Layer Information Similarity Concerned Network Embedding

Ruili Lu, Pengfei Jiao, Yinghui Wang, Huaming Wu, and Xue Chen

1Tianjin International Engineering Institute, Tianjin University, Tianjin 300072, China
2Law School of Tianjin University, Tianjin 300072, China
3College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
4Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

Correspondence should be addressed to Xue Chen; xuechen@tju.edu.cn

Received 10 June 2021; Accepted 17 August 2021; Published 26 August 2021

Copyright © 2021 Ruili Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Great achievements have been made in network embedding based on single-layer networks. However, there are a variety of scenarios and systems that can be presented as multiplex networks, which can reveal more interesting patterns hidden in the data compared to single-layer networks. In the field of network embedding, in order to project the multiplex network into the latent space, it is necessary to consider richer structural information among network layers. However, current methods for multiplex network embedding mostly focus on the similarity of nodes in each layer of the network, while ignoring the similarity between different layers. In this paper, for multiplex network embedding, we propose a Layer Information Similarity Concerned Network Embedding (LISCNE) model considering the similarities between layers. Firstly, we introduce the common vector for each node shared by all layers and layer vectors for each layer where common vectors obtain the overall structure of the multiplex network and layer vectors learn semantics for each layer. We get the node embeddings in each layer by concatenating the common vectors and layer vectors with the consideration that the node embedding is related not only to the surrounding neighbors but also to the overall semantics. Furthermore, we define an index to formalize the similarity between different layers and the cross-network association. Constrained by layer similarity, the layer vectors with greater similarity are closer to each other and the aligned node embedding in these layers is also closer. To evaluate our proposed model, we conduct node classification and link prediction tasks to verify the effectiveness of our model, and the results show that LISCNE can achieve better or comparable performance compared to existing baseline methods.

1. Introduction

In the past few decades, network embedding has obtained remarkable achievements. The basic idea is converting a node into a low-dimensional space in which the network structure and properties can be preserved effectively. In the early period, traditional models such as MDS [1], Isomap [2], LLE [3], and LE [4] are mainly based on dimensionality reduction technologies. These models are not suitable for large networks due to their computational complexity. As Word2Vec [5] plays a vital role in the field of natural language processing, random walk-based methods that regard nodes in the network as words are proposed, such as DeepWalk [6] and Node2Vec [7]. In recent years, with the continuous development of deep learning, SDNE [8], DNGR [9], and GCNs [10] have developed neural networks into network embedding models.

The methods mentioned above are all designed for single-layer networks. Figure 1(a) shows an example of a single-layer network, through which we can see that there is only one relation in the network. However, there are still many complex scenarios in the real world that cannot be described by single-layer networks. For example, the same set of individuals in social networks may participate in Twitter, Facebook, or Weibo for different purposes. Interactions in different social networks can be represented by a single-layer network. Each layer of the network has a specific relationship and specific semantics. However, these single-layer networks do not operate in isolation and there are always connections between them. Instead, these complex
Scenes can be represented as a multiplex network, which is also a multilayer network in which layers share the same set of nodes. Each layer in a multiplex network represents a particular relationship of nodes, and the structure of each layer is typically associated. Figure 1(b) illustrates an example of an undirected multiplex network, and it has a unique structure in different layers while there also exist correlations between layers. Unlike the single-layer network, there are three relationships among a set of nodes, each of which describes a unique interaction in the given network structure. Multiplex relationships cannot be captured using single-layer methods. Therefore, it is necessary to conduct in-depth research on multiplex network embedding.

Compared with the single-layer network, one of the challenges for multiplex network embedding is how to aggregate the diverse types of structure in different layer networks without destroying their unique properties. To solve this problem, MCGE [11], MANE [12], and MVNE [13] use the tensor factorization to concurrently capture the main local structure and correlations between different layers. MNE [14] and MGCN [15] define one common vector shared by all layers to capture the shared information in all layer networks and low-dimensional node vectors in each layer to capture the unique properties. In addition to introducing the common vector, CrossMNA [16] also introduces a layer vector to extract the semantic meaning. One2Multi [17] uses one encoder to encode the most informative network from which we can extract the shared information and multiple decoders to reconstruct all layers learning the specific structure in each layer. DMNE [18] and MrMine [19] take advantage of the links between subgraphs or communities to learn the cross-network relationships.

While each layer in the multiplex network is constructed from different semantics and makes the structure of each layer different, the varying relatedness between different semantics leads to diverse structural similarities between different layers. For example, we can observe from Figure 1(b) that layer 2 and layer 3 have more of the same edges between nodes compared to layer 1, that is, the structure of layer 2 and layers 3 is more similar than that of layer 1. Also, the similarity of any two layers is always different, which leads to the divergence in different layers of network analysis. It has been proved that considering inter-layer similarity can significantly improve the performance of link prediction [20] and community detection [21] in multiplex networks. Hence, it is an essential feature that should not be ignored in multiplex network embedding. However, the existing methods can obtain embedded representations of a multiplex network, and most of them fail to consider the similarities between different layers which is an important characteristic in the multiplex network.

To incorporate layer similarities when learning node vectors or layer vectors, we propose a novel model, Layer Information Similarity Concerned Network Embedding (LISCNE), and our model takes advantage of the common and local features in multiplex networks and exploits layer similarity at the same time.

Specifically, we firstly obtain node embeddings by concatenating common vector for each node shared by all layers and layer vector for each layer. Common vectors capture characteristics shared by cross layer by merging all the networks into a new single-layer network and training the common vector for each node in the new network. In addition, layer vectors learn the overall semantics for each layer. Then, to model the layer similarity, we define an index to formalize the similarity between different layers. With the constraint of layer similarities, we force the vectors with greater similarity to be closer.
The major contributions are summarized as follows:

(i) After investigating the existing multiplex network embedding methods, we find that the methods consider the node connectivity among layers but ignore the inter-layer similarities.

(ii) We propose a novel Layer Information Similarity Concerned Network Embedding (LISCNE) model, which effectively exploits the overall and local structure in multiplex networks and combines the concept of layer vectors with layer similarity at the same time.

(iii) We conduct experiments to evaluate the proposed method using several real-world datasets on link prediction and node classification tasks. Compared with existing benchmark methods, LISCNE can achieve better or comparable performance.

2. Related Work

In this section, we review related work from two main aspects, namely, single-layer network embedding and multiplex network embedding.

2.1. Single-Layer Network Embedding. By assuming that the more similar the structure of nodes is, the closer their representation vectors are, the network embedding can learn latent low-dimensional representations for the nodes or links in a network. Earlier studies [2, 3, 22–24] were mainly based on matrix factorization. Isomap [2] obtained the shortest path $d_{ij}$ between node $i$ and node $j$ by constructing a neighborhood graph with connectivity algorithms and then obtained the vector presentation by minimizing the function of $(d_{ij} - ||u_i - u_j||)^2$. GraRep [24] defined a node transition probability and preserved k-order proximity. Inspired by Word2Vec [5], new types of methods [6, 7, 25, 26] using skip-gram model [27] have gradually emerged. The goal of the skip-gram model is to maximize the co-occurrence probability based on the context in a sentence:

$$\max \prod_{v_i \in V} \prod_{v_j \in \text{context}(v_i)} \Pr(\phi(v_i)|\phi(v_j)).$$  \hspace{1cm} (1)

DeepWalk [6] regarded each vertex in the network as a word. It applied the Depth-First Sampling (DFS) strategy to obtain walk sequences when conducting random walks and performed the skip-gram algorithm for training the sequences. Node2Vec [7] employed a biased random walk strategy when getting the walk sequence. It defined two parameters $p$ and $q$ to adjust between BFS and DFS during random walks. Topo2Vec [26] used a greedy goal-based searching strategy to generate the node context and obtain the local and global topologically proximal nodes in a network. While these random walk-based methods cannot model the nonlinear structural information, some methods based on deep neural networks [8–10, 28–30] have been proposed. Both SDNE [8] and DNGR [9] used deep autoencoders, where SDNE used the encoder to preserve the first- and second-order proximity of nodes, while DNGR captured higher-order proximity by using PPMI matrix which is indirectly transformed by the probabilistic co-occurrence matrix created by random surffing. GCNs [10] iteratively aggregated previous node embeddings and their neighbor embeddings to learn the new node embeddings. VGEA [28] was an inference model parameterized by a two-layer GCN. Pedronete and Latecki [31] proposed rank-based self-training to improve the accuracy of GCNs on semisupervised classification tasks. Recently, some novel algorithms [32–34] in the field of Contrastive Self-Supervised Learning have yielded good results. The core is to measure the similarities of sample pairs in a representation space, and the similarity between positive samples is much greater than the negative samples. These models are performed on the single-layer network. More discussion and methods for network embedding can be found in [35–38].

2.2. Multiplex Network Embedding. To better represent the multiplex networks used to describe the real-world data, there also exist various works for multiplex network embedding.

MCGE [11] applied tensor factorization and defined a multiview kernel tensor to obtain common latent factors that capture the global structure information. Random walks have been applied in network embedding [14, 19, 39–42]. MNE [14] learned two vectors for a node at the same time, i.e., a common vector sharing by all layers and a lower-dimensional vector for an individual layer. Then, it introduced a transformation matrix to align these two vectors. PMNE’s [39] network aggregation and result aggregation are essentially single-layer approaches. Considering the interactions between layers, the co-analysis method can traverse between layers with a probability $r$ when taking a random walk. GATNE [43] proposed a unified framework to address the problem of embedding learning for attributed multiplex heterogeneous networks, and GATNE-T was a generalization of MNE [14] when training edge embeddings directly. MrMine [19] simultaneously learned the multinetwork representation at three resolutions of network, subgraph, and nodes, and it further constructed cross-resolution including network-subgraph, subgraph-node, node-node context. HMNE [44] defined a heuristic 3D interactive walk and sampled sequences of node cross layers. It preserved cross-layer neighborhood of nodes and learned information of multitype relations into a unified embedding space.

MVE [45] learned the robust representation by promoting the collaboration of different layers and different weights which were assigned to layers during voting. CrossMNA [16] defined a network vector extracting the semantic meaning of the network and an inter-vector reflecting the common features of the anchor nodes in different networks. Then, these two vectors were added to form an intra-vector, which preserved the specific structural feature for a node in its selected network. MGCN [46] extended GCN to multiplex networks, which defined a general vector and dimension-specific vector to capture the common and individual layer information. TCMGC [47] developed a multilayer GCN to...
capture the structure and multiview information. DMNE [18] used an encoder for all individual networks and regularized the cross-network embeddings through two types of loss functions to penalize the embedding inconsistency. DMGI [48] was an unsupervised model based on DGI [49]. In an individual layer, it performed the DGI algorithm to get the relation-type specific embedding and then took advantage of the multiplicity of the network by introducing consensus regularization and multiheaded attention mechanisms. MEGAN [50] was a multiplex GAN that designed a multilayer generator to model multilayer connectivity to generate fake samples and a node pair discriminator to enforce the generator to more accurately fit the distribution of multilayer network connectivity. One2Multi [17] used the network with the most information as the input of encoder to learn the shared information of all the networks and then used a multidecoder to reconstruct the multiplex network from the shared information.

All the single-layer models mentioned above are effective for single network embedding; however, they do not consider the correlation in the multiplex network. In addition, the GCN-based multiplex network embedding models only consider the local information in the network, while other models ignore the similarities between layers. Our model combines the similarities between the layers and can simultaneously capture the local and global information in the network and the multiplex relationships between layers.

3. Notations and Problem Formulation

We begin with a formal definition of multiplex network, followed by the problem formulation. For the sake of clarity, the main notations are summarized in Table 1.

**Definition 1** (multiplex network). A multiplex network consists of a set of $N$ nodes $V = \{v_1, v_2, \ldots, v_N\}$ and $L$ layer. All layers share the same set of nodes $V$ and the nodes form diverse structures in each layer. The structure layer $l$ can be represented as $\mathbf{e}_l$. We denote this multiplex network as $G = \{G^1, G^2, \ldots, G^L\} = \{V, \mathbf{e}\}$, where $\mathbf{e} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_L\}$. Given such a multiplex network with $L$ layers, the goal of our work is to learn low-dimensional embeddings $Z^l_i \in \mathbb{R}^d$ for each node $v_i$ on each individual network $G^l$, where $d$ is the dimension of the embedding. The learned representations can be used as features in a variety of applications such as node classification and visualization, relationship mining, and link prediction. In our experiments, we perform both link prediction and node classification tasks to verify the effectiveness of the learned embedding.

4. Layer Information Similarity Concerned Network Embedding

As the nodes in each layer of the multiplex network are same, they shared the common information and the same node may show some similar features among layers. However, the structure among nodes in each layer is formed by different semantics and thus leads to quite diverse local structures of this node in each layer, and the varying relatedness between different semantics also leads to diverse structural similarities between different layers. In this paper, we propose LISCNE which models the common and local features in multiplex networks and exploits layer similarity at the same time.

Figure 2 illustrates the framework of LISCNE for a three-layer multiplex network. The architecture contains two components. The first part is modeling the common vector for all nodes that are shared by the counterpart nodes among different layers. The second part is learning the node embedding in each layer by integrating the common layer and layer vector introduced to capture distinct semantic information of different networks. The last part is describing the process of training layer vectors with layer similarities. The embedding for node $v_i$ in layer $l$ is defined as

$$Z^l_i = f(u_i + r^l).$$

where $f$ is the map function integrating common vector and layer vector to get the final node presentation. In our model, we use concatenation as the map function. LISCNE specifies the relationship of different networks by the layer similarities, i.e., $S^{12}$ indicates the index of structural similarity of network $G_1$ and network $G_2$. By adding layer similarities to the layer vector, it can associate within-network and cross-network structure information.

Next, we will describe our model LISCNE in detail and introduce it in three parts: common feature modeling, learning node embedding in each layer, and integrating the similarity between layers.

4.1. Common Feature Modeling. In this part, we learn the common feature shared by the counterpart nodes among different layer networks in the multiplex network. Firstly, we use a network aggregation method to aggregate all layers into a new single-layer network, where multiple edges are not allowed. Specifically, we set the new network as $G^{new} = \{V, \mathbf{e}_{new}\}$, and for the edge in $\mathbf{e}_j \in \{e_1, e_2, \ldots, e_L\}$, we add the edge in $e_{new}$. The process is shown in Figure 3. Then, over the obtained new network, we learn the common vector

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>The number of layers in the multiplex network</td>
</tr>
<tr>
<td>$G^l$</td>
<td>The network for layer $l$ in the multiplex network</td>
</tr>
<tr>
<td>$N$</td>
<td>The number of nodes</td>
</tr>
<tr>
<td>$V$</td>
<td>The node set of the multiplex network</td>
</tr>
<tr>
<td>$\epsilon_l$</td>
<td>The edge set of $l$-th network</td>
</tr>
<tr>
<td>$r^l$</td>
<td>The layer vector for $l$-th network</td>
</tr>
<tr>
<td>$u_i$</td>
<td>The common vector for node $v_i$</td>
</tr>
<tr>
<td>$U$</td>
<td>The common vector matrix for all nodes</td>
</tr>
<tr>
<td>$Z^l_i$</td>
<td>The embedding vector for node $v_i$ in network $G^l$</td>
</tr>
<tr>
<td>$d_1$</td>
<td>The dimension of common vector</td>
</tr>
<tr>
<td>$d_2$</td>
<td>The dimension of layer vector</td>
</tr>
<tr>
<td>$d$</td>
<td>The dimension of final node vector</td>
</tr>
<tr>
<td>$S$</td>
<td>The similarity matrix between networks</td>
</tr>
<tr>
<td>$S^{\theta}$</td>
<td>The similarity between networks $G^\theta$ and $G^\theta$</td>
</tr>
</tbody>
</table>
matrix $U$ for all nodes. We take node $v_i$ as an example; to get the common vector $u_i$, our goal is to maximize the probability of its neighbors’ context in the given walk sequence:

$$\text{max } P(v_i-w, \ldots, v_i+w; u_i),$$

where $w$ is half of the window size and the neighbors of $v_i$ are $v_i-w, \ldots, v_i+w$.

Based on the assumption of conditional independence and using the logarithmic probability, it can be further factorized as

$$L_1 = \sum_{v_i \in V} \sum_{i-c \leq j < c+i} \log P(v_j|v_i),$$

where $P(v_j|v_i)$ can be defined with a softmax function as

$$P(v_j|v_i) = \frac{\exp(u_j^T u_i)}{\sum_{k \in V} \exp(u_k^T u_i)},$$

where $u_i$ and $u_j$ are the common vectors for the input node $v_i$ and context $v_j$, respectively.

### 4.2. Learning Node Embedding in Each Layer

As discussed before, each layer in a multiplex network has distinct information, and to capture the specific structure for an individual network, we introduce the layer vector that maps single layers into a latent space, i.e., the layer vector $r_i$ for the individual graph $G^l_i$. To obtain the overall structure of the multiplex network and layer vectors and learn semantics for each layer simultaneously, we get the node embedding for each layer by concatenating them. For a random node $v_i$, the embedding in layer $G^l_i$ can be defined as $V_i^l = u_i^l r_i^l$.

To preserve local neighborhoods of nodes in each layer, our goal is to maximize the probability of specific neighbors' context in each individual layer:

$$L_2 = \sum_{l} \sum_{v_i \in V} \sum_{v_j \in C^l(v_i)} \log P(v_j|v_i; Z_{ij}^l),$$

where $C_i(v_i)$ is the context of node $v_i$ in layer $G^l_i$ and $P(v_j|v_i; Z_{ij}^l)$ can be defined as

Figure 2: The simple framework of LISCNE.

Figure 3: The process of aggregating the multiplex network into a single-layer network.
\[ P(v_j | v_i; V') = \frac{\exp(Z_i^T z_j^j)}{\sum_{k \in V} \exp(Z_k^T z_j^j)} \tag{7} \]

where \( V_i \) and \( V_j \) are the node embeddings for the input node \( v_i \) and context \( v_j \), respectively.

### 4.3. Integrating the Similarity between Layer-Networks.

The layer vector learned above can capture the distinct structure information within the layer, while in the multiplex network, there is another essential characteristic, which is the similarity between layers varying from layer to layer. Najari et al. [20] testified that incorporating the inter-layer similarities can improve the link prediction performance. Therefore, inspired by their study, we thought of using similarities to enhance embedding capabilities. We added constraints for layer vectors with the similarities between this layer and other layers. Through integrating into the layer similarity, we made the layer vector capture the cross-layer and within-layer information simultaneously.

Firstly, in our model, we used the Global Overlap Rate (GOR) algorithm to measure the similarity among layers in multiple networks. In detail, given two layers \( \alpha \) and \( \beta \) in a multiplex network, an overlap edge means that the same node pair simultaneously exist in both networks. The global overlap between layers \( \alpha \) and \( \beta \) is denoted by \( S^{\alpha \beta} \), which represents the total number of overlapping edges observed in layers \( \alpha \) and \( \beta \). It can be formulated as

\[ S^{\alpha \beta} = \frac{|e_\alpha \cap e_\beta|}{|e_\alpha \cup e_\beta|} \tag{8} \]

where \( e_\alpha \) is the total number of edges in layer \( \alpha \). The range of \( S^{\alpha \beta} \) is in \([0,1]\), and the higher the value, the more the similarity between layers. Particularly, \( S^{\alpha \beta} = 0 \) represents that there are no overlapping edges between layers, indicating that the layers are not related; otherwise, \( S^{\alpha \beta} = 1 \) means that the layers are completely correlated.

The similarity \( S^{\alpha \beta} \) between layer \( \alpha \) and \( \beta \) is the same as similarity \( S^{\beta \alpha} \) of layer \( \beta \) and \( \alpha \), and this can also been seen from equation (8).

After illustrating the definition, the next problem we should address is how to incorporate it into the model. To address this issue, we assume that if the structures of the two layers are more similar, their representation in the vector space should be closer. We force the following equation to obtain the minimum value:

\[ L_3 = \sum_{\alpha \beta} \| r^\alpha - r^\beta \|^2 S^{\alpha \beta}. \tag{9} \]

From equation (9), we can employ stochastic gradient descent to minimize \( L_3 \) function as follows:

\[ r^j = r^j - \sum_{\beta} \frac{r^j - r^\beta}{\| r^j - r^\beta \|^2} S^{\alpha \beta}. \tag{10} \]

### 4.4. Time Complexity Analysis.

Our loss function includes two components. The first part is maximizing the probability-specific neighbors’ context in each individual layer to learn the node embedding in each layer, where the main processes of time consumption include getting random walk sequences and skip-gram training, just as the ordinary random walk algorithm. Assuming that the number of nodes is \( N \), the number of edges in each layer is \( M \), the walking length is \( T \), and the number of walking sequences per node is \( t \), the complexity of sampling all sequences is \( O(M) + O(N^Tt^t) \). Besides, the complexity of optimization of \( N^Tt \) sequences with the skip-gram model is \( O(N \log N) \). Therefore, the time complexity of learning the node embedding in each layer is \( O(M) + O(N) + O(N \log N) \). The second part is integrating the similarity between layer-networks. In this part, we exploit the structural similarity between pairs of two layers, and the time complexity is \( O(L^*(L-1)) \). In real-world network data, the number of layers of \( L \) is often very small. The time complexity of this part is relatively insignificant compared to that of the first part of learning node embedding in each layer. So, the overall time complexity of our model is \( L^*(O(M) + tOnq) + hO_1(N \log N) \).

### 5. Experiments

In this section, we conduct experiments to validate the proposed LISCNE. To compare our model with some state-of-the-art single-layer embedding methods and multiplex network embedding methods, we perform link prediction and node classification tasks on several datasets with different types of networks.

#### 5.1. Datasets.

We employ five real-world multinetwork datasets from three different fields: social, co-authorship, and genetic. The basic statistical information of the datasets is presented in Table 2.

All these datasets are downloaded from the CoMuNe lab’s website (https://comunelab.fbk.eu/data.php). The detailed descriptions are as follows:

(i) CKM [51]: by asking the physicians in Illinois, Bloomington, Quincy, and Galesburg three questions, this dataset is classified into three types of relationships. Its ground truth is related to node labels; therefore, we also use this dataset to perform the node classification task.

(ii) PIERRE [52]: this dataset maps layers to different working tasks within the Pierre Auger Collaboration. Based on the keywords and contents of all submissions between 2010 and 2012, the multiplex network is divided into 16 layers.

(iii) ARABIDOPSIS [53, 54]: based on BioGRID, this multiplex network considers genetic interactions of different types of organisms. The multiplex network used in the paper makes use of the following layers: direct interaction, physical association, additive genetic interaction defined by inequality, suppressive genetic interaction defined by inequality, synthetic genetic interaction defined by inequality, association, and colocalization.
(iv) MUS [53, 54]: the dataset is also based on BioGRID. The layers in this dataset are physical association, association, direct interaction, colocalization, additive genetic interaction defined by inequality, synthetic genetic interaction defined by inequality, and suppressive genetic interaction defined by inequality.

(v) Arxiv [52]: choosing papers with “networks” in the title or abstract up to May 2014 in arxiv, the dataset is divided into 13 layers corresponding to different categories with 14,489 nodes.

5.2. Baseline Models. To show the performance of our model, the following six baseline models are implemented for comparison, which can be classified into single-layer network embedding and multiplex network embedding.

(i) DeepWalk [6]: this is a classic single-layer network embedding method, which applies a random walk to get walk sequences and then conducts the skip-gram algorithm on the sequences to train the model.

(ii) Node2Vec [7]: this is also a typical single-layer network embedding model, which utilizes two parameters to take control of the traverse probability in taking the random walk strategies.

(iii) PMNE [39]: this is a multiplex network embedding model that consists of three methods, where network aggregation and result aggregation simply merge all networks or the embedding results of all networks into one, while co-analysis takes the interaction among layers. PMNE_n, PMNE_r, and PMNE_c are used to denote the network aggregation, result aggregation, and co-analysis, respectively.

(iv) MNE [14]: this is a multiplex network embedding model that defines two different dimensional vectors for a node to capture the common information in the whole network and the specific features in a single layer, respectively.

(v) CrossMNA [16]: this is a multiplex network embedding model and also a model for network alignment. It learns simultaneously inter-vector sharing by the anchor nodes in different networks and a network vector for each single layer.

5.3. Experimental Setting. For our model, we set both the common vector dimension and layer vector dimension to 100, and thus after concatenation, the final node embedding vector dimension is 200. For the sake of fairness, we set all the dimensions of final vectors compared with our models as 200. Additionally, for DeepWalk, we set the walk to 20 and the walk length to 80 for each node taking a random walk. For Node2Vec, we empirically set $p = 2$ and $q = 0.5$. For PMME, we follow the default setting in the original paper, which sets $\alpha$, $p$, and $q$ to 0.5. For MNE, we set the additional vector dimension to 10 and the common vector dimension to 200. For CrossMNA, according to the original paper, we set the dimension of the inter-layer vector to 200 and the dimension of the network vector to 100.

5.4. Evaluation Metrics. We perform link prediction and node classification tasks to validate the efficiency of our model. For the link prediction task, we execute experiments in each layer and take the average as the final results. Then, we randomly divide datasets into testing sets and training sets. When predicting each positive edge, we also randomly sample unconnected node pairs as a negative edge. We adopt the ROC-AUC evaluation metric to test model performance, that is, the higher the value of AUC is, the better the model performs. For the node classification task, we train all data to get node embeddings of individual layers through our model and baseline models, get the average node embedding of all layers, and then inject the embeddings into a classifier to evaluate the effect. In our experiment, we select a logistic regression classifier and choose the $F_1$ (weighted) and precision (weighted) as evaluation metrics.

5.5. Performance on Link Prediction. For single-layer methods, we train the node embedding for each layer and use it to predict links in the corresponding layer. For the three methods of PMNE, which take different strategies to aggregate the representations of all layers into one, we take the final node embedding to predict links in all layers. For all models, we average the AUC values of all relation types as final results. In experiments, we take five-fold cross-validation for all datasets.

The results are shown in Table 3, from which we can draw the following observations:

(i) The proposed LISCNE model can stably outperform or achieve comparable performance with all the baseline methods. The results show that merging the layer similarity into models can exactly improve the performance.

(ii) The multiplex network models almost perform better than single-layer models. Meanwhile, these single-layer models in different datasets vary a lot, e.g., in PIERRE dataset, DeepWalk and Node2Vec

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Network type</th>
<th>Layers</th>
<th>Nodes</th>
<th>Edges</th>
<th>Directed/undirected</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKM</td>
<td>Social</td>
<td>3</td>
<td>246</td>
<td>1,551</td>
<td>Directed</td>
</tr>
<tr>
<td>PIERRE</td>
<td>Co-authorship</td>
<td>16</td>
<td>514</td>
<td>7,153</td>
<td>Undirected</td>
</tr>
<tr>
<td>ARABIDOPSIS</td>
<td>Genetic</td>
<td>7</td>
<td>6,980</td>
<td>18,654</td>
<td>Directed</td>
</tr>
<tr>
<td>MUS</td>
<td>Genetic</td>
<td>7</td>
<td>7,747</td>
<td>19,842</td>
<td>Directed</td>
</tr>
<tr>
<td>Arxiv</td>
<td>Co-authorship</td>
<td>13</td>
<td>14,489</td>
<td>59,026</td>
<td>Undirected</td>
</tr>
</tbody>
</table>

Table 2: Statistics of datasets.
perform poorly. In other words, considering the intersection across layers is essential.

(iii) LISCNE, CrossMNA, and MNE are more effective than PMNE’s three methods. PMNE models learn an overall vector for each node by aggregating all layers, while LISCNE, CrossMNA, and MNE all simultaneously define a vector to capture the common information and another vector to capture the distinct information about each specific layer.

5.6. Performance over Common Vector Embedding Dimension. Figure 4 shows the performance of our model as the embedding dimension of the common vector increases. It can be clearly seen from the figure that the larger the dimension, the better the prediction effect. When the dimension reaches 10, the curve tends to stabilize. Here, for the sake of both accuracy and computational complexity, we set the common vector dimension $d_1$ to 100.

5.7. Performance on Node Classification. In the node classification task, we choose the CKM dataset with reliable node labels to conduct the experiment and take the companies as the classification label. In addition, the ones injected into the classifier are average node vectors for node embeddings in individual layers. For single-layer network methods, we train all nodes in each layer and get the average of node vectors in each layer. For MNE, CrossMNA, and our model, we also get the average of node vectors of intra-vector in individual layers. Then, all the node representations and corresponding node labels in each layer are divided into training and testing datasets to train the classifier. In our experiment, we use a logistic classifier and evaluate the classification performance with the metrics accuracy, precision, and $F_1$, respectively, which can be defined as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$F_1 = \frac{2TP}{2TP + FN + FP}$$

As shown in Figure 5, the results prove the effectiveness of our model, where our model LISCNE can provide the best performance in terms of $F_1$ and precision and achieve comparable accuracy with PMNE_n and PMNE_c. However, the effectiveness of multiplex network embedding models like CrossMNA on the link prediction task is not obvious. This may be because every model injected into the classifier is the average of node embedding in all layers. The effect of average is somewhat like aggregation and gets the shared information in all layers.

6. Conclusion and Future Work

In this paper, we propose an effective method called LISCNE for multiplex network embedding. LISCNE defines a common vector for all counterpart nodes in the multiplex network and also introduces a layer vector for each layer. Moreover, when learning layer vectors, it first merges the layer similarities to simultaneously capture intra-layer information and cross-network information. We have performed link prediction and node classification tasks to test LISCNE and conducted extensive experiments.
to verify the effectiveness of our proposed model. This model is applicable to aligned networks and certain networks, in which one node in some network is only connected to one node in another network.

Unfortunately, this kind of network cannot cover lots of scenarios in the real world, e.g., the association between a collaboration graph of researchers and a citation graph of papers, where an author can cite papers on multiple topics. In the future, we will extend our model to more manifold networks, for example, one node in some network is connected to several nodes in another network through different weights.

**Data Availability**

The datasets used to support the results of this study can be available from https://comunelab.fbk.eu/data.php.

**Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

**Acknowledgments**

This study was supported by the National Natural Science Foundation of China (61902278).

**References**


