Multi-View Self-Supervised Heterogeneous Graph Embedding

Jianan Zhao[1][0000−0002−9743−7588], Qianlong Wen[1][0000−0003−3812−8395],
Shiyu Sun[1][0000−0002−0225−5053], Yanfang Ye (✉)[1][0000−0002−6038−2173], and
Chuxu Zhang (✉)[2][0000−0002−8349−7926]

[1] Case Western Reserve University, OH, USA
[2] Brandeis University, MA, USA
{jxz1244, qxw294, sxs2293, yanfang.ye }@case.edu, chuxuzhang@brandeis.edu

Abstract. Graph mining tasks often suffer from the lack of supervision from labeled information due to the intrinsic sparseness of graphs and the high cost of manual annotation. To alleviate this issue, inspired by recent advances of self-supervised learning (SSL) on computer vision and natural language processing, graph self-supervised learning methods have been proposed and achieved remarkable performance by utilizing unlabeled information. However, most existing graph SSL methods focus on homogeneous graphs, ignoring the ubiquitous heterogeneity of real-world graphs where nodes and edges are of multiple types. Therefore, directly applying existing graph SSL methods to heterogeneous graphs can not fully capture the rich semantics and their correlations in heterogeneous graphs. In light of this, we investigate self-supervised learning on heterogeneous graphs and propose a novel model named Multi-View Self-supervised heterogeneous graph Embedding (MVSE). By encoding information from different views defined by meta-paths and optimizing both intra-view and inter-view contrastive learning tasks, MVSE comprehensively utilizes unlabeled information and learns node embeddings. Extensive experiments are conducted on various tasks to show the effectiveness of the proposed framework.

Keywords: Self-Supervised Learning · Heterogeneous Graph Embedding · Graph Neural Network

1 Introduction

With the proliferation of real-world interaction systems, graph mining has been a popular topic with many real-world applications such as node classification, graph classification, and recommendation. Due to the ubiquitous sparseness of graphs and the deficiency of label supervision, it is vital to fully utilize the unlabeled information on graphs. However, the current state-of-the-art algorithms, which are mostly based on Graph Neural Networks (GNNs) [24, 36, 41], mainly utilize unlabeled information by simply aggregating their features and cannot thoroughly take advantage of the abundant unlabeled data [20]. Recently, aiming to fully exploit the unlabeled information for GNNs, self-supervised learning

(SSL) is naturally harnessed for providing additional supervision and achieves impressive improvements on various graph learning tasks [27].

The existing graph SSL methods fall into two categories, generative and contrastive [27]. However, they mainly focus on designing self-supervised tasks on homogeneous graphs, overlooking the ubiquitous heterogeneity and rich semantics in graphs. Unlike homogeneous graphs, a heterogeneous graph [34] is composed of multiple types of nodes and edges. To illustrate, consider a bibliography graph with its network schema shown in Figure 1 (a), where four types of nodes: Author (A), Paper (P), Venue (V), and Term (T) along with three types of edges: an author writes a paper, a paper is published in a venue, and a paper contains a term.

To fully capture the rich heterogeneity and complex semantics inside heterogeneous graph data, we are motivated to study the problem of self-supervised learning on heterogeneous graphs. However, this is a non-trivial task as there are several challenges to be addressed. Above all, *how to deal with the intrinsic heterogeneity of heterogeneous graphs?* Different from homogeneous graphs, heterogeneous graph contains rich semantics for each node. For example, in the example bibliography graph mentioned above, we can introduce two meta-paths APA and APVPA to capture the co-author and co-venue semantics respectively. Therefore, how to design self-supervised tasks to fully capture the rich semantic information is a critical yet challenging problem. What's more, *how to effectively model the complex correlations between these different semantics?* Previous works mainly focus on discriminating the heterogeneous context instances [3, 5, 2, 12], e.g. whether two authors have a co-author relationship, preserving the intra-context proximity [43]. However, the complex correlations between these contexts (inter-context), e.g. whether two authors with co-venue relationships have co-author relationships, remain less explored. Modeling these interactions not only encourages the embedding to preserve these interactions between semantics, pushing the model to extract useful information and encode them in node embeddings, but also alleviates the negative impact of the intrinsic sparseness of heterogeneous graphs [49].

To address the challenges mentioned above, we study self-supervised learning on heterogeneous graphs and focus on comprehensively encode the semantics and their correlations into node embeddings. In particular, we propose a novel model named **M**ulti-**V**iew **S**elf-supervised heterogeneous graph **E**mbedding (MVSE). MVSE firstly samples semantic subgraphs of different views defined by meta-paths. Then, each semantic subgraph is encoded to its own semantic latent space and further decoded to other semantic spaces to capture the semantic correlations. Finally, the embeddings are optimized by a contrastive loss preserving both intra-view, and inter-view interactions of semantic contexts. Our major contributions are highlighted as follows:

- We propose a novel self-supervised heterogeneous graph embedding model, in which some delicate designs, e.g., heterogeneous context encoding and multi-view contrastive learning are proposed to comprehensively learn good heterogeneous graph embeddings. Our work is among the earliest works that study self-supervised learning on heterogeneous graphs.

– While intra-semantic relationships are widely utilized, few works have attempted to model the correlations between the semantics in heterogeneous graphs. We design self-supervised learning tasks that preserve both intra- and inter-semantic information in node embeddings.
– We conduct extensive experiments on three real-world datasets to validate the effectiveness of MVSE compared with state-of-the-art methods. Through parameter analysis and ablation study, we further demonstrate that though often overlooked, preserving inter-view interactions is beneficial for heterogeneous graph embedding.

2 Related Work

2.1 Self-Supervised Learning on Graphs

To fully exploit the ample unlabeled information, self-supervised learning (SSL) on graphs has become a promising research topic and achieved impressive improvements on various graph learning tasks [20]. Existing graph SSL methods design generative or contrastive tasks [27] to better harness the unlabeled graph data. On the one hand, generative graph SSL models learn graph embedding by recovering graph structure and attributes. For example, VGAE [23] applies GCN-based variational auto-encoder [22] to recover the adjacency matrix of the graph by measuring node proximity. GraphRNN [44] uses a graph-level RNN and reconstructs adjacency matrix iteratively. GPT-GNN adopts GCNs [24] to reconstruct both graph structure and attribute information. On the other hand, contrastive graph SSL models learn graph embedding by discriminating positive and negative samples generated from graphs. To illustrate, Context Prediction and Attribute Mask [15] are proposed to preserve the structural and attribute information. DGI [37] contrasts local (node) and global (graph) embedding via mutual information maximization. MVGRL [8] contrasts embeddings from first-order and high-order neighbors by maximizing mutual information. GCC [32] performs subgraph instance discrimination across different graphs. Though graph SSL works have achieved significant performance improvements, most of the existing Graph SSL works focus on homogeneous graphs and can not address the complex semantics of heterogeneous graphs.

2.2 Heterogeneous Graph Embedding

Our work is also related to heterogeneous graph embedding (HGE), which encodes nodes in a graph to low-dimensional representations while effectively preserving the heterogeneous graph structure. HGE methods can be roughly divided into three categories [43]: proximity-preserving methods, relation learning methods, and message passing methods. The proximity-preserving HGE methods [3, 47, 18, 47, 10] are mostly random walk [31] based and optimized by (heterogeneous) skip-gram. The relation-learning HGE methods [1, 26, 40, 35, 42, 28] construct head, tail, and relation triplets and optimize embedding by a relation-specific scoring function that evaluates an arbitrary triplet and outputs a scalar to measure the acceptability of this triplet. Recently, with the proliferation of graph neural networks [24, 36, 7], message-passing HGE methods are brought forward and have achieved remarkable improvements on series of applications [38, 14, 25, 13, 4, 38]. These message-passing HGEs learn graph embedding by aggregating and transforming the embeddings of the original neighbors [46, 14, 17, 48, 11] or metapath-based neighbors [39, 45, 6].

Nevertheless, most of the existing HGE methods follow a unified framework [43] which learns embedding by minimizing the distance between the node embeddings of target node and its context nodes, preserving the heterogeneous semantics. However, the underlying rich correlations [49] between these rich semantics are seldom discussed and explored.

3 The Proposed Model

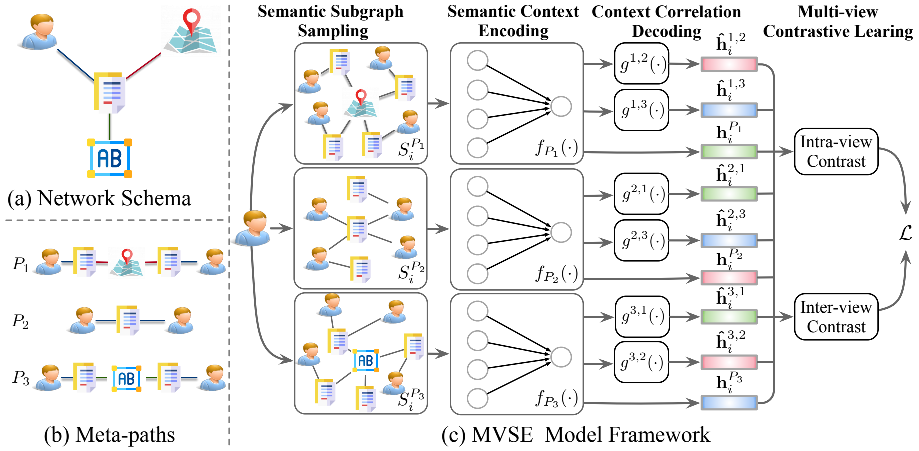
3.1 Model Framework

Consider a heterogeneous graph $G = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ composed of a node set $\mathcal{V}$, an edge set $\mathcal{E}$, and a feature matrix $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times d_F}$ (d_F: feature dimension) along with the node type mapping function $\phi : \mathcal{V} \to \mathcal{A}$, and the edge type mapping function $\psi : \mathcal{E} \to \mathcal{R}$, where $\mathcal{A}$ and $\mathcal{R}$ denotes the node and edge types, and $|\mathcal{A}| + |\mathcal{R}| > 2$. The task of heterogeneous graph embedding is to learn the representation of nodes $\mathbf{Z} \in \mathbb{R}^{|\mathcal{V}| \times d}$, where d is the dimension of representation.

The key idea of MVSE is to capture the rich heterogeneous semantics and their correlations by self-supervised contrastive learning. As shown in Figure 1 (c), given a node in heterogeneous graph, MVSE firstly samples several metapath-based semantic subgraphs and encodes them to its semantic space by semantic-specific encoders. Then, the semantic embeddings are further decoded to other semantic spaces to model the correlations between different semantics. Finally, the semantic embeddings and the decoded embeddings are optimized by intra-view, and inter-view contrastive learning losses.

3.2 Heterogeneous Context Encoding

A node in heterogeneous graph is associated with rich semantic information defined by meta-paths, providing different views of node property. Therefore, it is vital to encode the metapath-based neighbor information into node embeddings. Inspired by the recent advances of contrastive learning [8, 32], we propose heterogeneous subgraph instance discrimination as our self-supervised contrastive learning task. In this section, we elaborate how MVSE constructs multi-view heterogeneous subgraphs and encodes them as heterogeneous context embeddings.



(a) Network Schema

(b) Meta-paths

(c) MVSE Model Framework

Fig. 1. (a) The network schema of an example bibliography heterogeneous graph with four types of nodes: Author (A), Paper (P), Venue (V), and Term/keyword (T) along with three types of edges: an author writes a paper, a paper is published in a venue, and a paper contains a term. (b) The template meta-paths (views of semantics) APVPA, APA, APTPA. (c) The model framework of the proposed model MVSE.

Semantic Subgraph Sampling. Given a node v_i in heterogeneous graph G and a meta-path set $\mathcal{P}$, MVSE samples a subgraph instance set $\mathcal{S}_i^{\mathcal{P}} = \{S_i^P, P \in \mathcal{P}\}$ and further encodes them to semantic embeddings. In the homogeneous graph, an effective way of generating subgraph instances for contrastive learning is to apply RWR (random walk with restart), by iteratively generating subgraph structure via random walk with a restart probability γ [32]. Therefore, a straightforward idea to construct heterogeneous subgraph instances would be applying meta-path constrained RWR.

However, this straightforward extension can not well preserve the metapath-based context for heterogeneous graphs due to the intrinsic lack of high-order neighbors preservation of RWR. Specifically, each random walk trace is a Bernoulli trial with probability $(1-\gamma)^k$ sampling k-hop neighbors. Therefore, the number of time n_s that k-hop neighbors is sampled after n_{RW} number of restart time, is a binomial distribution:

$$P(n_s|k, n_{RW}) \sim B(n_{RW}, (1-\gamma)^k), \tag{1}$$

Hence, we can obtain that the expectation of number of times that k-hop neighbors are sampled a subgraph sampled by RWR :

$$E(n_s|k, n_{RW}) = n_{RW}(1-\gamma)^k, \tag{2}$$

which decreases exponentially when k increases, harming the high-order preservation. Specifically, with the recommended setting [32, 33], i.e. $\gamma = 0.8$, the probability of at least one 4-hop neighbor (which is the maximum depth of commonly used meta-paths e.g. APVPA, APTPA) is sampled in subgraphs within

20 trials is approximately 0.0315. In other words, the RWR sampled subgraph instances are composed of mostly low-order neighbors, harming the high-order semantics of meta-paths.

To address this issue, we instead propose to sample meta-path constrained subgraphs by a fixed-depth random walk subgraph sampling approach. Specifically, given a center node v_i and a meta-path P, MVSE samples a subgraph with the probability proportional to the edge weight of meta-path constrained relation for each walk. The walk stops when it reaches the maximum depth k_P. The overall subgraph S_i^P is constructed from all the nodes sampled in n_{RW} walks. Since the walks are of fixed length, it is guaranteed to preserve at least one k_P-hop neighbor in each semantic subgraph. Moreover, by adjusting the depth of subgraph via specifying k_P, users are able to control the receptive field of semantic relationships. To illustrate, the semantic subgraphs of meta-path APA with $k_{APA} = 4$ will preserve the "co-authors' co-author" semantic for an author.

Subgraph Context Encoding Here we introduce how to encode the semantic subgraphs to obtain a semantic embedding for each node. Specifically, given a node v_i and the sampled semantic subgraph set $\mathcal{S}_i^{\mathcal{P}} = \{S_i^P, P \in \mathcal{P}\}$, the task is to encode subgraphs into multi-view embeddings $\mathcal{H}_i = \{\mathbf{h}_i^P \in \mathbb{R}^{1\times d_s}, P \in \mathcal{P}\}$, where d_s stands for the hidden dimension of subgraph embeddings.

To fully capture the heterogeneity of different semantics [2, 28], we propose to use a semantic-specific encoder for each meta-path. Therefore, the semantic embedding of node v_i in the view of meta-path P denoted as $\mathbf{h}_i^P$ is obtained by:

$$\mathbf{h}_i^P = f_P(S_i^P), \tag{3}$$

where $f_P(\cdot)$ stands for the semantic-specific encoder for meta-path P. The choice of encoder can be any graph neural networks [24]. We adopt the Graph Isomorphism Network (GIN) [41] as the graph encoder. Hence, the semantic embedding is calculated by:

$$\begin{aligned}
\mathbf{h}_i^P &= \mathrm{CONCAT}\left(\mathrm{SUM}(\left\{\mathbf{h}_v^{P,(l)} \mid v \in S_i^P\right\}) \mid l = 0, 1, \ldots, L\right), \\
\mathbf{h}_v^{P,(l)} &= \mathrm{MLP}^{P,(l)}\left((1+\epsilon)\cdot \mathbf{h}_v^{P,(l-1)} + \sum_{u \in \mathcal{N}_i^P(v)} \mathbf{h}_u^{P,(l-1)}\right),
\end{aligned} \tag{4}$$

where $\mathrm{MLP}^{P,(l)}$ stands for the semantic-specific encoder for meta-path P at l-th layer, $\mathcal{N}_i^P(v)$ stands for the neighbors of node v in S_i^P, $\mathbf{h}_v^{P,(l)}$ is the l-th layer node representation of node v in semantic subgraph S_i^P, and the input is set as the node feature, i.e. $\mathbf{h}_v^{P,(0)} = \mathbf{x}_v$, ϵ is a fixed scalar.

3.3 Multi-view Contrastive Learning

At this point, we have obtained multi-view embeddings $\mathcal{H}_i$ of each node v_i. Here, we elaborate how to perform self-supervised contrastive learning on these embeddings to comprehensively learn the heterogeneous semantics and their correlations.

Preservation of Semantic Contexts We utilize MoCo [9] as the contrastive learning framework where a query node and a set of key nodes are contrasted in each epoch. MoCo maintains a dynamic dictionary of keys (nodes) and encodes the new keys on-the-fly by a momentum-updated encoder. In each epoch of MVSE, a query node is contrasted with K nodes, where K is the size of the dynamic dictionary. Here, as each node is encoded as multi-view embeddings $\mathcal{H}_i = \{\mathbf{h}_i^P \in \mathbb{R}^{1\times d_s}, P \in \mathcal{P}\}$, we perform multi-view contrastive learning on each view separately by the InfoNCE [30] loss, preserving the intra-semantic information:

$$\mathcal{L}_{intra} = \frac{1}{|\mathcal{P}|} \sum_{P\in\mathcal{P}} -\log \frac{\exp\left(\mathbf{h}_q^P \cdot \mathbf{h}_{k+}^P/\tau\right)}{\sum_{j=0}^{K} \exp\left(\mathbf{h}_q^P \cdot \mathbf{h}_{k_j}^P/\tau\right)}, \tag{5}$$

where $\mathbf{h}_q^P$ is the query node's semantic embedding of metapath P calculated by Equation 3, $\mathbf{h}_k^P$ stands for the key node's semantic embedding encoded by momentum encoders [9], $k+$ stands for the positive key in the dictionary, τ is the temperature hyper-parameter. Thus, by minimizing $\mathcal{L}_{intra}$, MVSE is able to distinguish subgraph instances of different nodes using each meta-path in $\mathcal{P}$.

Preservation of Semantic Correlations As discussed in the Introduction, most existing HGE methods focus on discriminating the heterogeneous context instances, e.g. whether two authors have a co-author relationship, preserving the intra-context relationships. However, few works have explored the complex interactions (inter-context) [49] between these contexts, e.g. whether two authors with co-venue relationships have co-author relationships.

In light of this, we propose to explicitly capture these correlations by inter-view contrastive learning. Specifically, for each semantic embedding $\mathbf{h}_i^P$ of node v_i of meta-path P, we model the correlations between semantics by decoding them to other semantic embeddings:

$$\hat{\mathbf{h}}_i^{s,t} = g^{s,t}(\mathbf{h}_i^{P_s}) \tag{6}$$

where $g^{s,t}(\cdot)$ stands for the decoder that decodes the semantic embedding from source view P_s to target view P_t. $\hat{\mathbf{h}}_i^{s,t}$ stands for the semantic embedding of target view P_s decoded from source view P_t. In this way, the correlation between source view and target view is preserved. For example, if we set source view as APVPA and target view as APA, the decoder attempts to predict the co-author relationships using the co-venue relationships, modeling the interactions between these two semantics. Hence, the complex correlations between semantics can be well preserved by the inter-view contrastive loss defined as follows:

$$\mathcal{L}_{inter} = \frac{1}{|\mathcal{P}| * (|\mathcal{P}|-1)} \sum_{P_s,P_t\in\mathcal{P},s\neq t} -\log \frac{\exp\left(\hat{\mathbf{h}}_i^{s,t} \cdot \mathbf{h}_{k+}^{P_t}/\tau\right)}{\sum_{j=0}^{K} \exp\left(\hat{\mathbf{h}}_i^{s,t} \cdot \mathbf{h}_{k_j}^{P_t}/\tau\right)}, \tag{7}$$

Finally, MVSE optimizes the overall loss $\mathcal{L}$ to comprehensively learn representations considering both the intra-view and inter-view semantics:

$$\mathcal{L} = \alpha \mathcal{L}_{intra} + (1 - \alpha)\mathcal{L}_{inter} \tag{8}$$

where α is the hyper-parameter for balancing different loss functions.

4 Experiment

To demonstrate the effectiveness of our proposed model, we conduct comprehensive experiments on three public benchmark heterogeneous graph datasets. We firstly evaluate our model on two downstream tasks (node classification and link prediction). Then, we perform ablation study to further demonstrate the effectiveness of the designs in MVSE. Visualization experiments are also conducted to show the effectiveness of our model intuitively.

4.1 Experimental Setup

Datasets. We employ the following real-world heterogeneous graph datasets to evaluate our proposed model.
DBLP [28]: We extract a subset of DBLP which includes 4,057 authors (A), 20 conferences (C), 14,328 papers (P) and four types of edges (AP, PA, CP, and PC). The target nodes are authors and they are divided into four areas: database, data mining, machine learning, and information retrieval. The node features are the terms related to authors, conferences and papers respectively.
ACM [45]: We extract papers published in KDD, SIGMOD, SIGCOMM, MobiCOMM, and VLDB and construct a heterogeneous graph which includes 5,912 authors (A), 3,025 papers (P), 57 conference subjects (S) and four types of edges (AP, PA, SP, and PS). The target nodes are papers and they are divided into three classes according to their conferences: database, data mining, and wireless communication. The node features are the terms related to authors, papers and subjects respectively.
IMDB [39]: We extract a subset of IMDB which includes 4,461 movies (M), 2,270 actors (A), 5,841 directors (D), and four types of edges (AM, MA, DM, and MD). The target nodes are movies labeled by genre (action, comedy, and drama). The movie features are bag-of-words representation of plot keywords.
Baselines. To comprehensively evaluate our model, we compare MVSE with ten graph embedding methods. Based on their working mechanisms, these baselines can be divided into three categories: The unsupervised representation learning methods, i.e.DeepWalk [31], MP2Vec [3], DGI [37], and HeGAN [12]; the semi-supervised representation learning methods, i.e. GCN [24], GIN [41], HAN [39], and GTN [45], and the self-supervised learning methods, i.e. GCC [32] and GPT-GNN [16]. For unsupervised baselines, the embeddings are learned without label supervision and then fed into a logistic classifier to perform the downstream tasks. The semi-supervised methods are optimized through an end-to-end supervised manner, e.g. cross entropy loss in node classification tasks. The self-supervised methods are firstly pre-trained to fully encode the unlabeled information and then fine-tuned by labeled information via cross entropy loss in node

classification.

Implementation Details. Here, we briefly introduce the experimental settings. For MVSE, in each epoch, we construct semantic subgraphs by performing 3 times of the meta-path constrained fixed-depth random walk (in Section 3.2) with the maximum depth set as twice the depth of meta-path, i.e. $n_{RW} = 3, k_P = 2|P|$, where $|P|$ stands for the depth of meta-path P. The decoders for modeling the semantic correlations are 2-layer MLPs. We use Adam [21] optimizer with learning rate set as 0.005. The semantic embedding dimension d_s is set as 64, therefore the dimension d of final node embedding $\mathbf{Z}$ is $64|\mathcal{P}|$. For MoCo-related settings, the dynamic dictionary size K is set as 4096 with $\tau = 0.07$. For all GNN related models, we use 2-layer GCNs [24] with weight decay set as 1e-5. The code and data to reproduce our results is publicly available at Github[3].

4.2 Node Classification

As a common graph application, node classification is widely used to evaluate the performance of the graph embedding algorithms. Given a graph with some labeled nodes, the task of node classification is to predict the labels of unlabeled nodes. Here, we evaluate the performance of node classification on the three datasets mentioned above in this section. For each dataset, the percentage of training labeled nodes are set as 1%, 3%, and 5%, and the rest of the labeled nodes are used as test nodes. We adopt Macro-F1 and Micro-F1 as metrics and report the node classification performance on the test set. The results (in percentage) of the three datasets are shown in Table 1, Table 2 and Table 3, respectively, from which we have the following observations: (1) By comprehensively preserving the rich semantics and their correlations inside heterogeneous graphs, our proposed MVSE outperforms other baselines, demonstrating the effectiveness of our proposed model. (2) Most self-supervised learning models (MVSE and GPT-GNN) generally achieve better performance than other baselines, since the pre-training of self-supervised tasks extract robust embedding with rich semantics and structural information and provide a better initialization for the fine-tuning process. The performance improvement is more significant when the ratio of labeled information is low. (3) Since the node features are of vital importance in node classification tasks, GNN-based models generally outperform the random walk-based models due to their ability to utilize node features.

4.3 Link Prediction

The objective of link prediction is to predict unobserved edges using the observed graph. To evaluate the effectiveness of semantic preservation, we use metapath-based link prediction task [19] on three datasets and evaluate the metapath-based link prediction performance on 2-hop symmetric meta-paths. Specifically, in each task, the meta-path instances are firstly randomly splitted as training and test

[3] https://github.com/Andy-Border/MVSE

Models	DBLP (1%)		DBLP (3%)		DBLP (5%)	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
GCN [24]	72.63	70.86	79.36	78.22	82.08	81.17
GIN [41]	71.22	67.26	88.23	88.28	88.31	88.32
HAN [39]	85.65	85.24	89.42	88.83	89.72	89.36
GTN [45]	85.99	85.45	88.66	88.13	89.73	89.37
DeepWalk [31]	86.58	87.31	87.48	87.95	87.72	88.34
DGI [37]	87.73	86.44	90.01	89.33	91.22	89.57
MP2Vec [3]	86.33	85.87	88.16	87.82	88.91	88.63
HeGAN [12]	79.12	77.73	81.66	80.25	83.78	82.44
GPT_GNN [16]	86.61	86.33	90.62	89.26	90.91	89.43
GCC [32]	78.92	77.94	81.78	81.11	82.67	81.89
MVSE	**90.46**	**89.27**	**91.57**	**90.97**	**91.96**	**89.83**

Table 1. Performance of node classification experiment on DBLP dataset in percentage (Micro-F1 and Macro-F1), MVSE outperforms the baselines in all the settings.

Models	ACM (1%)		ACM (3%)		ACM (5%)	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
GCN [24]	85.78	85.87	88.97	89.01	89.55	89.62
GIN [41]	78.40	77.99	84.62	84.69	87.09	87.17
HAN [39]	85.77	85.92	87.41	87.62	88.37	88.58
GTN [45]	80.08	79.54	84.56	84.16	88.71	88.24
DeepWalk [31]	79.02	79.28	80.75	81.03	80.15	80.57
DGI [37]	84.99	85.21	88.93	89.06	89.36	89.50
MP2Vec [3]	80.74	80.36	82.42	81.87	82.63	82.12
HeGAN [12]	78.23	78.67	80.84	81.35	81.95	82.52
GPT_GNN [16]	84.62	84.86	88.90	89.22	89.27	89.54
GCC [32]	80.14	78.84	83.91	82.35	84.72	83.17
MVSE	**86.14**	**86.17**	**89.43**	**89.44**	**89.74**	**89.64**

Table 2. Performance of node classification experiment on ACM dataset in percentage (Micro-F1 and Macro-F1), MVSE outperforms the baselines in all the settings.

set with 1:1 ratio. Then, the self-supervised/unsupervised models are applied to learn the node representations. Finally, the embeddings are fed to logistic regression classifiers and predict whether the test edges exist by training edges. We use F1 and AUC-ROC as our evaluation metrics, the results in percentage are shown in Table 4, from which we have the following observations: (1) MVSE consistently outperforms other baselines on all the metapath-based link prediction tasks. The reason is that MVSE is able to capture the correlations between meta-paths, thus alleviates the impact of intrinsic sparseness in graphs [49], e.g. APA link prediction can be enhanced by APCPA relationship, and further improve the link prediction performance. (2) Models that consider heterogeneity show better performance than their counterparts since they are able to extract the rich semantic contexts from the different meta-paths.

Models	IMDB (1%)		IMDB (3%)		IMDB (5%)	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
GCN [24]	51.71	42.76	55.06	45.88	58.05	50.69
GIN [41]	48.87	43.24	53.70	48.38	58.03	53.42
HAN [39]	50.76	43.47	52.87	48.46	56.43	51.25
GTN [45]	51.66	45.87	57.83	49.31	59.49	53.58
DeepWalk [31]	53.92	49.34	54.44	49.85	54.48	49.88
DGI [37]	53.02	44.61	56.62	48.29	57.93	50.15
MP2Vec [3]	54.50	**49.82**	55.10	50.34	56.97	52.79
HeGAN [12]	47.70	41.47	49.98	44.29	51.04	46.46
GPT_GNN [16]	55.17	48.30	58.78	52.69	61.24	56.74
GCC [32]	52.33	47.29	53.68	48.82	53.85	49.08
MVSE	**55.61**	44.25	**60.15**	**53.95**	**63.32**	**58.42**

Table 3. Performance of node classification experiment on IMDB dataset in percentage (Micro-F1 and Macro-F1), MVSE outperforms the baselines in most of the settings.

Models	DBLP (APA)		ACM (PAP)		ACM (PSP)		IMDB (MAM)		IMDB (MDM)	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC
DeepWalk [31]	79.08	77.72	72.99	72.73	78.65	69.85	72.62	75.37	59.16	60.37
DGI [37]	80.59	81.28	73.31	72.41	84.87	74.4	70.13	69.54	61.37	59.24
MP2Vec [3]	81.64	80.66	75.64	74.82	82.88	75.16	82.55	81.06	64.39	63.78
HeGAN [12]	80.75	80.26	78.67	78.51	81.25	71.02	80.27	80.11	67.75	68.13
GPT-GNN [16]	86.84	86.02	80.94	80.55	86.31	78.25	91.88	91.31	68.54	68.12
GCC [32]	79.15	78.63	73.62	72.98	74.22	68.17	81.63	82.37	64.71	63.85
MVSE	**88.09**	**87.92**	**81.18**	**80.73**	**87.72**	**79.22**	**98.25**	**98.23**	**69.51**	**69.54**

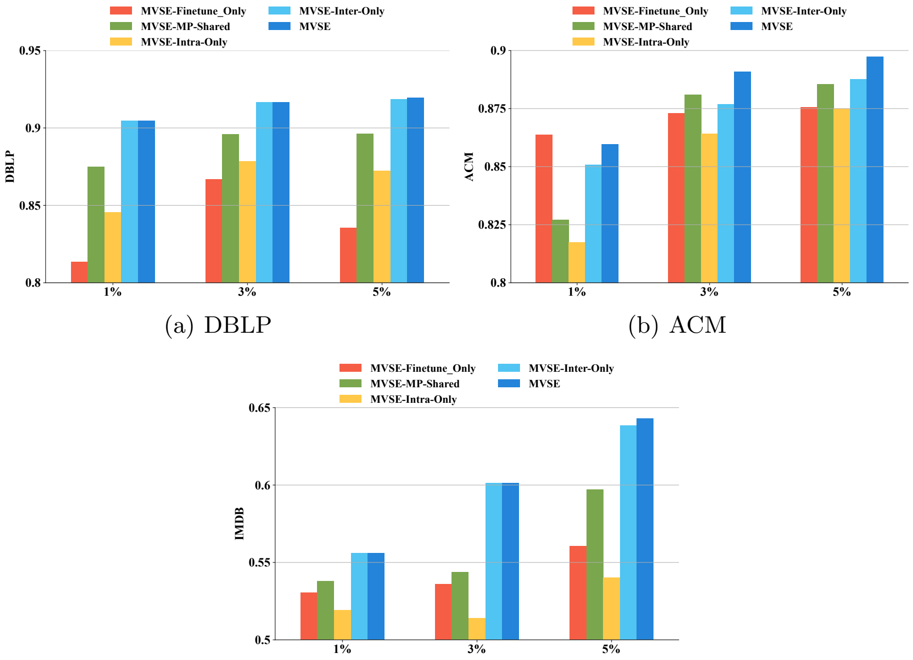
Table 4. Performance of link prediction experiment on different datasets and meta-paths in percentage (Micro-F1 and ROC-AUC), MVSE outperforms the baselines on all the datasets and meta-paths.

4.4 Ablation Study

In order to verify the effectiveness of the delicate designs in MVSE, we design five variants of MVSE and compare their node classification performance against MVSE on three datasets. The results in terms of Micro-F1 are shown in Figure 2 (a), Figure 2 (b) and Figure 2 (c), respectively.

Effectiveness of Heterogeneous Context Encoders. As discussed in Section 3.2, to capture the intrinsic heterogeneity [2, 28] in different metapath-based semantics, MVSE use semantic-specific encoders to embed the contexts of different meta-paths. To verify the effectiveness of semantic specific encoders, we propose a variant of MVSE which uses metapath-shared encoders, namely MVSE-MP-Shared. The results in Figure 2 show MVSE outperforms the variant on the three datasets since MVSE-MP-Shared ignores the heterogeneity of semantics by modeling them using an unified (homogeneous) model. This phenomenon further demonstrates the importance of considering heterogeneity in heterogeneous graph contrastive learning.

Effectiveness of Multi-View Contrastive Learning. As discussed in Section



(a) DBLP

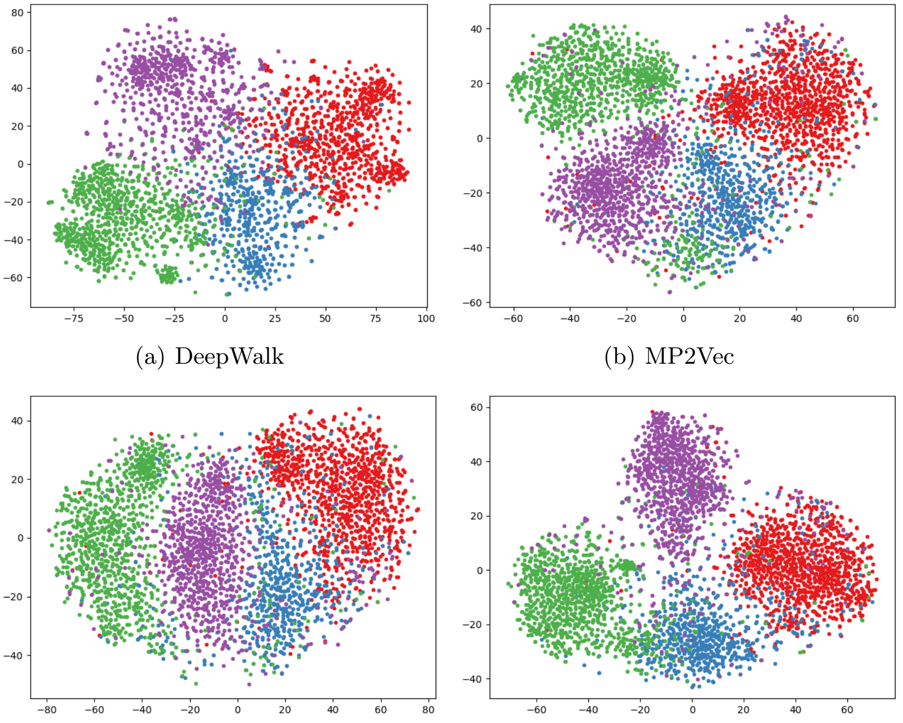
(b) ACM

(c) IMDB

Fig. 2. Performance of MVSE variants on different datasets (Micro-F1), MVSE outperforms the variants over all the datasets and settings. MVSE-Inter-Only has better performance than other variants, which demonstrates the importance of preserving semantic correlations.

3.3, MVSE comprehensively learns the heterogeneous semantics and their correlations by intra-view and inter-view contrastive learning tasks. To investigate the effects of these contrastive learning tasks, we propose two variants of MVSE which only consider intra-view (MVSE-Intra-Only) and inter-view (MVSE-Inter-Only) respectively and evaluate their node classification performance. From the results shown in Figure 2, we can find that MVSE beats all variants on every task, which indicates the effectiveness of performing multi-view contrastive learning by optimizing both intra- and inter-semantic SSL tasks. Besides, the phenomenon that MVSE-Inter-Only outperforms the other two variants further demonstrates the importance of preserving semantic correlations.

Effectiveness of Unlabeled Data Utilization. To investigate the ability of utilizing unlabeled information, we propose MVSE-Finetune-Only which skips the pre-training process and trains the model from scratch. As shown in Figure 2, MVSE consistently outperforms this variant in all tasks since MVSE-Finetune-Only cannot fully utilize the unlabeled information by optimizing objective that considers labeled information only [20]. The self-supervised pre-training strategy provides a better start point than the random initialization and further improves the classification performance. In addition, the performance improvement of MVSE over MVSE-Finetune-Only is generally more significant when the percentage of labeled nodes is low, demonstrating the superiority of SSL in tasks with little label supervision.



(a) DeepWalk

(b) MP2Vec

(c) GCC

(d) MVSE

Fig. 3. Node embedding visualization of different methods on DBLP dataset. Each point indicates one author and its color indicates the research area. MVSE has least overlapping area and largest cluster-wise distance.

4.5 Visualization

To examine the graph representation intuitively, we visualize embeddings of author nodes in DBLP using the t-SNE [29] algorithm. Here, we choose DeepWalk, MP2Vec, and GCC as the representatives of homogeneous embedding, heterogeneous embedding, and self-supervised based embedding methods, respectively. The visualization results are shown in Figure 3, from which we can find that although all of the baselines can roughly embed the authors with same research fields into same clusters, the heterogeneous models generate more distinct boundaries and less overlapping area between clusters. What's more, among all of the unsupervised graph learning algorithms, MVSE generates embeddings with the largest cluster-wise distance, indicating better embeddings are learned.

5 Conclusion

In this paper, we study self-supervised learning on heterogeneous graphs and propose a novel model named MVSE. MVSE samples and encodes semantic subgraphs of different views defined by meta-paths and captures the intra- and inter-view semantic information comprehensively by contrastive self-supervised learning. Our extensive experiments demonstrate the effectiveness of our proposed model and the necessity of preserving cross-view interactions for learning heterogeneous graph embeddings.

Acknowledgements

This work is partially supported by the NSF under grants IIS-2107172, IIS-2027127, IIS-2040144, CNS-2034470, IIS-1951504, CNS-1940859, CNS-1814825, OAC-1940855, and the NIJ 2018-75-CX-0032.

References

1. Bordes, A., Usunier, N., Garc´ıa-Dur´an, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS. pp. 2787–2795 (2013)
2. Chen, H., Yin, H., Wang, W., Wang, H., Nguyen, Q.V.H., Li, X.: PME: projected metric embedding on heterogeneous networks for link prediction. In: KDD. pp. 1177–1186 (2018)
3. Dong, Y., Chawla, N.V., Swami, A.: metapath2vec: Scalable representation learning for heterogeneous networks. In: KDD. pp. 135–144 (2017)
4. Fan, S., Zhu, J., Han, X., Shi, C., Hu, L., Ma, B., Li, Y.: Metapath-guided heterogeneous graph neural network for intent recommendation. In: KDD. pp. 2478–2486 (2019)
5. Fu, T.y., Lee, W.C., Lei, Z.: Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In: CIKM. pp. 1797–1806 (2017)
6. Fu, X., Zhang, J., Meng, Z., King, I.: MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding. In: WWW. pp. 2331–2341 (2020)
7. Hamilton, W.L., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: NIPS. pp. 1024–1034 (2017)
8. Hassani, K., Ahmadi, A.H.K.: Contrastive multi-view representation learning on graphs. In: ICML. pp. 4116–4126 (2020)
9. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.B.: Momentum contrast for unsupervised visual representation learning. In: CVPR. pp. 9726–9735 (2020)
10. He, Y., Song, Y., Li, J., Ji, C., Peng, J., Peng, H.: Hetespaceywalk: A heterogeneous spacey random walk for heterogeneous information network embedding. In: CIKM. pp. 639–648 (2019)

11. Hong, H., Guo, H., Lin, Y., Yang, X., Li, Z., Ye, J.: An attention-based graph neural network for heterogeneous structural learning. In: AAAI. pp. 4132–4139 (2020)
12. Hu, B., Fang, Y., Shi, C.: Adversarial learning on heterogeneous information networks. In: KDD. pp. 120–129 (2019)
13. Hu, B., Zhang, Z., Shi, C., Zhou, J., Li, X., Qi, Y.: Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism. In: AAAI. pp. 946–953 (2019)
14. Hu, L., Yang, T., Shi, C., Ji, H., Li, X.: Heterogeneous graph attention networks for semi-supervised short text classification. In: EMNLP-IJCNLP. pp. 4820–4829 (2019)
15. Hu, W., Liu, B., Gomes, J., Zitnik, M., Liang, P., Pande, V.S., Leskovec, J.: Strategies for pre-training graph neural networks. In: ICLR (2020)
16. Hu, Z., Dong, Y., Wang, K., Chang, K.W., Sun, Y.: Gpt-gnn: Generative pre-training of graph neural networks. In: KDD. pp. 1857–1867 (2020)
17. Hu, Z., Dong, Y., Wang, K., Sun, Y.: Heterogeneous graph transformer. In: WWW.pp. 2704–2710 (2020)
18. Hussein, R., Yang, D., Cudr´e-Mauroux, P.: Are meta-paths necessary?: Revisiting heterogeneous graph embeddings. In: CIKM. pp. 437–446 (2018)
19. Hwang, D., Park, J., Kwon, S., Kim, K.M., Ha, J.W., Kim, H.J.: Self-supervised auxiliary learning with meta-paths for heterogeneous graphs. arXiv preprint arXiv:2007.08294 (2020)
20. Jin, W., Derr, T., Liu, H., Wang, Y., Wang, S., Liu, Z., Tang, J.: Self-supervised learning on graphs: Deep insights and new direction. arXiv preprint arXiv:2006.10141 (2020)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
22. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: ICLR (2014)
23. Kipf, T.N., Welling, M.: Variational graph auto-encoders. arXiv preprint arXiv:1611.07308 (2016)
24. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR (2017)
25. Li, A., Qin, Z., Liu, R., Yang, Y., Li, D.: Spam review detection with graph convolutional networks. In: CIKM. pp. 2703–2711 (2019)
26. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: AAAI. pp. 2181–2187 (2015)
27. Liu, X., Zhang, F., Hou, Z., Wang, Z., Mian, L., Zhang, J., Tang, J.: Self-supervised learning: Generative or contrastive. arXiv preprint arXiv:2006.08218 (2020)
28. Lu, Y., Shi, C., Hu, L., Liu, Z.: Relation structure-aware heterogeneous information network embedding. In: AAAI. pp. 4456–4463 (2019)
29. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. JMLR 9(11), 2579–2605 (2008)
30. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
31. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: KDD. pp. 701–710 (2014)
32. Qiu, J., Chen, Q., Dong, Y., Zhang, J., Yang, H., Ding, M., Wang, K., Tang, J.: GCC: graph contrastive coding for graph neural network pre-training. In: KDD. pp. 1150–1160 (2020)

33. Qiu, J., Tang, J., Ma, H., Dong, Y., Wang, K., Tang, J.: DeepInf: social influence prediction with deep learning. In: KDD. pp. 2110–2119 (2018)
34. Sun, Y., Han, J., Yan, X., Yu, P.S., Wu, T.: Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. VLDB 4(11), 992–1003 (2011)
35. Trouillon, T., Welbl, J., Riedel, S., Gaussier, ´E., Bouchard, G.: Complex embeddings for simple link prediction. In: ICML. pp. 2071–2080 (2016)
36. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Li'o, P., Bengio, Y.: Graph attention networks. In: ICLR (2018)
37. Velickovic, P., Fedus, W., Hamilton, W.L., Li'o, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. In: ICLR (2019)
38. Wang, X., Bo, D., Shi, C., Fan, S., Ye, Y., Yu, P.S.: A survey on heterogeneous graph embedding: Methods, techniques, applications and sources. arXiv preprint arXiv:2011.14867 (2020)
39. Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S.: Heterogeneous graph attention network. In: WWW. pp. 2022–2032 (2019)
40. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: AAAI. pp. 1112–1119 (2014)
41. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: ICLR (2019)
42. Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: ICLR (2015)
43. Yang, C., Xiao, Y., Zhang, Y., Sun, Y., Han, J.: Heterogeneous network representation learning: A unified framework with survey and benchmark. TKDE (2020)
44. You, J., Ying, R., Ren, X., Hamilton, W.L., Leskovec, J.: Graphrnn: Generating realistic graphs with deep auto-regressive models. In: ICML. pp. 5694–5703 (2018)
45. Yun, S., Jeong, M., Kim, R., Kang, J., Kim, H.J.: Graph transformer networks. In: NIPS. pp. 11960–11970 (2019)
46. Zhang, C., Song, D., Huang, C., Swami, A., Chawla, N.V.: Heterogeneous graph neural network. In: KDD. pp. 793–803 (2019)
47. Zhang, C., Swami, A., Chawla, N.V.: SHNE: representation learning for semanticassociated heterogeneous networks. In: WSDM. pp. 690–698 (2019)
48. Zhao, J., Wang, X., Shi, C., Liu, Z., Ye, Y.: Network schema preserving heterogeneous information network embedding. In: IJCAI. pp. 1366–1372 (2020)
49. Zhao, K., Bai, T., Wu, B., Wang, B., Zhang, Y., Yang, Y., Nie, J.: Deep adversarial completion for sparse heterogeneous information network embedding. In: WWW. pp. 508–518 (2020)