hypergraphs, we refer to the existing work by Valdivia et al. [173] and summarize only the main aspects here. Our technique supports not all traditional tasks in set analysis [5], and in dynamic network analysis [2, 14, 84, 97], summarized in [20]. However, it provides support for several additional tasks relevant to our driving application. These include the clustering of related groups independently of their temporal connection, the inspection of shared attributes of connections, the following of temporal evolutions, while both retaining an overview and simultaneously being able to explore details. In short, the experts are interested in connectivity information involving both graph topology as well as attribute values, which can be separated between time ranges. One main requirement is the need to include external (domain) knowledge that is not directly available as raw data and includes conceptualized topics in line with their mental categorization. These tasks are not sufficiently described or supported by existing taxonomies, as they neglect the additional complexity incorporated by hypergraphs and the domain knowledge integration.

Given the sparse research in hypergraph visualization, it is unsurprising that there is no prior work on bridging both fields; this is the gap we aim to fill: offering a technique that addresses the shortcomings discussed above, enabling the exploration and refinement of hypergraph models using interactive visualization, closing the visual analytics loop.

4.3 EXTENSION OF MACHINE LEARNING TO HYPERGRAPHS

In the following two sections, we describe the overall workflow of our approach, shown in Figure 10. We begin with an exemplary description of one geometric deep learning model, adapted to a task relevant for our law enforcement domain experts: the temporal prediction and analysis of patterns in communication data. It acts as a blueprint for problem-specific temporal hypergraph models. In Section 4.4, we then discuss the interactive exploration using visual analytic principles.

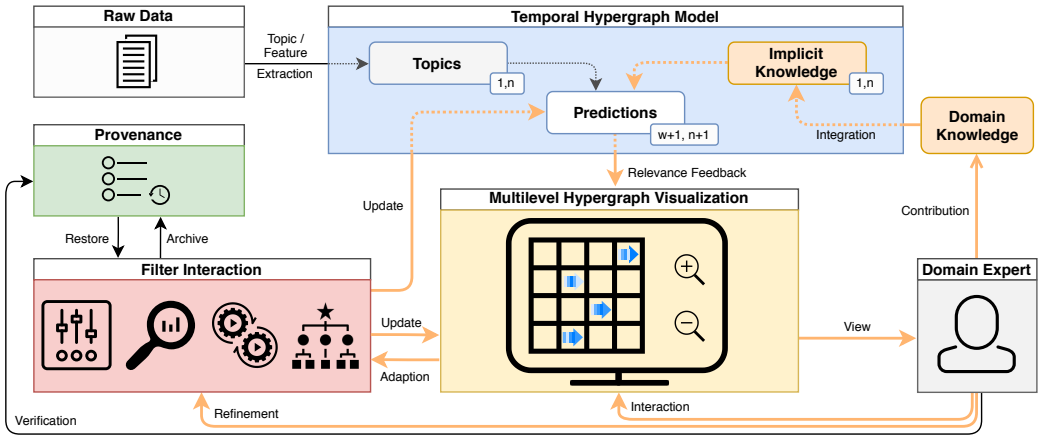


Figure 10: High-level workflow of our technique, showcasing the main components and the interaction flow for the exploration and refinement of temporal hypergraph models, adapted to use case A in Section 4.5. The workflow begins with ***raw data*** *extraction and the generation of a* ***temporal hypergraph model****. The model state is visualized using a matrix-based* ***multilevel hypergraph visualization*** *, allowing for various exploration and* ***filter schemes****, including search and filters , a dynamically modifiable partition hierarchy , and matrix-reordering techniques . The* ***domain expert*** *can interact with the model by either refining the filter schemes or by contributing domain knowledge, which both update the model. The model feedback can then be explored and accepted, closing the* ***visual analytics loop*** *→. The chronology of interactions and contributions are available for recovery or verification as a* ***provenance history*** *, facilitating accountability.*

4.3.1 *Notation and Formulation of a Temporal Hypergraph*

In set theory, an undirected hypergraph $H = (V, E)$ is defined as an ordered pair, where $V = \{v_1, .., v_n\}$ represents the n vertices (hypernodes) and subsets of these vertices $E = \{e_1, .., e_m\}$ constitute the m distinct hyperedges. H is represented by the incidence matrix $\mathbb{I}{=}|V|{\times}|E|$, with entries $i(v_j,e_k)$ 1 if $v_j{\in}e_k$ and 0 otherwise. We define the neighborhood of v_j as the set $N(v_j)$ of nodes within the same hyperedges as v_j.

In adapting a generic temporal hypergraph model to our use case, we follow our previous work [11], representing the relationship between internet forum users and their behavioral characteristics (both "explicit" and "implicit"). The available metadata (in particular forum category) forms the explicit characteristic of a user, while their topics of discussion outline the implicit communication characteristic. Thereby, we construct two separate hypergraphs depicting the connection of users with these explicit and implicit behavioral characteristics. To model the temporal component, let us define a temporal hypergraph by $H_{[t]}$, at a given time t, where each user is represented as a node, and each type of explicit/implicit characteristic is represented as a separate hyperedge. We denote the explicit and implicit hypergraphs, at any given time t, by $H^0_{[t]}$ and $H'_{[t]}$, respectively. Consequently, in $H'_{[t]}$, each topic is depicted as a separate hyperedge and users (nodes) who adhere to a common topic of interest are connected by it. Thus, forecasting the evolution of users' topics of interest for time $t{+}1$ becomes equivalent to the task of finding new relations over the existing relations in hypergraph $H'_{[t]}$.

4.3.2 *Relevance Feedback to Deep Learning Model*

As indicated, the underlying model for forecasting future interests of internet forum users is based on predicting links in temporal hypergraphs. The task of link prediction on a hypergraph $H_{[t]}$ with a fixed set of edges E aims at updating the set e_k. This link prediction can be formulated as missing value imputation or a matrix completion task on $\mathbb{I}$. In the following, we extend our previous work [11, 13], to allow for the incorporation of feedback in HYPER-MATRIX. Therefore, we first reconsider the module for training a geometric deep-learning model. Then, we formulate how feedback from the user can be employed to update the model.

TRAINING MODULE Let $\mathbb{I}_{[t]}$ denote the incidence matrix of $H'_{[t]}$ at time t which we can factorize as $\mathbb{I}_{[t]}=X_t Y_t^T$ with X_t and Y_t the row and column matrices, respectively. Hypergraph $H^0_{[t]}$ will be utilized as an auxiliary set of explicit information between users for predicting links in the implicit hypergraph $H'_{[t]}$. The information in $H^0_{[t]}$ is encoded by extracting its Laplacian denoted by Δ_0. The Laplacian Δ_0 gives a measure for the relatedness between any pair of users [32]. Using such a similarity measure can significantly enhance the user-topic link prediction outcomes by reducing extraneous noise and thus smoothing the model output.

To train the model, we employ a semi-supervised learning setup, hence the predictive loss is backpropagated by using a small set (around 5–8 %) of known links in $H'_{[t+1]}$. These known links create an upper bound for the number of timesteps the model can predict in $\hat{\mathbb{I}}_{[t+1]}$. Details can be found in [11, 13]. For training, we take the incidence matrix $\mathbb{I}_{[t]}$ at time t and use the hypergraph link prediction model $\mathbb{H}_{GDL}$ to learn the best parameter set $\Phi[t]$ for predicting the incidence matrix $\mathbb{I}_{[t+1]}$ at time $t+1$:

$$\hat{\mathbb{I}}_{[t+1]}, \Phi_{[t]} = \mathbb{H}_{GDL}(\mathbb{I}_{[t]}, \Delta_0) \tag{4.1}$$

FEEDBACK MODULE In order to integrate domain knowledge into the underlying model, we propose a novel interactive learning formulation to incorporate feedback from the domain expert. These feedbacks are assumed to contain definitive implicit information about the topic of interest for certain users in the dataset. Instead of just updating the information by directly changing the topic (hyperedge) of the respective users (nodes), these feedbacks should also create a "ripple effect" on the overall connections in the hypergraph $H'_{[t]}$. That is, if the feedback $f_{[t]}$ at time t involving the single user (u_j) denoted by node v_j in the hypergraph $H'_{[t]}$, then incorporating $f[t]$ will entail a twofold operation: *1. Update*: Topics for user u_j are updated, i.e., add/remove v_j to/from the respective hyperedges $E = \{e_1, .., e_m\}$ corresponding to $f_{[t]}$. *2. Predict*: Change topics for users in close communication with u_j based on their relatedness to u_j, i.e., re-calculate the connection strength for vertices in $N(v_j)$ with the hyperedges $E = \{e_1, .., e_m\}$. The first operation is a straightforward updating of the matrix $\mathbb{I}_{[t+1]}$ by updating new values corresponding to nodes and edges suggested in the feedbacks $f_{[t]}$. The change in the neighborhood connections are calculated by using the updated matrix $\mathbb{I}_{[t+1]}+f_{[t]}$ as input to our link prediction model $\mathbb{H}_{GDL}$. However, in the feedback module, instead of learning parameters through an iterative process, the learned parameters $\Phi_{[t]}$ are used as initialization of the already trained model $\mathbb{H}_{GDL}$. This ensures the model converges in

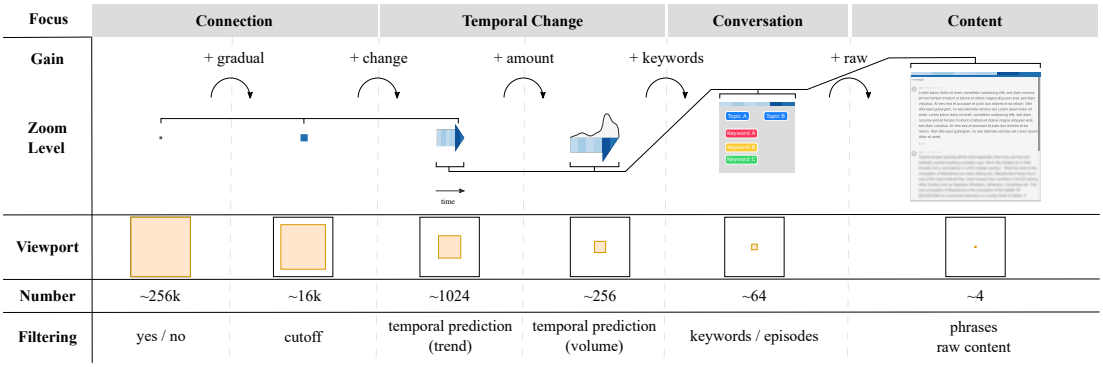


Figure 11: Semantic zoom levels and the different filtering levels (cf. Section 4.4.1). At each zoom step, the analyst gains another type of information about the model, filtering a different layer of complexity. As the focus becomes more detailed, the visualization takes up more space (zoom level and viewport as shown not to scale), while the number of visible entities decreases accordingly. The temporal predictions are shown in different forms throughout all levels (see fine grey line), with the detailed temporal evolution first shown in Level 3 and continuing down to Level 6.

far less time after incorporating the feedbacks $f_{[t]}$ than when learned from scratch. The following equation shows the representation of the feedback module in symbolic form:

$$\hat{\mathbb{I}}_{[t+1]} = \mathbb{H}_{GDL}(\mathbb{I}_{[t+1]} + f_{[t]}, \Delta_0, \Phi_{[t]}) \tag{4.2}$$

4.4 INTERACTIVE HYPERGRAPH MODEL EXPLORATION

In this section, we focus on the visualization and interaction with the temporal hypergraph model, providing a tight coupling between the data manipulation and display (see Figure 10). We begin by describing how the model state can be depicted using a matrix-based visualization that provides drill-down capabilities across multiple levels via semantic zoom. Drill-down is thereby defined as the seamless zooming through the different levels during exploratory analysis, starting from a general overview to increasingly more focused and detailed information, as highlighted in Figure 11. To facilitate the interactive exploration, we present user-steering based on classical filters for standard search tasks, a dynamically modifiable partition hierarchy to include user-based structuring, and various matrix reordering techniques for the focused analysis of connections and groups. We then specify the interactions that allow domain knowledge to be incorporated into the machine learning model via relevance feedback and highlight how the updated predictions can be reflected in the existing visualization. This workflow facilitates the explainability of the underlying model, thus enabling the domain experts to provide more meaningful feedback. Finally, we describe how all interactions, domain knowledge input, and model output are stored in a provenance history, providing accountability and making the decision-making processes more transparent.

4.4.1 *Model Visualization*

As discussed above, the complexity of temporal hypergraphs makes them difficult to visualize. Hence, we propose a multi-level matrix-based approach, specifically tailored to the hyper-dimensionality as well as the temporal component. The visualization (see Figure 9) consists of a menu bar on top, controlling the interaction concepts discussed later, and, for the main part, a matrix-like viewport, showing nodes as rows and hyperedges as columns, with corresponding row and column headers. This viewport provides freely pan-able and zoom-able drill-down capabilities across six levels of semantic zoom, shown in Figure 11, increasing or decreasing the information detail: from an overview of model predictions down to contents. For this purpose, we use three different level types: cells, arrows, and content boxes. Colored cell visualizations are used in Levels 1 and 2. An arrow-like representation reflecting a timeline is used in Levels 3 and 4. The base of the arrow represents the past, while the head reflects predictions. As the predictions become more uncertain with time, the arrowhead becomes smaller, reflecting the increased uncertainty and thus the decreased relevancy of the prediction. Levels 5 and 6 add text-based elements like keywords or raw content. Level 3 and beyond all contain the temporal aspect.

The visualization depends on the zoom state of the viewport. During drill-down, the focus shifts from a general structure overview over the temporal evolution to the raw content, providing the expert with more and more detailed information. Before we start with the description of this process, we define some necessary terms. As the feedback model outputs probabilities for the connections (see Section 4.3.2), gradual differences can be analyzed. When setting a minimum threshold for a connection to be meaningful, this allows for a *binary choice*. Showing a color encoding of the connection strength allows for a more expressive representation of the *gradual differences*. Setting a *cutoff* threshold can still be used to avoid cluttering with low-probability entries. The drill-down shifts the focus of the analysis. It starts at the (binary) connectivity information, extends to gradual connection strength (Level 2), to the temporal change represented as an arrow (Level 3), to the temporal change encoded using position instead of only color (Level 4), then to information summarizing the underlying content for the predictions, in this case, keywords (Level 5), and, at last, to the raw data (Level 6). The design choice for an arrow glyph representation in Levels 3 and 4 is based on five reasons: (1) The principal idea of an arrow glyph was previously published [157] and found to be beneficial. Then, (2) given the target audience, a representation as an *arrow of time* is closely related to everyday experience. Further, (3) the separation into arrow base and head allows a clear distinction between past data and model predictions, which is very important for the target audience. The arrowhead also allows to visually reflect the decreasing prediction accuracy by becoming smaller. In terms of (4) visual advantages, an arrow provides a distinct shape, while, e.g., a cell is easily perceived to merge with neighboring cells, which is undesired. The choice also comes with disadvantages, introducing white space and can sometimes lead to distracting patterns. Finally, (5) a design study on combining timeline and graph visualization by Saraiya et al. [145] shows that our approach—simultaneously overlaying the timeline—is best suited for detecting outliers. This is one of the main tasks for these levels, given the focus on change. The study also supports the design choice of showing only a single timestep in Levels 1 and 2, as the focus is on the topological

structure. However, different visual representations like horizon graphs might be better suited when focusing on a continuous analysis. The seamless changes between levels speed up navigating through large models while eliminating demanding context switches. Moreover, at each step, the information becomes more complex, requiring more screen space to visualize. For a regular HD screen, we give rough guidance ($O(n)$) on the number of elements that can be usefully shown on-screen, amounting to around 256k grid cells of connectivity information and around four for the raw content.

4.4.2 *Interactive Exploration and Drill-Down*

To facilitate the interactive exploration, we contribute a user-steering based on classical concepts and filters for standard search tasks, a dynamically modifiable partition hierarchy to include user-controlled structuring and various matrix reordering techniques for the focused analysis of connections and groups. All these interactions concepts are reactive, and the visualization can smoothly and instantly update ($<$ 100 ms), except for the domain knowledge integration in Section 4.4.3.

INTERACTION AND FILTER CONCEPTS Standard methods available in an interactive visualization are included, like (1) highlighting selected rows or columns, (2) highlighting hovered cells, (3) tooltip-based menus, (4) marking (i.e., starring) individual entries to highlight them for tracking and follow-up, (5) adding textual notes, and (6) showing additional meta-information. Modal views allow to (7) control the partition hierarchy (see details below), while setting an (8) overall cutoff threshold allows controlling the confidence threshold of the underlying model. A (9) global search function provides the ability to search for node- and edge information as well as content and highlights the matching components. At last, the menu bar allows (10) controlling the matrix reordering (see detail below).

DYNAMICALLY MODIFIABLE PARTITION HIERARCHY To allow domain experts to articulate their mental categorization to the model, the experts can create (nested) groups of different nodes or hyperedges, creating hierarchies. The nodes or hyperedges hereby relate to the leaves of the dendrogram. The groups can be expanded or contracted either directly from the node or hyperedge headers, visually indicated by color, or by editing them inside the partition hierarchy viewer in a modal overlay. The viewer shows a dendrogram-based representation with freely reorderable entries. Each branch of this dendrogram can be independently collapsed or expanded, i.e., the abstraction level is local to each branch and not globally set. For example, it is possible to collapse a large, uninteresting sub-branch, including the nested nodes it contains, while simultaneously having one branch fully expanded and another only up to the penultimate level. This is also independent of the overall visualization level, similar in concept to multiple fixed magic lenses, visually supporting different analysis paths. The hierarchy allows, for example, to group complementing entities together, to build meta-entities, and even hierarchies of entities.

(a) Multi-step hierarchy *(b) Size* *(c) First occurrence*

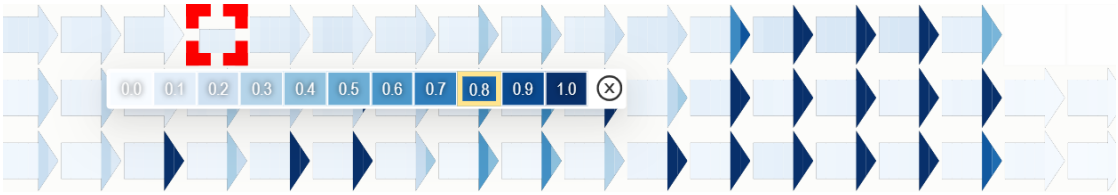
Figure 12: Comparison of different matrix reordering techniques to facilitate the detection of similar groups and connections. Compared to the unordered state and the slightly improved ordering by size, the adoption of a default multi-step, dendrogram-based reordering, modified and adapted from [21], enhances the clustering by similarity.

MATRIX REORDERING AND SORTING To support the tasks relevant for our driving application (see Section 4.2.3), a matrix reordering is desirable such that related users and topics appear close to each other. Due to the independent and often conflicting interpretations of both axes and the sparseness of the underlying matrix, the direct application of standard 2D numeric sorting algorithms (e.g., Multi-scale-, Chen-, or Travelling salesman problem ordering) [21] often leads to unsatisfactory results, as they are mainly applicable to pairwise comparison matrices.

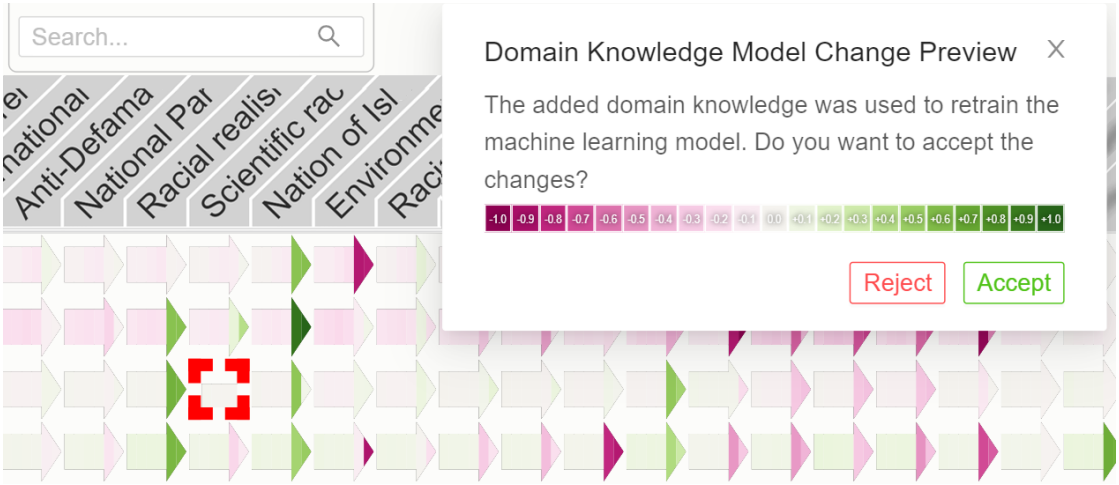
As part of the visualization, we offer three main different reordering strategies, as shown in Figure 12: (a) matrix-reordering (default), (b) sorted by size (connectivity), (c) first occurrence (original). The reordering is applied individually for each axis, as the requirement may differentiate between search tasks, not always favoring a block-like clustering. It also provides more flexibility for adopting other sorting methods in domain adaptions of our technique. The underlying sorting principles build upon a dendrogram-based serial matrix reordering discussed by Behrisch et al. [21]. It forms a multi-step process, combining the sorting of node and edge similarity vectors. Supported dendrogram methods are ward-, single-, average-, and complete linkage, combined with any pairwise distance function like Euclidean, cosine, or Jaccard. We refrain from discussing individual choices, which can vary strongly on domain adaption. For our case study, the Jaccard and cosine distance provide consistent results.

4.4.3 *Visual Analytics for Model Updates*

To increase the traceability of domain knowledge integration and explainability of the resulting model changes, we propose an interactive change feedback visualization, that seamlessly integrates with our visualization. The two-step process is shown in Figure 13. An expert can integrate domain knowledge by selecting a cell and setting a new connection strength (Figure 13a), thereby complement missing or override model *input* data. This input is used to partly retrain the model and refine its predictions as described in Section 4.3.2, leading to a ripple effect. Thereby, the model has prediction authority, i.e., the user cannot manually fix the ultimate output to guarantee model authenticity. A spinner indicates the few seconds long operation. The resulting *changes* are displayed inside the same view (Figure 13b). A diverging color scale is used, showing changes instead of predictions. Through two visually distinct scales, it is immediately



(a) Manually input a connection strength.



(b) Resulting changes in the model predictions.

Figure 13: Resulting changes in the model prediction (ripple effect) from an input are visualized by a diverging color scale (from negative to no to positive change). They can be explored and rejected or accepted. This allows for model verification and multiple, different analysis paths.

apparent if predictions or changes in the predictions are shown. The view integration allows for consistency, reducing the mental workload, and improving mental mapping.

Changes can be inspected on all levels of the visualization. The exploration is *not* restricted to just the current viewport, finding even weak connections. Change detection is facilitated, allowing rejection if deemed implausible or acceptance if convincing, enabling the followup of multiple analysis paths. By iteratively and interactively queering the model and see how it responds to domain knowledge integration, experts can discern better how connections and processes in the model are related, improving understanding and increasing explainability.

Experts in many applications are interested in their analytical progress and must reproducibly document the steps. We address this by a re-loadable provenance, storing the interaction sequence, domain knowledge input, model output, and fixed RNG seeds. This allows for inspection, verification, and traceability while providing accountability and making decision processes transparent. The provenance history allows undoing analysis steps, preventing dead-ends, revisiting and explaining past steps, but also bridging off to diverging analysis trails.

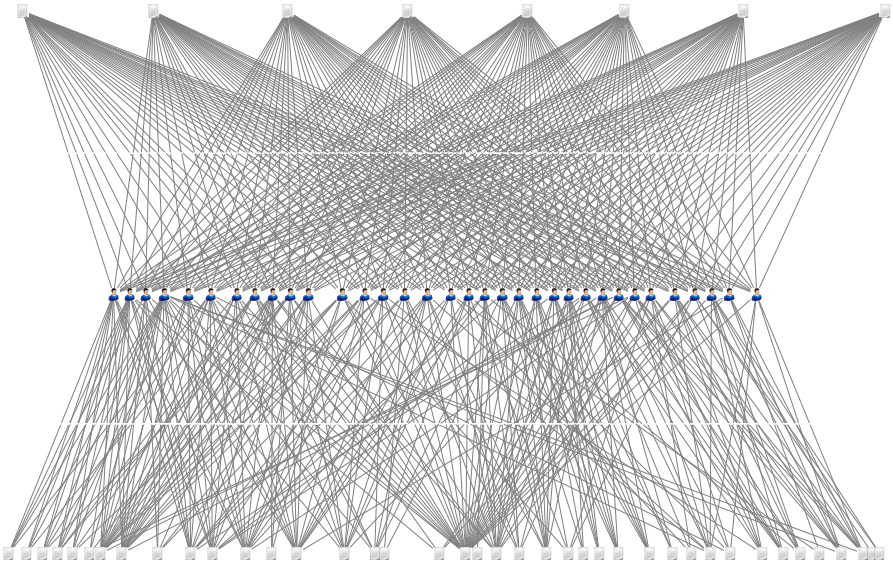
4.5 CASE STUDY: INTERNET FORUM COMMUNICATION DATA

To demonstrate the visual exploration of temporal hypergraph models in HYPER-MATRIX, we conduct a case study, showing the applicability of our technique and improvements compared to existing approaches.

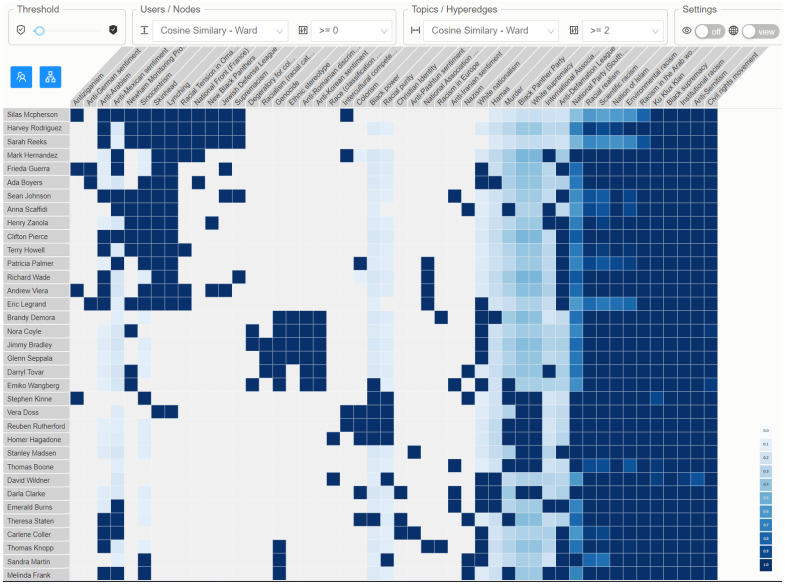
The communication data was collected from an internet forum well-known to law enforcement. It contains 335 188 text posts from 4904 users. We pre-processed the data using standard NLP methods to extract 158 topics, based on a domain-specific ontology. As described in Section 4.3, users are associated with nodes and topics become dynamic hyperedges. To allow for a reasonable side-by-side comparison with the existing approaches, shown in Figure 14, we had to restrict to a subset, consisting of 35 users, 65 topics, and six timesteps. This is around four times more than conventional approaches are designed for. We confirmed that our prototype works for significantly larger networks (cf. Section 4.7). Our prediction model is fed with four years (timesteps) of historical data and then predicts the evolution of the next two years as two timesteps. Almost any real-world data is noisy and may miss some relationships. Consequently, some of the conclusions drawn here may be inaccurate. However, we focus on demonstrating the concepts and benefits of the visual analysis process HYPER-MATRIX provides.

The task we want to focus on in this case study is the identification of related groups and missing links, common in criminal investigations. To identify users discussing the same topics and topics discussed by the same group, the matrix reordering and connectivity information in Level 2 can be used to see structures, as shown in Figure 14b. Their spatial closeness acts as primary identification criterion, as similar row/column vectors are grouped closely. From this, their spatial closeness, describing the multi-step alignment, supports discovering related users or topics discussed simultaneously, but also latent connections. Distinct orderings can be applied separately to nodes and hyperedges, for example, to either favor overall similarity (cosine) or matching parts (Jaccard). For other requirements, it is also possible to include different metrics. To reduce noise and exclude weak connections, the top menu allows to set a threshold for the connection strength for historical and predicted data. A flag controls the ordering mode to either respect the filtered or the full dataset (including filtered elements). To further structure the view, the experts can manually click and select to group users and topics to reflect their mental categorization of users and topics. This allows to reflect domain-specific ontologies (e.g., similar concepts) or represent known formations of users.

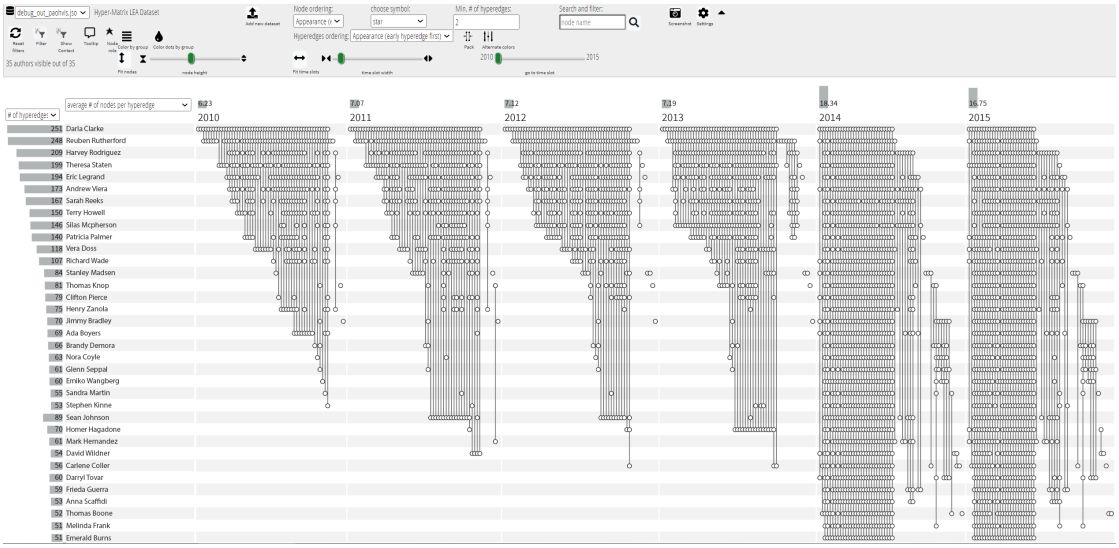
Zooming to the lower visualization levels shows the temporal development. Compared to existing approaches (see Figure 14c) our technique (Figure 14d) increases the scalability and comparability for dense temporal evolution. Compared to the industry standard Figure 14a, presenting the temporal evolution as a timeline-like arrow within each cell reduces comparison distances. Levels 5 and 6 allow an expert to understand the actual data on which a predicted connection is based: The main keywords of the relevant text fragments and, respectively, the actual raw text fragments (cf. Figure 11). This ability allows the expert to verify predictions and detect shortcomings as, for example, irony and coded synonyms are still difficult to be detected automatically. If the expert has identified shortcomings on any level, e.g., missing connections or wrong attribution of an ambiguous term, the technique allows for the inclusion of this additional domain knowledge. To externalize knowledge, the expert selects the corresponding connection



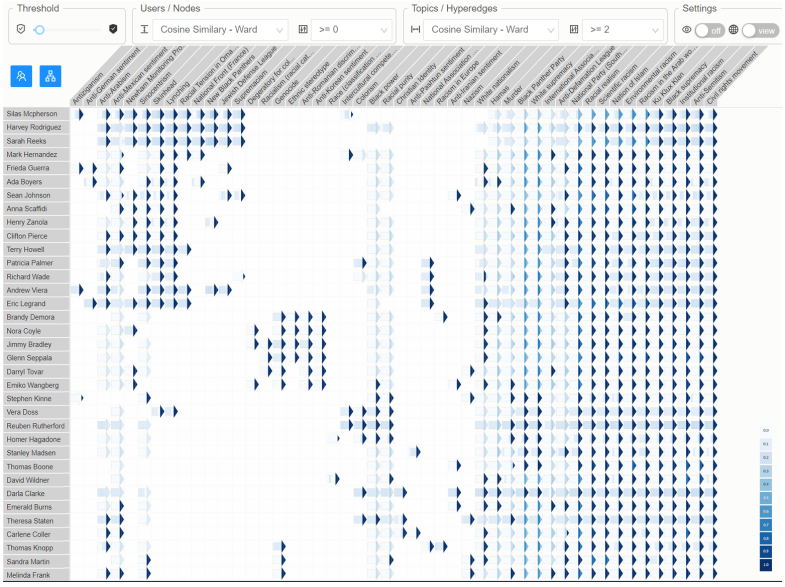
(a) IBM i2 Analyst's Notebook. Automatically generated graph representation from the hypergraph model displaying the connections (labels removed) for the furthest predicted year using a modified bipartite representation. Data-wise, this can be compared to the connectivity information in our Levels 1 and 2. Clutter and occlusion prevent a meaningful global analysis, and while individual users and topics can be explored, this is only possible slowly, not without difficulty, and likely requires moving entities around to identify connections safely.



(b) Our technique at Level 2, showing the same predicted connectivity *information as Analyst's Notebook in Figure 14a. Clusters and related users/topics can be pinpointed more easily. The color scheme and filtering settings in the top menu bar also facilitate to identify the prediction strength, which can be estimated by using the overlayed legend in the bottom right corner. The blue buttons allow to access the partition hierarchy modifier to view a dendrogram view of the grouped entities.*



(c) PAOHvis [172]. The temporal hypergraph evolution shows the individual hyperedges, allowing to find connected users and topics. However, the hypergraph size is at the upper limit for a feasible visualization, already leading to some cluttering. Also, due to the temporal splitting, the comparability between years is hindered for such complex, non-sparse hyperedges compared to our technique, but better suited for comparing topics in the same year.



(d) Our technique at Level 3, showing the same temporal evolution *information as PAOHvis in Figure 14c. The scalability is increased, showing no occlusion and the comparability of trends (important for the case study) is improved. This is due to retained cell ordering and short comparison distance. The downside is a reduced comparability between topics in the same year. The nature of the predictions depend on the model.*

Figure 14: Case study comparison of different approaches using the same internet forum hypergraph model dataset and exactly the same data view (connection strength > 0.1*, min. 2 hyperedges). Compared are the state-of-the-art industry solution IBM i2 Analyst's Notebook (Figure 14a), PAOHvis (Figure 14c) against our technique, showing the information at two different levels of abstraction (Figures 14b and 14d). Further, both external approaches only support a fixed network while our technique allows for an interactive refinement and domain knowledge integration.*

and specifies the proposed strength on a scale between 0 and 1. This translates to definite knowledge about no and guaranteed connection, respectively. More nuanced values like .7 allow the expert to reflect his own uncertainty. This allows them to try out hunches while simultaneously preserving some model flexibility. For this reason, the change preview (cf. Figure 13b) is extremely relevant for the domain experts, as it allows them to see directly how their knowledge transforms the model prior to accepting the changes. They can explore the consequences by zooming and panning through all levels and correlate their findings with their intuition or other facts. If unsatisfied, they can go back. Otherwise, they can continue and repeat this visual analytics loop multiple times. This rapid feedback supports the expert in refining the model without being blind to the resulting consequences, but being able to control and explore the latest model state at all times. As the domain experts focus is on exploratory analysis the iterative refinement supports finding connections and missing links faster. With domain knowledge that is difficult to be integrated a priori, step-by-step changes are more understandable.

4.6 FORMATIVE EVALUATION

We performed formative evaluation sessions involving three domain experts (P1-P3). P1 is a criminal investigator working for a European law enforcement agency, having more than 30 years of experience, 20 years spent in digital and criminal investigations. His expertise includes communication and network analysis, familiarity with commercial systems like IBM i2 Analyst's Notebook [78], the graph visualization tool Gephi [17], as well as the large network analyzer Pajek [18, 19]. P2 works at the same agency in a different division, and has more than 20 years of experience in criminal investigations, specialized in group structure and content analysis. P3 is a senior project lead at a governmental research institute, studying analytical raw data analysis for more than ten years.

4.6.1 *Study Procedure*

The formative evaluation was conducted individually via remote screen sharing, taking about 60 minutes. For later review of these remote screen sharing sessions, they were recorded after receiving the formal consent of the experts. In the first 10 minutes a demo presented how to perform the visual analysis, explore and refine data and processes, and integrate domain knowledge in the search process and in the machine learning model. The next 30 minutes were spent between the experts using the system and providing feedback, as well as additional on-demand demonstrations. The tasks the experts performed include overview, the identification of the most promising leads, and the drill-down through the different zoom-levels down to the actual raw content, in this case, communication data. Further, we demonstrated and debated the different interaction techniques, like cutoff values and thresholds, matrix sorting and reordering strategies, and the dynamically modifiable partition hierarchy, as well as the machine learning feedback process.

In the last 20 minutes, the authors interviewed the experts asking 32 prepared questions (see annexes). During each of the formative evaluation sessions, the experts engaged actively, trying out concepts, asking questions, commenting on the features, and pointing

out issues. If an expert already partially gave comments during the 30 minutes session, they were offered to extend their answer. For example, when an aversion or surprising idea was mentioned, we additionally focused on these aspects. The interview was designed to elicit aspects of our technique that the experts find relevant for their work or confusing or misinterpretable, as well as opinions on the individual approaches.

4.6.2 *Findings and Lessons Learned*

The **main observations** during the study are that our approach can effectively support most analytical requirements of the experts and that the experts favor both the **rapid exploration** of large datasets at different levels as well as the ability to integrate and contribute with their **domain knowledge**. This matches with their need to identify general trends in single combinations of users and topics and simultaneously identify co-occurrences. For this, the general prediction is more important than being able to identify differences between entities in the same year (cf. Figure 14). The underlying model we built upon [11] has proven to perform sufficiently well in this prediction task with an AUC (area under curve) of the ROC (receiver operating characteristic) of .88 and a recall value of .81. Excluded from the requirements are concepts outside the design scope, like purely mathematical capabilities as, for example, general centrality calculations, for which algorithms exist and could be included. In the following, we structure and summarize the main findings based on the expert's interactions and comments.

The domain experts agree that our approach of structuring information in **multiple levels** of details, using a **matrix-based approach**, is novel and therefore is not used in practice in their domain. For example, so far P1 has worked with either text-based or graph-based tools, and thinks our approach can "perfectly complement" existing workflows. The experts highlight the ability to effortlessly explore so much information (cf. P3), thereby "saving time" (P1), enabling a "quick analysis" (P3), while providing a "great overview ... with much details, ... but without overloading" (P2) the analyst, with an ease that is unexpected, given previous experience with this amount of data (cf. P2). We observed, that the experts often switch between the levels for targeting (upper levels) and then exploration and confirmation (lower levels). As P2 notes, this increases the size limit of the visually analyzable graph models, enhancing upon existing systems. "Together with the search capability" (P1), this allows for a very flexible workflow, enabling a good overview even for larger datasets.

The initial overview visualizations (Levels 1 and 2) are welcomed for providing a fast overview (cf. P1). The **color scheme** in Level 2 is regarded as comprehensible without explanation and aligning with expectations (cf. P1). It helps to provide guidance "where to start" (P1), and supports analysts in "planing their actions" (P3). To make the color scheme absolutely comparable, P3 requested the addition of a color legend. The **glyphs** are appreciated for providing details on the temporal distribution and future predictions (P2, P3). The glyph-based arrow representation in Levels 3 and 4 is appreciated for providing details on the temporal distribution (cf. P2, P3) interesting to the experts, and, most importantly, "the future predictions" (P2) in context of the historical data. Depicting future predictions in the arrowhead and the past data in the shaft, and seeing both together was described as "helpful" (P3). The alignment by fixed timesteps, like years, is regarded

as precise and practicable (cf. P1) by the experts. In comparison, the distribution as line chart in Level 4 received mixed responses, with P1 and P3 finding it beneficial for their understanding to get a better, absolute reading, while P2 feels "it does not add much". The **keyword visualization** (Level 5) is regarded as fine for an abstract summary of the content visualization but could be extended (cf. P3). This layer, representing the "main connection" (P1) to the actual raw data, is important (cf. P1), and only shown when relevant in high zoom levels, "where the text content is relevant" (P1).

The ability to **search** through all underlying textual data and highlight matches in the views was received enthusiastically by all experts, as they can also transfer and fulfill some of their existing workflow, e.g., content- and text-based workflows, with our technique. It allows to explore global tendencies while enabling to query locally (cf. P2), not being distracted by other matches "not relevant at the moment" (P2).

While the visualization alone helps them already some ways, providing them "with improved degree of detail ... unknown so far" (P1), all the experts also agree that the **interaction concepts** constitute an essential and relevant part of the approach, "helping them with strategical and operational decision" (P1). The **matrix reordering strategies** significantly improving the visual clarity of the overview, are regarded as "very interesting" (P2), and enable the experts to detect "groups" (P1) as well as connections easily, allowing them to "quickly identify hotspots" (P2), while putting less emphasis on weak connections. This is regarded as very supportive, being rarely supported in analysis systems (cf. P1), "saving costs and time" (P1). We observed that the experts use this as system guidance. The **partition hierarchy** is regarded by all experts as "essential" (P2), with P3 describing it as a "core functionality". It allows grouping different model parts into physical concepts, applying structure comparable to existing mental models (cf. P2), improving the mental mapping. It "makes decision easier" (P3) and allows to "connect things" (P3).

The experts further describe that with existing tools, one major problem is that their **mental concepts** and models can "not [be integrated] enough" (P2) in the exploration, making it harder and less comprehensible. They notice that our approach supports them in three ways not present in existing tools: (1) the **interactive exploration** allowing to follow their instinct, (2) the modifiable **partition hierarchy** to express and capture their mental concepts, and, "most importantly" (P1), (3) the ability to integrate their **domain and external knowledge** directly in the model. While the experts wished that they could already "generate a report [... and] export single entries" (P2) as commercial systems do, they note the enormous conceptional benefits of our technique. They regard them as "optimal" (P1), as there "are concepts and knowledge that cannot be modeled with machine learning [alone]" (P1) and are not "available" (P1) in the data. This knowledge then "cannot be integrated so far" (P1), is often documented in the head of the domain expert or "on a post-it note on the desk" (P1), leading to a high risk of the knowledge being "lost" (P1) or not leveraged. According to P1, the **knowledge integration** is performed iteratively during exploration, which we also observed as the experts adding knowledge intermittently, beginning with their main suspects and then expanding, adding knowledge when necessary either from post-its or when reading a name triggers a memory. The experts think that our **feedback loop** contributes to their analysis (cf. P1), replacing and "perfectly complementing" (P1) existing workflows. They regard the ability to *interactively* insert their knowledge as versatile. P1 noted that inserting all

knowledge beforehand would be error-prone and "practically impossible" for larger datasets. To see "**validation** [possibilities] on changes" (P3) is especially important for vetting, and the change view is regarded as "very clear" (P1), allowing them a first glance, beneficial for prefiltering, steering and follow-up search guidance (cf. P1) to better divide their time for exploration. For improved usability P1 suggested to enable clicking to jump directly to the raw data in the change preview mode for validation. P1 regards the ability for a **global accept/reject** as sufficient for now, conceding that a partial accept could be explored in the future, although he does not see an immediate benefit. They state that the **0–1 scale** is "understandable and usable" (P1), but note that using the "5x5x5 system" (P1)—a commonly used police system based on letters A–E and 1–5 for source and intelligence evaluation [129]—would be immediately understood and universally accepted in the target domain. The approach allows them to integrate their domain knowledge on multiple levels, together with the ability to perform a "quick analysis" (P3) of "large amounts of information" (P2) "in a targeted" (P1), non-overloading manner. From the observations of the experts, we derived a set of **tentative tasks**, relevant in law enforcement: (1) finding linked users/topics, (2) connecting users which share related topics to identify co-conspirators, (3) using classical text-based search in the raw data to identify users, (4) finding and judging an in/decrease of user activity for a topic, (5) finding a temporal co-occurrence between topics and users, (6) adding domain knowledge to a specific user and specific topic and judging the implications, (7) transfer raw data patterns and identify related users, and (8) confirming the model predictions by cross-validation plausibility with the raw data texts.

4.7 DISCUSSION AND FUTURE WORK

During the evaluation, we received multiple proposals on how our approach could be extended further, including by mathematical analysis methods and industry-grade interfaces. In the following, we discuss the limitations and broader applicability of our approach, also in the context of future work. For our prototype approach, we adapted the generic blueprint of a machine learning model to the case study. This use case has its own limitation, requiring structured data with time and author information, and dependent on advanced topic extraction models. We tested our prototype successfully with 1 000 users, 800 topics, and 15 timesteps on an HD screen, typically the upper size for large investigations. In terms of data type, the technique can cope both with sparse and none-sparse matrix structures. For the former, the matrix reordering allows to prioritize more relevant connections and order them further on the top left, reducing the required screen usage for the main parts. Of course, a *homogeneous* sparse matrix does not benefit from that. In this case, and for none-sparse matrices, the different zoom levels shift the size limitations. Nevertheless, they do not scale infinitely. Scrolling would be needed when scaling further, even for the overview level. According to domain expert P1, there the primary concern would be the number of users (y-axis), but using the partition hierarchy and matrix reordering could partially mitigate the issue. When increasing the number of time steps, the arrow becomes more detailed, shifting from blocks to a more continuous stream, becoming less distinguishable. For our use case, this fine-grained time is not primarily relevant because the experts aim at seeing who has recently been interested in

a topic. However, it might become an issue when the task requires to extract detailed timestamps. Therefore one could use hovering, magnification on demand, or a more specialized visualization. Also, the visualization presented is better at analyzing trends and connectivity tasks on an overview level. Comparing the same time step in Levels 3 and beyond between two non-aligned nodes, however, becomes harder. For further work, we envision an adaptable overview layer showing a specific time point, allowing cross-cell comparability. When adapting to different use cases, some of the filtering methodology likely has to be changed. For example, when supporting biochemical process analysis, the raw attributes are not texts anymore, which (1) would need a different visualization for the content in the two lowest display level, but would also impact (2) the search functionality, which would need to be adapted to search and filter for biological and chemical properties instead of text. The discussed visualization components serve only as examples for the visual analytics workflow presented. When adapting to a different field, there exist manifold possibilities for extensions, by integrating domain-specific visualization components. We provision this by a modular view architecture, supporting independent layer modules. Further enhancements are multiple magic lenses to allow for simultaneous drill-down to different levels.

In the future, we envision improvements to the feedback system, for example, showing how domain knowledge propagates not only between two model states, i.e., before and after adding knowledge but also explaining the effects of previously introduced knowledge, for example, by interactively highlighting the individual influences on hover. This is supported by our architecture, but the computation time scales linearly with the number of domain knowledge inputs, which leads to computation times of several minutes and more, making it infeasible in an interactive environment for fast iterations. We hope to improve this by enhanced engineering, reducing the model setup and reloading times by advanced ways of updating the hypergraph model.

4.8 CONCLUSION

Many processes are difficult to describe using traditional graph-based concepts and benefit from more precise yet more complex modeling as temporal hypergraphs. We address this challenge by using a geometric deep learning approach and extend it to hypergraphs. However, such deep learning models typically do not incorporate domain knowledge, usually unavailable in the data. This is not least because domain experts struggle to articulate their knowledge without rapid, iterative feedback and intuitive representations matching their mental models, alternatively requiring a detailed a priori understanding of the problem. Hence, domain expertise is often not leveraged to its full potential.

We contribute a technique, named HYPER-MATRIX, to make temporal hypergraph model exploration more accessible for domain experts by enabling the integration of domain knowledge into the process and support their mental models through a multi-level matrix-based visualization architecture. The technique enables the interactive evaluation and seamless refinement of such models while providing a tight coupling and rapid, iterative feedback cycles to the underlying machine learning model. Model changes in response to the integration of domain knowledge are visualized transparently by a change

preview, allowing experts to foster a more detailed understanding of how the underlying model works while externalizing their knowledge to teach the machine.

The approach allows to swiftly explore vast search spaces while maintaining focus and eliminating demanding context switches. Drill-down capabilities across multiple levels allow studying details and model contents on demand while retaining the overview. This architecture facilitates a focused analysis of relevant model aspects, allowing experts to detect patterns more rapidly and accurately. It is complemented by interactive filtering and search, various matrix reordering techniques, and a dynamically modifiable partition hierarchy, allowing the integration of domain knowledge in the visualization layers.

We evaluate our approach in one case study and through formative evaluation with law enforcement experts using real-world communication data. The results show that our approach surpasses existing solutions in terms of scalability and applicability, enabling the incorporation of domain knowledge and allowing fast and targeted search-space traversal. While we focused on topic prediction for law enforcement as driving application, the interactions and concepts work with any temporal hypergraph, being model agnostic and applicable more generically to a wider variety of domains. With our technique, we hope to pave the way for domain experts to a more interactive exploration and refinement of temporal hypergraph models, enabling them to use their knowledge not only for steering but also to articulate it into the machine learning model.

5

HYPERLEARN: A DISTRIBUTED APPROACH FOR REPRESENTATION LEARNING IN DATASETS WITH MANY MODALITIES

Multimodal datasets contain an enormous amount of relational information, which grows exponentially with the introduction of new modalities. Learning representations in such a scenario is inherently complex due to the presence of multiple heterogeneous information channels. These channels can encode both (a) inter-relations between the items of different modalities and (b) intra-relations between the items of the same modality. Encoding multimedia items into a continuous low-dimensional semantic space such that both types of relations are captured and preserved is extremely challenging, especially if the goal is a unified end-to-end learning framework. The two key challenges that need to be addressed are: 1) the framework must be able to merge complex intra and inter relations without losing any valuable information and 2) the learning model should be invariant to the addition of new and potentially very different modalities. In this paper, we propose a flexible framework which can scale to data streams from many modalities. To that end we introduce a hypergraph-based model for data representation and deploy Graph Convolutional Networks to fuse relational information within and across modalities. Our approach provides an efficient solution for distributing otherwise extremely computationally expensive or even unfeasible training processes across multiple-GPUs, without any sacrifices in accuracy. Moreover, adding new modalities to our model requires only an additional GPU unit keeping the computational time unchanged, which brings representation learning to truly multimodal datasets. We demonstrate the feasibility of our approach in the experiments on multimedia datasets featuring second, third and fourth order relations.

5.1 INTRODUCTION

The field of multimedia has been slowly, but steadily growing beyond simple combining of diverse modalities, such as text, audio and video, to modeling their complex relations and interactions. These relations are commonly perceived, and therefore, modelled as only pair-wise connections between two items, which is a major drawback in the majority of the existing techniques. Going beyond pair-wise connections to encode higher-order relations can not only discover complex inter-dependencies between items but also help in removing ambiguous relations. For instance: in the task of social image-tag refinement, conventional approaches focus on exploiting the pairwise tag-image relations, without considering the user information which has been proven extremely useful in resolving tag ambiguities and closing the semantic gap between visual representation and semantic

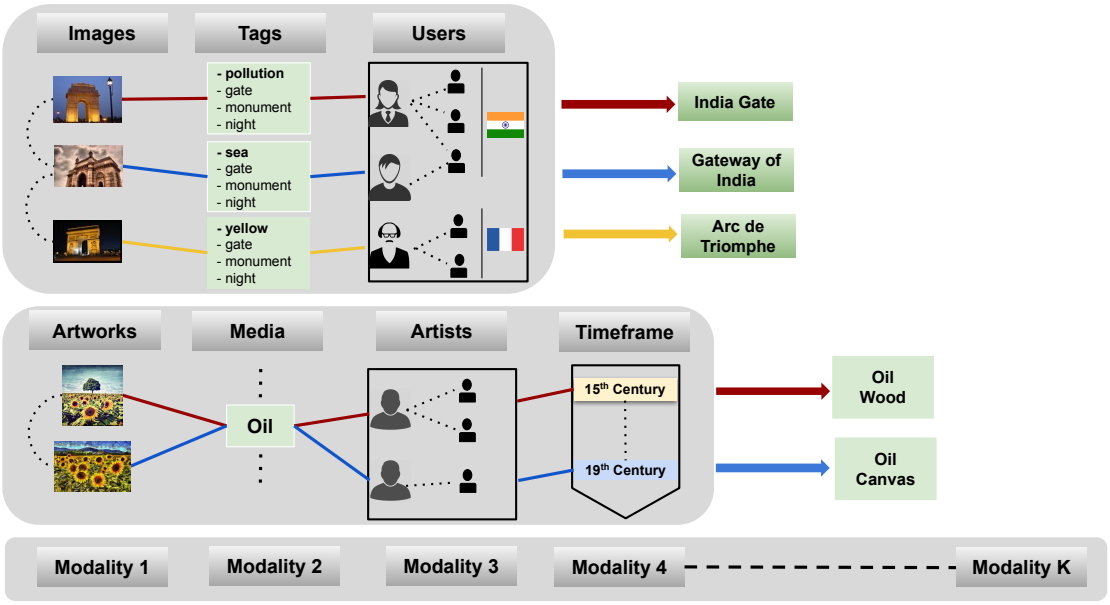


Figure 15: Example showing importance of capturing ternary relations (images-tags-users) in a social network dataset and quaternary relations (artworks-media-artists-timeframe) in artistic dataset. HyperLearn exploits such relations to learn complex representations for each modality. At the same time, HyperLearn provides a distributed learning approach, which makes it scalable to datasets with many modalities

meaning [42, 108, 165, 166]. It is hence an interesting, but far more challenging problem in multimedia to exploit and learn higher-order relations to be able to (a) learn a better representation for each item, (b) improve pairwise retrieval tasks and (c) discover far more complex relations which can be ternary (3^{rd} order), quaternary (4^{th} order), quinary (5^{th} order) or even beyond. As examples, figure 15 shows the importance of modeling higher-order relations in social networks and in artistic analysis respectively. In the upper example from Figure 15, textual annotations and information about user demographics is utilized for disambiguation between landmarks with very similar visual appearance. Similarly, the second example illustrates quaternary relations formed by the artworks, media, artists and the time-frame in which they were active. Capturing such complex relations is of utmost importance in a number of tasks performed by the domain experts, such as author attribution, influence and appreciation analysis.

Learning representations in multimodal datasets is an extremely complex task due to the enormous amount of relational information available. At the same time most of these relations have an innate property of 'homophily', which is the fact that similarity breeds connections. Exploiting this property of similarities can help in immensely simplifying the understanding of these relations. These similarities can be derived from both intra relations between items of the same modality and inter-relations between items across different modalities. Unifying the two types of relations in a complementary manner has the potential to bolster the performance of practically any multimedia task. Thus, in this work we propose an efficient learning framework that can merge information generated by both intra as well as inter-relations in datasets with many modalities. We conjecture that such an approach can pave the way for a generic methodology for learning

representations by exploiting higher-order relations. At the same time, we introduce an approach that makes our framework scale to multiple modalities.

We focus on learning a low-dimensional representation for each multimodal item using an unsupervised framework. The unsupervised methods utilize relational information both within as well as across modalities to learn common representations for a given multimodal dataset. The co-occurrence information simply means that two items from different modalities are semantically similar if they co-exist in a multimedia collection. For example, the textual description of a video often describes the events shown in the visual channel. Many of the multimedia tasks revolve around this compact latent representation of each multimodal entity [130, 139]. The major challenge lies in bridging the learning gap between the two types of relations in a way that they can be semantically complementary in describing similar concepts. Learning representation is usually extremely expensive, both in computational time and required storage as even a relatively small multimedia collection normally contains a multitude of complex relations.

Handling a large amount of relations requires a framework with a flexible approach to training across multiple pipelines. Most of the existing algorithms fail in parallelizing their framework into separate pipelines [109, 194], resulting in large time and memory consumption. Thus, in the proposed framework we can parallelize the training process for different modalities into separate pipelines, each requiring just an additional GPU core. By doing so, we facilitate joint multimodal representation learning on highly heterogeneous multimedia collections containing an arbitrarily large number of modalities, effectively hitting an elusive target sought after since the early days of multimedia research. The points below highlight the contributions of this paper:

- We address the challenging problem of multimodal representation learning by proposing HyperLearn, an unsupervised framework capable of jointly modeling relations between the items of the same modality, as well as across different modalities.
- Based on the concept of geometric deep learning on hypergraphs, our HyperLearn framework is effective in extracting higher-order relations in multimodal datasets.
- In order to reduce prohibitively high computational costs associated with multimodal representation learning, in this work we propose a distributed learning approach, which can be parallelized across multiple GPUs without harming the accuracy. Moreover, introducing a new modality into HyperLearn framework requires only an additional GPU, which makes it scalable to datasets with many modalities.
- Extensive experimentation shows that our approach is task-independent, with a potential for deployment in a variety of applications and multimedia collections.

5.2 RELATED WORK

The core challenge in multimodal learning revolves around learning representations that can process and relate information from multiple heterogeneous modalities. Most of existing multimodal representation learning methods can be split into two broad categories

– multimodal network embeddings and tensor factorization-based latent representation learning. In this section we reflect on the representative approaches from these two categories. Since, in this work we extend the notion of graph convolution networks for multimodal datasets, we also touch upon some of the existing techniques that aim to deploy deep learning on graphs.

5.2.1 *Multimodal Network Embedding*

A common strategy for representation learning is to project different modalities together into a joint feature space. Traditional methods [119, 142, 168] focus on generating node embeddings by constructing an affinity graph on the nodes and then finding the leading eigenvectors for representing each node. With the advent of deep learning, neural networks have become a popular way to construct combined representations. They owe their popularity to the ability to jointly learn high-quality representations in an end-to-end manner. For example, Srivastava and Salakhutdinov proposed an approach for learning higher-level representation for multiple modalities, such as images and texts using Deep Boltzmann Machines (DBM) [156]. Since then a large number of multimodal representation learning methods based on deep learning have been proposed. Some of these methods attempt to learn a multimodal network embedding by combining the content and link information [34,74,103,109,164,194,199]. Other set of methods focuses on modeling the correlation between multiple modalities to learn a shared representation of multimedia items. An example of such coordinated representation is Deep Canonical Correlation Analysis (DCCA) that aims to find a non-linear mapping that maximizes the correlation between the mapped vectors from the two modalities [192]. Ambiguities often occur while using network embedding methods to learn multimodal relations due to sub-optimal usage of available information. This is mostly because these methods assume relations between items to be pairwise which often leads to loss of information [12, 32, 100].

5.2.2 *Tensor Factorization Based Latent Representation Learning*

Decoupling a multidimensional tensor into its factor matrices has been proven successful in unraveling latent representations of their components in an unsupervised manner [93, 96, 128]. Most existing approaches aim to embed both entities and relations into a low-dimensional space for tasks such as link prediction [170], reasoning in knowledge bases [153] or multi-label classification problems [117]. Recent methods on social image understanding incorporate user information as the third modality for tag based image retrieval and image-tag refinement problems [162, 165, 166]. Even though most of these approaches are suitable for large datasets, one of the main disadvantages of using a factorization based model is the lack of flexibility when scaling to highly multidimensional datasets. Additionally, most of the tensor decomposition methods are based on the optimization with a least squared criterion, which severely lacks robustness to outliers [88].

In this work, we first overcome the issues of network embedding methods by using a hypergraph-based learning method. Secondly, we introduce a scalable approach to

tensor decomposition for scaling representation learning to many modalities. Finally, we can combine the advantages of rich information from network structure with the unsupervised nature of tensor decomposition in one single end-to-end framework.

5.2.3 *Geometric Deep Learning on graphs*

Geometric deep learning [30] brings the algorithms that can help learn from non-euclidean data like graphs and 3D objects by proposing an ordering of mathematical operators that is different from common convolutional networks. The aim of Geometric Deep Learning is to process signals defined on the vertices of an undirected graph $\mathbb{G}(V,E,W)$, where V is the set of vertices, E is set of edges, and $W \in \mathbb{R}^{|V|\times|V|}$ is the adjacency matrix. Following [43, 149], spectral domain convolution of signals x and y defined on the vertices of a graph is formulated as:

$$x \circledast y = \Phi(\Phi^T x).(\Phi^T y) = \Phi(\mathcal{F}(x).\mathcal{F}(y)) \tag{5.1}$$

Here, $\Phi^T x$ corresponds to Graph Fourier Transform and $\mathcal{F}(.)$ represents Fourier Transform; the eigen functions Φ of the graph laplacian play the role of Fourier modes; the corresponding eigenvalues Λ of the graph laplacian are identified as the frequencies of the graph. Recent applications of graph convolutional networks range from computer graphics [29] to chemistry [48]. The spectral graph convolutional neural networks (*GCN*), originally proposed in [31] and extended in [43] were proven effective in classification of handwritten digits and news texts. A simplification of the *GCN* formulation was proposed in [91] for semi-supervised classification of nodes in a graph. In the computer vision community, *GCN* has been extended to describe shapes in different human poses [118], perform action detection in videos [177] and for image and 3D shape analysis [125]. However, in the multimedia field there have been considerably less examples of using deep learning on graphs for modeling highly multimodal datasets with [12, 143] as notable exceptions.

In this paper, we propose an approach that introduces the application of graph convolutional networks on multimodal datasets. We deploy Multi-Graph Convolution Network (MGCNN) originally proposed by [125] for the matrix completion task using row and column graphs as auxiliary information. It aims at extracting spatial features from a matrix by using information from both the row and column graphs. For a matrix $X \in \mathbb{R}^{N_1\times N_2}$, MGCNN is given by

$$\widetilde{X} = \sum_{j,j'=0}^{q} \theta_{jj'} T_j(\mathbb{L}_r) X T_{j'}(\mathbb{L}_c) \tag{5.2}$$

where, $\Theta = \theta_{jj'}$ is $(q+1)\times(q+1)$ dimensional matrix which represents the coefficients of the filters, $T_j(.)$ denotes the Chebyshev polynomial of degree j and $\mathbb{L}_r$, $\mathbb{L}_c$ are the row and column Graph Laplacians respectively. Using Equation 5.2 as the convolutional layer of MGCNN, it produces q output channels ($N_1 \times N_2 \times q$) for matrix $X \in \mathbb{R}^{N_1\times N_2}$ with a single input channel. In this way, one can extract q dimensional features for each item in matrix X by combining information from row and column graphs, which can correspond to e.g. individual modalities.

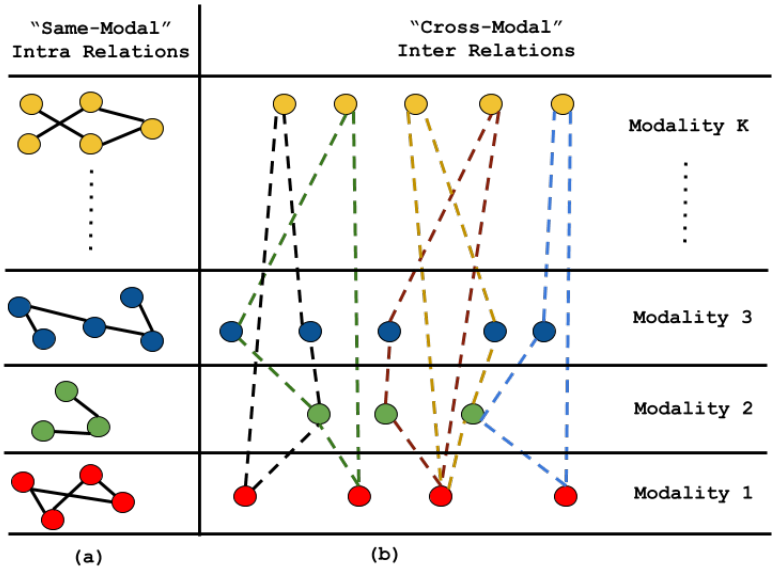


Figure 16: (a) Pair-wise relationship among the items of the same modality in K-modal data. (b) Complex higher-order heterogeneous relationships between entities of different modalities using a Hypergraph representation.

5.3 THE PROPOSED FRAMEWORK

In this section, we propose a novel distributed learning framework that can simultaneously exploit both intra and inter-relations in multimodal datasets. We depict these inter-relations on a hypergraph and conjecture that this way of representing higher-order relations reduces any loss of information contained within the multimodal network structure [12,100,206]. Mathematically, a hypergraph is depicted by its adjacency tensor [16]. A simple tensor factorization on this adjacency tensor can disentangle modalities into their compact representations. However, this kind of representation lacks information from the intra-relation of items belonging to the same modality. Subsequently, we therefore incorporate intra-relations among entities as auxiliary information to facilitate flow of within-modal relationship information.

5.3.1 *Notations*

We use boldface underlined letters, such as $\underline{\boldsymbol{X}}$, to denote tensors and simple upper case letters, such as U, to denote matrices. Let $\odot$ represent the "Khatri-Rao" product [85] defined as

$$U \odot V = (U_{ij} \otimes V_{ij})_{ij} \tag{5.3}$$

where, $U \in \mathbb{R}^{L \times R}$ and $V \in \mathbb{R}^{M \times R}$ are arbitrary matrices and $\otimes$ is the Kronecker Product. The resulting matrix $U \odot V$ is an expanded matrix of dimension $LM \times R$ on the columns of U and V.

5.3.2 *Representing Cross-Modal Inter-Relations using Hypergraphs*

Hypergraphs have been proven extremely efficient in depicting higher-order and heterogeneous relations. A hypergraph is the most efficient way to represent complex relationships between a multitude of diverse entities, as it minimizes any loss of available information [12, 32, 185]. Given multimodal data, we construct a unified hypergraph $\mathcal{H}(V, E)$ by building hyperedges (E) around each of the individual multimodal items which are represented on a set of nodes (V). These hyperedges correspond to the cross-modal relations between items of different modalities as illustrated in Figure 16.

A more formalised mathematical interpretation of this unified hypergraph is given by its adjacency tensor $\underline{\boldsymbol{X}}$, where the number of components of the tensor is equal to the number of modalities in the hypergraph. Further, each hyperedge corresponds to an entry in the tensor whose value are the weights of the hyperedge. For simplicity, in this work we focus on unweighted hypergraphs.

Thus, a multimodal data with K modalities is depicted on a tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{N_1 \times N_2 ..... \times N_K}$, where each component N_θ $(1 \leq \theta \leq K)$ of this tensor represents one of the K heterogeneous modalities. A single element $\underline{\boldsymbol{x}}$ of $\underline{\boldsymbol{X}}$ is addressed by providing its precise position by a series of indices $n_1, n_2, .., n_K$ i.e.

$$\underline{\boldsymbol{x}}_{n_1 n_2 .. n_K} \equiv \underline{\boldsymbol{X}}_{n_1 n_2 .. n_K}; \quad 1 \leq n_1 \leq |N_1|, ..., 1 \leq n_K \leq |N_K| \tag{5.4}$$

Further, a hyperedge around a set of nodes can be represented as binary values such that $\underline{\boldsymbol{x}}_{n_1 n_2 .. n_K} = 1$ if the relation $(n_1, n_2, ..., n_K)$ is known i.e. if there exists a mutual relation between the K modalities for that instance. For example, in the social network use case, with a possible corresponding image-tag-user associated tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, the images (n_1) are represented on rows, users (n_2) on columns and tags (n_3) on tubes. If the l^{th} image uploaded by the m^{th} user is annotated with the n^{th} tag, then $\underline{\boldsymbol{x}}_{lmn} = 1$ and 0 otherwise.

5.3.3 *Representing Intra-Relations Between the Items of the Same Modality*

Relationships between items of the same modality are dependent on the nature/properties of the modality. For instance, relationships between users in a social network is defined based on their common interests. To make our framework flexible, each modality $(\theta; 1 \leq \theta \leq K)$ is represented on a separate graph $\boldsymbol{G_\theta}$ whose connections can be defined independently. For example: relations among images can be established based on their visual features, for tags it can be calculated based on their co-occurrence and for users it can very well be based on their mutual likes/dislikes. We denote the adjacency matrix of $\boldsymbol{G_\theta}$ by Λ_θ where each of its entries $\Lambda_\theta^{i,j} = 1$, if there exists a relation between the i^{th} and j^{th} element and 0 otherwise. The corresponding normalized graph laplacians $(\mathbb{L}_\theta)$ are given by

$$\mathbb{L}_\theta = D_\theta^{\frac{1}{2}} \Lambda_\theta D_\theta^{-\frac{1}{2}} \tag{5.5}$$

where, $D_\theta = diag(\sum_{j \neq i} \Lambda_\theta^{i,j})$ is known as the degree matrix.

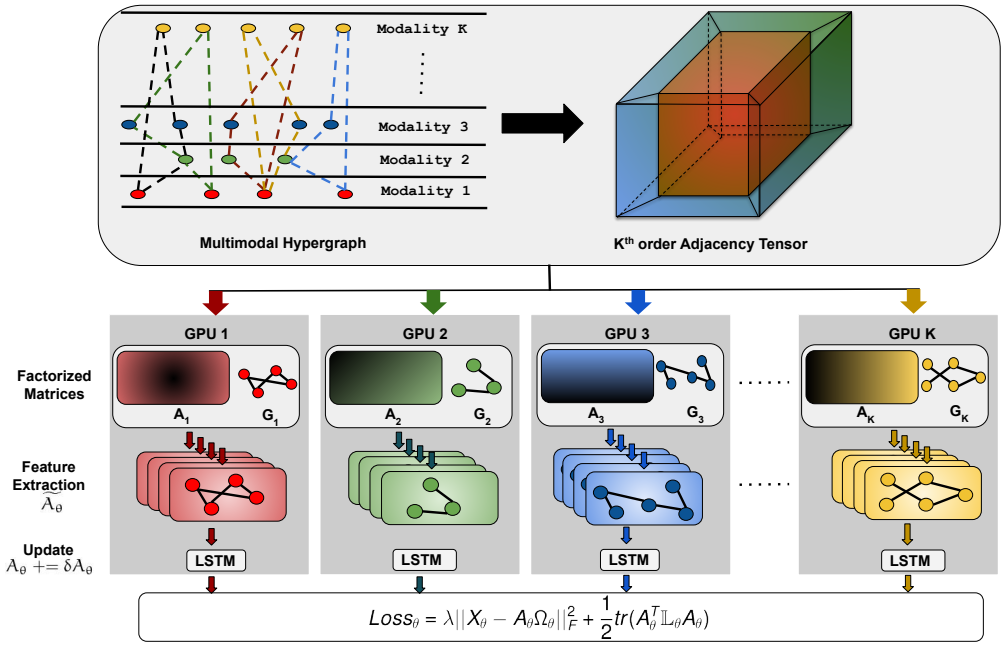


Figure 17: Proposed HyperLearn framework deployed on K modalities with a distributed learning approach

5.3.4 *Combined Inter-Intra Relational Feature Extraction*

Tensor $\underline{\boldsymbol{X}}$ can be factorized using Candecomp/Parafac(CP) - decomposition [67] which decomposes a tensor into a sum of outer products of vectors $(a_r^{(\theta)})$.

The CP-decomposition $\widetilde{\underline{\boldsymbol{X}}}$ of $\underline{\boldsymbol{X}}$ is defined as

$$\widetilde{\boldsymbol{X}} = \sum_{r=1}^{R} a_r^{(1)} \circ a_r^{(2)} \circ ... \circ a_r^{(K)} \tag{5.6}$$

$$= \boldsymbol{I} \times_1 A_1 \times_2 A_2 \times_3 ... \times_K A_K \tag{5.7}$$

where $\circ$ is the outer product and $\times_i$ represents mode-i multiplication (Tensor matrix product). Matrices $A_\theta \in \mathbb{R}^{|N_\theta| \times R}$ are called factor matrices of rank R and $\boldsymbol{I}$ is an R^{th} order identity tensor. Matrices A_θ are essentially the latent lower dimensional representations for each of the N_θ components of the tensor and therefore, for each of the K modalities.

Subsequently, we introduce an approach that can learn robust representations A_θ by combining intra relational information. We extract spatial features that merge information from each of the graphs $\boldsymbol{G_\theta}$ with the latent representation matrices A_θ using Multi-Graph Convolutional Network (MGCNN) layers given by

$$\widetilde{A_\theta} = \sum_{j,j'=0}^{q} \theta_{jj'} T_j(\mathbb{L}_\theta) A_\theta \tag{5.8}$$

where, the output $\widetilde{A_\theta} \in \mathbb{R}^{|N_\theta| \times R \times q}$ has q output channels. Similar to [125], we use an $LSTM$ to implement the feature diffusion process which essentially iteratively predicts accurate changes δA_θ for the matrix A_θ. Due to its ability to keep long-term internal

states, this *LSTM* architecture is highly efficient in learning complex non-linear diffusion processes.

5.3.5 *Loss Function Incorporating Cross-Modality Inter-Relations and Within-Modality Intra Relations*

In standard CP decomposition of a tensor, its factor matrices are approximated by finding a solution to the following equation

$$\min_{A_1,..,A_K} \|\underline{\boldsymbol{X}} - (\boldsymbol{I} \times_1 A_1 \times_2 A_2 \times_3 ... \times_K A_K)\|_F^2 \tag{5.9}$$

This equation essentially tries to find low dimensional factor matrices A_θ such that their combination is as close as possible to the original tensor $\underline{\boldsymbol{X}}$. Further, to add relational information among items within each of these A_θ, we extend the "within-mode" regularization term introduced in [105] for matrices and [128] for third order tensors to generic K^{th} order tensors. The basic idea is to add a regularization term to Equation 5.9 such that it can force two similar objects in each modality to have similar factors, so that they operate similarly. Thus, the combined loss function is given by:

$$\min_{A_1,..,A_K} \frac{1}{2}(tr \sum_{\theta=1}^{K} A_\theta^T \mathbb{L}_\theta A_\theta) + \lambda \|\underline{\boldsymbol{X}} - (\boldsymbol{I} \times_1 A_1 \times_2 .. \times_K A_K)\|_F^2 \tag{5.10}$$

where, $tr(.)$ returns the trace of a matrix. In Equation 5.10, the first term ensures closeness between items of the same modalities and the second term consolidates the relative similarities between items across modalities. Minimizing Equation 5.10 is a non-convex optimization problem for a set of variables $A_1, .., A_K$. Apart from being an NP-hard problem, computationally it is also expensive to perform even simple operations like element wise product on a K^{th} order tensor. To get a more robust solution, we introduce an alternating method to tensor decomposition similar to [88, 93]. The key insight of such a method is to iteratively solve one of the K components of the tensor while keeping the rest fixed. We exploit this kind of alternating optimization solution to parallelize our framework across multiple GPUs, by placing each modality on one of them. This creates an independent pipeline for all of the K modalities as shown in Figure 17 which summarizes our distributed learning framework for multimodal datasets.

5.3.6 *Distributed Training Approach for Learning Latent Representations*

The separable feature extraction process for each modality makes our methodology unique and scalable to multiple modalities. These separate pipelines are combined by a joint loss function. Consider solving Equation 5.10 by keeping all other components except N_{θ_0} as constant. Since, all but one component of the tensor is a variable, unfolding original tensor $\underline{\boldsymbol{X}}$ into a matrix along the N_{θ_0} component results in matrix $X_\theta^{(0)}$ with dimensions $|N_{\theta_0}| \times |N_1 N_2 .....N_\theta|$ (where $1 \leq \theta \leq K \quad s.t. \quad \theta \neq \theta_0$). So, the loss function in Equation 5.10 can be rewritten for each of the K components (N_θ) as

Table 4: Table showing the total number of intra and inter-relations between items on MovieLens, MIR Flickr and OmniArt datasets.

Movie Lens	$\mathscr{R}$(U) 12,594	$\mathscr{R}$(M) 28,928	$\mathscr{R}$(U-M) 100,000		
MIR Flickr	$\mathscr{R}$(I) 93,695,167	$\mathscr{R}$(T) 25,170	$\mathscr{R}$(U) 9,900,716	$\mathscr{R}$(I-T-U) 48,760	
Omni Art	$\mathscr{R}$(I) 4,628,009	$\mathscr{R}$(A) 849,482	$\mathscr{R}$(M) 21,178	$\mathscr{R}(T_f)$ 144	$\mathscr{R}$(I-A-M-T_f) 28,399

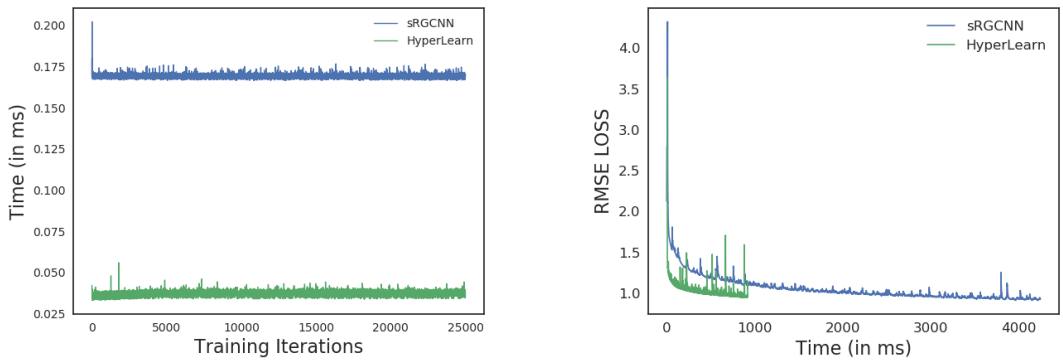
$$Loss_\theta = \lambda \|X_\theta - A_\theta \Omega_\theta\|_F^2 + \frac{1}{2} tr(A_\theta^T \mathbb{L}_\theta A_\theta) \tag{5.11}$$

, where $\Omega_\theta = A_1 \odot A_2 \odot .. \odot A_{\theta-1} \odot A_{\theta+1} .. \odot A_K$ and $\odot$ represents the "Khatri-Rao" product.

5.4 EXPERIMENTS

We start our experimental evaluation showing the performance of our approach on a 2-dimensional standard matrix completion task and then extend it to 3 and 4 dimensional cases. For 2D, 3D and 4D case, we use MovieLens [121], MIR Flickr [76] and OmniArt [158] datasets respectively. We conjecture that our framework can be generalized to datasets with even more modalities. Table 4 summarizes the number of inter and intra relations for the three above mentioned cases. Here, $\mathscr{R}(.)$ represents the number of relations. As seen from the table, even relatively small datasets feature a multitude of relations, which makes learning them even more challenging.

5.4.1 *Task 1: Matrix Completion on Graphs*



(a) Time (in ms) taken for each training iteration

(b) Convergence rate of RMSE Loss over time

Figure 18: Illustration of the convergence rate of HyperLearn against sRGCNN. Our method clearly requires a much lower training time per iteration and also converges much faster than sRGCNN.

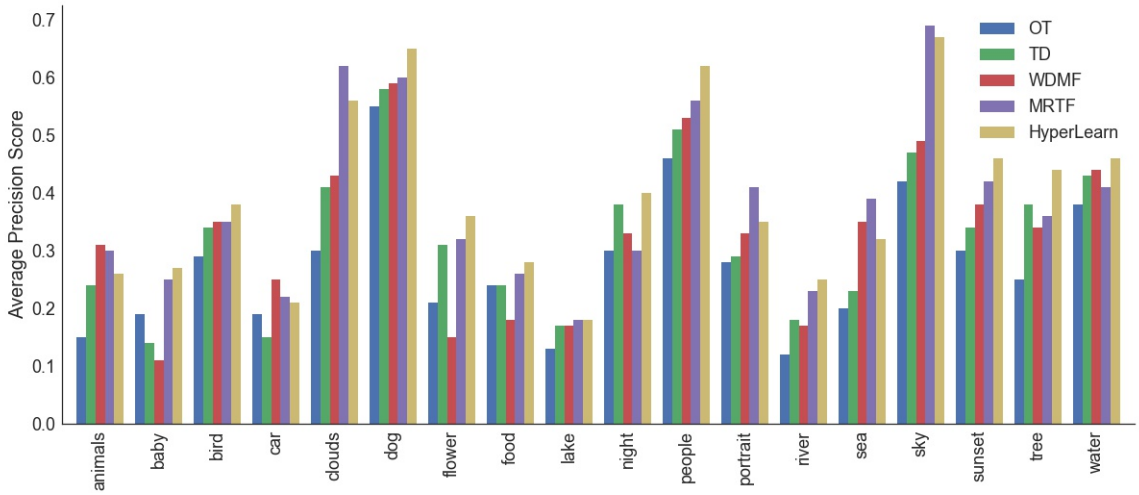


Figure 19: Detailed performance comparison in terms of Average Precision over 18 concepts on the MIRFlickr dataset

We show the computational advantage of our approach against a matrix completion method that makes use of side information as a baseline. For this, we use the standard MovieLens 100K dataset [121], which consists of 100,000 ratings on a scale of 0 to 5 corresponding to 943 users (U) and for 1,682 movies (M). We follow the experimental setup of Monti et.al. [125] for constructing the respective user and movie intra-relation graphs as unweighted 10-nearest neighbor graphs.

We compare the performance of HyperLearn with separable Recurrent Graph Convolutional Networks (sRGCNN) as proposed in [125]. As can be seen from Figure 18, our approach attains comparable performance to the state of the art alternative, while being much faster. The feature extraction approach alternating between movie and item graphs reduces the time complexity (although not linearly) considerably as can be seen from Figure 18(a) which in turn increases the rate of convergence for the algorithm as depicted in Figure 18(b). However, due to continuous alternating loss calculations, sometimes the back-propagated gradients tend to get biased towards one of the modalities resulting in some higher peaks for HyperLearn in Figure 18(b).

5.4.2 *Task 2: Social Image Understanding*

In this experiment, we test the performance of our model on a 3rd order multimodal relational dataset. We apply our method to uncover latent image representations by jointly exploring user-provided tagging information, visual features of images and user demographics. We conduct experiments on the social image dataset: MIR Flickr [76]. The MIR Flickr dataset consists of 25,000 images (I) from Flickr posted by 6,386 users (U) with over 50,000 user-provided tags (T) in total. Some tags are obviously noisy and should be removed. Tags appearing at least 50 times are kept and the remaining ones are removed as in [108, 166]. To include user information, we crawl the groups joined by each user through the Flickr API. Some images have broken links, or are deleted

by their users. We remove such images from our dataset which leaves us with 15,662 images, 6,618 users and 315 tags. The dataset also provides manually-created ground truth image annotations at the semantic concept level. For this filtered dataset, there are 18 unique concepts such as animals, bird, sky etc. for the images which we adopt to evaluate the performance. We create an intra-relation graph for images by taking 10-nearest neighbors based on their widely used standard SIFT features. For users, we create edges between them if they joined the same groups and for tags a graph is created based on their co-occurrence.

To empirically evaluate the effectiveness of our proposed method, we present the performance of the latent representation of images in classifying them into 18 concepts. We compare our model with the following methods:

- **OT**: The user-provided tags from Flickr as baseline.
- **TD**: The conventional CANDECOMP/PARAFAC (CP) tensor decomposition [67]
- **WDMF**: Weakly-supervised Deep Matrix Factorization for Social Image Understanding [106]
- **MRTF**: Multi-correlation Regularized Tensor Factorization approach [144]

Table 5: Comparison of the training times (in hours) on MIR Flickr dataset

Model	Training Time (in hours)
WDMF	4.2 ± 0.4
MRTF	2.7 ± 0.3
HyperLearn	1.8 ± 0.3

These methods, to the best of our knowledge, cannot provide the flexibility of performing distributed training for each modality using multiple GPUs. We report Average Precision (AP) scores for comparing our HyperLearn approach against all of these methods. Average Precision (AP) is the standard measure used for multi-label classification benchmarks. It corresponds to the average of the precision at each position where a relevant image appears. Figure 19 shows the comparative performance for all the 18 concepts. We also compare HyperLearn with MRTF and WDMF in terms of the training time and report the results in Table 5. As can be seen from this table, HyperLearn executes faster than MRTF and WDMF while its performance is at par or even better for most of the concepts in the multi-label classification task shown in Figure 19.

Through this experiment we show that - (a) the performance of our approach is at par with the existing methods in understanding social image relationships (b) by introducing a distributed approach we can cut down training time of the model significantly.

5.4.3 *Task 3: OmniArt*

In the last experiment, we show the performance of our model in learning relations that go beyond 3rd order of connections. For this we require a highly multimodal dataset containing complex relations that are hard to interpret. One such dataset is OmniArt, a

large-scale artistic benchmark collection consisting of digitized artworks from museums across the world [158, 159]. OmniArt comprises millions of artworks coupled with extensive metadata such as genre, school, material information, creation period and dominant color. This makes the dataset extremely multi-relational and, at the same time, very challenging to perform learning tasks.

For the purpose of comparison with related work, we first perform the artist attribution task in which we attempt to determine the creator of an artwork based on his/her inter-relations with artworks, media (e.g., oil, watercolor, canvas etc.) and creation period (timeframe), along with their intra-relations. To this end we select artworks corresponding to the most common artists in the collection. Considering each of these data streams - artworks (I), artist (A), media (M) and their timeframe (T_f) in centuries as a separate modality, we create the inter-relation hypergraph between them. Subsequently, intra-relation graphs are created for each of the 4 modalities in the following way:

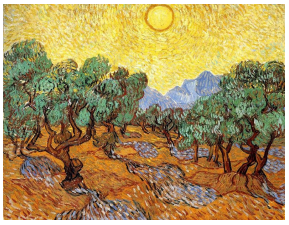
- $\boldsymbol{G_I}$: Based on color palettes similarity
- $\boldsymbol{G_A}$: Based on the schools the artist belongs to
- $\boldsymbol{G_M}$: Based on the co-occurrence in all artworks
- $\boldsymbol{G_{T_f}}$: Based on the style and genre prevalent in that century

We take a sub-sample of the OmniArt dataset consisting of 10,000 artworks from 2,776 artists in the time period ranging all the way from 8^{th} to 20^{th} century along with 63 prominent media types. On this sampled dataset, we achieve an accuracy of 61.7% for the artist attribution task. The performance of our model is at par with the benchmark accuracy of 64.5% [158]. In addition, we conjecture that Hypergraph has an important advantage – the ability to learn even higher order relations, i.e. 5^{th}, 6^{th} and beyond, something that we intend proving in future work.

In the particular case of OmniArt, such higher-order relations would include information about e.g., artist, school, timeframe, medium, dominant colour use, semantics and (implicit) social network. For example, Figure 20 shows the well-known "Olive Trees with Yellow Sky and Sun" painted by Vincent van Gogh in 1889 and Claude Monet's masterpiece "Marine View with a Sunset" from 1875. As nicely portrayed by these two examples, while the two artists exhibit many stylistic similarities, sharing motives and a time period, their materialization is very different. Influenced by Monet, Van Gogh changed both his colour palette and coarseness of brushstrokes, so technically, his work became closer to the French Impressionism. Detecting "tipping points" in the artist's opus would require multimedia representations capable of capturing information about e.g. colour, texture and semantic concepts depicted in the paintings, but also information about school, social network, relevant locations and timeframe and historical context. We believe that our proposed framework is a significant and brave step forward in ultimately deploying multimedia analysis for solving such complex tasks.

5.5 CONCLUSION AND FUTURE WORK

In this paper we propose HyperLearn, a hypergraph-based framework for learning complex higher-order relationships in multimedia datasets. The proposed distributed



(a) Vincent van Gogh – Olive Trees with Yellow Sky and Sun, 1889

(b) Claude Monet – Marine View with a Sunset, 1875

Figure 20: Van Gogh (a) and Monet (b) have many stylistic similarities, but their materialization is different. Capturing their similarities, differences and influences requires the ability to model higher-order relations.

training approach makes this framework scalable to many modalities. We demonstrate benefits of our approach with regards to both performance and computational time through extensive experimentation on MovieLens and MIRFlickr datasets with 2 and 3 modalities respectively. To show the flexibility of HyperLearn in encoding a larger number modalities, we perform experiments on 4^{th} order relations from the OmniArt dataset. In conclusion, on the examples of very different datasets, domains and use cases, we demonstrate that HyperLearn can be extremely useful in learning representations that can capture complex higher order relations within and across multiple modalities. For future work we plan to test the approach on even higher number of heterogeneous modalities and further extend this approach to much larger datasets by solving sub-tensors derived from slicing hypergraph into multiple smaller hypergraphs.

6

ADAPTIVE NEURAL MESSAGE PASSING FOR INDUCTIVE LEARNING ON HYPERGRAPHS

Graphs are the most ubiquitous data structures for representing relational datasets and performing inferences in them. They model, however, only pairwise relations between nodes and are not designed for encoding the higher-order relations. This drawback is mitigated by hypergraphs, in which an edge can connect an arbitrary number of nodes. Most hypergraph learning approaches convert the hypergraph structure to that of a graph and then deploy existing geometric deep learning methods. This transformation leads to information loss, and sub-optimal exploitation of the hypergraph's expressive power. We present HYPERMSG, a novel hypergraph learning framework that uses a modular two-level neural message passing strategy to accurately and efficiently propagate information within each hyperedge and across the hyperedges. HYPERMSG adapts to the data and task by learning an attention weight associated with each node's degree centrality. Such a mechanism quantifies both local and global importance of a node, capturing the structural properties of a hypergraph. HYPERMSG is inductive, allowing inference on previously unseen nodes. Further, it is robust and outperforms state-of-the-art hypergraph learning methods on a wide range of tasks and datasets. Finally, we demonstrate the effectiveness of HYPERMSG in learning multimodal relations through detailed experimentation on a challenging multimedia dataset.

6.1 INTRODUCTION

Modelling the intrinsic properties of many real-world datasets, such as their structure and connections between the data points, requires relational data structures. Graphs are a popular data structure for discovering useful information in relational datasets due to their capability to combine object-level information with the underlying inter-object relations [187]. Data encountered in many real-world scenarios, however, contain relationships among objects which are not dyadic (pairwise) but triadic, tetradic or higher. As an example, consider a research community, where authors publish papers in groups of more than two. Representing such groups of collaborators using just pairwise edges specific to ordinary graphs inevitably leads to information loss. In such cases, the interactions among the objects can be fully modelled by including higher-order relations instead of pairwise relations only [185]. Some example domains where graphs are already shown to be insufficiently expressive are social networks [161], video surveillance [193] and neuroscience [64]. In all such domains the information at group level contributes to understanding the data and solving tasks. Group level and other higher-order relationships

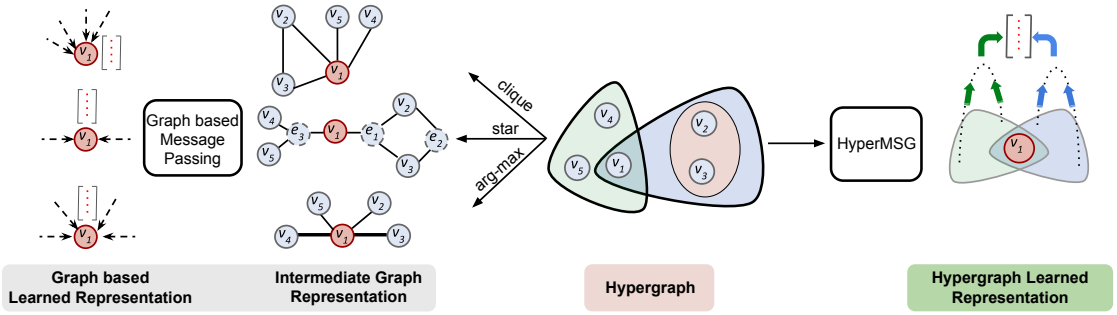


*Figure 21: Illustration of the difference between traditional methods that require an intermediate graph representation and our method (*HYPERMSG*). The left side shows the reduction of a hypergraph to a graph using clique, star and functional metric based (arg-max) [53] [191] expansion methods. The clique expansion loses the unique information associated with the hyperedge defined by the set of nodes $\{v_2, v_3\}$, and it cannot distinguish it from the hyperedge defined by the nodes $\{v_1, v_2, v_3\}$. Star expansion creates a heterogeneous graph that is difficult to handle using most well-studied graph methods [195]. Functional metric (in particular arg-max) based expansion [191] does not exploit the full structural information within hypergraph. On the other hand,* HYPERMSG *formulates neural message passing framework on hypergraphs without any reduction to graphs and directly learns node representations. Hence, it provides a much better basis for hypergraph representation learning.*

can be readily represented with *hypergraphs* [26], where a *hyperedge* can connect an arbitrary number of nodes as opposed to just two nodes in a graph.

Hypergraphs have already been shown to provide a flexible and natural framework to represent higher-order relations within homogeneous data [206]. Real-world datasets, however, often consist of objects with multiple modalities. Traditional graph- or hypergraph-based representations for multimodal data model relations in each modality independently, ignoring the heterogeneous relations between objects. Yet, such relations in the multimodal data can provide complementary information revealing fundamental characteristics of objects and their context. For instance, consider a social media platform which has users, tweets, hashtags, and images. Representing multimodal relations in the collective information generated when a user posts a tweet with an image containing multiple hashtags is infeasible using binary pairwise relations [100] or by considering tweets, users, and images as non-related channels. Hypergraphs efficiently represent multimodal objects with higher-order relations inside a channel and among channels in various domains such as visual arts [10], discussion forums [11], visual question answering systems [87], music recommender systems [32] and protein-protein interaction networks [92].

Apart from being multimodal, in real-world datasets the objects and relations among them are dynamically evolving. That is, during training time there will be nodes which are unseen and relations which are partially observed. Such instances often exist in scenarios such as finding the most promising target audience for a marketing campaign or making movie recommendations with new movies appearing all the time. Performing inference on unseen nodes is challenging and requires an inductive learning framework.

Making inferences in relational datasets represented on a graph or hypergraph requires means to accurately propagate information across the nodes. Recent advances in geomet-

ric deep learning [30] resulted in several techniques to do so by using spectral or spatial convolution on graphs. Spectral graph convolution is based on spectral graph theory [31] and provides a well-founded mathematical framework for designing translation-invariant operators (filters). It requires, however, computing eigenvectors of large matrices, which is computationally expensive. Another shortcoming is that these filters are not localized i.e. to compute the output of a node, the convolution operation does not consider a limited number of neighborhood hops. Hence, its complexity grows with the size of the graph and therefore the method is not scalable to large datasets. This led to works applying approximation to the spectral graph convolution using K-localized filters which use a Kth-order polynomial, like in *ChebNet* [43] and *GCN* [91]. Such filter localization techniques can be interpreted as spatial methods that perform message passing on local neighborhoods which are a certain number of steps away from a node. Due to their efficiency, spatial graph convolution approaches have received most attention in recent years. The resulting neural network model performs message passing operations informed by the structure of the graph to distribute encoded feature information among connected nodes [62] [90]. Message passing has proven to be effective for graph inference, but has yet to be researched for hypergraphs.

Several approaches for devising message passing in hypergraphs have been proposed aiming to exploit their expressive power [53, 190, 191]. A common implicit assumption they make is that a hypergraph is a specific type of ordinary graph. If the assumption holds, reducing the hypergraph learning problem to that of a graph should suffice. Strategies to reduce a hypergraph to a graph include transforming the hyperedges into multiple edges using clique expansion [53] [79] [201], converting to a heterogeneous graph using star expansion [1], and replacing every hyperedge with a weighted edge created using a certain predefined metric on the functional properties of the node [191]. By using a graph as an intermediate (proxy) representation, these approaches make existing graph-based message passing methods [62, 187] applicable to hypergraphs, as can be seen in the left part of Fig. 21. However, a hypergraph is not a special case of graph. The opposite is true, graphs are simply a specific type of hypergraph [26], where the hyperedges are a superset of the pairwise edges. The complex relations in a hypergraph cannot be viewed as an instantiation of an ordinary graph, thus reducing the hypergraph problem to that of a graph cannot fully utilize the available information [47]. To address tasks in datasets with higher-order relations, a truly hypergraph-based message passing formulation is needed that ensures information propagation within and across hyperedges.

Another major limitation of the existing hypergraph representation learning frameworks is their inherently transductive nature [191] [15], thus they are inapplicable for making any inferences on unseen nodes. Often, the injection of new unseen nodes can lead to the introduction of noisy features in the message passing framework [55]. We introduce a probabilistic neighborhood sampling approach which acts as a regularizer and facilitates the proposed inductive learning framework. In a hypergraph, message passing operation on a node is different from regular graphs. This is because in a hypergraph the message propagation from the neighborhood of a node is performed at two levels - within a hyperedge and across hyperedges. Thus unlike regular non-weighted graphs, in a hypergraph not all neighborhood nodes hold equal importance. To quantify the role of a node, we identify that the uniqueness of a hypergraph lies in the detailed structure of

its hyperedges as well as the distribution of nodes across them. In a graph, each edge is only shared between two nodes. A hyperedge, however, can contain a large set of nodes, and each node could have different contributions within the hyperedge based on its relations with neighboring nodes. This is often defined by empirical measures, such as the degree centrality [83] [23]. Degree centrality captures how popular or active a node is in a hypergraph. Since such a paradigm is sensitive to the choice of dataset as well as the task, choosing an empirical measure is not optimal. We introduce a deterministic neighborhood attention mechanism which quantifies the importance of a node in terms of its neighborhood and the number of hyperedges to which it belongs. The proposed attention mechanism is self-adaptive to the choice of task, dataset and to any variation in the hypergraph structure. Combining such an attention mechanism with our message passing framework, we propose an inductive learning framework that exploits the full structure of hypergraphs, without any hypergraph-to-graph conversion, to perform several tasks such as node classification, link prediction, and hypergraph classification.

The points below highlight the contributions of this paper.

- We present HYPERMSG, a hypergraph learning framework with a two-level message passing scheme that jointly captures the relations within a hyperedge and across the hyperedges.
- HYPERMSG is inductive in nature, and facilitates probabilistic sampling of both seen and unseen nodes, based on their importance in message passing.
- HYPERMSG adapts to the dataset and task by implicitly learning the importance of neighborhood nodes in representation learning.
- We demonstrate that HYPERMSG outperforms the state-of-the-art methods by remarkable margins on standard benchmarks consisting of citation networks. In addition, we show that HYPERMSG is highly efficient in exploring the complex heterogeneous interactions in multimodal hypergraphs to perform tasks such as multi-label classification, link prediction and recommendation. The robustness and general applicability of HYPERMSG is further validated on the task of hypergraph (brain) classification in an extremely noisy neuroimaging dataset.

6.2 RELATED WORK

Message passing on hypergraphs aims at learning low-dimensional representations for signals (features) defined on the nodes. Recently, several methods have been proposed for message passing on graphs, and deployed for modeling physical systems, learning molecular fingerprints, predicting protein interfaces, and classifying diseases [207]. However, there is a lack of an accurate message passing framework for hypergraphs. The biggest challenge in hypergraph-based learning is posed by the high variation in hyperedge cardinality i.e. the number of nodes in each hyperedge, which limits the accurate and efficient information propagation from one node to another along the hyperedges. For better understanding of message passing neural networks, we first provide a brief overview of some popular graph-based message passing neural networks. We will then discuss the existing techniques for representation learning on hypergraphs that are related to our proposed framework HYPERMSG.

Message Passing in Graphs The general idea behind a graph convolutional neural network is to define convolution on graphs, where the input is a graph instead of e.g. a 2-D image grid. Methods for performing convolution on graphs can be broadly classified into spatial (message passing) or spectral methods [30]. As mentioned in the introduction, the major drawbacks of spectral approaches is their rigidity in generalizing to new graph structures and their high computational complexity [204, 207]. In this work, we focus on spatial approaches that use message passing neural networks. The basic framework of any message passing network takes in a graph with signals (i.e. features) on each of its nodes as input and learns embeddings for each node by aggregating information from its neighbors [189]. Message passing neural networks outline a general message passing framework for learning such an aggregation mechanism on graphs. It passes information (messages) from one node to another along edges and repeats it in K-steps to let information propagate through the graph. Several variants of this general approach have been proposed, such as [62, 63, 66, 91, 107]. Recently, it has been further extended to heterogeneous graphs in which nodes (and edges) are typed, allowing graphs to incorporate auxiliary information. Some of these ideas include using attention-based mechanism such as typed attention [112], neighbour attention [198], vertex-level and semantic-level attention [178] across different types of nodes or using metapath based aggregations [57]. Unlike graph-based models with only binary relations, hypergraph learning models need to explore the higher-order relations in the data [197].

Message Passing in Hypergraphs Since the introduction of learning with hypergraphs [206], several such methods have been introduced [59], and successfully deployed in various tasks, such as link prediction [100], community detection [37] and visual object tracking [184]. In spectral theory of hypergraphs, methods have been proposed that fully exploit the hypergraph structure using non-linear Laplacian operators [33, 68]. However, these methods have similar drawbacks to spectral graph methods with regard to their computational complexity and scalability. Learning on the hypergraph can also be seen as the process of message passing along the hypergraph structure in analyzing the structured data. Emulating a graph-based message passing framework for hypergraphs is not straightforward since a hyperedge involves more than two nodes which makes the interactions inside each hyperedge complex. Hypergraphs are mathematically represented using either an incidence matrix or an adjacency tensor. Incidence matrix based representations of hypergraphs are rigid in describing the structures of higher order relations [101]. On the other hand, formulating message passing on a higher-dimensional representation of a hypergraph using adjacency tensors makes it computationally expensive and restricted to small datasets and uniform hypergraphs [203].

To circumvent the above issues, [53] and [15] reduce a hypergraph to a graph using clique expansion and perform graph convolutions on them. Further, [79] assume the initial hypergraph structure is weak, and extend the work of [53] to construct dynamic hypergraphs. Recently proposed HyperGCN replaces a hyperedge with pair-wise weighted edges between vertices called mediators [191]. The weights are calculated by comparing the functional properties of neighboring nodes (*e.g.*, using `arg-max` over them) and hence lose the structural information within a hypergraph. With the use of mediators, HyperGCN can be interpreted as an improved variant of clique expansion, and to the best of our knowledge, is also the state-of-the-art method for all the hypergraph representation learning methods, where a graph based message passing neural network is eventually

used. However, it still suffers from the same limitation as the clique expansion [47]. These limitations are further discussed on the example of a Fano plane in the Appendix. Furthermore, these approaches are inherently transductive and thus, as indicated earlier, cannot perform inference on unseen nodes. In [136] a generalized hypergraph learning framework is presented that uses random walks, however its message passing framework is not robust for heterogeneous hypergraphs and its validity for inductive setting is still unexplored. None of the above approaches utilize the complete structural information in the hypergraph, leading to sub-optimal learning performance.

In our preliminary work we introduced HyperSAGE [7], an inductive representation learning framework for hypergraphs that can exploit their full structure by aggregating messages in a two-stage procedure. It managed to achieve near state-of-the art performance on benchmark datasets. However, it suffers from lack of adaptability to datasets and structural variation in hypergraphs and poor parallelism. Huang et al. [75] further built over our proposed HyperSAGE framework to circumvent these issues and generalize the aggregation approach. Recently, [46] proposed message passing in hypergraphs using a dual attention mechanism for classifying text in documents by constructing multiple types of hyperedges (sequential, syntactic and semantic). The proposed message passing framework combines both types of neighborhood information (within hyperedges and across hyperedges) separately and thus mitigates the shortfall of hypergraph to graph conversion. However, the proposed dual attention mechanism uses the functional properties (i.e. node features) of the hypergraph. This leads to insufficient information propagation as it fails to quantify the importance of a node with respect to its neighbors within and across hyperedges, which is an important aspect for capturing the structural properties of a hypergraph. In this work, we propose HYPERMSG, which eliminates matrix- or tensor-based formulations in its neural message passing scheme for hypergraphs. It is further inductive and utilizes all the available information in a hypergraph.

6.3 HYPERGRAPH PRELIMINARIES

In this section, we first introduce some of the preliminary definitions and notations of hypergraphs which will be used to formulate the HYPERMSG framework.

Definition 1 (Hypergraph). *A hypergraph $\mathcal{H}$ is represented as $\mathcal{H} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$, where $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ denotes a set of N nodes and $\mathcal{E} = \{e_1, e_2, ..., e_K\}$ a set of hyperedges, with each hyperedge comprising a non-empty subset from $\mathcal{V}$. $X \in \mathbb{R}^{N \times d}$ denotes the feature matrix, with the feature vector x_i corresponding to the respective v_i column.*

The cardinality of any hyperedge e_l is the number of nodes contained in that hyperedge, given by $|e_l|$. Unlike in a graph, the hyperedges of $\mathcal{H}$ can contain different number of nodes i.e. $1 \leq |e_i| \leq |\mathcal{V}|$. Definition 1 makes it clear that graphs are simply a special case of hypergraphs with a fixed cardinality of 2 for all the edges.

We define three different types of neighborhoods as follows.

Definition 2 (Intra-edge neighborhood). *The intra-edge or local neighborhood of a node $v_i \in \mathcal{V}$ for any hyperedge $e \in \mathcal{E}$ is defined as the set of nodes v_j belonging to e and is denoted by $\mathcal{N}(v_i, e)$.*

The intra-edge neighborhood of a node captures the higher order relationships and provides localized group level information to it. Further, let $E(v_i) = \{e \in \mathcal{E} | v_i \in e\}$ be the set of hyperedges containing node v_i. The degree of node v_i is thus given by $|E(v_i)|$.

Definition 3 (Inter-edge neighborhood). *The inter-edge or global neighborhood of a node $v_i \in \mathcal{V}$, is defined as the neighborhood of v_i spanning across the set of hyperedges $E(v_i)$ and is given by $\mathcal{N}(v_i) = \bigcup_{e \in E(v_i)} \mathcal{N}(v_i, e)$.*

The inter-edge or global neighborhood of a node captures its global positioning and gathers information from hyperedges similar to the node. Finally, the condensed neighborhood parameterized by α gives a subset of nodes within a hyperedge.

Definition 4 (Condensed neighborhood). *The condensed neighborhood of any node $v_i \in \mathcal{V}$ is defined as the sampled set $\mathcal{N}(v_i, e; \alpha)$ comprising of α nodes from a hyperedge $e \in E(v_i)$, if $\alpha < |e|$, or all nodes in e if $\alpha >= |e|$.*

6.4 PROPOSED MODEL

The main concept behind devising a message passing neural network on hypergraphs is to aggregate feature information from the neighborhood of a node which spans across multiple hyperedges with varying cardinality. In this section we propose HYPERMSG, a framework that performs message passing at two levels for a hypergraph. Further, we discuss our adaptive framework that implicitly learns the importance of each node in the representation learning process. Our approach inherently allows inductive learning, which makes it also applicable on hypergraphs with unseen nodes.

6.4.1 *Two-level Message Passing Framework*

We propose to interpret the propagation of information in a given hypergraph as a two-level aggregation problem, where the neighborhood of any node is divided into its *intra-edge* neighbors and *inter-edge* neighbors. This information is present in the form of signals on each node often referred to as message as mentioned in the previous section. For message aggregation, we define the aggregation function as a permutation invariant set function on a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E}, X)$ that takes as input a countable unordered message set and outputs a reduced or aggregated message of the same dimension as the original message. Further, for two-level aggregation, let $\mathcal{F}_1(\cdot)$ and $\mathcal{F}_2(\cdot)$ denote the intra-edge and inter-edge aggregation functions, respectively. Schematic representation of the two functions is provided in Fig. 22. Similar to X, we also define $Z \in \mathbb{R}^{N \times l}$ as the aggregated feature matrix built using the outputs z_i with dimension l from the aggregation functions. Message passing at node v_i can then be expressed as

$$s_1 \leftarrow \{x_j \,|\, v_j \in \mathcal{N}(v_i, e; \alpha)\}, \tag{6.1}$$

$$s_2 \leftarrow \{\mathcal{F}_1^{(e)}(s_1) \;|\; e \in E(v_i)\}, \tag{6.2}$$

$$z_i \leftarrow x_i + \mathcal{F}_2(s_2), \tag{6.3}$$

where s_1 and s_2 denote the unordered sets of feature vectors and intra-edge aggregations, respectively.

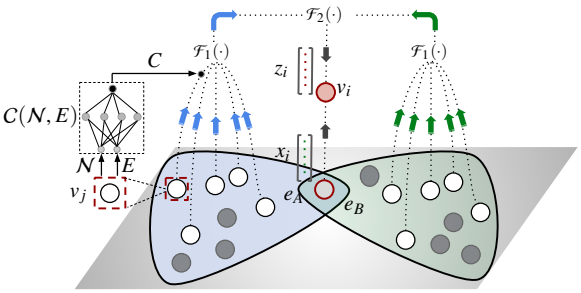


Figure 22: Schematic representation of the two-level message passing scheme of HYPERMSG*, with aggregation functions $\mathcal{F}_1(\cdot)$ and $\mathcal{F}_2(\cdot)$. It shows information aggregation from two hyperedges e_A and e_B, where the intra-edge aggregation is from sampled sets of 5 nodes ($\alpha = 5$) for each hyperedge. The function $C(\mathcal{N},E)$ is used to compute attention weight C for quantifying the importance of node v_j during intra-edge aggregation on e_A. For node v_i, x_i and z_i denote the input and aggregated feature vector, respectively.*

To ensure that the expressive power of a hypergraph is preserved or at least the information loss is minimized, the choice of aggregation function should comply with certain properties. First, it should capture the features of neighborhood nodes in a manner that is invariant to permutation of nodes and hyperedges within its neighborhood. This ensures the final message propagation to a node does not depend on the sequence in which we pick up its neighbors. Many graph-based methods use aggregation functions, such as `mean` and `max` functions [91]. Mean and max-pooling aggregators are well-defined multiset functions because they are permutation invariant and these functions learn different attributes of the neighborhood (`max` learns distinct elements and `mean` learns distributions). These aggregations have proven to be successful for node classification problems [66, 91]. However, they are not injective and hence have their limitations in learning unique representations in various cases such as when the nodes have repeating features [189] or when the features are continuous [40]. There are other non-standard neighbor aggregation schemes that we do not cover, e.g., weighted average via attention [176], LSTM pooling [66, 127] and stochastic aggregations [181]. We emphasize that our theoretical framework is general enough to characterize the representational power using a family of generalized mean aggregation functions. In the future, it would be interesting to apply our framework to analyze and understand other aggregation schemes.

Property 1 (Hypergraph Isomorphic Equivariance). *A message aggregation function $\mathcal{F}(\cdot)$ is equivariant to hypergraph isomorphism, if for two isomorphic hypergraphs $\mathcal{H} = (\mathcal{V},\mathcal{E},X)$, $\mathcal{H}^* = (\mathcal{V}^*,\mathcal{E}^*,X^*)$, and permulation operator σ, given that $\mathcal{H}^* = \sigma \bullet \mathcal{H}$, and Z and Z^* represent the aggregated feature matrices obtained using $\mathcal{F}(\cdot)$ on $\mathcal{H}$ and $\mathcal{H}^*$ respectively, the condition $Z^* = \sigma \bullet Z$ holds.*

Secondly, the aggregation function should also preserve the global neighborhood invariance at the 'dominant nodes' of the graph. Here, dominant nodes refer to nodes that act as hubs in power-law networks, possessing many more connections than their neighbors [3]. The aggregation function should ideally be insensitive to whether the provided hypergraph contains a few large hyperedges, or a large number of smaller ones obtained from splitting them. Generally, a hyperedge would be split in a manner that the dominant nodes are shared across the resulting hyperedges. In such cases,

global neighborhood invariance would imply that the aggregated output at these nodes before and after the splitting of any associated hyperedge stays the same. Otherwise, the learned representation of a node will change significantly with each split. Based on these considerations, we define the following properties for a generic message aggregation function.

Property 2 (Global Neighborhood Invariance). *A message aggregation function $\mathcal{F}(\cdot)$ satisfies global neighborhood invariance at any node $v_i \in \mathcal{V}$ for a given hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E}, X)$, if for any hyperedge $e \in E(v_i)$ being split into multiple hyperedges without changing $\mathcal{N}(v_i)$, and z_i and z_i^* denoting the aggregated feature vectors obtained before and after splitting, the condition $z_i^* = z_i$ holds.*

Aggregation using Generalized Means. One major advantage of our strategy is that the message passing module is decoupled from the choice of the aggregation method itself. This allows our approach to be used with a broad set of generalized means aggregation functions, a powerful and rich family of aggregation functions that have been shown to perform well for graph representation learning [102]. The permutation invariant nature of generalized means makes them satisfy Property 1. Further, we show that with appropriate combinations of the intra-edge and inter-edge aggregations, Property 2 is also satisfied.

Mathematically, generalized means can be expressed as $M_p = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}$, where n refers to the number of elements to aggregate, and p denotes its power. The choice of p allows providing different interpretations to the aggregation function. For example, $p = 1$ denotes arithmetic mean aggregation, $p = 2$ refers to mean squared estimate and a large value of p approximates max pooling from the group. Similarly, M_p can be used for approximating geometric means with $p \to 0$. We use generalized means for intra-edge as well as inter-edge aggregation. The two functions $\mathcal{F}_1(\cdot)$ and $\mathcal{F}_2(\cdot)$ as stated in Section 6.4.1, for aggregation at node v_i are then defined as

$$\mathcal{F}_1^{(e)}(s_1) = \left(\frac{1}{|\mathcal{N}(v_i)|}\sum_{v_j \in \mathcal{N}(v_i,e)} w_j x_j^p\right)^{\frac{1}{p}}, \tag{6.4}$$

$$\text{where, } w_j = \frac{1}{|\mathcal{N}(v_i,e)|} * \left(\sum_{m=1}^{|E(v_i)|} \frac{1}{|\mathcal{N}(v_i,e_m)|}\right)^{-1},$$

$$\mathcal{F}_2(s_2) = \left(\frac{1}{|E(v_i)|}\sum_{e \in E(v_i)} {s_2}^p\right)^{\frac{1}{p}}. \tag{6.5}$$

For Property 2 to hold in in Eq. 6.4 and Eq. 6.5 above, power term p needs to be the same for $\mathcal{F}_1$ and $\mathcal{F}_2$. Also, the scaling term w_j needs to be added to balance the bias in the weighting introduced in intra-edge aggregation due to varying cardinality across the hyperedges. The related mathematical proof is presented in the Appendix.

6.4.2 *Learning the Importance of Nodes*

Commonly, the contributions from the neighboring nodes within a hyperedge are weighted equally through the choice of simple aggregation functions such as `mean` and `max`, among others. Thus, the importance of any node is defined by only its functional characteristics - the values that are contained in its feature vector. However, hypergraphs also possess a complex structure, and this component is almost unused in the learning process.

We hypothesize that analyzing the structural information can reveal the importance of each node and enhance the message passing process. The structural information of a node v in a hypergraph can be primarily defined by two terms: its global neighbourhood set $\mathcal{N}(v)$ and its hyperedge set $E(v)$. To quantify this information for a node in any graph, [56] introduced a degree centrality measure which is simply the total number of edges incident on that node. However, compared to graphs where $|\mathcal{N}(v)| = |E(v)|$, these two terms can be significantly different for a hypergraph depending on its structure. Defining empirical degree centrality functions to measure the importance of any node in hypergraphs was shown to be effective, however, such functions are sensitive to the choice of dataset and task [83].

We introduce a learnable function to quantify the importance of a node which is expressed in terms of $\mathcal{N}$ and E as $C(\mathcal{N}, E)$. For hypergraphs, node importance function $C(\mathcal{N}, E)$ is a complex non-linear mapping. Thus, rather than empirically choosing a single function, we learn the importance value for each node using a small fully-connected neural network comprising two hidden layers. For any node v, the output of $C(\cdot,\cdot)$, denoted by C, can be interpreted as an attention weight that defines the importance of that node in the message passing process (see Fig. 22). Let C_j denote the attention weight for node $v_j \in \mathcal{N}(v_i, e)$. We conjecture that using C as a weighting term for each node during message passing can improve the learned node embeddings in a hypergraph irrespective of dataset or task.

6.4.3 *Inductive Learning on Hypergraphs*

Inductive learning of nodes is a challenging problem for hypergraphs as it requires the model to generalize on previously unseen nodes with a diverse set of features and associated sub-hypergraphs. Most existing approaches are inherently transductive as they make predictions on nodes in a single, fixed hypergraph. These approaches directly optimize the node representations using matrix-factorization-based objectives, and thus do not generalize to unseen data. HYPERMSG tackles this challenge through the use of a neural message passing framework as described in section 4.1 that learns to generate embeddings by sampling and aggregating features from a node's local neighborhood. Our approach uses a neural network comprising L layers, and feature-aggregation is performed at each of these layers, as well as across the hyperedges. Algorithm 1 describes the forward propagation mechanism, which implements the aggregation function introduced above. At each iteration, nodes first aggregate information from their neighbors within a specific hyperedge. This is repeated over all the hyperedges across all the L layers. The trainable weight matrices W^l with $l \in L$ are used to aggregate information across

the feature dimension and propagate it through the hypergraph. The representation on any unseen node can then be obtained by an aggregation process similar to that on seen nodes.

Algorithm 1: HYPERMSG Inductive Message Passing

Input : $\mathcal{H} = (\mathcal{V}, \mathcal{E}, X)$; depth L; weight matrices W^l for $l = 1 \dots L$; non-linearity σ; intra-edge aggregation function $\mathcal{F}_1(\cdot)$; inter-edge aggregation function $\mathcal{F}_2(\cdot)$; node importance function $C(\cdot,\cdot)$

Output : Node embeddings $z_i \mid v_i \in \mathcal{V}$

$h_i^0 \leftarrow x_i \in X \mid v_i \in \mathcal{V}$

for $l = 1 \dots L$ **do**

  $h_i^l \leftarrow h_i^{l-1}$; $s_2 = \{\emptyset\}$

  for $e \in E(v_i)$ **do**

    $s_1 \leftarrow \{C(\mathcal{N}(v_j), E(v_j)) \odot x_{j,l-1} \mid v_j \in e\}$

    $s_2 \leftarrow \mathcal{F}_1(s_1)$

  end

  $h_i^l \leftarrow h_i^l + \mathcal{F}_2(s_2)$

  $h_i^l \leftarrow \sigma(W^l(h_i^l / \|h_i^l\|_2))$

end

$z_i \leftarrow h_i^L \mid v_i \in \mathcal{V}$

Probabilistic Sampling-based Aggregation. The modular framework of HYPERMSG provides flexibility in adapting the message aggregation module to fit a desired computational memory. This is achieved through aggregating information from only a condensed neighborhood set (Definition 4) $\mathcal{N}(v_i, e; \alpha)$ instead of the full neighborhood $\mathcal{N}(v_i, e)$. We propose to apply sub-sampling only on the nodes from the training set, and use information from the full neighborhood for the test set. The advantages of this are twofold. First, a reduced number of samples per aggregation at training time reduces the memory capacity requirement. Second, similar to dropout [155], it serves to add regularization to the optimization process. Using the full neighborhood on test data avoids randomness in the test predictions, and generates consistent output.

For the sampling process, we choose to perform probabilistic selection of the nodes for aggregation based on their importance in the hypergraph using the learned attention weights as described in section 4.2. The probability P_j that the node v_j gets selected in the sampled subset at any iteration is then given by

$$P_j = \frac{C_j}{\sum_{j=0}^{|\mathcal{N}(v_i,e)|} C_j} \tag{6.6}$$

implying that the neighbor node judged as more important for v_i should be sampled more often.

Time complexity analysis. The time complexity for training HYPERMSG is $\mathcal{O}(T|\mathcal{E}|(1 + h(d + c)))$ where, T is the total number of training iterations, d denotes the dimensionality of the input feature vector, h is the number of hidden layers, and c is the number of classes. Thus, the time complexity of HYPERMSG is at par with HGNN [53] and HyperGCN [191], with an added advantage of probabilistic sampling by using α that can

reduce the memory constraints while also acting as a regularizer. Additional details are provided in the supplementary material.

6.5 EXPERIMENTS

We perform a variety of experiments to evaluate HYPERMSG and compare its performance with other hypergraph based learning methods. Firstly, the performance of HYPERMSG is evaluated on the task of semi-supervised node classification in hypergraphs through several experiments on representative benchmark datasets. This experiment is performed to test the node-level representative learning capability of HYPERMSG and analyze all its possible variants. The results are compared with the state-of-the-art hypergraph representation learning methods. Secondly, we study the stability of HYPERMSG by proposing the task of hypergraph classification on an extremely noisy brain neuroimaging dataset. Finally, to show the efficiency of HYPERMSG in performing multimodal learning, we compare its performance with recent hypergraph learning methods on a social multimedia dataset. We further show the advantage of using our inductive framework to perform node classification of previously unseen nodes in the multimedia dataset.

6.5.1 *Semi-supervised Node Classification in an Academic Network*

Experimental setup. We use the standard co-citation and co-authorship network datasets: CiteSeer, PubMed, Cora [146], DBLP [141] and arXiv [38] for this experiment where the task is to predict the topic to which a document belongs (multi-class classification). The input feature vector x_i corresponds to a bag of words, where $x_{i,j} \in x_i$ is the normalized frequency of occurrence of the j^{th} word. Further, for all experiments, we use a two-layered neural network. All models are implemented in Pytorch and trained using Adam optimizer [89]. Additional implementation details are provided in the Appendix.

Significance of learning the importance of nodes. For a better understanding on the significance of learning a node importance function for hypergraphs, we show example distributions of $|\mathcal{N}(v)|/|E(v)|$ for CORA co-citation and co-authorship datasets (see Fig. 23). For the CORA co-citation dataset, $|\mathcal{N}(v)|/|E(v)|$ is mostly close to 1. For this data, we found that the performance gain with message passing using HYPERMSG over the graph-based model is relatively small. On the contrary, for CORA co-authorship, the distribution of $|\mathcal{N}(v)|/|E(v)|$ is right-skewed, and has values significantly higher than 1 as well. For this data, HYPERMSG outperforms graph-based models for node classification by more than 5%. This experiment illustrates that even with the same functional characteristics (features) of nodes, the structural information encoded within a hypergraph plays a key role in learning better node embeddings. Additional analysis is presented in the Appendix.

Effect of generalized mean aggregations. We study here the effect of different choices of p in the proposed two-level aggregation function on the performance of HYPERMSG. Fig. 24 shows the accuracy scores obtained for 4 different choices of p and α on DBLP and Pubmed datasets. Across all values of p, we observe that $p = 1$ works best for both datasets. For other choices of p, the performance of the model is

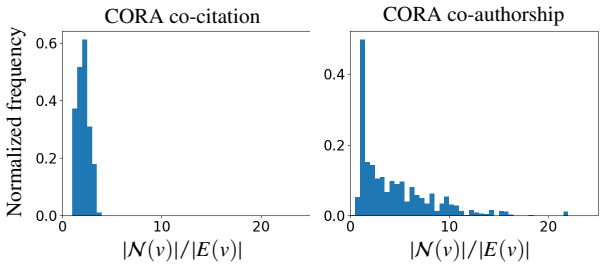


Figure 23: Distribution of $|\mathcal{N}(v)|/|E(v)|$ values across CORA co-citation and CORA-co-authorship datasets.

Table 6: Performance scores (in terms of accuracy %) for various hypergraph learning methods on co-authorship or co-citation datasets. The term 'Hom.' denotes homogeneous networks that use either of co-citation and co-authorship datasets, and 'Het.' refers to those using the combination of both datasets.

Data	**Method**	Cora	DBLP	arXiv	Pubmed	Citeseer
Hom.	MLP + HLR [191]	63.1 ± 1.8	61.6 ± 2.1	61.7 ± 2.3	69.1 ± 1.5	62.3 ± 1.6
	HGNN [53]	66.3 ± 2.8	73.8 ± 2.1	68.1 ± 2.7	68.1 ± 3.5	62.6 ± 1.6
	HyperGCN [191]	69.7 ± 3.7	74.2 ± 5.2	68.2 ± 3.6	73.4 ± 3.8	62.7 ± 4.6
	UniGAT [75]	75.0 ± 1.1	87.8 ± 1.1	77.2 ± 1.3	74.6 ± 1.2	63.8 ± 1.5
	UniGCN [75]	75.3 ± 1.3	**88.0 ± 1.1**	77.3 ± 1.8	74.1 ± 1.0	63.7 ± 1.5
Het.	HetGNN [198]	72.6 ± 1.3	77.9 ± 2.0	74.5 ± 2.0	-	-
	HAN [178]	72.8 ± 1.9	77.9 ± 1.4	75.0 ± 2.2	-	-
	MAGNN [57]	73.3 ± 1.5	78.3 ± 1.8	75.8 ± 1.8	-	-
	MPNN-R [190]	74.7 ± 1.5	78.6 ± 1.7	77.7 ± 1.7	-	-
Hom.	HYPERMSG (non-adaptive, **ours**)	74.9 ± 1.0	80.4 ± 1.1	78.6 ± 1.4	76.2 ± 1.6	66.4 ± 1.8
	HYPERMSG (**ours**)	**77.7 ± 1.2**	85.7 ± 1.1	**79.1 ± 1.6**	**77.1 ± 1.2**	**66.8 ± 1.6**

reduced. For $\alpha = 2$, performance of the model seems to be independent of the choice of p for both the datasets. A possible explanation could be that the number of neighbors is very small, and change in p does not affect the propagation of information significantly.

Effect of probabilistic sampling. We study here the effect of number of samples per aggregation on the performance of the model (Fig. 24). For DBLP, model performance increases with increasing value of α. However, for Pubmed, we observe that performance improves up to $\alpha = 8$, but then a slight drop is observed for larger sets of neighbors. Note that for Pubmed, the majority of the hyperedges have cardinality less than or equal to 10. This means that for $\alpha = 16$, information will be aggregated from almost all the neighbors, thereby involving almost no random sampling. Stochastic sampling of nodes can serve as a regularization mechanism and reduce the impact of noisy hyperedges. This is possibly the reason that performance for $\alpha = 8$ is higher than for 16.

Performance comparison with existing methods. In Table 6, we compare the results of HYPERMSG with state-of-the-art hypergraph learning methods. Among these, the homogeneous networks use either the co-citation or co-authorship datasets. The heterogeneous networks combine information from both co-citation and co-authorship datasets. For HYPERMSG, we use homogeneous network for a fair comparison. We

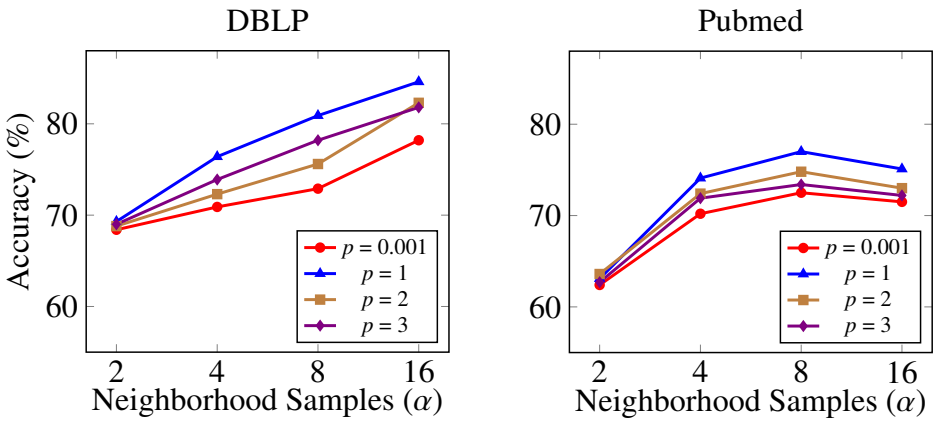


Figure 24: Performance of HYPERMSG *for different choices of generalized means (p) and neighborhood samples (α).*

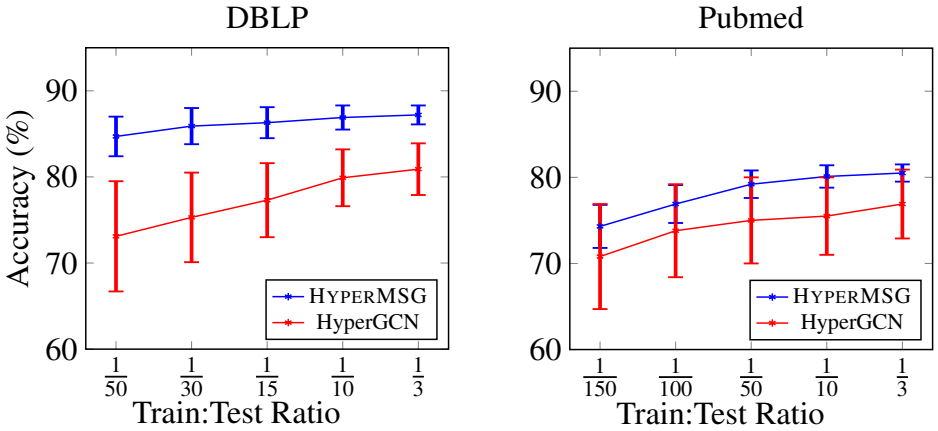


Figure 25: Accuracy scores for HYPERMSG *and HyperGCN obtained for different train-test ratios.*

report the results with arithmetic mean $p = 1$ using the complete neighborhood i.e., $\alpha = |e|$. To show the significance of learning importance of nodes, we report the results for HYPERMSG as well as its non-adaptive variant in Table 6. For all models, 10 data splits over 8 random weight initializations are used, totalling 80 experiments per method for every dataset. The data splits are the same as in HyperGCN and are described in the Appendix. Note that for Pubmed and Citeseer dataset, the co-authorship information does not exist and hence, the heterogeneous models are not applicable.

From Table 6, we observe that both variants of HYPERMSG outperform the homogeneous networks by remarkable margins. Interestingly, our implementations of HYPERMSG with only homogeneous information are able to outperform the heterogenous networks that use information from two different sources. This clearly demonstrates that HYPERMSG exhibits strong representative power and is able to extract information from the hypergraphs at levels beyond the existing baselines.

Stability analysis. We further study the stability of our method in terms of the variance observed in performance for different train-test split ratios. We compare the results with HyperGCN under similar settings, as it is the state-of-the-art method of a broad set of hypergraph learning methods which are based on some sort of hypergraph

Table 7: Performance of HYPERMSG *and its variants on nodes which were part of the training hypergraph (seen) and nodes which were not part of the training hypergraph (unseen).*

	DBLP		Pubmed		Citeseer		Cora (citation)	
Method	Seen	Unseen	Seen	Unseen	Seen	Unseen	Seen	Unseen
MLP + HLR	64.5±2.5	58.7±3.1	66.8±2.4	62.4±3.5	60.1±1.2	58.2±1.9	65.7±2.3	64.2±2.5
UniGCN	**88.5±1.2**	**82.6±2.2**	83.7±1.1	83.3±1.3	71.2±1.2	**70.6±1.9**	74.3±2.3	71.5±2.5
HYPERMSG ($\alpha = 4$)	84.7±2.8	72.2±2.9	79.2 ± 3.1	70.4 ± 2.6	69.3±1.9	68.8±2.9	74.8 ± 1.5	73.2 ± 2.0
HYPERMSG ($\alpha = 8$)	87.5±1.9	77.7±1.1	**84.4±1.4**	**83.5±1.2**	70.8±1.8	69.6±1.6	**75.4 ± 0.8**	74.6 ± 1.7
HYPERMSG ($\alpha = 16$)	88.0±1.3	81.5± 1.1	83.6±0.8	81.7±0.9	**71.7±0.8**	**70.6±1.0**	75.1 ± 0.5	**74.8 ± 1.3**

to graph conversion. Fig. 25 shows results for the two learning methods on 5 different train-test ratios. We see that the performance of both models improves when a higher fraction of data is used for training, and the difference in their performances decreases at the train-test ratio of 1/3. However, for smaller ratios, we see that HYPERMSG outperforms HyperGCN across all datasets. Further, the standard deviation for the predictions of HYPERMSG is lower than that of HyperGCN. Clearly, this implies that HYPERMSG is able to better exploit the information contained in the hypergraph compared to HyperGCN, and can thus produce more accurate and stable predictions. Results of the experiment on Cora and Citeseer, which exhibit a similar trend, can be found in the Appendix.

Classification of unseen nodes. To test the performance of HYPERMSG on unseen nodes, we create inductive learning datasets for DBLP, Pubmed and Citeseer. To do so, we split each dataset into a ratio of 1:3:1 for the training, validation (seen) and test (unseen) sets, respectively. To obtain the unseen test set, we break the hypergraph into two sub-hypergraphs. Note that this splitting leads to several node connections being disregarded as well as a relatively sparse test hypergraph. This can induce noise in the learning process, which we tackle by employing our probabilistic sampling mechanism in message passing. We show that by capturing all the structural and functional properties of a hypergraph, HYPERMSG performs better than the other models.

Table 7 presents the performance scores of HYPERMSG for different choices of α. To the best of our knowledge, no competitive baseline inductive learning method existed before our preliminary work HyperSAGE [7] which introduced a two-level message aggregation framework. Recently, [75] further extended the two-level message aggregation framework and proposed a set of inductive learning methods. In this experiment, we compare the performance of HYPERMSG with the base MLP+HLR approach and UniGCN [75] as a reference. A general observation is that HYPERMSG works well for unseen nodes and significantly outperforms the other models except in the case of DBLP. Further, the difference between performance scores for seen and unseen nodes is stable across all datasets. Based on these observations, it can be concluded that HYPERMSG works well in an inductive setting, i.e. on unseen nodes, as well.

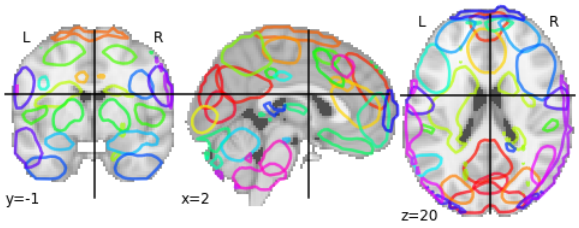


Figure 26: Example slices of a brain sample from the autism neuroimaging data [41] for the three planes showing the construction of a hypergraph.

Table 8: Performance scores for HYPERMSG *and baseline hypergraph learning methods for the task of brain classification for autism spectral disorder based on neuroimaging data.*

Method	AUC-ROC
HGNN [53]	62.5 ± 4.9
HyperGCN [191]	59.2 ± 5.8
HYPERMSG (non-adaptive, **ours**)	64.3 ± 4.4
HYPERMSG (**ours**)	**67.2** ± **5.3**

6.5.2 *Hypergraph classification on neuroimaging data*

To demonstrate the generality of HYPERMSG, we use it on the task of hypergraph classification. For performing hypergraph classification, the model needs to integrate the learned features from each node in a way that the combined features are significantly dissimilar across hypergraphs. This makes hypergraph classification a challenging task as compared to the previous node classification tasks. Further, to also study the robustness of the proposed model under noisy scenarios, we choose an extremely noisy brain neuroimaging dataset [41], where each brain image is a hypergraph. It involves classification of control subjects for autism spectrum disorder (ASD) using 4D resting-state functional magnetic resonance imaging (fMRI) data. Typically, an fMRI sample comprises about 20,000 voxels with 300 time points, making it extremely high-dimensional with a significant level of inherent noise [51, 134].

Some earlier methods have sought to mitigate this problem through aggregating information along one of the dimensions, thus leading to a 3D volume [49, 169]. Other approaches involve summarization of the data as cross-correlation matrix between macro regions of the brain [9, 135]. Although these solutions simplify data handling, they come at the expense of significant information loss. Hypergraphs can be used to handle this task without any such approximations. An example representation of a brain hypergraph is shown in Fig. 26. More details about the processing steps and data preparation are provided in the Appendix.

Table 8 shows the performance scores for the various hypergraph-based methods. Due to class imbalance in the dataset, we use AUC-ROC to measure the performance. We see that both variants of HYPERMSG outperform HGNN, HyperGCN and UniGCN methods, with our adaptive variant showing an improvement of around 5% over HGNN. The results show the robustness and stability of HYPERMSG in utilizing the information

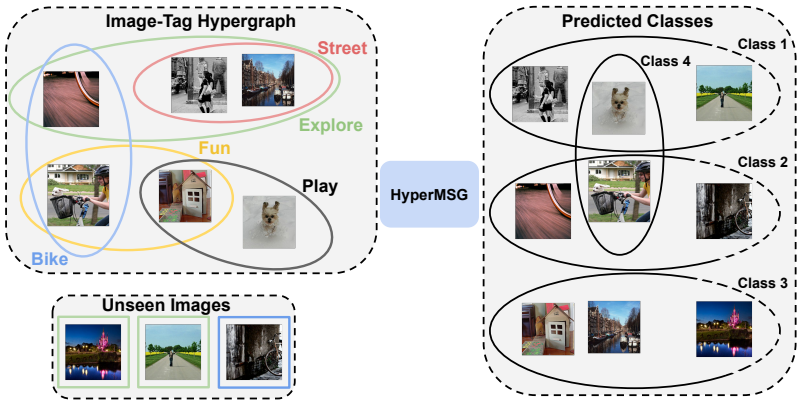


Figure 27: An example of the proposed multimodal hypergraph analysis on the Flickr dataset. The input to the model is an image-tag hypergraph (left) where each image is represented on a node and the hyperedges correspond to the tags associated with the images. The goal is to classify an image into its respective class which can be the ground truth labels in case of Task 1, users in case of Task 2 and groups in case of Task 3. Unlike previous methods such as [12], HYPERMSG *can perform inferences on unseen images as well.*

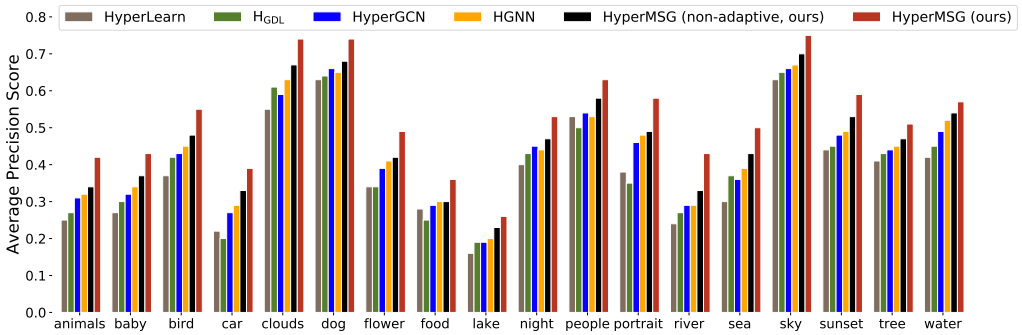


Figure 28: Task 1 (Multi-Label Image Classification) Detailed performance comparison in terms of Average Precision over 18 concepts on the MIRFlickr dataset

within hypergraphs. The large variance observed in performance for all the methods can be attributed to the low signal-to-noise ratio of the dataset.

6.5.3 *Multimodal Hypergraph Analysis*

Social networks are rich with multimodal information and learning an effective representation for the entities of interest in them, such as users, images and text, has gained a great attention in different applications. Examples include node classification [27, 119], link prediction [100, 110], community detection [11, 98], and network visualization [54, 163]. In this section, we evaluate the performance of HYPERMSG on MIR Flickr [76], a social multimedia dataset commonly used for multimodal learning on hypergraphs [10, 12, 188].

Experimental setup. The MIR Flickr dataset contains heterogeneous entities and relationships among them. In particular, the dataset consists of 25,000 images (I) from

(a) Task 2: Image-User Link Prediction *(b) Task 3: Group Recommendation*

Figure 29: Receiver Operating Characteristics (ROC) curve showing the performance of the models on (a) Image-User Link Prediction and (b) Group Recommendation. On both tasks HYPERMSG *consistently outperforms alternative multimodal hypergraph learning methods.*

Flickr posted by 6,386 users (U) to 10,575 groups (G) and annotated with over 50,000 user-provided tags (T). The dataset also provides manually-created ground truth image annotations at the semantic concept level. There are 25 such unique concepts in the dataset. Similar to the related multimedia works [10, 108], we remove noisy tags occurring less than 50 times and obtain a vocabulary consisting of 457 tags. Accordingly, 18 concepts are preserved and utilized to validate the performance in our experiments. In this work, we follow [12] as a testbed for demonstrating multimodal learning capabilities of HYPERMSG by performing 3 types of tasks: *Task 1*: Multi-Label Image Classification, *Task 2*: Image-User Link Prediction and *Task 3*: Group Recommendation.

Hypergraph Construction We create a hypergraph where each node represents an image and the hyperedges (relations between images) are formed using the tags, similar to our previous work [12]. For all the three tasks hyperedges correspond to the image-tag relationships i.e. a hyperedge is constructed among images sharing a common tag. Subsequently, each image is associated with ground truth labels, users and groups respectively for the corresponding three tasks. Figure 27 shows the image-tag hypergraph (left) and the predicted classes (right) i.e. the ground-truth labels in case of *Task 1*, users in case of *Task 2* and groups in case of *Task 3*.

Performance comparison with existing methods. We compare the performance of our proposed HYPERMSG model with the following hypergraph learning methods: HyperGCN [191], HGNN [53], HyperLearn [10] and H_{GDL} [12]. For a fair and thorough comparison, we conduct the experiments using only structural information and hence we assign identity features to each image (node of the hypergraph). In the case of *Task 1*, we provide a detailed performance analysis on each of the concepts. We report the performance of our proposed HYPERMSG approach and the alternatives in terms of Average Precision (AP), a standard evaluation measure used for multi-label image classification benchmarks. For the other two tasks, we combine the classification scores for all the users/groups and plot their respective Receiver Operating Characteristics (ROC) curves. The ROC curve depicts how well a model is able to predict the presence/absence of a class among images with the corresponding metadata. Figure 28 shows that both variants of HYPERMSG consistently outperform the other hypergraph learning methods in *Task 1*, i.e. classifying images into the 18 concepts. The results of this experiment suggest that our proposed approach is more effective in modeling social image-tag

Table 9: Average precision of HYPERMSG *and baseline hypergraph learning methods for the task of multi-label image classification on MIR Flickr dataset on both seen and unseen nodes.*

Method	Seen	Unseen
MLP + HLR	0.64 ± 1.3	0.61 ± 1.4
UniGCN	0.70 ± 1.1	0.67 ± 1.3
HYPERMSG (non-adaptive, **ours**)	0.70 ± 1.1	0.68 ± 1.3
HYPERMSG (**ours**)	**0.73 ± 0.9**	**0.69 ± 1.1**

relationships than the state-of-the-art alternatives. The advantage of HYPERMSG in learning multimodal relations as compared to the other hypergraph methods is further confirmed by the ROC curves shown in Figure 29 for *Task 2* and *Task 3*. In all three experiments we observe that the adaptive version of HYPERMSG utilizing attention-based mechanism comes out as the best performer. We hypothesize that this can be attributed to the significance of global and local neighborhoods in social multimedia networks where the nodes that have many connections (high-degree nodes) tend to be connected to the other nodes with many connections, while they are surrounded by many small clusters of low-degree nodes [123].

Inductive Learning. To evaluate the performance of HYPERMSG on unseen nodes, we perform multi-label image classification on the MIR Flickr dataset. For this experiment, we include the features of the images as well, which we extract from the penultimate layer of a pre-trained VGG-16 network [151]. Similar to Section 6.5.1, we compare our method with inductive learning methods, MLP+HLR and UniGCN and report the average precision (AP) scores averaged over all the classes. We split the dataset into the training, validation (seen) and test (unseen) samples according to a $1:3:1$ ratio. To obtain the unseen test set, we break the hypergraph into two sub-hypergraphs, ensuring that the training set contains at least one image from each class. Table 9 shows that the HYPERMSG performs better than the alternatives on both seen and unseen nodes, while also producing a lower standard deviation in the performance level.

6.6 CONCLUSION

In this paper, we have presented HyperMSG, a two-level neural message passing framework for inductive learning on hypergraphs. HyperMSG fully utilizes the inherent higher-order relations in a hypergraph structure without reducing it to an intermediate graph representation. It adaptively learns the importance of each node during the learning process, thereby improving the message passing process. Through experiments on several representative datasets, we have shown that HyperMSG outperforms the existing methods for hypergraph learning. We have demonstrated that HyperMSG generates stable predictions for very sparse sampling as well as when the nodes are unseen during the training process. From the results on the challenging task of brain image classification for autism, we conclude that HyperMSG yields accurate and robust results in hypergraph classification even on extremely noisy time-series datasets. Finally, HyperMSG is suitable for performing mutimodal representation learning tasks and can outperform all existing hypergraph based learning methods on information rich multimedia data.

BIBLIOGRAPHY

[1] S. Agarwal, K. Branson, and S. Belongie. Higher order learning with graphs. In *Proceedings of the 23rd international conference on Machine learning*, pages 17–24, 2006.

[2] J. Ahn, C. Plaisant, and B. Shneiderman. A Task Taxonomy for Network Evolution Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):365–376, 2014.

[3] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[4] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. J. Rodgers. Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. In *Proceedings of the Eurographics Conference on Visualization (EuroVis)*, 2014.

[5] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. J. Rodgers. The State-of-the-Art of Set Visualization. *Computer Graphics Forum*, 35(1):234–260, 2016.

[6] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data*. Springer-Verlag, Berlin/Heidelberg, 2006.

[7] D. Arya, D. K. Gupta, S. Rudinac, and M. Worring. Hypersage: Generalizing inductive representation learning on hypergraphs. *arXiv preprint arXiv:2010.04558*, 2020.

[8] D. Arya, D. K. Gupta, S. Rudinac, and M. Worring. Adaptive neural message passing for inductive learning on hypergraphs. *arXiv preprint arXiv:2109.10683*, 2021.

[9] D. Arya, R. Olij, D. K. Gupta, A. El Gazzar, G. Wingen, M. Worring, and R. M. Thomas. Fusing structural and functional MRIs using graph convolutional networks for autism classification. In *Medical Imaging with Deep Learning*, pages 44–61. PMLR, 2020.

[10] D. Arya, S. Rudinac, and M. Worring. Hyperlearn: a distributed approach for representation learning in datasets with many modalities. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2245–2253, 2019.

[11] D. Arya, S. Rudinac, and M. Worring. Predicting behavioural patterns in discussion forums using deep learning on hypergraphs. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6. IEEE, 2019.

[12] D. Arya and M. Worring. Exploiting relational information in social networks using geometric deep learning on hypergraphs. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, pages 117–125, 2018.

[13] D. Arya and M. Worring. Exploiting Relational Information in Social Networks Using Geometric Deep Learning on Hypergraphs. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, pages 117–125, 2018.

[14] B. Bach, E. Pietriga, and J.-D. Fekete. GraphDiaries: Animated Transitions and Temporal Navigation for Dynamic Networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2014.

[15] S. Bai, F. Zhang, and P. H. Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, page 107637, 2020.

[16] A. Banerjee, A. Char, and B. Mondal. Spectra of general hypergraphs. *Linear Algebra and its Applications*, 518:14–30, 2017.

[17] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, ICWSM, pages 361–362. AAAI, 2009.

[18] V. Batagelj and A. Mrvar. Pajek - Program for Large Network Analysis. *Connections*, 21(2):47–57, 1998.

[19] V. Batagelj and A. Mrvar. Pajek - Analysis and Visualization of Large Networks. In *Graph Drawing*, Lecture Notes in Computer Science, pages 77–103. Springer, Berlin, Heidelberg, 2002.

[20] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A Taxonomy and Survey of Dynamic Graph Visualization. *Computer Graphics Forum*, 36(1):133–159, 2017.

[21] M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.-D. Fekete. Matrix Reordering Methods for Table and Network Visualization. *Computer Graphics Forum*, 35(3):693–716, 2016.

[22] M. Behrisch, J. Davey, F. Fischer, O. Thonnard, T. Schreck, D. A. Keim, and J. Kohlhammer. Visual Analysis of Sets of Heterogeneous Matrices Using Projection-Based Distance Functions and Semantic Zoom. In *Computer Graphics Forum*, volume 33, pages 411–420, 2014.

[23] A. R. Benson. Three hypergraph eigenvector centralities. *SIAM Journal on Mathematics of Data Science*, 1(2):293–312, 2019.

[24] K. Benzi, V. Kalofolias, X. Bresson, and P. Vandergheynst. Song recommendation with non-negative matrix factorization and graph total variation. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 2439–2443. Ieee, 2016.

[25] C. Berge. *Graphs and Hypergraphs (North-Holland mathematical library; v. 6)*. Elsevier, 1973.

[26] C. Berge and E. Minieka. *Graphs and Hypergraphs*. North-Holland mathematical library. North-Holland Publishing Company, 1976.

[27] S. Bhagat, G. Cormode, and S. Muthukrishnan. Node classification in social networks. In *Social network data analytics*, pages 115–148. Springer, 2011.

[28] D. G. Bobrow and T. Winograd. An overview of KRL, a knowledge representation language. *Cognitive science*, 1(1):3–46, 1977.

[29] D. Boscaini, J. Masci, E. Rodolà, and M. Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 3189–3197, 2016.

[30] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[31] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.

[32] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He. Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 391–400, 2010.

[33] T.-H. H. Chan, A. Louis, Z. G. Tang, and C. Zhang. Spectral properties of hypergraph laplacian and approximation algorithms. *Journal of the ACM (JACM)*, 65(3):1–48, 2018.

[34] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. 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. ACM, 2015.

[35] G. Chen, J. Zhang, F. Wang, C. Zhang, and Y. Gao. Efficient multi-label classification with hypergraph regularization. In *Proceedings of Computer Vision and Pattern Recognition, 2009. CVPR*, pages 1658–1665. IEEE, 2009.

[36] J. Chen, H. Gao, Z. Wu, and D. Li. Tag co-occurrence relationship prediction in heterogeneous information networks. In *Parallel and Distributed Systems (ICPADS), 2013 International Conference on*, pages 528–533. IEEE, 2013.

[37] I. Chien, C.-Y. Lin, and I.-H. Wang. Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 871–879. PMLR, 2018.

[38] C. B. Clement, M. Bierbaum, K. P. O'Keeffe, and A. A. Alemi. On the use of arxiv as a dataset. *arXiv preprint arXiv:1905.00075*, 2019.

[39] M. Conway. Terrorist's use of the internet and fighting back. *Information and Security*, 19:9, 2006.

[40] G. Corso, L. Cavalleri, D. Beaini, P. Liò, and P. Veličković. Principal neighbourhood aggregation for graph nets. *arXiv preprint arXiv:2004.05718*, 2020.

[41] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7, 2013.

[42] P. Cui, S.-W. Liu, W.-W. Zhu, H.-B. Luan, T.-S. Chua, and S.-Q. Yang. Social-sensed image search. *ACM Transactions on Information Systems (TOIS)*, 32(2):8, 2014.

[43] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29:3844–3852, 2016.

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.

[45] C. P. Diehl, G. Namata, and L. Getoor. Relationship identification for social network discovery. In *AAAI*, volume 22, pages 546–552, 2007.

[46] K. Ding, J. Wang, J. Li, D. Li, and H. Liu. Be more with less: Hypergraph attention networks for inductive text classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4927–4936, 2020.

[47] Y. Dong, W. Sawin, and Y. Bengio. HNHN: Hypergraph networks with hyperedge neurons. *arXiv preprint arXiv:2006.12278*, 2020.

[48] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in neural information processing systems*, pages 2224–2232, 2015.

[49] A. El Gazzar, L. Cerliani, G. van Wingen, and R. M. Thomas. Simple 1-d convolutional networks for resting-state fMRI based classification in autism. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2019.

[50] N. Elmqvist, T.-N. Do, H. Goodell, N. Henry, and J.-D. Fekete. ZAME: Interactive Large-Scale Graph Visualization. In *2008 IEEE Pacific Visualization Symposium*, pages 215–222, 2008.

[51] O. Esteban, D. Birman, M. Schaer, O. O. Koyejo, R. A. Poldrack, and K. J. Gorgolewski. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PloS one*, 12(9):e0184661, 2017.

[52] European Union. Terrorism situation and trend report, 2012.

[53] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3558–3565, 2019.

[54] M. T. Fischer, D. Arya, D. Streeb, D. Seebacher, D. A. Keim, and M. Worring. Visual analytics for temporal hypergraph model exploration. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[55] J. Fox and S. Rajamanickam. How robust are graph neural networks to structural noise? *arXiv preprint arXiv:1912.10206*, 2019.

[56] L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

[57] X. Fu, J. Zhang, Z. Meng, and I. King. Magnn: metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of The Web Conference 2020*, pages 2331–2341, 2020.

[58] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao. Tag-based social image search with visual-text joint hypergraph learning. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1517–1520. ACM, 2011.

[59] Y. Gao, Z. Zhang, H. Lin, X. Zhao, S. Du, and C. Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[60] M. Ghoniem, G. Shurkhovetskyy, A. Bahey, and B. Otjacques. VAFLE: Visual Analytics of Firewall Log Events. In P. C. Wong, D. L. Kao, M. C. Hao, and C. Chen, editors, *Visualization and Data Analysis 2014*, SPIE Proceedings, page 901704. SPIE, 2014.

[61] R. Gibson and S. Ward. A proposed methodology for studying the function and effectiveness of party and candidate web sites. *Social science computer review*, 18(3):301–319, 2000.

[62] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[63] M. Gori, G. Monfardini, and F. Scarselli. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pages 729–734. IEEE, 2005.

[64] S. Gu, M. Yang, J. D. Medaglia, R. C. Gur, R. E. Gur, T. D. Satterthwaite, and D. S. Bassett. Functional hypergraph uncovers novel covariant structures over neurodevelopment. *Human brain mapping*, 38(8):3823–3835, 2017.

[65] D. K. Gupta, D. Arya, and E. Gavves. Rotation equivariant siamese networks for tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12362–12371, 2021.

[66] W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034, 2017.

[67] R. A. Harshman et al. Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970.

[68] M. Hein, S. Setzer, L. Jost, and S. S. Rangapuram. The total variation on hypergraphs-learning on hypergraphs revisited. In *Advances in Neural Information Processing Systems*, pages 2427–2435, 2013.

[69] B. Heintz and A. Chandra. Beyond Graphs. *ACM SIGMETRICS Performance Evaluation Review*, 41(4):94–97, 2014.

[70] R. Helms and K. Buijsrogge. Knowledge Network Analysis: A Technique to Analyze Knowledge Management Bottlenecks in Organizations. In *16th International Workshop on Database and Expert Systems Applications (DEXA'05)*, pages 410–414, 2005.

[71] M. Henaff, J. Bruna, and Y. LeCun. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015.

[72] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[73] L. Hu, J. Cao, G. Xu, J. Wang, Z. Gu, and L. Cao. Cross-domain collaborative filtering via bilinear multilevel analysis. In *International Joint Conference on Artificial Intelligence*. IJCAI/AAAI, 2013.

[74] F. Huang, X. Zhang, C. Li, Z. Li, Y. He, and Z. Zhao. Multimodal network embedding via attention based multi-view variational autoencoder. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, pages 108–116. ACM, 2018.

[75] J. Huang and J. Yang. Unignn: a unified framework for graph and hypergraph neural networks. *arXiv preprint arXiv:2105.00956*, 2021.

[76] M. J. Huiskes and M. S. Lew. The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pages 39–43, 2008.

[77] T. Hwang, Z. Tian, R. Kuangy, and J.-P. Kocher. Learning on Weighted Hypergraphs to Integrate Protein Interactions and Gene Expressions for Cancer Outcome Prediction. In *2008 Eighth IEEE International Conference on Data Mining*, pages 293–302, 2008.

[78] IBM. i2 Analyst's Notebook, 2020.

[79] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao. Dynamic hypergraph neural networks. In *IJCAI*, pages 2635–2641, 2019.

[80] L. Jin, Y. Chen, T. Wang, P. Hui, and A. V. Vasilakos. Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine*, 51(9):144–150, 2013.

[81] J. Johnson, L. Ballan, and L. Fei-Fei. Love thy neighbors: Image annotation by exploiting image metadata. In *Proceedings of the IEEE international conference on computer vision*, pages 4624–4632, 2015.

[82] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst. Matrix completion on graphs. *arXiv preprint arXiv:1408.1717*, 2014.

[83] K. Kapoor, D. Sharma, and J. Srivastava. Weighted node degree centrality for hypergraphs. In *2013 IEEE 2nd Network Science Workshop (NSW)*, pages 152–155. IEEE, 2013.

[84] N. Kerracher, J. Kennedy, and K. Chalmers. A Task Taxonomy for Temporal Graph Visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 2015.

[85] C. Khatri and C. R. Rao. Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 167–180, 1968.

[86] B. Kim, B. Lee, and J. Seo. Visualizing Set Concordance with Permutation Matrices and Fan Diagrams. *Interacting with computers*, 19(5):630–643, 2007.

[87] E.-S. Kim, W. Y. Kang, K.-W. On, Y.-J. Heo, and B.-T. Zhang. Hypergraph attention networks for multimodal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14581–14590, 2020.

[88] H.-J. Kim, E. Ollila, V. Koivunen, and C. Croux. Robust and sparse estimation of tensor decompositions. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 965–968. IEEE, 2013.

[89] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *The International Conference on Learning Representations (ICLR)*, 2015.

[90] T. N. Kipf. Deep learning with graph-structured representations. *PhD Thesis*, 2020.

[91] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *ICLR*, 2016.

[92] F. Klimm, C. M. Deane, and G. Reinert. Hypergraphs for predicting essential genes using multiprotein complex data. *Journal of Complex Networks*, 9(2):cnaa028, 2021.

[93] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[94] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[95] F. Kuhn and R. Oshman. Dynamic Networks: Models and Algorithms. *ACM SIGACT News*, 42(1):82–96, 2011.

[96] T. Lacroix, N. Usunier, and G. Obozinski. Canonical tensor decomposition for knowledge base completion. In *International Conference on Machine Learning*, pages 2869–2878, 2018.

[97] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task Taxonomy for Graph Visualization. In *Proceedings of the 2006 AVI workshop on Beyond time and errors novel evaluation methods for information visualization - BELIV '06*, page 1, 2006.

[98] J. Leskovec, K. J. Lang, and M. Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, pages 631–640, 2010.

[99] J. Leskovec and J. J. Mcauley. Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, pages 539–547, 2012.

[100] D. Li, Z. Xu, S. Li, and X. Sun. Link prediction in social networks based on hypergraph. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 41–42, 2013.

[101] G. Li, L. Qi, and G. Yu. The z-eigenvalues of a symmetric tensor and its application to spectral hypergraph theory. *Numerical Linear Algebra with Applications*, 20(6):1001–1029, 2013.

[102] G. Li, C. Xiong, A. Thabet, and B. Ghanem. Deepergcn: All you need to train deeper GCNs. *arXiv preprint arXiv:2006.07739*, 2020.

[103] H. Li, H. Wang, Z. Yang, and M. Odagaki. Variation autoencoder based network representation learning for classification. In *Proceedings of ACL 2017, Student Research Workshop*, pages 56–61, 2017.

[104] R. Li, C. Wang, and K. C.-C. Chang. User profiling in an ego network: co-profiling attributes and relationships. In *Proceedings of the 23rd international conference on World wide web*, pages 819–830. ACM, 2014.

[105] W.-J. Li and D.-Y. Yeung. Relation regularized matrix factorization. In *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.

[106] X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. Snoek, and A. D. Bimbo. Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval. *ACM Computing Surveys (CSUR)*, 49(1):14, 2016.

[107] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*, 2015.

[108] Z. Li and J. Tang. Weakly supervised deep matrix factorization for social image understanding. *IEEE Transactions on Image Processing*, 26(1):276–288, 2016.

[109] Z. Li, J. Tang, and T. Mei. Deep collaborative embedding for social image understanding. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[110] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031, 2007.

[111] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, and A. Kelliher. Metafac: community discovery via relational hypergraph factorization. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 527–536. ACM, 2009.

[112] H. Linmei, T. Yang, C. Shi, H. Ji, and X. Li. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4823–4832, 2019.

[113] D. Liu, N. Blenn, and P. Van Mieghem. A social network model exhibiting tunable overlapping community structure. *Procedia Computer Science*, 9:1400–1409, 2012.

[114] M. Liu, Y. Gao, P.-T. Yap, and D. Shen. Multi-Hypergraph Learning for Incomplete Multimodality Data. *IEEE Journal of Biomedical and Health Informatics*, 22(4):1197–1208, 2017.

[115] Y.-F. Liu, J.-M. Guo, and L. An. Multimedia classification using bipolar relation graphs. *IEEE Transactions on Multimedia*, 19(8):1860–1869, 2017.

[116] Q. Luo and D. Zhong. Using Social Network Analysis to Explain Communication Characteristics of Travel-related Electronic Word-of-Mouth on Social Networking Sites. *Tourism Management*, 46:274–282, 2015.

[117] K. Maruhashi, M. Todoriki, T. Ohwa, K. Goto, Y. Hasegawa, H. Inakoshi, and H. Anai. Learning multi-way relations via tensor decomposition with neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[118] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45, 2015.

[119] J. McAuley and J. Leskovec. Image labeling on a network: using social-network metadata for image classification. In *European conference on computer vision*, pages 828–841. Springer, 2012.

[120] W. Meulemans, N. H. Riche, B. Speckmann, B. Alper, and T. Dwyer. KelpFusion: A Hybrid Set Visualization Technique. *IEEE Transactions on Visualization and Computer Graphics*, 19(11):1846–1858, 2013.

[121] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl. Movielens unplugged: experiences with an occasionally connected recommender system. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 263–266. ACM, 2003.

[122] D. Milne and I. H. Witten. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 509–518. ACM, 2008.

[123] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, 2007.

[124] F. Monti, D. Boscaini, and J. Masci. Geometric deep learning on graphs and manifolds using mixture model CNNs. In *Proceedings of Computer Vision and Pattern Recognition, CVPR*, 2017.

[125] F. Monti, M. Bronstein, and X. Bresson. Geometric matrix completion with recurrent multi-graph neural networks. In *Advances in Neural Information Processing Systems*, pages 3700–3710, 2017.

[126] A. A. Moreira, D. R. Paula, R. N. Costa Filho, and J. S. Andrade Jr. Competitive cluster growth in complex networks. *Physical Review E*, 73(6):065101, 2006.

[127] R. L. Murphy, B. Srinivasan, V. Rao, and B. Ribeiro. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. *arXiv preprint arXiv:1811.01900*, 2018.

[128] A. Narita, K. Hayashi, R. Tomioka, and H. Kashima. Tensor factorization using auxiliary information. *Data Mining and Knowledge Discovery*, 25(2):298–324, 2012.

[129] National Policing Improvement Agency, editor. *Guidance on the Management of Police Information - Appendix 2*. 2nd edition, 2010.

[130] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

[131] Z. Niu, G. Hua, X. Gao, and Q. Tian. Semi-supervised relational topic model for weakly annotated image recognition in social media. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4233–4240, 2014.

[132] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész. Analysis of a Large-Scale Weighted Network of One-to-One Human Communication. *New Journal of Physics*, 9(6):179, 2007.

[133] W. Pan, N. N. Liu, E. W. Xiang, and Q. Yang. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 2318, 2011.

[134] S. Parisot, B. Glocker, S. I. Ktena, S. Arslan, M. D. Schirmer, and D. Rueckert. A flexible graphical model for multi-modal parcellation of the cortex. *Neuroimage*, 162:226–248, 2017.

[135] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert. Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical image analysis*, 48:117–130, 2018.

[136] J. Payne. Deep hyperedges: a framework for transductive and inductive learning on hypergraphs, 2019.

[137] N. Pržulj. Protein-Protein Interactions: Making Sense of Networks Via Graph-Theoretic Modeling. *Bioessays*, 33(2):115–123, 2011.

[138] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon. Collaborative filtering with graph information: Consistency and scalable methods. In *Advances in neural information processing systems*, pages 2107–2115, 2015.

[139] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos. A new approach to cross-modal multimedia retrieval. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 251–260. ACM, 2010.

[140] E. Ravasz and A.-L. Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.

[141] R. A. Rossi and N. K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.

[142] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

[143] S. Rudinac, I. Gornishka, and M. Worring. Multimodal classification of violent online political extremism content with graph convolutional networks. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, pages 245–252. ACM, 2017.

[144] J. Sang, J. Liu, and C. Xu. Exploiting user information for image tag refinement. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1129–1132. ACM, 2011.

[145] P. Saraiya, P. Lee, and C. North. Visualization of Graphs with Associated Timeseries Data. In J. Stasko and M. O. Ward, editors, *Proceedings of the IEEE Symposium on Information Visualization*, InfoVis, pages 225–232. IEEE, 2005.

[146] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

[147] F. Shi, J. G. Foster, and J. A. Evans. Weaving the Fabric of Science: Dynamic Network Models of Science's Unfolding Structure. *Social Networks*, 43:73–85, 2015.

[148] Y. Shi, M. Larson, and A. Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1):3, 2014.

[149] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine, vol. 30, no. 3, pp. 83 – 98*, 2013.

[150] M. Simonovsky and N. Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of Computer Vision and Pattern Recognition, 2017. CVPR*, 2017.

[151] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[152] C. Snoek, K. Van De Sande, D. Fontijne, A. Habibian, M. Jain, S. Kordumova, Z. Li, M. Mazloom, S. Pintea, R. Tao, et al. Mediamill at trecvid 2013: Searching concepts, objects, instances and events in video. In *NIST TRECVID Workshop*, 2013.

[153] R. Socher, D. Chen, C. D. Manning, and A. Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pages 926–934, 2013.

[154] N. Srebro, J. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. In *Advances in neural information processing systems*, pages 1329–1336, 2005.

[155] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[156] N. Srivastava and R. R. Salakhutdinov. Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems*, pages 2222–2230, 2012.

[157] D. Streeb, D. Arya, D. A. Keim, and M. Worring. Visual Analytics Framework for the Assessment of Temporal Hypergraph Prediction Models. In *Proceeedings of the Set Visual Analytics Workshop at IEEE VIS 2019*, 2019.

[158] G. Strezoski and M. Worring. Omniart: multi-task deep learning for artistic data analysis. *arXiv preprint arXiv:1708.00684*, 2017.

[159] G. Strezoski and M. Worring. Omniart: A large-scale artistic benchmark. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):88, 2018.

[160] L. Sun, S. Ji, and J. Ye. Hypergraph spectral learning for multi-label classification. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 668–676. ACM, 2008.

[161] S. Tan, J. Bu, C. Chen, B. Xu, C. Wang, and X. He. Using rich social media information for music recommendation via hypergraph model. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 7(1):1–22, 2011.

[162] J. Tang, Z. Li, M. Wang, and R. Zhao. Neighborhood discriminant hashing for large-scale image retrieval. *IEEE Transactions on Image Processing*, 24(9):2827–2840, 2015.

[163] J. Tang, J. Liu, M. Zhang, and Q. Mei. Visualizing large-scale and high-dimensional data. In *Proceedings of the 25th international conference on world wide web*, pages 287–297, 2016.

[164] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077. International World Wide Web Conferences Steering Committee, 2015.

[165] J. Tang, X. Shu, Z. Li, Y.-G. Jiang, and Q. Tian. Social anchor-unit graph regularized tensor completion for large-scale image retagging. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[166] J. Tang, X. Shu, G.-J. Qi, Z. Li, M. Wang, S. Yan, and R. Jain. Tri-clustered tensor completion for social-aware image tag refinement. *IEEE transactions on pattern analysis and machine intelligence*, 39(8):1662–1674, 2017.

[167] B. Tangirala, I. Bhandari, D. Laszlo, D. K. Gupta, R. M. Thomas, and D. Arya. Livestock monitoring with transformer. *arXiv preprint arXiv:2111.00801*, 2021.

[168] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[169] R. M. Thomas, S. Gallo, L. Cerliani, P. Zhutovsky, A. El-Gazzar, and G. van Wingen. Classifying autism spectrum disorder using the temporal statistics of resting-state functional MRI data with 3d convolutional neural networks. *Frontiers in Psychiatry*, 11:440, 2020.

[170] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080, 2016.

[171] P. Valdivia, P. Buono, and J.-D. Fekete. Hypenet: Visualizing Dynamic Hypergraphs. *EuroVis 2017-19th EG/VGC Conference on Visualization*.

[172] P. Valdivia, P. Buono, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Using Dynamic Hypergraphs to Reveal the Evolution of the Business Network of a 17th Century French Woman Merchant. In *VIS 2018 - 3rd Workshop on Visualization for the Digital Humanities*, Berlin, Germany, 2018.

[173] P. Valdivia, P. Buono, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2019.

[174] F. van Ham. Using Multilevel Call Matrices in Large Software Projects. In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*, pages 227–232. IEEE, 19-21 Oct. 2003.

[175] C. Vehlow, F. Beck, and D. Weiskopf. Visualizing Group Structures in Graphs: A Survey. *Computer Graphics Forum*, 36(6):201–225, 2017.

[176] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. *ICLR*, 2018.

[177] X. Wang and A. Gupta. Videos as space-time region graphs. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 399–417, 2018.

[178] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, pages 2022–2032, 2019.

[179] X. Wang and G. Sukthankar. Link prediction in multi-relational collaboration networks. In *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 1445–1447. ACM, 2013.

[180] X. Wang and G. Sukthankar. Link prediction in heterogeneous collaboration networks. In *Social network analysis-community detection and evolution*, pages 165–192. Springer, 2014.

[181] Y. Wang and T. Karaletsos. Stochastic aggregation in graph neural networks. *arXiv preprint arXiv:2102.12648*, 2021.

[182] Y.-J. Wang, M. Xian, J. Liu, and G.-y. Wang. Study of Network Security Evaluation Based on Attack Graph Model. *Journal of China Institute of Communications*, 28(3):29, 2007.

[183] G. Weimann. *www. terror. net: How modern terrorism uses the Internet*, volume 116. DIANE Publishing, 2004.

[184] L. Wen, D. Du, S. Li, X. Bian, and S. Lyu. Learning non-uniform hypergraph for multi-object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8981–8988, 2019.

[185] M. M. Wolf, A. M. Klinvex, and D. M. Dunlavy. Advantages to modeling relational data using hypergraphs versus graphs. In *2016 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–7. IEEE, 2016.

[186] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 2014.

[187] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[188] J. Xu, V. Singh, Z. Guan, and B. Manjunath. Unified hypergraph for image ranking in a multimodal context. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2333–2336. IEEE, 2012.

[189] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? *International Conference on Learning Representations (ICLR)*, 2019.

[190] N. Yadati. Neural message passing for multi-relational ordered and recursive hypergraphs. *Advances in Neural Information Processing Systems*, 33, 2020.

[191] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. In *Advances in Neural Information Processing Systems*, pages 1511–1522, 2019.

[192] F. Yan and K. Mikolajczyk. Deep correlation for matching images and text. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3441–3450, 2015.

[193] Y. Yan, J. Qin, J. Chen, L. Liu, F. Zhu, Y. Tai, and L. Shao. Learning multi-granular hypergraphs for video-based person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2899–2908, 2020.

[194] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Chang. Network representation learning with rich text information. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[195] C. Yang, R. Wang, S. Yao, and T. Abdelzaher. Hypergraph learning with line expansion. *arXiv preprint arXiv:2005.04843*, 2020.

[196] Z. Yin, M. Gupta, T. Weninger, and J. Han. Linkrec: a unified framework for link recommendation with user attributes and graph structure. In *Proceedings of the 19th international conference on World wide web*, pages 1211–1212. ACM, 2010.

[197] S.-e. Yoon, H. Song, K. Shin, and Y. Yi. How much and when do we need higher-order information in hypergraphs? a case study on hyperedge prediction. In *Proceedings of The Web Conference 2020*, pages 2627–2633, 2020.

[198] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 793–803, 2019.

[199] D. Zhang, J. Yin, X. Zhu, and C. Zhang. User profile preserving social network embedding. In *IJCAI*, pages 3378–3384, 2017.

[200] M. Zhang and Y. Chen. Link prediction based on graph neural networks. In *Advances in Neural Information Processing Systems*, pages 5165–5175, 2018.

[201] M. Zhang, Z. Cui, S. Jiang, and Y. Chen. Beyond link prediction: Predicting hyperlinks in adjacency space. In *AAAI*, volume 1, page 6, 2018.

[202] M. Zhang, Z. Cui, M. Neumann, and Y. Chen. An end-to-end deep learning architecture for graph classification. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[203] S. Zhang, Z. Ding, and S. Cui. Introducing hypergraph signal processing: theoretical foundation and practical applications. *IEEE Internet of Things Journal*, 7(1):639–660, 2019.

[204] Z. Zhang, P. Cui, and W. Zhu. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[205] Y.-L. Zhao, Q. Chen, S. Yan, T.-S. Chua, and D. Zhang. Detecting profilable and overlapping communities with user-generated multimedia contents in lbsns. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 10(1):3, 2013.

[206] D. Zhou, J. Huang, and B. Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in neural information processing systems*, pages 1601–1608, 2007.

[207] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020.

7

CONCLUSIONS

7.1 THESIS SUMMARY

This thesis investigates the potential of hypergraphs for capturing higher-order relations between objects in a multimodal dataset. These relations are often sub-optimally represented by pairwise connections used in a graph. Hence, in order to unlock the full potential of relational information within a multimodal dataset, this thesis proposes several geometric deep learning approaches for capturing and learning higher-order relations.

Chapter 2 focuses on determining the degree to which deep learning on hypergraphs can exploit higher-order relations in a multimodal dataset and make predictions solely by capturing relational information among the objects. This is a challenging problem as the significance of capturing these relations grows rapidly with the increase in the number of objects. Traditional graph-based representation of these relations leads to a loss in information, as it does not capture higher-order relations. The chapter investigates the advantage of using a hypergraph over a graph in representing multimodal datasets. To this end we propose a geometric deep learning framework which truly captures relational information between objects and learns representations of these objects. The proposed representation learning approach explores the potential of hypergraph topology, without looking at the content of the nodes. We demonstrate through experiments that a hypergraph-based representation is the most efficient way to build a model for learning the same volume of information in a relational dataset as compared to a graph.

Chapter 3 proposes a framework that amalgamates the content of an object into the hypergraph learning framework proposed in Chapter 2. We show that incorporating content-based analysis can further enhance the representation learning capabilities of a hypergraph model and provide flexibility in performing several tasks. This chapter paves the way forward for combining the content of an object with its relations using a hypergraph. As a case study, we perform an analysis of a violent online political extremism discussion forum, which poses major challenges for domain experts. These challenges arise from the complexity of relations and interactions between the users of the forum and the difficulty in understanding the content they share in the form of text, images and videos. We demonstrate the generalizability and flexibility of our hypergraph learning framework in jointly modeling content with relations and conducting extensive experimentation around four practical use cases.

Chapter 4 focuses on the increasingly relevant challenge of interactive deep learning. It aims at designing algorithms that can facilitate training of deep learning models with a human expert in the loop. We propose a hypergraph-based methodology HYPER-MATRIX, that can overcome the rigidity of geometric deep learning models in adapting to any structural change in terms of addition or removal of edges in a hypergraph structure. In real world scenarios, these structural changes can occur with evolving relations between objects over time. We evaluate our approach in a case study and through formative evaluation with law enforcement experts using real-world

communication data. The results show that our approach surpasses existing solutions in terms of scalability, interactivity and accuracy.

Chapter 5 examines the problem of scaling hypergraph learning models to large-scale datasets with many modalities. In this chapter, we propose HYPERLEARN, a hypergraph-based framework which uses a distributed training approach for learning complex higher-order relations. Adding new modalities to HYPERLEARN requires only an additional GPU unit, keeping the computational time unchanged, which brings representation learning to truly multimodal datasets. Through extensive experiments, we demonstrate benefits of our approach with regards to both accuracy and computational time through extensive experimentation.

Finally, Chapter 6 addresses the major limitation of existing hypergraph learning frameworks - their inherently transductive nature. This implies that these methods can only perform inference on objects that were present in the hypergraph at training time, and fail to infer on previously unseen objects. The transductive nature limits the ability of existing approaches to adapt to any structural change in terms of addition or removal of nodes and edges in hypergraphs. In this chapter, we propose HYPERMSG that paves the way for an inductive learning solution for learning on hypergraphs. HYPERMSG can perform inferences on previously unseen nodes, and can thus be used to model evolving hypergraphs. It comprises a message passing scheme which is capable of jointly capturing the intra-relations (within a hyperedge) as well as the inter-relations (across hyperedges) between objects (nodes). Through quantitative experiments on several representative datasets such as citation networks, social multimedia networks and a neuroimaging dataset, we show that HYPERMSG outperforms the existing methods for hypergraph learning. Further, we conclude that HYPERMSG yields accurate and robust results for multiple tasks and is thus suitable for performing multimodal representation learning on constantly evolving real world datasets.

7.2 REFLECTION AND FUTURE WORK

Given that the field of geometric deep learning is still in its early years, we view this thesis as early work on the topic of deep learning on hypergraphs. We have made the following three distinct contributions to the topic:

- An *iterative learning framework* that connects matrix and tensor completion with deep learning for capturing higher-order relations using a hypergraph. We introduced methods to exploit relational information between objects in multimodal datasets and perform several tasks such as classification, link prediction and recommendation. We show the advantage of using hypergraphs over graphs to capture both the pair wise and high-order relations among objects, yielding more and without any information loss compared to graphs.

- An *interactive learning framework* for incorporation of expert knowledge and seamless refinement of hypergraph models. The proposed approach can be directly used in relevance feedback and active learning scenarios in multimedia analytics systems, and can accommodate any structural changes of a hypergraph.

- An *inductive learning framework* that uses a modular two-level neural message passing strategy on hypergraphs to accurately and efficiently propagate information within a hypergraph. The proposed approach allows inference on previously unseen nodes. It can accurately quantify both the local and global importance of a node, thus capturing the structural properties of a hypergraph.

Bearing in mind these contributions, we believe the thesis constitutes the first complete framework for multimodal deep learning on hypergraphs: we can now deploy an efficient,

scalable, generic and inductive learning approach for a multimodal dataset represented on a hypergraph.

Much of this thesis involves multiple modalities. In particular, this thesis emphasized the importance of higher-order relations among objects in multimedia datasets. These relations are defined by either using the common intrinsic (content-based) characteristics of the objects or by grouping them based on their metadata derived from another modality. Many questions are still open, and we believe geometric deep learning for multimedia analytics is the way forward.

The first major question that we will like to investigate in future research is extending the message passing mechanism to objects belonging to different types of modalities. This is a challenging task as objects (nodes) belonging to different modalities contain features with dissimilar nature and distributions. This makes the flow of message across objects non-uniform and unstable. Information flow by message passing between objects belonging not only to the same modality, but also between modalities, will further improve the learned representations, making the model versatile and a truly intelligent multimedia analytics tools.

Along other lines, we can turn ourselves to the future of using hypergraphs for video analytics. Even though many recent works have introduced hypergraphs for capturing long term spatio-temporal dependencies for multi-object tracking in videos [184, 193], further research on the use of hypergraphs for video intelligence in scenarios with overlapping relations between objects is still inadequate. Some of the prominent directions of research where a hypergraph structure can be used are - (a) in grouping of points over time in point cloud object tracking for visually similar objects encountered in video surveillance applications, (b) in combining actions, images and background scenes for action recognition especially in ambiguous videos and (c) in aiding long-term multi-object trackers with valuable long-term group based relations among objects.

Overall, while the road of multimodal learning is still wide open, as is the role of hypergraphs in the journey on it, we believe this thesis significantly contributes to the first kilometers in developing a multimodal deep learning framework for hypergraphs. We hope that the comprehensive frameworks presented in this thesis empower further research advancing the capabilities of hypergraphs in multimodal learning for insight gain, in turn also advancing the field of geometric deep learning.

SAMENVATTING

MULTIMODAL DEEP LEARNING ON HYPERGRAPHS

Wij vatten dit proefschrift samen door de belangrijkste bevindingen en onderzoeksvragen van de vorige hoofdstukken nogmaals lands te gaan.

Hoofdstuk 2 richt zich op het bepalen van de mate waarin deep learning op hypergrafen hogere-orde relaties in een multimodale dataset kan benutten en voorspellingen kan doen, uitsluitend door relationele informatie tussen de objecten vast te leggen. Dit is een uitdagend probleem, aangezien het belang van het vastleggen van deze relaties snel groeit wanneer het aantal objecten toeneemt. Traditionele, op graaf-gebaseerde weergave van deze relaties leidt tot een verlies aan informatie omdat het geen hogere-orde relaties vastlegt. Dit hoofdstuk onderzoekt het voordeel van het gebruik van een hypergraaf boven een graaf voor het weergeven van multimodale datasets. Hiertoe stellen we een geometrisch deep learning-raamwerk voor dat echt relationele informatie tussen objecten vastlegt en representaties van deze objecten leert. De voorgestelde benadering voor het leren van representaties verkent het potentieel van hypergrafietopologie, zonder naar de inhoud van de knooppunten te kijken. We laten door middel van experimenten zien dat een hypergraaf-gebaseerde representatie de meest efficiënte manier is om een model te bouwen voor het leren van dezelfde hoeveelheid informatie in een relationele dataset als in een graaf.

Hoofdstuk 3 stelt een raamwerk voor dat de inhoud van een object samenvoegt in het hypergraaf-leerraamwerk dat in hoofdstuk 2 is voorgesteld. We laten zien dat het opnemen van op inhoud gebaseerde analyse de representatieleermogelijkheden van een hypergraaf- model verder kan verbeteren en flexibiliteit kan bieden bij het uitvoeren van verschillende taken. Dit hoofdstuk maakt de weg vrij voor het combineren van de inhoud van een object met zijn relaties met behulp van een hypergraaf. Als case study voeren we een analyse uit van een gewelddadig online discussieforum over politiek extremisme, dat grote uitdagingen vormt voor domeinexperts. Deze uitdagingen komen voort uit de complexiteit van relaties en interacties tussen de gebruikers van het forum en de moeilijkheid om de inhoud die ze delen in de vorm van tekst, afbeeldingen en video's te begrijpen. We demonstreren de generaliseerbaarheid en flexibiliteit van ons hypergraaf-leerraamwerk door gezamenlijk inhoud te modelleren met relaties en uitgebreide experimenten rond vier praktische use- cases.

Hoofdstuk 4 richt zich op de steeds relevantere uitdaging van interactief deep learning. Het is gericht op het ontwerpen van algoritmen die de training van deep learning-modellen kunnen vergemakkelijken met een menselijke expert in het proces. We stellen een hypergraaf-gebaseerde methodologie HYPER-MATRIX voor, die de starheid van geometrische deep learning-modellen kan overwinnen bij het aanpassen aan elke structurele verandering in termen van toevoeging of verwijdering van randen in een hypergraaf-structuur. In real- world scenario's kunnen deze structurele veranderingen optreden met veranderende relaties tussen objecten in de tijd. We evalueren onze aanpak in een case study en door formatieve evaluatie met wetshandhavingsexperts met behulp van real-world communicatiegegevens. De resultaten laten zien dat onze aanpak bestaande oplossingen overtreft op het gebied van schaalbaarheid, interactiviteit en nauwkeurigheid.

Hoofdstuk 5 onderzoekt het probleem van het schalen van hypergraaf-leermodellen naar grootschalige datasets met veel modaliteiten. In dit hoofdstuk stellen we HYPERLEARN voor, een op hypergrafen gebaseerd raamwerk dat een gedistribueerde trainingsbenadering gebruikt voor het leren van complexe hogere-orderelaties. Het toevoegen van nieuwe modaliteiten aan

HYPERLEARN vereist slechts een extra GPU-eenheid, waardoor de rekentijd ongewijzigd blijft, wat representatieleren naar echt multimodale datasets brengt. Door middel van uitgebreide experimenten demonstreren we de voordelen van onze aanpak met betrekking tot zowel nauwkeurigheid als rekentijd door middel van uitgebreide experimenten.

Tot slot gaat hoofdstuk 6 in op de belangrijkste beperking van bestaande hypergraaf- leerraamwerken - hun inherent transductieve aard. Dit houdt in dat deze methoden alleen geëvalueerd kunnen worden op objecten die aanwezig waren in de hypergraaf tijdens de trainingstijd, en er niet in slagen om conclusies te trekken over voorheen onzichtbare objecten. De transductieve aard beperkt het vermogen van bestaande benaderingen om zich aan te passen aan elke structurele verandering in termen van toevoeging of verwijdering van knooppunten en randen in hypergrafen. In dit hoofdstuk stellen we HYPERMSG voor dat de weg vrijmaakt naar een inductieve leeroplossing voor het leren op hypergrafen. HYPERMSG kan gevolgtrekkingen uitvoeren op voorheen onzichtbare knooppunten en kan dus worden gebruikt om evoluerende hypergrafen te modelleren. Het omvat een schema voor het doorgeven van berichten dat in staat is om de onderlinge relaties (binnen een hyperrand) en de onderlinge relaties (over hyperranden) tussen objecten (knooppunten) gezamenlijk vast te leggen. Door kwantitatieve experimenten op verschillende representatieve datasets zoals citatienetwerken, sociale multimedianetwerken en een neuroimaging-dataset, laten we zien dat HYPERMSG beter presteert dan de bestaande methoden voor hypergraaf-leren. Verder concluderen we dat HYPERMSG nauwkeurige en robuuste resultaten oplevert voor meerdere taken en dus geschikt is voor het uitvoeren van multimodaal representatieleren op constant evoluerende real-world datasets.

Hoewel de weg naar multimodaal leren nog steeds openligt, evenals de rol van hypergrafen in de reis hierover, geloven we dat dit proefschrift een significante bijdrage levert aan de eerste kilometers bij het ontwikkelen van een multimodaal deep learning-raamwerk voor hypergrafen. Dit proefschrift vormt een compleet raamwerk voor multimodaal deep learning op hypergrafen: door een efficiënte, schaalbare, generieke en inductieve leerbenadering voor te stellen voor een multimodale dataset weergegeven op een hypergraaf. We hopen dat de uitgebreide kaders die in dit proefschrift worden gepresenteerd, verder onderzoek mogelijk maken dat de mogelijkheden van hypergrafen in multimodaal leren voor het verkrijgen van inzichten bevordert, en op hun beurt ook het gebied van geometrisch deep learning bevorderen.

ACKNOWLEDGMENTS

The process of earning a doctorate was long and arduous - coming all the way from far east, it was obvious to land into puddles, often literally, during the course of my Ph.D. It certainly could not have been accomplished without the help, guidance, support and prod from so many people. This thesis would not have taken the desired shape without their contributions. I would like to recognize these otherwise unseen forces.

Firstly, I am extremely grateful to my supervisor, Marcel Worring for his invaluable advice, continuous support, and patience during my Ph.D. The journey started when I interviewed with you and Stevan in September 2016, and within three weeks I resigned from my previous job and applied for the Dutch visa. Looking back that was almost a leap of faith. I will always be thankful to you for providing me with this opportunity and making me understand the nuances of academic research. As a mentor, you helped me structure my thoughts, kept me grounded in times of extreme optimism, helped me develop a scientific mindset and made me a better time planner ahead of deadlines (which is still a work in progress). The project that we were associated with during this Ph.D. gave me a lot of exposure in the field of forensic research, AI model deployment and in exploring Europe. It was a great experience with occasional frustrations which I had the freedom to openly share with Marcel. I was glad I got to learn so much about European culture from you through our meetings in several countries. Your timely advice, meticulous scrutiny, and uncountable reviews have helped me the most in finishing this thesis.

My gratitude extends to my co-supervisor Stevan Rudinac, who made my PhD journey smoother and much more enjoyable with his honest feedback and at times crazy ideas. You could be the one person whom I could approach for any suggestions at any time or place whether it was at 4 in the morning or whether you were in the middle of a vacation somewhere in Serbia. I learnt a lot from you about writing an academic paper, being open to any type of criticism and most importantly about always keeping a positive outlook in life with a smile. Your enthusiasm is infectious and always encouraged me in everything no matter how misplaced my confidence might have been.

I would express my sincere gratitude to the members of my Ph.D. defense committee, namely Prof. Alan Hanjalic, Dr. Thomas Kipf, Prof. Cees Snoek, Prof. Zeno Geradts and Dr. Evangelos Kanoulas. It is a pleasure and an honor to have you on my committee.

Life during the PhD was made much easier by my office mates and the many friends I made over the years. All my lab mates with whom I had innumerable sessions of dinners, borrels, volleyball, squash and Mill. The research group and the fish tank changed a lot over the course of time, however the first year would always be closest to my memory. Pascal Mettes and Spencer Capallo made the initial months easier and enjoyable with their knowledge and wisdom about UvA and beyond. Sarah Ibrahimi, thank you for being such a great colleague, conversation partner, my all-things-Dutch-guru and for organizing all the events during this period. Gjorgji Strezoski, the man, the friend, the paranymph and a true inspiration. It was an absolute pleasure to share the same problems, solutions and sometimes same societal issues with you. I feel lucky that my PhD journey overlapped with yours. It was and still is a learning experience to have you around. I would like to thank all of my colleagues and labmates at Science Park for being a part of this journey, in particular Mehmet, Andrew, Shuo, Yunlu, Inske, Sadaf, Nour, Jia-Hong, Mert, Berkay, Anil, Maurice, Shuai, Andrea, Fida, William Thong, Jorn, Tom, Zenglin, David, Jiaojiao Zhao, Riaan. Life at Science Park would have not been so colorful without you all.

Similar to any other thesis from our group, it could not have been done without the help of Dennis Koelma. It is no secret that none of our experiments would have finished in time without his constant vigilance on the servers. I hope I have unreserved all my nodes in DAS5! Virginie Mess and Petra Venema, for keeping all of us in check and for always being available to help us. Morris Franken, thanks for being the squash partner over the years, we still have a lot of squash left in us. Kandan Ramakrishnan for helping me out in the initial months - explaining all the subtleties of PhD life in Amsterdam and keeping my homemade idly cravings in check.

I would like to especially thank Deepak and Rajat for all the collaborative efforts and invaluable mentorship. Deepak helped me immensely in brainstorming of ideas and writing papers. He made sure all the coffee breaks were worth taking during the day. I highly appreciate all the tips Rajat provided during this time and thankful for introducing me to some of the most interesting topics ranging from AI in neuroscience to livestock monitoring which further extended into cooking south Indian food, tasting varieties of beer and discussing cricket.

I was lucky to have Kuldeep and Vedant as my flatmates during this period. With them, it never felt staying far from homeland and we could make Havikshorst an extension of our IIT Kanpur corridor. To Daroga (atleast in this thesis no one will object his name), there was and will never be any formal words or pleasantries shared, not even in this section of the thesis. You were and will always be our go-to man. Living in Amsterdam was made more homely by Debjani, Pushpa and Saurabh by sharing many breakfast, lunch and dinner, celebrating birthdays, riding boats and taking care of everything whenever I was away. Finally, to making me feel a part of their family, I would really like to thank Rajat and Teresa for rejoicing with me when times were good and sharing my sorrow when they were not.

Life in Amsterdam would not be the same if the ‘cricketer’ within me was not satisfied. Before I could have chosen a cricket club in Netherlands the club chose me. ACC provided me with the same weekend ups and downs, excitement and thrills which I thought had finished the day I left India. There were many memories created, if only we could have won more matches, they would all have been worth remembering. Nonetheless, I thank all the members at ACC (it will be a long list if I start naming them) especially Taku and Raza who made sure to keep my competitive spirit ignited every week on the cricket field and improve my game every season.

After joining Storeshippers, I still had to spend many nights and weekends working on the thesis. They kept my learning curve sharp and rising with new wisdoms everyday during our delicious lunches. I am grateful to the colleagues for their understanding and support in this period.

I would like to thank my parents - Devendra and Seema who made innumerable sacrifices for me. To my sister - Divya who is always been supportive and extremely caring. And to all my family members who have contributed in many ways, I owe a debt of gratitude to them without whom I would have not come this far. I will always be obliged to Navrose for advising me over the years whenever I had a muddled mind. And to Subroto sir for teaching all the necessary life lessons and explaining them in the most simplest of words.

Lastly, I thank my wife, Akanksha. The courage she showed in graciously accepting my decision to suddenly move outside India was highly commendable. She took a leap of faith in later moving to a country closer to me, so can spend the maximum possible time together. And while she never wrote a line of code nor fixed any GPU related issues, she has been a significant contributor to this work; through all the obstructions, deadlines, flight travels, the pandemic and all the unseen, unexpected complications that don’t make it into the text.