Structure Fusion Based on Graph Convolutional Networks for Node Classification in Citation Networks

Guangfeng Lin *[ID]**, Jing Wang, Kaiyang Liao, Fan Zhao and Wanjun Chen**

Department of Information Science, Xi'an University of Technology, Xi'an 710048, China;
wangjing63@xaut.edu.cn (J.W.); liaokaiyang@xaut.edu.cn (K.L.); vcu@xaut.edu.cn (F.Z.);
wjchen@xaut.edu.cn (W.C.)
* Correspondence: lgf78103@xaut.edu.cn

check for
updates

Abstract: Suffering from the multi-view data diversity and complexity, most of the existing graph convolutional networks focus on the networks' architecture construction or the salient graph structure preservation for node classification in citation networks and usually ignore capturing the complete graph structure of nodes for enhancing classification performance. To mine the more complete distribution structure from multi-graph structures of multi-view data with the consideration of their specificity and the commonality, we propose structure fusion based on graph convolutional networks (SF-GCN) for improving the performance of node classification in a semi-supervised way. SF-GCN can not only exploit the special characteristic of each view datum by spectral embedding preserving multi-graph structures, but also explore the common style of multi-view data by the distance metric between multi-graph structures. Suppose the linear relationship between multi-graph structures; we can construct the optimization function of the structure fusion model by balancing the specificity loss and the commonality loss. By solving this function, we can simultaneously obtain the fusion spectral embedding from the multi-view data and the fusion structure as the adjacent matrix to input graph convolutional networks for node classification in a semi-supervised way. Furthermore, we generalize the structure fusion to structure diffusion propagation and present structure propagation fusion based on graph convolutional networks (SPF-GCN) for utilizing these structure interactions. Experiments demonstrate that the performance of SPF-GCN outperforms that of the state-of-the-art methods on three challenging datasets, which are Cora, Citeseer, and Pubmed in citation networks.

Keywords: structure fusion; graph convolutional networks; node classification; citation networks

---

1. Introduction

As an efficient representation of data distribution, the graph plays an important role for describing the intrinsic structure of data. Therefore, many existing works have constructed a significant theory and method depending on the graph structure of data in pattern recognition, such as the graph cut building energy function for the semantic segmentation task [1], graph-based learning system constructing the accurate recommendations for the interaction of the different objects [2,3], graph modeling molecules' bioactivity for drug discovery [4,5], and graphs simulating the link connection of a citation network for the different group classification [4–6]. In fact, we can usually observe data and their distribution relationship (This distribution relationship is defined as the data of structure, which often can be described by a graph. Each datum is the node of the graph, and the distance metric of data pairs is the weight between nodes in the graph.) from multiple views, which provide a more abundant and complete information for object recognition. Learning on a multi-graph (In each view, we can use a graph for describing the distribution relationship of the observation data, and in multiple views,

we can obtain multiple graphs, which are called multi-graphs. In the multi-graphs, we can learn their relevance based on some optimization functions, and this process is called learning on the multi-graph.) can effectively mine multiple relationships (this includes specificity and commonality in Section 4. The specificity relationship is the linear relationship between the embedding matrix of multiple structures, while the commonality relationship is the distance metric between the embedding matrix of multiple structures.) to discriminate the different data objects.

The recent learning methods for mining multi-graph structure information mainly have tow categories. One is structure fusion [7–17] or diffusion on the tensor product graph [18–23] based on the complete data, which include each view observation datum. Another is graph convolutional networks for the salient graph structure preservation [6] or node information fusion [24,25] based on incomplete data, which lacks some view observation data. For example, the link relationship among nodes can be extracted by the hyperlink direction in citation networks, but it cannot be described by the corresponding node feature computation. In other words, these link relationships exist, while the corresponding node features are lost in citation networks. Therefore, the method based on graph convolutional networks usually ignores the complete complementary of the different observation structures based on the incomplete multi-view data. To analyze this issue, we attempt to construct structure fusion based on graph convolutional networks for classifying the different nodes in citation networks in a semi-supervised way. Figure 1 shows the overall flow diagram of structure fusion based on graph convolutional networks (SF-GCN).
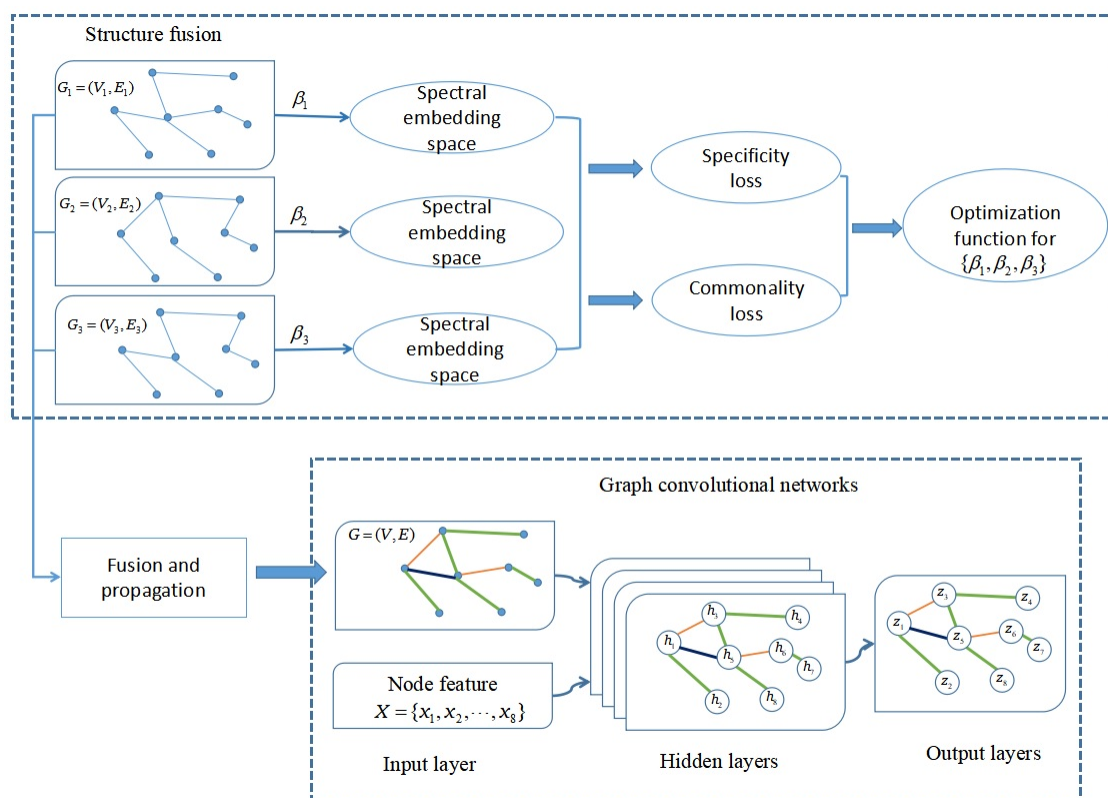

Figure 1. The diagram of structure fusion based on graph convolutional networks (SF-GCN), in which three graphs indicating the structure of the multi-view data and eight nodes (the different color connecting lines mean the various connecting weights) expressing the multi-nodes in these graphs; $\beta = [\beta_1 \; \beta_2 \; \beta_3]$ (the linear coefficient between the multi-graph structure) can be learned for complementary fusion.

The inspiration of SF-GCN comes from multi-GCN in the literature [6], but there are two different points of comparison with multi-GCN. One is that SF-GCN considers the different roles of multiple structures by learning different weights, while multi-GCN only deals with their relationship by the same weights. The other is that SF-GCN focuses on the contributions of all nodes' structure in the fusion structure, while multi-GCN only emphasizes the salient structure of the part nodes. In the classification sense, the strong and weak link relationships between nodes both considered for complementing the structure are more fit to the intrinsic structure of the data for node classification in citation networks.

Our contributions can be summarized as follows. (a) We present a novel structure fusion based on graph convolutional networks (SF-GCN) that discriminates the different classes by optimizing the linear relationship of multiple observation structures with balancing the specificity loss and the commonality loss. (b) In three citation datasets with the sparse document feature and document link relationship, the proposed SF-GCN outperforms the state-of-the-art methods for node classification. (c) Our model generalizes the different multi-graph fusion methods for evaluating the performance of the proposed SF-GCN.

2. Related Works

In this section, we mainly review recent related works about structure fusion, graph neural networks, and node classification based on GCN.

2.1. Structure Fusion

The structure fusion initially proposed in [7] can merge multiple structures for shape classification. In the follow-up works, the extend methods can be divided into three categories according to the different fusion methods.

The first kind of method tried to find the optimized linear relationship of multiple observation structures based on the different manifold learning methods [8] or statistics model analysis [9] for encoding the importance degree of the different structure in multiple views. These methods can consider the different weights of the homogeneous structure for capturing the more complete distribution relationship.

The second kind of method attempts to mine the nonlinear relationship of the heterogeneous feature structure based on the global feature [10,11] or the local feature encoding [12] for finding the complementary of the different structures in multiple representations. These methods can bridge the gap between heterogeneous structures for the uniform fusion feature representation.

The third kind of methods can capture the dynamic changes of multiple structures for semi-supervised classification [13] or the structure propagation method for zero-shot learning [14–17] for transferring the structure information of the different data objects. These methods can find the dynamic rule of the structure change to help understand the cross-domain data.

From the above, existing methods emphasize the completeness of data and their relationship based on data mapping, while graph convolutional networks focus on the transformation and evolution of the data structure by deep learning frameworks. Therefore, we expect to draw support from structure fusion based on the structure metric and graph convolutional networks for processing the incomplete view data and find the evolution law of the fusion structure with the consideration of their specificity and commonality.

2.2. Graph Neural Networks

Graph neural networks can discover the potential data relationship by computation based on graph nodes and links. Especially, the computation is defined as convolution for graph data, and graph convolution networks (GCN) have become a promising direction in pattern recognition. In terms of the different node representation, graph convolution networks include spectral-based GCN and spatial-based GCN.

Spectral-based GCN can define the graph Fourier transform based on the graph Laplacian matrix for projecting the graph signal into the orthonormal space. The difference of these methods is the selection of the filter, which may be the learned parameters set [26], Chebyshev polynomial [4], or the first-order Chebyshev polynomial [27,28].

Spatial-based GCN regards an image as a special graph with a pixel describing a node. To avoid the storage of all states, these methods have presented improved training strategies, such as sub-graph training [29] or stochastically asynchronous training [30]. Furthermore, some complex networks' architecture can utilize the gating unit to control the selection of the node neighborhood [31], or design two graph convolution networks with the consideration of the local and global consistency on the graph [25], or adjust the receptive field of the node on graph by hyper-parameters [32].

We need to process the image as a node of the graph, so spatial-based GCN is not suitable for node classification in citation network. Spectral-based GCN can explicitly construct the learning model on the graph structure, which can easily be separated from GCN architecture. Therefore, this point provides an independent way for processing multiple structures. In this paper, we focus on the important role of the graph (structure) from multi-view data and attempt to mine the plentiful information from multiple structures for spectral-based GCN input.

2.3. Node Classification Based on GCN

According to the different node information processing method, the recent node classification methods based on GCN are divided into two categories. One is node neighbor information exploiting for GCN, and the other is node information fusion based on GCN.

Node neighbor information exploiting attempts to capture the distribution structure of the node neighbor for obtaining the stable graph structure representation. For example, graph attention networks (GAT) can specify different weights to different nodes in a neighborhood [33]; stochastic training of graph convolutional networks (StoGCN) allows sampling an arbitrarily small neighbor size [34]; deep graph infomax (DGI) can maximize mutual information between different levels of subgraphs centered around nodes of interest (the different way for considering neighbor information) [35].

Node information fusion tries to mine the information from the multi-view node description or multiple structures for complementing the difference of multi-view data. For instance, large-scale learnable graph convolutional networks (LGCN) can fuse neighboring nodes' features by ranking selection to transform the graph data into grid-like structures in 1D format [24]; dual graph convolutional networks (DGCN) can consider local and global consistency for fusing different view graphs of raw data [25]; multi-GCN can extract and select the significant structure form the multi-view structure by manifold ranking [6].

The proposed method belongs to the node information fusion method for exploiting neighbor information, and the difference compared with the above methods focuses on the complementary of multiple structures by mining their commonality, specificity, and interactive propagation.

3. Graph Convolutional Networks

In terms of the multiplication of convolution in the Fourier domain, graph convolution is defined as the multiplication between the signal $s \in R_n$ and the filter g_η [26]. Further, graph convolution can also be approximated by first-order Chebyshev polynomials [27] as follows.

$$
\begin{aligned}
g_\eta * s &= U g_\eta U^T s \\
&\approx \sum_{k=0}^{1} \eta_k' T_k(\tilde{L}) s \\
&\approx \eta (I + \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}) s
\end{aligned}
\tag{1}
$$

where U is the eigen-decomposition of the normalized Laplacian $L = I - D^{-1/2}WD^{-1/2}$ (I is the identity matrix; D is the degree matrix of graph G); $\tilde{L} = I - \tilde{D}^{-1/2}\tilde{W}\tilde{D}^{-1/2}$ ($\tilde{D}$ and $\tilde{W}$ respectively are the rescaled degree and adjacent matrix by $\tilde{W} = W + I$); T_k expresses the Chebyshev polynomials; $\eta = \eta_0' = -\eta_1'$.

Fusion structure W (details in Section 4.3) can directly be input into the above graph convolutional networks. The forward propagation based on two layers of graph convolutional networks can be indicated as follows.

$$Z = softmax(\tilde{W}ReLU(\tilde{W}S\Theta^0)\Theta^1) \tag{2}$$

where Z is the output of networks; S is the representation matrix of each node; Θ^0 and Θ^1 respectively are the first and second layer filter parameters; $ReLU$ and $softmax$ are the different types of activation function located in the various layers.

To obtain the complete graph information as the input of GCN, we deal with multiple structures' fusion in the next section.

4. Multiple Structures' Fusion

To the best of our knowledge, existing structure fusion methods usually construct the optimizing function for feature projection, in which feature data and the corresponding structure jointly participate in the computation. Because of the possible data loss and the structure preservation of multi-view data, we expect to build a novel structure fusion by the structure metric, in which the optimizing function involves multiple structures for avoiding the negative effect of the data lost. The relationships of multiple structures include not only the specificity relationship, but also the commonality relationship. Therefore, we also anticipate that a novel structure fusion can be constrained by these characteristics of multiple structures. Figure 1 demonstrates the internal mechanism of structure fusion in SF-GCN. First, we construct the specificity loss based on spectral embedding method with the consideration of a multiple structure linear relationship. Second, we measure the commonality loss between multiple structures based on the distance metric in the Grassmann manifold. Finally, we jointly exploit the structure fusion based on two losses and input the fusion structure into GCN for node classification.

4.1. Specificity Loss of Multiple Structures

Given an object set with m multi-views, we can use graph G_i to describe the observation distribution of data on each view. Therefore, the graph G_i is the representation of the observation structure, and $G = \{G_i | i = 1, 2, ..., m\}$ can indicate multiple structures of data from multi-view observations. Because multiple structures detail the same object set, each G_i includes the same vertex set V or the possible different edges set E_i. If W_i is the adjacency matrix of G_i and is the numerical expression of the structure in the ith view, in terms of spectral embedding [36], we can obtain the following optimization function on the embedding matrix $Y_i \in R^{n \times k}$ (n is the number of samples, and k is the dimension of the embedding space) of each view. The data embedding shows the distribution characteristic of the data, and the different data embedding can be solved for each view datum as follows.

$$Y_i = \arg\min tr(Y_i^T L_i Y_i), \quad s.t. \ Y_i^T Y_i = \mathbf{1} \tag{3}$$

where tr stands for the trace of the square matrix, $L_i = D_i - W_i$ is the Laplacian matrix of G_i, and D_i is the degree matrix for G_i. Therefore, L_i can still describe the characteristic of the structure on graph G_i. We can compute the embedding matrix Y_i by optimizing Equation (3), which is equivalent to an eigenvalue solution problem. When all eigenvalues are solved, eigenvectors corresponding to the smallest eigenvalues can build the embedding matrix Y_i, which is a mapping from the original nodes into the low-dimensional spectral space [36]. We can regard $tr(Y_i^T L_i Y_i)$ as the specificity loss

of the structure on graph G_i, and then, we can reformulate the specificity loss of multiple structures as follows.

$$Loss_s = tr(Y^T LY) \tag{4}$$

where Y is the embedding matrix of multiple structures in graph G and closely approximates Y_i. Suppose fusion structure W is the linear combination of W_i, then L and L_i have the same linear relationship $L = \sum_{i=1}^{m} \beta_i L_i$, in which β_i is the linear coefficient to encode the importance of multiple structures.

4.2. Commonality Loss of Multiple Structures

To measure the commonality loss of multiple structures, we need the metric for the distance between Laplacian matrices L_i and L. According to the solution of Equation (3), we can obtain Equation (5) for describing the internal connection between embedding matrix Y_i and the corresponding Laplacian matrix L_i.

$$L_i = Y_i \lambda_i Y_i^T \tag{5}$$

where λ_i is the diagonal matrix, in which diagonal values are the smallest eigenvalues of L_i. Y_i can be explained as a subspace for preserving the smaller variance of the column in L_i, that is for reserving the bigger variance of the column in the structure W_i. In other words, Y_i can maintain the greater discrimination of the data. Similarly, Y has the same sense in the multiple observation structure. Therefore, we can replace the distance between Laplacian matrices L_i and L by the distance between Y_i and Y for indirectly computing the commonality loss of multiple structures. This point is consistent in the specificity loss of the assumption, which is that Y_i approximates Y between each view and multi-views.

In terms of Grassmann manifold theory [37,38], the orthonormal matrix $Y_i \in R^{n \times k}$ can be regarded as the column of Y_i spanning a unique subspace, which can be projected into a unique point on Grassmann manifold $\mathcal{G}(n, k)$. Similarly, Y also can be mapped into a unique point on this Grassmann manifold. Therefore, the principle angles $\{\theta_j\}_{j=1}^{k}$ between these subspaces can represent the distance between Y_i and Y. Furthermore, this distance can be reformulated as follows [39].

$$d^2(Y, Y_i) = \sum_{j=1}^{k} \sin^2 \theta_j = k - tr(YY^T Y_i Y_i^T) \tag{6}$$

In multiple structures, we can define the commonality loss $Loss_c$ as the distance between Y and $\{Y_i\}_{i=1}^{m}$ as follows.

$$Loss_c = \sum_{i=1}^{m} d^2(Y, Y_i) = km - \sum_{i=1}^{m} tr(YY^T Y_i Y_i^T) \tag{7}$$

4.3. Structure Fusion by Structure Metric Losses

As two structure metric losses, specificity loss can balance the contribution of the structure in each view, while commonality loss can consider the similarity of multiple structures in multi-views. These structure metric losses can both constrain the linear relationship $\{\beta_i\}_{i=1}^{m}$ of multiple structures. Therefore, we combine these structure metric losses as a total loss for encoding the importance of multiple structures. The total loss can be reformulated as follows.

$$Loss = loss_s + \alpha loss_c \tag{8}$$

where α is the regularization parameter. From Equation (8), we can construct the object optimization function as follows.

$$
\begin{aligned}
\{Y, \{\beta_i\}_{i=1}^m\} &= \arg\min(loss_s + \alpha loss_c) \\
&= \arg\min tr(Y^T LY) + \alpha(km - \sum_{i=1}^m tr(YY^T Y_i Y_i^T)) \\
&= \arg\min tr(Y^T \sum_{i=1}^m \beta_i L_i Y) + \alpha(km - \sum_{i=1}^m tr(YY^T Y_i Y_i^T)) \\
s.t. \ Y^T Y &= \mathbf{1}, \ \sum_{i=1}^m \beta_i = 1, \ \alpha > 0
\end{aligned}
\tag{9}
$$

In commonality loss, constant term km cannot influence the loss trend change, so we may remove this term for conveniently computing. Equation (9) is reformulated as Equation (10) for balancing parameter $\{\beta_i\}_{i=1}^m$ between zero and one.

$$
\begin{aligned}
\{Y, \gamma\} &= \arg\min(tr(Y^T \sum_{i=1}^m \beta_i L_i Y) - \alpha \sum_{i=1}^m tr(YY^T Y_i Y_i^T)))^2 \\
&= \arg\min(tr(Y^T (\sum_{i=1}^m \beta_i L_i - \alpha \sum_{i=1}^m Y_i Y_i^T)Y))^2 \\
s.t. \ Y^T Y &= \mathbf{1}, \ \sum_{i=1}^m \beta_i = 1, \ \alpha > 0, \ \gamma = \{\{\beta_i\}_{i=1}^m, \alpha\}
\end{aligned}
\tag{10}
$$

Equation (10) is a nonconvex optimization problem, and we can solve this problem by Y and γ alternated optimization. If γ is fixed, Equation (10) can be transformed as an eigenvalue solving problem as follows.

$$
\begin{aligned}
\{Y\} &= \arg\min tr(Y^T MY) \\
s.t. \ Y^T Y &= \mathbf{1}, \ \alpha > 0, \ M = (\sum_{i=1}^m \beta_i L_i - \alpha \sum_{i=1}^m Y_i Y_i^T), \ \gamma = \{\{\beta_i\}_{i=1}^m, \alpha\}
\end{aligned}
\tag{11}
$$

Equation (11) is equivalent to an eigenvalue solution problem. When all eigenvalues of M are solved, eigenvectors corresponding to the smallest eigenvalues can build the fusion embedding matrix Y. If Y is fixed, Equation (10) can be converted into a quadratic programming problem as follows.

$$
\begin{aligned}
\{\gamma\} &= \arg\min(tr(Y^T (\sum_{i=1}^m \beta_i L_i - \alpha \sum_{i=1}^m Y_i Y_i^T)Y))^2 \\
s.t. \ \sum_{i=1}^m \beta_i &= 1, \ \alpha > 0, \ \gamma = \{\{\beta_i\}_{i=1}^m, \alpha\}
\end{aligned}
\tag{12}
$$

By alternate solving between Equations (11) and (12), we can obtain fusion embedding matrix Y and the linear relationship γ of multiple structures. Furthermore, the fusion structure (fusion adjacent matrix) can be computed by $W = \sum_{i=1}^m \beta_i W_i$.

Algorithm 1 shows the pseudocode for the fusion structure of multiple structures. In this algorithm, there are three steps. The first step (Line 1) initializes the linear relationship of multiple structures. The second step (from Line 2 to Line 3) computes the Laplacian matrix and the spectral embedding in each view. The third step (from Line 4 to Line 6) alternately optimizes the spectral embedding fusion and the linear relationship of multiple structures. The last step (Line 8) calculates the fusion structure by the linear combination of each structure. Therefore, the complexity of this algorithm is $O(mn^3 + mn^2 kT + k^{3.5}l^2 T)$, in which m represents multi-views; n is the sample number;

k is the dimension of the selected eigenvectors; T is the iterative times of optimization; l is the number of bits in the input of the algorithm.

Algorithm 1 Fusion structure of multiple structures.

Input: $\{W_i\}_{i=1}^m$: $n \times n$ adjacency matrices of graph $\{G_i\}_{i=1}^m$; α: regularization parameter of the total loss; T: the iteration times

Output: W: fusion structure of multiple structures

1: Initializing the linear relationship $\{\beta_i\}_{i=1}^m$ of multiple structures and regularization parameter α
2: Computing the Laplacian matrix L_i of G_i
3: Computing the spectral embedding Y_i of the structure in each view by Equation (3)
4: **for** $1 < t < T$ **do**

5:     Computing the spectral embedding fusion Y of multiple structures in multi-views by Equation (11)

6:     Updating the linear relationship γ of multiple structures by Equation (12)
7: **end for**
8: Computing the fusion structure by $W = \sum_{i=1}^m \beta_i W_i$

---

5. Experiments

For evaluating the proposed SF-GCN, we carried out the experiments from four aspects. Firstly, we conducted the comparing experiment between the proposed SF-GCN and the baseline methods, which included graph convolutional networks (GCN) [27] with the combination view and multi-GCN [6]. Secondly, we utilized the different multi-graph fusion methods for analyzing the intrinsic mechanism of the proposed SF-GCN. Thirdly, we showed the experimental results between the proposed SF-GCN and the state-of-the-art methods for the node classification in citation networks. Finally, we implemented the proposed SF-GCN method of the lost structure for demonstrating the importance of the complete structure.

5.1. Datasets

We used three paper citation networks in the experiments. These popular datasets usually utilized in node classification respectively were Cora, Citeseer, and Pubmed. The Cora dataset had 7 classes that involved 2708 grouped publications about machine learning and their undirected graph. The Citeseer dataset included 6 classes that had 3327 scientific papers and their undirected graph. In these datasets, each publication stood for a node of the related graph and was represented by a one-hot vector, each element of which could indicate the presence and absence state of a word in the learned directory. The Pubmed dataset had 3 classes that contained 19,717 diabetes-related publications and their undirected graph. In this dataset, each paper (each node of the related graph) could be described by a term frequency-inverse document frequency (TF-IDF) [40]. Table 1 shows the statistics of these datasets. To obtain the structure of the second view from the publication description, we normalized the cosine similarity between these publication. If the similarity was greater than 0.8, we produced an edge for the corresponding nodes in the citation network. This configuration was the same in [6].

Table 1. Three datasets' statistics in citation networks.

Datasets	Nodes Number	Edges Number	Classes Number	Feature Dimension	Label Rate
Cora	2708	5429	7	1433	0.052
Citeseer	3327	4732	6	3703	0.036
Pubmed	19,717	44,338	3	500	0.003

5.2. Experimental Configuration

In the experiments, we followed the configuration of GCN [27], in which we trained a two-layer GCN for a maximum of 200 epochs and a test model in 1000 labeled samples. Moreover, we selected the same validation set of 500 labeled sample for hyper-parameter optimization (dropout rate for all layers, number of hidden units, and learning rate).

In the proposed SF-GCN, we initially set the linear relationship $\{\beta_i\}_{i=1}^m$ of multiple structures and regularization parameter α as 0.5 and then updated these parameters in the iteration optimization. The iteration time T of the algorithm was 5 according to the convergence degree.

5.3. Comparison with the Base-Line Methods

The proposed method (SF-GCN) could be constructed based on GCN [27] and attempted to mine the different structure information for completing the intrinsic structure in multi-view data. Therefore, two base-line methods (GCN and multi-GCN can find and capture the different structure information from the different considerations) were involved for processing multi-view data based on GCN. GCN for multi-views [27] could concatenate the different structure to build a sparse block-diagonal matrix where each block corresponded to the different structures (the adjacent matrix of different graphs). Multi-GCN [6] could preserve the significant structure of the different structures by manifold ranking. In contrast with these base-line methods, the proposed method (SF-GCN) could not only enhance the common structure, but also retain the specific structure by structure fusion.

Table 2 shows that the classification of SF-GCN outperformed that of the base-line methods, and the least improvement of SF-GCN respectively was 0.7% for Cora, 2.1% for Citeseer, and 0.6% for PubMed. However, GCN for multi-views was not superior to GCN for single-views, and it demonstrated that information mining of multi-view data was a key point for node classification. Therefore, SF-GCN attempted to mine the structure information from multi-view data for node classification and obtained better performance.

Table 2. Accuracy comparison of the structure fusion based on graph convolutional networks (SF-GCN) method with the base-line methods for node classification in citation network (View 1 stands for the graph structure from the original dataset, while View 2 indicates the graph structure from the cosine similarity of node representation).

Method	Cora	Citeseer	PubMed
GCN [27] for View 1	81.5	70.3	78.7
GCN [27] for View 2	53.6	50.7	69.5
GCN [27] for multi-view	80.4	70.7	78.2
Multi-GCN [6]	82.5	71.3	NA
SF-GCN	**83.3** $\pm$ 0.4	**73.4** $\pm$ 0.7	**79.3** $\pm$ 0.4

5.4. Structure Fusion Generalization

Structure fusion (SF) focuses on the complementation of the distribution structure from the different view data, and $W = \sum_{i=1}^m \beta_i W_i$ can be defined in Section 4.3. However, the diffusion [19,22] and propagation [14,16] of the different structure can also describe the complex relationship of the various structure,and become an important part of structure fusion. Therefore, we can define fusion structure W by the propagation fusion (PF) of the different structure as follows.

$$W = \prod_i^m W_i \tag{13}$$

The propagation fusion can exchange and interact with the relationship information between the various structures and mine the neighbor relationship of multiple structures. However, this propagation can effect the clustering performance of the original structure by high-order iteration multiplication. Therefore, we only considered zero-order (for example SF) and first-order (for instance PF) multiplication, that is structure propagation fusion (SPF), as follows.

$$W = \sum_{i=1}^{m} \beta_i W_i + \prod_{i}^{m} W_i \tag{14}$$

For evaluating structure fusion generalization, we compared structure fusion based on graph convolutional networks (SF-GCN), propagation fusion based on graph convolutional networks (PF-GCN), and structure propagation fusion based on graph convolutional networks (SPF-GCN). In Table 3, we observe that the performance of SPF-GCN was better than those of other methods, and the least improvement of SPF-GCN respectively was 0.2% for Cora, 0.1% for Citeseer, and 0.7% for PubMed, while the performance of SP was superior to that of PF-GCN, the improvement of SF-GCN respectively being 0.6% for Cora, 0.9% for Citeseer, and 0.2% for PubMed. Therefore, PF and SF both were benefited for further mining the structure information, and the role of SF was more important than that of PF.

Table 3. Structure fusion generalization classification accuracy in three methods, which are structure fusion based on graph convolutional networks (SF-GCN), propagation fusion based on graph convolutional networks (PF-GCN), and structure propagation fusion based on graph convolutional networks (SPF-GCN).

Method	Cora	Citeseer	PubMed
SF-GCN	83.3 ± 0.4	73.4 ± 0.7	79.3 ± 0.4
PF-GCN	82.7 ± 0.5	72.5 ± 0.4	79.1 ± 0.6
SPF-GCN	$\textbf{83.5} \pm 0.6$	$\textbf{73.5} \pm 0.8$	$\textbf{80.0} \pm 0.5$

5.5. Comparison with the State-of-the-Art

Because graph convolutional networks and structure fusion are basic ideas for constructing the proposed method SPF-GCN, we involved six related state-of-the-art methods (in Section 2.3) for evaluating SPF-GCN. Table 4 shows that SPF-GCN outperformed other state-of-the-art methods except DGCN on the Cora and PubMed datasets. Although SPF-GCN and DGCN reached the same performance on the Cora and PubMed datasets, SPF-GCN could preserve the higher computation efficiency of the original GCN because of the separable computation between structure fusion and GCN.

Table 4. Accuracy comparison of SF-GCN and SPF-GCN with state-of-the-art methods for node classification in the citation networks. GAT, graph attention networks; StoGCN, stochastic training of graph convolutional network; DGI, deep graph infomax (DGI); LGCN, learnable graph convolutional network; DGCN, dual graph convolutional network.

Method	Cora	Citeseer	PubMed
GAT [33]	83.0 ± 0.7	72.5 ± 0.7	79.3 ± 0.3
StoGCN [34]	82.0 ± 0.8	70.9 ± 0.2	78.7 ± 0.3
DGI [35]	82.3 ± 0.6	71.8 ± 0.7	76.8 ± 0.6
LGCN [24]	83.3 ± 0.5	73.0 ± 0.6	79.5 ± 0.2
DGCN [25]	**83.5**	72.6	**80.0**
Multi-GCN [6]	82.5	71.3	NA
SF-GCN	83.3 ± 0.4	73.4 ± 0.7	79.3 ± 0.4
SPF-GCN	$\textbf{83.5} \pm 0.6$	$\textbf{73.5} \pm 0.8$	$\textbf{80.0} \pm 0.5$

5.6. Incomplete Structure Influence

Structure fusion can capture the complementary information of multiple structures, and this complementary information can supply an efficient way for incomplete structure influence. The main reason for the incomplete structure may be the noise and data loss in practical situations. For evaluating the performance of the proposed methods under the condition of the incomplete structure, we designed an experiment in all datasets. In semi-supervised classification, the distribution structure of test datasets was more important than that of training datasets and could assure the performance of classification because of the transfer relation of the structure between training and test datasets. Therefore, we randomly deleted some structure of the test datasets to destroy this transfer relation for simulating an incomplete structure.

In detail, we proportionally set the adjacency matrix (graph structure from the original dataset) of elements (corresponding to test datasets) to zero from 10% to 60% and then respectively implemented GCN for multi-views [27], DGCN [25], SF-GCN, and SPF-GCN methods for all datasets. In Figure 2, we selected the structure loss degree from 10%, 20%, 30%, 40%, 50%, to 60% to construct the different classification models for evaluating the performance of the compared methods. Especially, there was a smaller descent of SPF-GCN classification accuracy with structure loss increasing from 10% to 60%, e.g., 83.3 to 82.5 on Cora, 73.4 to 73.0 on Citeseer, and 79.8 to 79.4 on PubMed. We could observe that the proposed SF-GCN and SPF-GCN were more stable and robust with the incomplete degree of the structure increasing than GCN for multi-views and DGCN. In this situation, the performance of SPF-GCN was better than that of SF-GCN, while the performance of GCN outperformed that of DGCN on the Cora dataset, and the performance of DGCN was superior to that of GCN on the Citeseer and PubMed datasets. The details of this reason can be analyzed in Section 5.7.
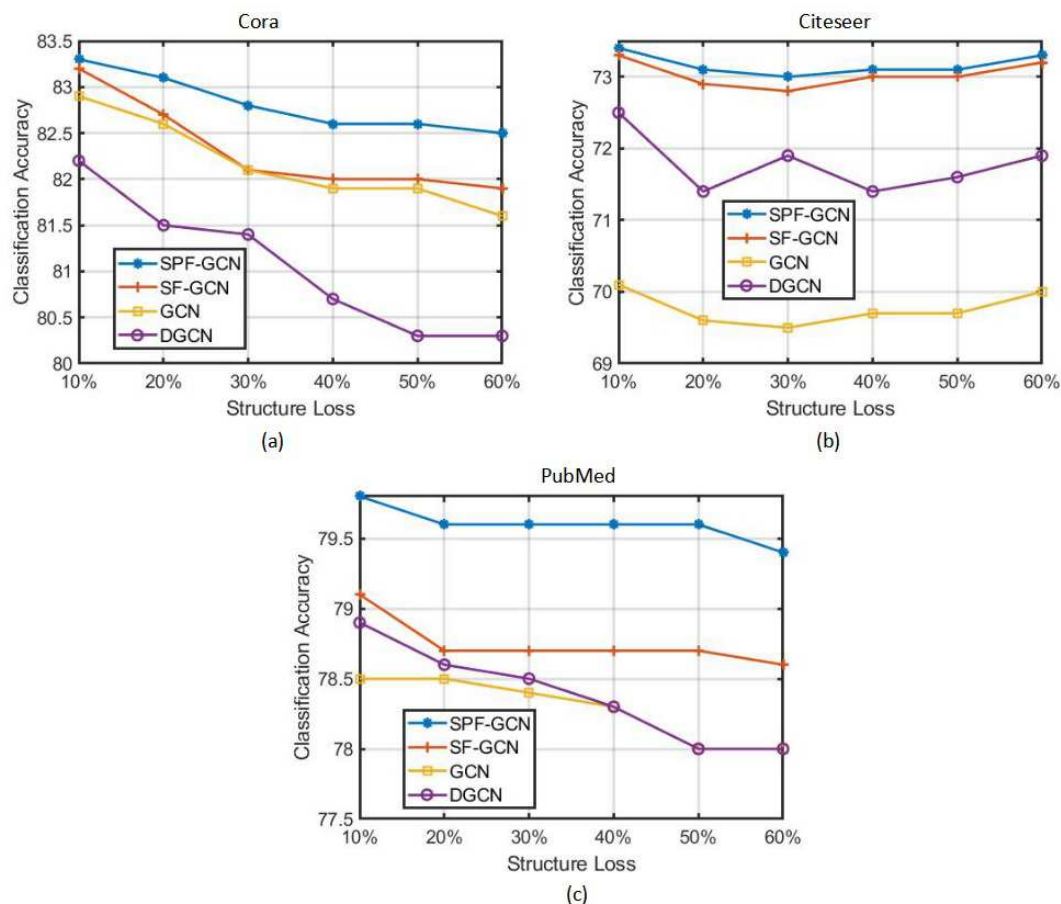

Figure 2. Impact of the structure loss on the classification accuracy for citation networks on the (**a**) Cora, (**b**) Citeseer, and (**c**) PubMed datasets.

5.7. Experimental Results' Analysis

In our experiments, we compared the proposed method with eight methods, which contained two kinds of base-line methods (multi-GCN [6], GCN [27] for multi-views, GCN [27] for View 1 and View 2 in Section 5.3), two kinds of structure fusion generalization methods (PF-GCN and SPF-GCN in Section 5.4), and six kinds of state-of-the-art methods (GAT [33], StoGCN [34], DGI [35], LGCN [24], DGCN [25], and multi-GCN [6] in Section 5.5). These methods could utilize the graph structure mining based on graph convolutional networks for semi-supervised classification by different ways. In contrast to other methods, the proposed SF-GCN and SPF-GCN methods focused on the complementary relationship of multiple structures by the consideration of their commonality and specificity. Moreover, the proposed SPF-GCN method could not only capture the optimization distribution of the fusion structure, but also emphasize the interactive propagation between the different structures. From the above experiments, we could observe several points as follows.

- The performance of SF-GCN was superior to that of the base-line methods (multi-GCN [6], GCN [27] for multi-views, and GCN [27] for View 1 and View 2 in Section 5.3). GCN [27] constructed a general graph convolutional architecture by the first-order approximation of spectral graph convolutions for greatly improving the computation efficiency of graph convolutional networks and also provided a feasible deep mining framework for effective semi-supervised classification.   For using multiple structures, GCN for multi-views could input a sparse block-diagonal matrix, each block of which corresponded to the different structures. Therefore, the relationship of each block (the different structure) was ignored for GCN, and this point lead to the poor performance (for some times, the performance of GCN for multi-view was worse than that of GCN for View 1) of GCN for multi-views. In contrast, multi-GCN [6] could capture the relationship of the different structures to preserve the significant structure of merging the subspace. However, multi-GCN [6] neglected optimizing the fusion relationship of the different structures, while the proposed SF-GCN focused on finding these relationships by jointly considering the commonality and specificity loss of multiple structures for obtaining better performance for node classification.
- SPF-GCN showed the best performance in structure fusion generalization experiments, whereas the performance of SF-GCN was better than that of SP-GCN. The main reason was that SF-GCN emphasized the complementary information by the optimizing fusion relationship of the different structures, while SP-GCN tended toward the interactive propagation by the diffusion influence between the different structures. The complement fusion played a more important role than the interactive propagation because of the specificity structure of individual view data, but both fusion and propagation could contribute the multiple structure mining for enhancing the performance of node classification.
- The performance improvements of SPF-GCN compared with six kinds of state-of-the-art methods were respectively different. A similar performance of SPF-GCN was shown in the comparison with LGCN and DGCN on Cora and PubMed. Except these situation, the better improvement of SPF-GCN could be demonstrated for other methods. The main reasons were that LGCN could emphasize neighboring nodes' feature fusion for the stable node representation and DGCN could correlate the local and global consistency for complementing the different structures. The proposed SPF-GCN was expected not only to capture the structure commonality for complementing the different information, but also to preserve the structure specificity for mining the discriminative information. Therefore, the proposed SPF-GCN could improve the classification performance in most experiments. In the least, the proposed SPF-GCN had a similar performance as the best performance of other methods in all experiments. In addition, the proposed SPF-GCN was based on GCN frameworks, so it had an efficient implementation like GCN. In the experiments, the computation efficiency of the proposed SPF-GCN was higher than that of the state-of-the-art methods (the details of the computation efficiency are in Section 4.2).

- The structure showed the distribution of the data and was very important for training the GCN model. The incomplete structure could evaluate the robustness of the related GCN model. We selected the classical GCN, the state-of-the-art DGCN, SF-GCN, and SPF-GCN for the robustness test. The proposed SPF-GCN showed the best performance on the three datasets. On Cora, the performance of GCN was better than DGCN, while the performance of GCN was worse than DGCN on Citeseer and PubMed. This showed that local and global consistency for fusing graph information in DGCN tended toward the unstable characteristic because of the tight constraint of incomplete structure consistency. The loose constraint of GCN for incomplete structure correlation led to the worse performance. The proposed SPF-GCN could compromise these constraints for balancing the incomplete structure information by optimizing the weight of multiple structures and also connect the different structures for complementing the different information. Therefore, the proposed SPF-GCN obtained the best performance in the experiments.
- The proposed SPF-GCN was expected to mine the commonality and the specificity of multiple structures. The commonality described the similarity characteristic of structures by the Grassmann manifold metric, while the specificity narrated the different characteristics of the structures by spectral embedding. In the proposed method, the commonality was constructed based on the specificity. Therefore, we only executed the ablation experiment for preserving the specificity loss by deleting the commonality loss from the total loss. This experiment obtained the following performance: 82.6% on Cora, 71.5% on Citeseer, and 78.9% on PubMed. These results obviously were worse than the performance of the proposed SF-GCN and SPF-GCN, which could balance the commonality and specificity for mining the suited weight of multiple structures.

6. Conclusions

We proposed structure fusion based on graph convolutional networks (SF-GCN) to address the multi-view data diversity and complexity for node classification. SF-GCN could not only adapt spectral embedding to preserve the specificity of the structure, but also model the relationship of the different structures to find the commonality of multiple structures by the manifold metric. Furthermore, the proposed structure propagation fusion based on graph convolutional networks (SPF-GCN) could combine the structure fusion framework with structure propagation to generate the complete structure graph for improving the performance of node classification. Finally, the optimization learning of the SF-GCN could obtain both the suitable weight for the different structure and merge the embedding space. For evaluating the proposed SF-GCN and SPF-GCN, we carried out the comparison experiments about the baseline methods, the different multi-graph fusion methods, the state-of-the-art methods, and the the lost structure analysis on the Cora, Citeseer, and Pubmed datasets. The experiment results demonstrated that SF-GCN and SPF-GCN could obtain promising results for node classification on citation networks.

From the above, it was shown that structure information mining was important for recognizing the data category. Especially, multi-graphs based on the observation data could provide rich information for reconstructing the intrinsic structure of data. It could be used for social networks' information analysis, knowledge networks' information construction, and virus networks' discovery and discrimination. Since structure information mining is so important for structure recovery, the structure may be directly learned from the original data in the future works.

Author Contributions: Conceptualization, G.L.; investigation, G.L., J.W., and F.Z.; methodology, G.L., K.L., and W.C.; project administration, G.L.; writing, original draft, G.L.; writing, review and editing, G.L. All authors read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Veksler, O. Efficient Graph Cut Optimization for Full CRFs with Quantized Edges. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. doi:10.1109/TPAMI.2019.2906204. [CrossRef]
2. Monti, F.; Bronstein, M.; Bresson, X. Geometric matrix completion with recurrent multi-graph neural networks. *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2017; pp. 3697–3707.
3. Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W.L.; Leskovec, J. Graph convolutional neural networks for web-scale recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, UK, 19–23 August 2018; ACM: New York, NY, USA, 2018; pp. 974–983.
4. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–8 December 2016; Curran Associates, Inc.: New York, NY, USA, 2016; pp. 3844–3852.
5. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural message passing for quantum chemistry. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1263–1272.
6. Khan, M.R.; Blumenstock, J.E. Multi-GCN: Graph Convolutional Networks for Multi-View Networks, with Applications to Global Poverty. *arXiv* **2019**, arXiv:1901.11213.
7. Lin, G.; Zhu, H.; Kang, X.; Fan, C.; Zhang, E. Multi-feature structure fusion of contours for unsupervised shape classification. *Pattern Recognit. Lett.* **2013**, *34*, 1286–1290. [CrossRef]
8. Lin, G.; Zhu, H.; Kang, X.; Fan, C.; Zhang, E. Feature structure fusion and its application. *Inf. Fusion* **2014**, *20*, 146–154. [CrossRef]
9. Lin, G.; Zhu, H.; Kang, X.; Miu, Y.; Zhang, E. Feature structure fusion modelling for classification. *IET Image Process.* **2015**, *9*, 883–888. [CrossRef]
10. Lin, G.; Fan, G.; Yu, L.; Kang, X.; Zhang, E. Heterogeneous structure fusion for Target Recognition in infrared imagery. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015; pp. 118–125.
11. Lin, G.; Fan, G.; Kang, X.; Zhang, E.; Yu, L. Heterogeneous feature structure fusion for classification. *Pattern Recognit.* **2016**, *53*, 1–11. [CrossRef]
12. Lin, G.; Fan, C.; Zhu, H.; Miu, Y.; Kang, X. Visual feature coding based on heterogeneous structure fusion for image classification. *Inf. Fusion* **2017**, *36*, 275–283. [CrossRef]
13. Lin, G.; Liao, K.; Sun, B.; Chen, Y.; Zhao, F. Dynamic graph fusion label propagation for semi-supervised multi-modality classification. *Pattern Recognit.* **2017**, *68*, 14–23. [CrossRef]
14. Lin, G.; Chen, Y.; Zhao, F. Structure propagation for zero-shot learning. *arXiv* **2017**, arXiv:1711.09513.
15. Lin, G.; Fan, C.; Chen, W.; Chen, Y.; Zhao, F. Class label autoencoder for zero-shot learning. *arXiv* **2018**, arXiv:1801.08301.
16. Lin, G.; Chen, Y.; Zhao, F. Structure Fusion and Propagation for Zero-Shot Learning. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Guangzhou, China, 23–26 November 2018; pp. 465–477.
17. Lin, G.; Chen, W.; Liao, K.; Kang, X.; Fan, C. Transfer feature generating networks with semantic classes structure for zero-shot learning. *arXiv* **2019**, arXiv:1903.02204.
18. Yang, X.; Latecki, L.J. Affinity learning on a tensor product graph with applications to shape and image retrieval. In Proceedings of the IEEE CVPR 2011, Providence, RI, USA, 20–25 June 2011; pp. 2369–2376.
19. Yang, X.; Prasad, L.; Latecki, L.J. Affinity learning with diffusion on tensor product graph. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 28–38. [CrossRef] [PubMed]
20. Bai, S.; Zhou, Z.; Wang, J.; Bai, X.; Latecki, L.J.; Tian, Q. Automatic Ensemble Diffusion for 3D Shape and Image Retrieval. *IEEE Trans. Image Process.* **2019**, *28*, 88–101. [CrossRef] [PubMed]
21. Li, Q.; An, S.; Li, L.; Liu, W. Semi-supervised Learning on Graph with an Alternating Diffusion Process. *arXiv* **2019**, arXiv:1902.06105.
22. Bai, S.; Bai, X.; Tian, Q.; Latecki, L.J. Regularized diffusion process for visual retrieval. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; AAAI Press: Palo Alto, CA, USA 2017; pp. 3967–3973.

23. Bai, S.; Zhou, Z.; Wang, J.; Bai, X.; Jan Latecki, L.; Tian, Q. Ensemble diffusion for retrieval. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 774–783.

24. Gao, H.; Wang, Z.; Ji, S. Large-Scale Learnable Graph Convolutional Networks. *arXiv* **2018**, arXiv:1808.03965.

25. Zhuang, C.; Ma, Q. Dual graph convolutional networks for graph-based semi-supervised classification. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, Lyon, France, 23–27 April 2018; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2018; pp. 499–508.

26. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2013**, arXiv:1312.6203.

27. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.

28. Chen, J.; Ma, T.; Xiao, C. Fastgcn: Fast learning with graph convolutional networks via importance sampling. *arXiv* **2018**, arXiv:1801.10247.

29. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive representation learning on large graphs. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: New York, NY, USA, 2017; pp. 1024–1034.

30. Dai, H.; Kozareva, Z.; Dai, B.; Smola, A.; Song, L. Learning steady-states of iterative algorithms over graphs. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1114–1122.

31. Liu, Z.; Chen, C.; Li, L.; Zhou, J.; Li, X.; Song, L.; Qi, Y. Geniepath: Graph neural networks with adaptive receptive paths. *arXiv* **2018**, arXiv:1802.00910.

32. Van Tran, D.; Navarin, N.; Sperduti, A. On Filter Size in Graph Convolutional Networks. *arXiv* **2018**, arXiv:1811.10435.

33. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2017**, arXiv:1710.10903.

34. Chen, J.; Zhu, J.; Song, L. Stochastic Training of Graph Convolutional Networks with Variance Reduction. *arXiv* **2017**, arXiv:1710.10568.

35. Veličković, P.; Fedus, W.; Hamilton, W.L.; Liò, P.; Bengio, Y.; Hjelm, R.D. Deep Graph Infomax. *arXiv* **2018**, arXiv:1809.10341.

36. Xia, T.; Tao, D.; Mei, T.; Zhang, Y. Multiview spectral embedding. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2010**, *40*, 1438–1446.

37. Lin, G.F.; Zhu, H.; Fan, C.X.; Zhang, E.H.; Luo, L. Multi-cluster Feature Selection Based on Grassmann Manifold. *Jisuanji Gongcheng/Comput. Eng.* **2012**, *38*, 178–181.

38. Turaga, P.; Veeraraghavan, A.; Srivastava, A.; Chellappa, R. Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2273–2286. [CrossRef]

39. Dong, X.; Frossard, P.; Vandergheynst, P.; Nefedov, N. Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds. *IEEE Trans. Signal Process.* **2013**, *62*, 905–918. [CrossRef]

40. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A comprehensive survey on graph neural networks. *arXiv* **2019**, arXiv:1901.00596.