# Machine learning tasks and representations for heterogeneous information networks

Fang, Y.

**Publication date**
2023

# 1
# Introduction

Complex information often involves multiple types of objects and relations. Such information can be represented via *heterogeneous information networks* (HINs) [83]. In a HIN different types of nodes (objects) are connected by edges (relations) [93]. Typical HINs involve bibliography networks, social networks, knowledge graphs etc. Previous research typically chooses to build homogeneous information networks to model relations in real-life networks, which only feature a single type of node. Such homogeneous information networks only describe part of the information in a real-life system, and ignore the heterogeneity of different nodes and relations, which will cause a significant loss of information loss that cannot easily be recovered.

In recent years, HINs have drawn more attention from the research community. HINs provide a more natural and complete abstraction of the real world than homogeneous information networks. HINs provide a richer modeling tool than homogeneous information networks. They naturally integrate different types of objects and interactions, while containing more fruitful structural and semantic features. Based on the above advantages, and compared to homogeneous information networks, HINs help to provide more effective solutions for many machine learning tasks, such as search, classification, and prediction tasks [8].

To be able to process a network in a machine learning context, *network representation learning* (NRL), also known as network embedding learning, has been investigated extensively. NRL is aimed at obtaining an embedding of a network in a low-dimensional space, Classical network embedding models like DeepWalk [77], LINE [95], and node2vec [33] have been devised for homogeneous networks, using random walks to capture the structure of networks. However, these methods lack the ability to capture a *heterogeneous* information network with multiple types of objects and relations. Hence, models designed specifically for HINs have been proposed [18, 28, 50]. A central concept here is that of a *metapath*, which is a sequence of node types with edge types in between. To leverage the relationship between nodes and metapaths, different mechanisms have been proposed, such as the heterogeneous SkipGram [18], proximity distance [50], and the Hardmard function [28]. However, because of the limited ability of metapaths to capture the neighborhood structure of a node, the performance of these network representation learning methods is limited.

Recently, graph neural networks (GNNs) have shown promising results on the task of modeling the structure of a network [55, 101, 115]. GNNs usually involve encoders that are able to explore and capture the neighborhood structure around a node, thus improving the performance on representing an HIN. GNNs propagate and aggregate the node features along the network topology, and are realized in an end-to-end semi-supervised manner.

In this thesis, we aim to explore different learning mechanisms for network representation learning to fit into different scenarios and different machine learning tasks of HINs. Specifically, we first study representation learning of dynamic HINs, as real-life networks are always evolving. Then, we pre-train a HIN without using labeled information, and the pre-trained GNN model can easily be adapted to different datasets and different downstream tasks. We also investigate few-shot learning of HINs, where only a handful of labels are given. Last, we make use of the auxiliary textual information to further facilitate the learning of HINs.

## 1.1 Research outline and questions

Throughout this thesis, we intend to answer the following research question, *how to comprehensively mine the information and features of HINs under different scenarios*? Different representation learning mechanisms are proposed in this thesis, so that the corresponding downstream machine learning tasks can be addressed. Specifically, four research aspects of HINs are considered in this thesis, and we formulate each of them as a research question.

**RQ1** How to learn the representation of dynamic HINs?

Real-life information networks are always evolving, with new nodes and edges coming and old nodes and edges deleted. Most traditional NRL models focus on static networks. Unlike static network embeddings, the techniques for dynamic HINs need to be incremental and scalable so as to be able to handle network evolutions effectively. This renders most existing static embedding models, which need to process the entire network step by step, unsuitable and inefficient.

To answer **RQ1**, we propose a novel dynamic HIN embedding model, named M-DHIN, which provides a scalable method to capture the features of a dynamic HIN via so-called *metagraphs*. We also propose an LSTM-based deep autoencoder mechanism to enable M-DHIN to predict the future network via history structure evolutions.

**RQ2** How to pre-train HINs?

Traditional NRL models such as GNNs need to be trained in an end-to-end manner with supervised information for a task, and the model learned on one dataset cannot easily be transferred to other, out-of-domain datasets. For different datasets and tasks, the methods listed above need to be re-trained all over. Additionally, in many real-life datasets, the amount of available labeled data is rarely sufficient for effective training. The shortcomings listed above could be addressed by pre-training techniques that have been widely used in natural language processing and computer vision. Then, how can we pre-train graph-like data?

To answer **RQ2**, we first design two pre-training tasks that are applied on large datasets mining self-supervision information. Then, for a specific downstream task on a specific dataset, we use fine-tuning techniques with few task-specific parameters, so that the pre-trained model could be fast adapted to new tasks and new datasets.

**RQ3** How to learn the representation of HINs in a few-shot setting?

During traditional representation learning processes, it is taken for granted that the majority of labels in the network is available, and the GNNs are trained in a supervised manner. In practice, however, it is common that only a handful of labels are given, which poses serious challenges to keeping up the performance. For few-shot learning, meta-learning approaches have been studied extensively in computer vision, so is it possible to apply meta-learning techniques on graph data?

To answer **RQ3**, we propose a unified meta-learning framework that takes subgraphs as training samples to form meta-training and meta-testing datasets. A heterogeous GNN module is used as a base model to fully capture heterogeneous information. We also adopt a GAN-based contrastive module to leverage unsupervised information, and a structure module to employ graph structural information. With the help of the meta-training framework, it can be applied across different tasks and graphs.

**RQ4** How to make use of the textual information when learning the representation of HINs?

Most HINs come with textual information, e.g., a title and an abstract of a paper node in an academic network, which can provide fruitful additional information for downstream tasks. Most current work on HINs ignores such textual information and maps the node of a graph into low-dimensional representations based only on structural information. Existing textual network embedding models are all designed for a supervised scenario, which requires abundant labeled data for training. In other words, they are not applicable to the few-shot learning setting. However, in real-life applications, it is common that only a handful of labels are available, which poses serious challenges to keeping up the performance. Second, these methods are all originally designed for homogeneous networks, with no prior work yet trying to solve few-shot learning issues for textual HINs.

To answer **RQ4**, we propose a learnable continuous prompt learning framework to solve the few-shot problem for textual HINs. Specifically, our proposal contains a text encoder to leverage textual information, a graph encoder that encodes structural and heterogeneous features, along with self-supervised information. We then introduce a contrastive learning mechanism to align the text and graph embeddings.

## 1.2 Main contributions

In this section, we summarize the main contributions of this thesis.

### 1.2.1 Algorithmic contributions

We propose several novel network embedding algorithms to fit into four different HIN representation scenarios.

(1) Representation learning for Dynamic HINs

   (a) A novel dynamic network embedding model, M-DHIN, that learns representations of every snapshot of a dynamic HIN via metagraph-based complex embeddings on dynamic datasets.
   (b) An LSTM-based deep autoencoder mechanism to enable M-DHIN to predict the future network via history structure evolutions.

(2) Pre-training on HINs

   (a) A pre-training and fine-tuning framework PF-HIN to mine information contained in a HIN; PF-HIN is transferable to different downstream tasks and to datasets of different domains.
   (b) A deep bi-directional transformer encoder to capture the structural features of a HIN; the architecture of PF-HIN is a variant of a GNN.
   (c) Two pre-training mechanisms, i.e., type-based masked node modeling and adjacent node prediction; both help PF-HIN to capture heterogeneous node features and relationships between nodes.

(3) Representation learning on few-shot HINs

   (a) A meta-learning model META-HIN to deal with few-shot leaning problems in HINs.
   (b) A sampling strategy which extracts subgraphs to be trained so that META-HIN is applicable to three tasks and transferable across different HINs.
   (c) A structural module, a heterogeneous module, and a GAN-based contrastive module to capture the structural information, heterogeneous features, and unlabeled information of a subgraph, respectively.

(4) Representation learning on textual HINs

   (a) A prompt learning framework P-HIN to leverage textual information in textual HINs, and to deal with few-shot leaning problems simultaneously.
   (b) A graph encoder that can capture structural and heterogeneous features of HINs, while preserving the node-level and edge-level self-supervised information.

### 1.2.2 Empirical contributions

To verify the effectiveness of our algorithms, we conduct extensive experiments to evaluate the proposed algorithms for the considered tasks.

(1) Representation learning on Dynamic HINs

   (a) An empirical evaluation of M-DHIN on six downstream tasks, and a comparison against state-of-the-art models.

(b) An analysis of the impact of different modules of M-DHIN.

(c) An analysis of the impact of parameters, i.e., the choice of dimension and the negative ratio when sampling.

(2) Pre-training on HINs

(a) An empirical evaluation of PF-HIN on four downstream tasks and four datasets, and a comparison against state-of-the-art models.

(b) An analysis of the impact of different modules of PF-HIN.

(c) An analysis of the impact of parameters, i.e., the maximum length of the input sequence and the dimension of the node embedding.

(3) Representation learning on few-shot HINs

(a) An empirical evaluation of META-HIN on three downstream tasks, and a comparison against state-of-the-art models.

(b) An analysis of the impact of different modules of META-HIN.

(c) An analysis of the impact of parameters, i.e., the maximum number of nodes of a subgraph, and the number of shots used for meta-learning.

(4) Representation learning on textual HINs

(a) An empirical evaluation of P-HIN on three downstream tasks, and a comparison against state-of-the-art models.

(b) An analysis of the impact of different modules of P-HIN.

(c) An analysis of the impact of parameters, i.e., the number of the text tokens, and the number of shots used for training.

## 1.3 Thesis overview

In this section, we present an overview of the remaining chapters. In Chapter 2 to 5, we introduce four algorithms to handle four different heterogeneous information network scenarios with different downstream tasks. Then we empirically analyze the performance of the proposed algorithms against state-of-the-art alternatives. Specifically, the thesis is organized as follows.

In Chapter 2, we propose a novel and scalable representation learning model, M-DHIN, to explore the evolution of a dynamic HIN. We regard a dynamic HIN as a series of snapshots with different time stamps. We first use a static embedding method to learn the initial embeddings of a dynamic HIN at the first time stamp. We describe the features of the initial HIN via metagraphs, which retains more structural and semantic information than traditional path-oriented static models. We also adopt a complex embedding scheme to better distinguish between symmetric and asymmetric metagraphs. Unlike traditional models that process an entire network at each time stamp, we build a so-called *change dataset* that only includes nodes involved in a triadic closure or opening process, as well as newly added or deleted nodes. Then, we utilize the above metagraph-based mechanism to train on the change dataset. As a result of this setup, M-DHIN is scalable to large dynamic HINs since it only needs to model the entire HIN once while only the changed parts need to be processed over time. Existing dynamic embedding models only express the existing snapshots and cannot predict the future network structure. To equip M-DHIN with this ability, we introduce an LSTM based deep autoencoder model that processes the evolution of the graph via an LSTM encoder and outputs the predicted graph. Finally, we evaluate the proposed model, M-DHIN, on real-life datasets and demonstrate that it significantly and consistently outperforms state-of-the-art models.

In Chapter 3, we propose a self-supervised pre-training and fine-tuning framework, PF-HIN, to capture the features of a heterogeneous information network. Unlike traditional network representation learning models that have to train the entire model all over again for every downstream task and dataset, PF-HIN only needs to fine-tune the model and a small number of extra task-specific parameters, thus improving model efficiency and effectiveness. During pre-training, we first transform the neighborhood of a given node into a sequence. PF-HIN is pre-trained based on two self-supervised tasks, masked node modeling and adjacent node prediction. We adopt deep bi-directional transformer encoders to train the model, and leverage factorized embedding parameterization and cross-layer parameter sharing to reduce the parameters. In the fine-tuning stage, we choose four benchmark downstream tasks, i.e., link prediction, similarity search, node classification, and node clustering. PF-HIN outperforms state-of-the-art alternatives on each of these tasks, on four datasets.

In Chapter 4, we design a meta-learning framework, called META-HIN, for few-shot learning problems on HINs. To the best of our knowledge, we are the first to design a specific algorithm to address the few-shot problem on HINs. Unlike most previous models, which focus on a single task on a single graph, META-HIN is able to deal with three tasks (i.e., node classification, link prediction, and anomaly detection) across multiple graphs. Subgraphs are sampled to build the support and query set. Before being processed by the meta-learning module, subgraphs are modeled via a structure module to capture structural features. Then, a heterogeneous GNN module is used as the base model to express features of subgraphs. We also design a GAN-based contrastive learning module that is able to exploit unsupervised information of the subgraphs. In our experiments, we fuse several datasets from multiple domains to verify META-HIN's broad applicability in a multiple-graph scenario. META-HIN consistently and significantly outperforms state-of-the-art alternatives on every task and across all datasets that we consider.

In Chapter 5, we propose a prompt-learning framework P-HIN that provides a new angle to leverage textual information and that fits few-shot problems. To the best of our knowledge, we are among the first to introduce and exploit the idea of prompt learning in the context of textual HINs. The proposed framework P-HIN is composed of a text encoder and a graph encoder, and utilizes contrastive learning to align the graph-text pair. This pre-training operation naturally fits our few-shot learning setting. For the graph encoder, we introduce two graph pre-training tasks, masked node modeling and edge reconstruction, to exploit self-supervised information. During optimization, instead of handcrafted prompts, we use a learnable continuous text that enables more efficient

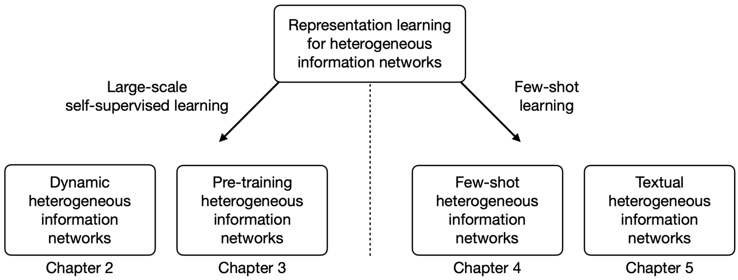

Figure 1.1: A visual map of the thesis.

and task-relevant transfer to downstream datasets. We consider a residual connection to use the context from the graph to prompt the text encoder. In experiments, P-HIN consistently and significantly outperforms state-of-the-art alternatives on all real-life datasets.

In Chapter 6, we summarize the thesis and provide some future research directions that build on the results in the thesis.

Figure 1.1 provides a graphical overview of the thesis. The chapters can be read independently of each other.

## 1.4 Origins

The main research chapters in the thesis are based on the following papers:

**Chapter 2** is based on the following paper:

- Y. Fang, X. Zhao, P. Huang, W. Xiao, and M. de Rijke. Scalable representation learning for dynamic heterogeneous information networks via metagraphs. *ACM Trans. Inf. Syst.*, 40(4):64:1–64:27, 2022.

YF designed the model, implemented it, ran experiments and did most of the writing. All authors contributed to design and discussion of the model. MdR and XZ contributed to the writing.

**Chapter 3** is based on the following paper:

- Y. Fang, X. Zhao, Y. Chen, W. Xiao, and M. de Rijke. PF-HIN: Pre-training for heterogeneous information networks. *IEEE Trans. Knowl. Data Eng.*, 35(8):8372–8385.

YF designed the model, implemented it, ran experiments and did most of the writing. All authors contributed to design and discussion of the model. MdR and XZ contributed to the writing.

**Chapter 4** is based on the following paper:

- Y. Fang, X. Zhao, W. Xiao, and M. de Rijke. Few-shot learning for heterogeneous information networks. *ACM Trans. Inf. Syst.*, Under review.

YF designed the model, implemented it, ran experiments and did most of the writing. All authors contributed to design and discussion of the model. MdR and XZ contributed to the writing.

**Chapter 5** is based on the following paper:

- Y. Fang, X. Zhao, W. Xiao, and M. de Rijke. Prompt learning for textual heterogeneous information networks. *IEEE Trans. Knowl. Data Eng.*, Under review.

YF designed the model, implemented it, ran experiments and did most of the writing. All authors contributed to design and discussion of the model. MdR and XZ contributed to the writing.

The writing of the thesis also benefited from work on the following publications:

- P. Huang, X. Zhao, M. Hu, Y. Fang, X. Li, and W. Xiao. Extract-select: A span selection framework for nested named entity recognition with generative adversarial training. In *Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 85–96. Association for Computational Linguistics, 2022.

- Q. Guo, X. Zhao, Y. Fang, S. Yang, X. Lin, and D. Ouyang. Learning hypersphere for few-shot anomaly detection on attributed networks. In *CIKM 2022: 31st ACM International Conference on Information and Knowledge Management*, pages 635–645. ACM, 2022.

In the following chapter, we will introduce how to learn the representation of dynamic HINs.