The Devil is in the Conflict: Disentangled Information Graph Neural Networks for Fraud Detection

Zhixun Li[1,†], Dingshuo Chen[2,3,†], Qiang Liu[2,3], Shu Wu[2,3,*]
[1]*School of Computer Science and Technology, Beijing Institute of Technology*
[2]*Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences*
[3]*School of Artificial Intelligence, University of Chinese Academy of Sciences*
lizhixun@bit.edu.cn, dingshuo.chen@cripac.ia.ac.cn, {qiang.liu, shu.wu}@nlpr.ia.ac.cn

2022 IEEE International Conference on Data Mining (ICDM) | 978-1-6654-5099-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICDM54844.2022.00131

***Abstract*—Graph-based fraud detection has heretofore received considerable attention. Owning to the great success of Graph Neural Networks (GNNs), many approaches adopting GNNs for fraud detection has been gaining momentum. However, most existing methods are based on the strong inductive bias of homophily, which indicates that the context neighbors tend to have same labels or similar features. In real scenarios, fraudsters often engage in camouflage behaviors in order to avoid detection system. Therefore, the homophilic assumption no longer holds, which is known as the inconsistency problem. In this paper, we argue that the performance degradation is mainly attributed to the inconsistency between topology and attribute. To address this problem, we propose to disentangle the fraud network into two views, each corresponding to topology and attribute respectively. Then we propose a simple and effective method that uses the attention mechanism to adaptively fuse two views which captures data-specific preference. In addition, we further improve it by introducing mutual information constraints for topology and attribute. To this end, we propose a Disentangled Information Graph Neural Network (DIGNN) model, which utilizes variational bounds to find an approximate solution to our proposed optimization objective function. Extensive experiments demonstrate that our model can significantly outperform state-of-the-art baselines on real-world fraud detection datasets.**

***Index Terms*—Graph Neural Networks, Fraud Detection, Information Theory**

I. Introduction

Graph-based fraud detection is a crucial task and has tremendous impact in various applications, such as opinion fraud detection [1], fake news detection [2], review spams [3] and financial fraud detection [4], [5]. In these scenarios, as graph can effectively model the correlations among entities, interactive activities on platform can be characterized as a graph, where users or objects are often treated as nodes, and transactions or relations between them are treated as edges.

Numerous techniques have been proposed to detect the fraudsters. Recently, driven by the powerful representation capability of graph structure and advances of Graph Neural Networks (GNNs) [6]–[8], many approaches try to harness GNNs for fraud detection on either homogeneous or heterogeneous graphs. The main idea is to leverage GNNs to learn expressive node representations with the goal of distinguishing abnormal nodes from the normal ones in the latent embedding space. Message-Passing GNNs (MP-GNNs) are mainstreaming in recent years, which aggregate neighbor node features and achieve local smoothing by stacking layers. Although MP-GNNs can obtain satisfactory performance on most of cases, the strong inductive bias of homophily limits their representative ability on heterophilic graphs. Some works [9] point out that plentiful GNNs can be seen as low-pass filters, so their generalization ability on high frequency graph signals are poor. In fraud detection task, fraudsters often imitate normal users in order to camouflage themselves, hence they will interact with normal users more frequently. For instance, normal users account for 81% of the fraudsters' neighbor nodes in YelpChi dataset (Figure 1). In other words, fraudsters' features are inconsistent with their behaviors (interactions, e.g., topological structure). Thus, recall that MP-GNNs do not work well on heterophilic graphs, they fail to tackle the inconsistency phenomenon in graph-based fraud detection and fraudsters could fool the detection system.

Recently, a few works have noticed this problem, and they employ aggregating weights to reduce the adverse impact of dissimilar neighbors, or set similarity-aware thresholds to select and re-link similar nodes. For instance, GraphConsis [10] computes consistent score between connected node pairs as the sampling probability. PC-GNN [5] combines label information and latent embeddings as distance function to measure similarity. Although such methods can alleviate the inconsistency problem in some extent, they discard a lot of information during filting dissimilar neighbors out, thus they may lead to sub-optimal performance.

In this paper, we analyze the inconsistency problem in graph-based fraud detection task, which has been obstructing a full understanding of this field. First, we clarify that the inconsistency problem is the bottleneck of graph fraud detection. According to [11], the underlying optimization process of GNNs is equivalent with minimizing the topology

[†]The first two authors contributed equally to this work.
[*]Corresponding author.

978-1-6654-5099-7/22/$31.00 ©2022 IEEE
DOI 10.1109/ICDM54844.2022.00131

and attribute constraints, and Yang et al. [12] indicates that the degradation of performance is imputed to the compromise between topology and attribute. Due to the camouflage behaviors (topology) of fraudsters, which are inconsistent with their essence (attribute), this conflict in fraud networks may injure the discriminative ability of GNNs. Second, the forefronts of different datasets are diverse, and most existing methods are not satisfactory in fusing topological structures and node attributes [13]. For example, fraudsters may possess distinguishable attribute on some platforms, but their deceptive behaviors can confuse the detection model. Therefore, we are motivated to explore a novel method that is able to minimize the conflict between topology and attribute and meanwhile effectively extract most task-relevant information from datasets.

We borrow the concept of multi-view learning problems to graph-based fraud detection task and propose a simple and effective model, **D**isentangled **I**nformation **G**raph **N**eural **N**etwork (DIGNN). Technically, we first disentangle fraud networks into topology and attribute views. Next, we employ attention mechanism to fuse two view embeddings adaptively for extracting task-relevant information. Surprisingly, we observe that this simple method surpasses all state-of-the-art baselines. This empirically proves that the conflict between topology and attribute causes the inconsistency problem. Besides, to further decrease the entanglement between topology and attribute and improve the performance, we design a new optimization objective based on information theory, which resorts to variational bounds to minimize mutual information between two views and maximize the mutual information between view embeddings and original inputs.

We conduct extensive experiments to compare our proposed model with existing graph-based fraud detection models, the results demonstrate the effectiveness of our model. In summary, the contributions of this paper can be summarized as follows:

- We analyze the cause of the inconsistency problem, and point out that it is mainly attributed to the conflict between topology and attribute. In light of this, we propose a simple yet effective model, DIGNN, which firstly disentangles fraud network into two views and fuses them by attention mechanism.
- We propose a novel optimization objective based on mutual information theory and theoretically derive its upper bound for tractable calculation.
- We verify the effectiveness of our model on real-world fraud detection datasets. It is shown that our model is able to significantly improve the performance in terms of all commonly adopted metrics.

II. Related Work

A. Graph-based Fraud Detection

The core idea of graph-based fraud detection task is taking the advantages of GNNs to get the discriminative node embeddings, and find out the malicious ones in the latent space.

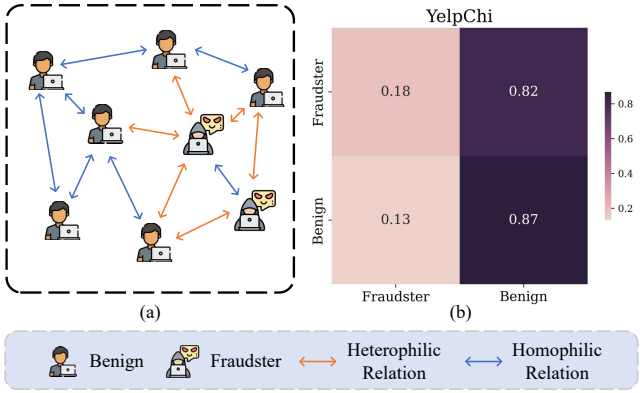

Fig. 1. (a) Illustration of graph-based fraud detection. (b) Neighbor distribution of fraudsters and benign users in YelpChi dataset.

Examples include [10], [14], [15] for review fraud detection, [2] for fake news detection and [4], [5], [16]–[18] for financial fraud detection. Ma et al. [19] provides a comprehensive investigation on graph-based fraud detection.

Most of existing GNNs methods holds homophilic assumption that neighbor nodes share same labels or similar features. However, fraudsters will try to conceal themselves, so that their features are inconsistent with their camouflage behaviors. Some graph-based fraud detection works have noticed this problem. GraphConsis [10] pioneers to formulate and tackle the inconsistency problem. They introduce three kinds of inconsistency phenomenon existing in fraud networks. CARE-GNN [14] devises a label-aware similarity measure to find informative neighboring nodes and utilizes reinforcement learning to select similar neighbors. FRAUDRE [20] aggregates difference between adjacent node pairs. PC-GNN [5] devises a choose operation to select beneficial neighbors based on feature similarity.

B. Multi-view on GNNs

Topology and attribute are two essential compositions of graphs. However existing state-of-the-art GNN models are disable to effectively fuse topological structure and node attributes. AM-GCN [13] uses k-nearest neighbor to construct feature graph and combine it with topological structure view and common embeddings. SCRL [21] designs a self-supervised approach to maximize the agreement of the embeddings in the topology graph and the feature graph. LINKX [22] processes node attributes and topological structure in an orthogonal manner. In this paper, we also follow this idea and extend it by proposing a novel architecture and optimization objective.

Information-theoretic methods have been gaining momentum in recent years, which take into consideration the mutual dependency of different views. MIB [23] extends the information bottleneck principle to unsupervised multi-view setting to discard superfluous information. DVIB [24] and CMIB [25] leverage mutual information constrains to better preserve shared and private information of multi-view learning. To

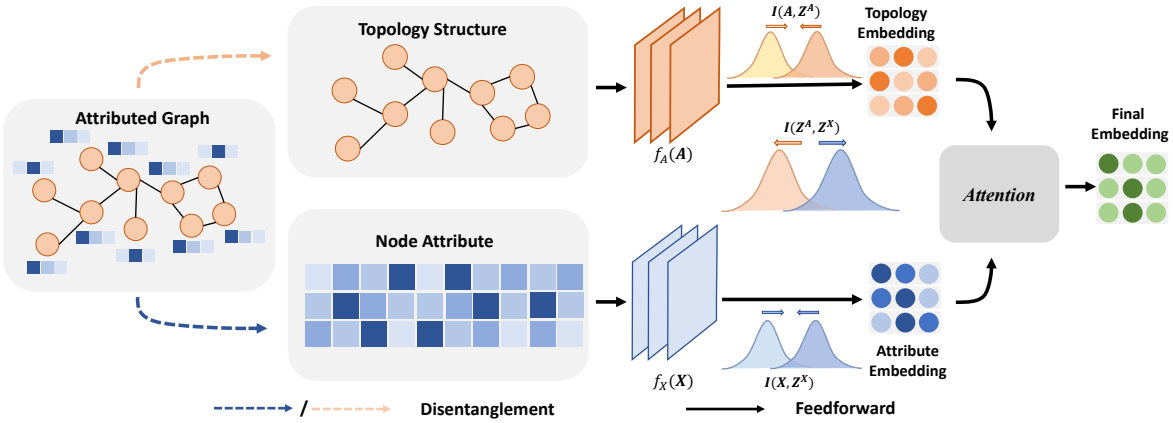

Fig. 2. The overview of our proposed DIGNN. The attributed fraud network is disentangled into topological structure and node attributes. DIGNN processes these two views in parallel and fuses them by attention mechanism. In addition, DIGNN minimizes the mutual information between topology embeddings and attribute embeddings, and maximizes the mutual information between embeddings and input data respectively.

cope with intractable computation of mutual information, these methods adopt variational inference to optimize objective lower and upper bounds.

III. PRELIMINARIES

Graph-based Fraud Detection. Given a fraud network $\mathcal{G} = (\mathcal{V}, \mathbf{A}, \mathbf{X})$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ is the set of nodes; $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the adjacency matrix, if v_i and v_j are connected, $\mathbf{A}_{ij} = 1$, otherwise, $\mathbf{A}_{ij} = 0$; $\mathbf{X} \in \mathbb{R}^{N\times D}$ denotes node feature matrix, each node v_i is associated with a D-dimensional feature vector $\mathbf{x}_i$ and a label $y_i \in \{0, 1\}$, where 0 denotes the node is a normal user (negative) and 1 indicates it is a fraudster (positive). The core idea of graph-based fraud detection is to learn discriminative node embeddings to detect the anomaly samples in latent space.

IV. METHOD

In this section, we will present our model, **D**isentangled **I**nformation **G**raph **N**eural **N**etwork (DIGNN). Figure 2 gives an overview of our model. It consists of three main objectives: 1. Disentangle attribute fraud network into topology and attribute views and fuse them by attention mechanism; 2. To further reduce the conflict between two views, we minimize the mutual information between them; 3. In order to maintain the semantic information from input space, we maximize the mutual information between view-specific embeddings and their original inputs.

A. View-specific Embedding

It is universally acknowledged that topology and attribute are of vital importance for graph learning. However, in graph fraud detection scenario, traditional message passing along neighboring nodes is inappropriate as graph signal smoothing makes fraudsters more indistinguishable. To alleviate the inconsistency problem, we disentangle the topology and attribute information and encode them in parallel.

Given an attributed fraud network $\mathcal{G}$, it can be disentangled into topology view $\mathbf{A}$ and attribute view $\mathbf{X}$. Here we provide two view encoders f_A, f_X for each input view, as shown in Figure 2. Specifically, we employ Multi-Layer Perceptron (MLP) as encoders to obtain view-specific embeddings $\mathbf{Z}^A, \mathbf{Z}^X \in \mathbb{R}^{N\times d}$:

$$\mathbf{Z}^A = f_A(\mathbf{A}),\ \mathbf{Z}^X = f_X(\mathbf{X}), \tag{1}$$

in which d is the embedding dimension. With these two embeddings, we need to fuse them to obtain final representation and extract task-relevant information.

B. Cross-view Fusion

Now we have two view-specific embeddings $\mathbf{Z}^A$ and $\mathbf{Z}^X$, we then perform cross-view fusion by utilizing attention mechanism. The attention value ω_i can be represented as:

$$\omega_i^j = q \cdot \tanh(\mathbf{W} \cdot (\mathbf{z}_i^j)^\top + b),\ j \in \{A, X\} \tag{2}$$

where q denotes the learnable attention vector, $\mathbf{W}$ is the weight matrix and b is bias vector. Thus, we can get the attention values ω_i^A and ω_i^X for view-specific embeddings $\mathbf{z}_i^A$ and $\mathbf{z}_i^X$, respectively. Then we normalize them via softmax function to get the final weight:

$$\alpha_i^j = \text{softmax}(\omega_i^j) = \frac{\exp(\omega_i^j)}{\exp(\omega_i^A) + \exp(\omega_i^X)},\ j \in \{A, X\} \tag{3}$$

Larger attention weight α_i implies that the corresponding embeddings is more important, and it is determined by specific dataset. Then the final output embedding $\mathbf{z}_i$ can be combined by two view-specific embeddings with its corresponding attention weight as:

$$\mathbf{z}_i = \alpha_i^A \cdot \mathbf{z}_i^A + \alpha_i^X \cdot \mathbf{z}_i^X. \tag{4}$$

And we put it into a linear classifier, while training by a cross-entropy loss function:

$$\mathcal{L}_{ce} = \sum_{v_i \in \mathcal{V}_{\text{train}}} -\log(y_i \cdot \sigma(\mathbf{W}' \cdot \mathbf{z}_i + b')) \tag{5}$$

in which $\mathbf{W}'$ and b' is the weight matrix and bias vector of linear classifier, σ is a softmax function, and $\mathcal{V}_{\text{train}}$ is the training node set.

C. Mutual Information Optimization

Up to now, we have discussed how to get view-specific embeddings and fuse them with attention mechanism. However, as mentioned in [12], the representative GNN models tend to deteriorate their expressive power due to interference between attribute and topology. By leveraging the information theory, we propose a novel optimization objective to alleviate the aforementioned problem. Furthermore, we derive the variational bound of our optimization objective and discuss the intrinsic effect and intuitive insight. Without loss of generality, we let $\mathbf{X}_1, \mathbf{X}_2$ to represent original views and $\mathbf{Z}_1, \mathbf{Z}_2$ to represent view-specific embeddings for ease of reading.

Optimization principles The first principle aims to induce model to learn mutual-exclusive embeddings, which ameliorates the compromise problem between attribute and topology. Considering that mutual information measures the mutual dependence of variables, we introduce the constraint term **min** $I(\mathbf{Z}_1, \mathbf{Z}_2)$ to our optimization objective. In this way, model is able to reduce the redundancy and enhance the ability on exploiting sufficient semantic information in embedding space with limited dimensionality.

Nevertheless, mutual-exclusive constraint is prone to impair the helpful shared information. For instance, in Amazon dataset, handcrafted features are highly correlated to social networks (topology), thus mutual-exclusive constraint will injure attribute semantics during training. The second principle builds the relationship between view-specific embeddings and their original inputs. In virtue of rich but distinct semantics inherent in the attribute and topology, it is necessary to extract useful features and meanwhile maintain respective information from input data space. We further introduce the constraint term **max** $I(\mathbf{Z}_i, \mathbf{X}_i)$ to our optimization objective to encode inputs with more view-specific information available. To sum up, our mutual information optimization objective can be summarized as follow:

$$\min\ I(\mathbf{Z}_1, \mathbf{Z}_2) - \sum_{i=1}^{2} I(\mathbf{Z}_i, \mathbf{X}_i) \tag{6}$$

We further make theoretical analysis to derive the lower bound of $I(\mathbf{Z}, \mathbf{X})$ and upper bound of $I(\mathbf{Z}_1, \mathbf{Z}_2)$ for tractable optimization objective, which is dubbed $\mathcal{L}_{rec}$ and $\mathcal{L}_{exc}$ respectively.

$$\mathcal{L}_{rec} = \sum_{i=1}^{2} \mathbb{E}_{p(x_i, z_i)} \log\ q(x_i|z_i) \tag{7}$$

$$\mathcal{L}_{exc} = \frac{1}{2}\left[\mathbb{E}_{p(z_1,x_2)} \log \frac{p_{x_2}(z_2|x_2)}{r(z_1)} + \mathbb{E}_{p(z_2,x_1)} \log \frac{p_{x_1}(z_1|x_1)}{r(z_2)}\right] \tag{8}$$

where $q(x_i|z_i)$ is the variational approximation of conditional distribution $p(x|z)$, and p_{x_1} and p_{x_2} represent encoders that encode information from original feature space. The bound will become tighter as the marginal distribution $r(z)$ approaches the priors $p(z)$. Eventually, the overall optimization objective is formulated as follow

$$\mathcal{L} = \mathcal{L}_{ce} + \alpha \cdot \mathcal{L}_{rec} + \beta \cdot \mathcal{L}_{exc} \tag{9}$$

where α and β are scalar factors. Moreover, it is worth noting that the second term reconstruction loss is equivalent to graph signal denoising but without signal smoothness, which is reasonable considering the inconsistency problem of graph anomaly detection. Intuitively, our loss function denoises the original graph signal and achieves mutual exclusion between attribute and topology together with supervised information.

V. Experiments

A. Experiment Setup

1) Datasets: Our proposed DIGNN model is evaluated on two real-world opinion fraud network datasets: YelpChi [26] and Amazon [27].

2) Baselines: We compare with several representative state-of-the-art models to verify the effectiveness of DIGNN in graph-based fraud detection. GCN [6], GAT [8], GraphSAGE [7], DR-GCN [28], CARE-GNN [14], FRAUDRE [20], PC-GNN [5]. And we also add two variants of DIGNN for ablation study.

3) Settings: The parameters of DIGNN are optimized with Adam optimizer, the train, valid, and test ratio are set to be 40%, 20%, and 40% respectively. We use Scikit-learn to implement train-test split, and the imbalance ratio is consistent in three sets. It is worth noting that to alleviate the influence of class imbalance, we employ down-sampling or re-weighting to train DIGNN.

For GCN, GAT, and GraphSage, they suffer from the class imbalance and inconsistency problem, and will always predict normal (negative) samples. Therefor, we follow PC-GNN to utilize threshold-moving strategy, and the classification threshold is set to be 0.2 for YelpChi and Amazon. For CARE-GNN, FRAUDRE, PC-GNN, we use the parameters introduced by authors.

4) Metrics: The fraud detection datasets display a skewed class distribution, so accuracy is not suitable to evaluate the effectiveness of fraud detection models. The evaluation metrics should have no bias to any class. Therefore, we use three common metrics, namely **F1-macro**, **AUC** and **GMean**.

B. Analysis of Attention Mechanism (RQ1)

We analyze the attention values and visualize them for investigating whether the attention values learned by our model is meaningful. The attention changing trends are shown in Figure 3. The x-axis is the number of training epochs and y-axis is the average attention value. With the training epoch increasing, the difference between the corresponding attention values of topology and attribute begin to be striking. We can observe that DIGNN pays more attention on attribute and topology on YelpChi and Amazon datasets respectively. It demonstrates our model has a strong capability to extract the task-relevant information from these two views.

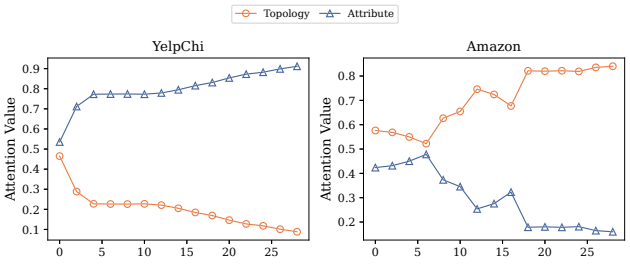

Fig. 3. The attention changing trends w.r.t epochs.

TABLE I
PERFORMANCE COMPARISON ON YELPCHI AND AMAZON.

Method	Dataset	Yelpchi			Amazon		
	Metric	F1-macro	AUC	GMean	F1-macro	AUC	GMean
Baselines	GCN	0.4929	0.6274	0.1886	0.5461	0.8328	0.2570
	GAT	0.4879	0.5715	0.1659	0.6464	0.8102	0.6675
	GraphSAGE	0.4405	0.5439	0.2589	0.6416	0.7589	0.5949
	DR-GCN	0.5523	0.5921	0.4038	0.6488	0.8295	0.5357
	CARE-GNN	0.6075	0.7713	0.7023	0.8875	0.9398	0.8848
	FRAUDRE	0.5841	0.7427	0.6654	0.8806	0.9272	0.8808
	PC-GNN	0.6130	0.7715	0.7068	0.8557	0.9482	0.8952
Ablation	DIGNN$_{\setminus S}$	0.5120	0.6120	0.5895	0.7308	0.8913	0.8088
	DIGNN$_{\setminus M}$	0.6994	0.8389	0.7348	0.9186	0.9645	0.9195
Ours	DIGNN	**0.7092**	**0.8526**	**0.7596**	**0.9189**	**0.9729**	**0.9281**

C. Performance Comparison (RQ2)

We compare the performance of DIGNN with state-of-the-art methods. The corresponding F1-macro, AUC and GMean scores are shown in Table I, we have the following two observations.

First, DIGNN significantly boosts the performance for all metrics on YelpChi and Amazon datasets than other SOTA baseline methods. We can observe that PC-GNN outperforms other baselines in most metrics, but our model can still surpass it by a significant margin. In Amazon dataset, graph-based fraud detection methods have already achieved high performance and the increasing room is limited. But our model can still get appreciable improvements.

Second, the compared baseline methods can be divided into two groups, traditional MP-GNNs and graph-based fraud detection methods. GCN, GAT, GraphSAGE are tradition GNN models, and DR-GCN is designed for imbalanced node classes. They do not consider the inconsistency problem so that we can observe these models get poor performance on YelpChi and Amazon datasets. CARE-GNN and PC-GNN are graph-based fraud detection methods, they both sample neighbors according to similarity measure, which can alleviate inconsistency problem to a certain degree. Therefore, they can perform better on these two datasets.

In general, DIGNN outperforms all baselines in F1-macro, AUC and GMean on YelpChi and Amazon datasets, which can demonstrate the effectiveness of our model.

D. Ablation Study (RQ3)

We compare DIGNN with two corresponding variants DIGNN$_{\setminus S}$ and DIGNN$_{\setminus M}$ to figure out how do different components of DIGNN contribute to performance improvement. The results of two datasets are shown in Table I. We can observe that DIGNN surpasses its variants in most of metrics. For DIGNN$_{\setminus M}$, its overall performance on Yelpchi and Amazon is inferior to complete model, which verifies the effectiveness of our proposed mutual information objective. For DIGNN$_{\setminus S}$, we can observe that DIGNN is evidently better than model without sampling strategy. We suppose it is caused by the noise information of the structure view. Sampling strategy plays a denoising effect on structural information to some extent.

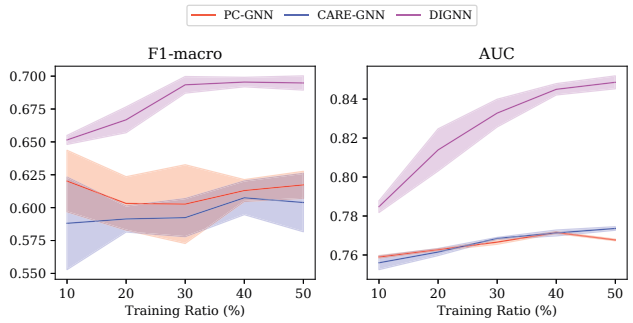

Fig. 4. Sensitivity analysis with respect to different training ratio on YelpChi dataset. The solid line represents the average score of 3 runs and the shadow indicates the standard deviation.

E. Sensitive Analysis (RQ4)

We further evaluate the performance of DIGNN with respect to the training ratio and hyperparameters α, β. For training ratio, we vary the percentage of training nodes from 10% to 50%, and compare DIGNN with other two baselines, CARE-GNN and PC-GNN. Figure 4 shows the performance of F1-macro and AUC on YelpChi dataset. We can observe that DIGNN always achieves best performance among the three models. When the training ratio is 10%, DIGNN still performs better than PC-GNN training on 50% samples. And DIGNN surpasses CARE-GNN and PC-GNN by a large margin in AUC. The result on GMean also presents similar tendency, but in order to save space, we won't show it.

For hyperparameters α and β, we vary these two hyper-parameters from 0 to 1, and the corresponding results are shown in Figure 5. Considering the limit space, we only present AUC performance on YelpChi and Amazon datasets. It can be observed that the optimal selection of these two hyper-parameters varies greatly on the different datasets. In the YelpChi dataset, higher AUC performance can be achieved by selecting larger β ($\beta \geq 0.8$). And in the Amazon dataset, larger α ($\alpha \geq 0.6$) and smaller β ($\beta \leq 0.4$) can get a better result.

VI. CONCLUSION

In this paper, we suggest that disentangling operation is beneficial to alleviate the inconsistency problem in fraud network. In order to decrease the conflict between topological

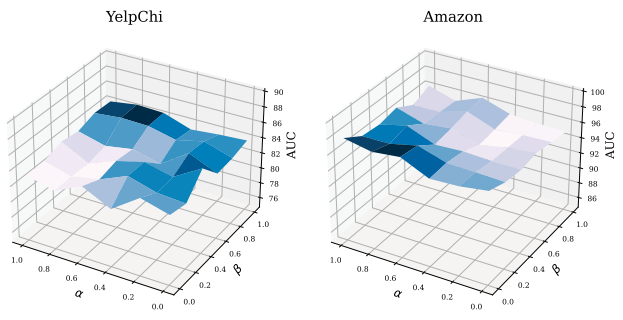

Fig. 5. AUC as the two hyper-parameters α and β varying from 0 to 1.

structure and node attribute, we propose a simple yet effective model named DIGNN. It firstly disentangles the attribute fraud network into topology and attribute two views. Then DIGNN fuses two kinds of view information adaptively by attention mechanism, which can effectively extract task-relevant information. Moreover, we design a novel optimization objective to further reduce the entanglement between these two view-specific embeddings and maintain their semantic information. Experiment results demonstrate that DIGNN outperforms state-of-the-art methods on two real-world graph fraud detection datasets.

VII. ACKNOWLEDGE

This work is jointly sponsored by National Natural Science Foundation of China (U19B2038, 62141608, 62206291) and CCF-AFSG Research Fund (20210001).

REFERENCES

[1] A. Li, Z. Qin, R. Liu, Y. Yang, and D. Li, "Spam review detection with graph convolutional networks," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 2703–2711.

[2] W. Xu, J. Wu, Q. Liu, S. Wu, and L. Wang, "Mining fine-grained semantics via graph neural networks for evidence-based fake news detection," *arXiv preprint arXiv:2201.06885*, 2022.

[3] L. Deng, C. Wu, D. Lian, Y. Wu, and E. Chen, "Markov-driven graph convolutional networksfor social spammer detection," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[4] D. Wang, J. Lin, P. Cui, Q. Jia, Z. Wang, Y. Fang, Q. Yu, J. Zhou, S. Yang, and Y. Qi, "A semi-supervised graph attentive network for financial fraud detection," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 598–607.

[5] Y. Liu, X. Ao, Z. Qin, J. Chi, J. Feng, H. Yang, and Q. He, "Pick and choose: a gnn-based imbalanced learning approach for fraud detection," in *Proceedings of the Web Conference 2021*, 2021, pp. 3168–3177.

[6] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[7] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[8] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *stat*, vol. 1050, p. 20, 2017.

[9] H. Nt and T. Maehara, "Revisiting graph neural networks: All we have is low-pass filters," *arXiv preprint arXiv:1905.09550*, 2019.

[10] Z. Liu, Y. Dou, P. S. Yu, Y. Deng, and H. Peng, "Alleviating the inconsistency problem of applying graph neural network to fraud detection," in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 2020, pp. 1569–1572.

[11] Y. Ma, X. Liu, T. Zhao, Y. Liu, J. Tang, and N. Shah, "A unified view on graph neural networks as graph signal denoising," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 1202–1211.

[12] L. Yang, W. Zhou, W. Peng, B. Niu, J. Gu, C. Wang, X. Cao, and D. He, "Graph neural networks beyond compromise between attribute and topology," 2022.

[13] X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, and J. Pei, "Am-gcn: Adaptive multi-channel graph convolutional networks," in *Proceedings of the 26th ACM SIGKDD International conference on knowledge discovery & data mining*, 2020, pp. 1243–1253.

[14] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, "Enhancing graph neural network-based fraud detectors against camouflaged fraudsters," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 315–324.

[15] Y. Wang, J. Zhang, S. Guo, H. Yin, C. Li, and H. Chen, "Decoupling representation learning and classification for gnn-based anomaly detection," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 1239–1248.

[16] X. Ao, Y. Liu, Z. Qin, Y. Sun, and Q. He, "Temporal high-order proximity aware behavior analysis on ethereum," *World Wide Web*, vol. 24, no. 5, pp. 1565–1585, 2021.

[17] T. Liang, G. Zeng, Q. Zhong, J. Chi, J. Feng, X. Ao, and J. Tang, "Credit risk and limits forecasting in e-commerce consumer lending service via multi-view-aware mixture-of-experts nets," in *Proceedings of the 14th ACM international conference on web search and data mining*, 2021, pp. 229–237.

[18] G. Zhang, Z. Li, J. Huang, J. Wu, C. Zhou, J. Yang, and J. Gao, "efraudcom: An e-commerce fraud detection system via competitive graph neural networks," *ACM Transactions on Information Systems (TOIS)*, vol. 40, no. 3, pp. 1–29, 2022.

[19] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[20] G. Zhang, J. Wu, J. Yang, A. Beheshti, S. Xue, C. Zhou, and Q. Z. Sheng, "Fraudre: Fraud detection dual-resistant to graph inconsistency and imbalance," in *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2021, pp. 867–876.

[21] C. Liu, L. Wen, Z. Kang, G. Luo, and L. Tian, "Self-supervised consensus representation learning for attributed graph," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 2654–2662.

[22] D. Lim, F. Hohne, X. Li, S. L. Huang, V. Gupta, O. Bhalerao, and S. N. Lim, "Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[23] M. Federici, A. Dutta, P. Forré, N. Kushman, and Z. Akata, "Learning robust representations via multi-view information bottleneck," *arXiv preprint arXiv:2002.07017*, 2020.

[24] F. Bao, "Disentangled variational information bottleneck for multiview representation learning," in *CAAI International Conference on Artificial Intelligence*. Springer, 2021, pp. 91–102.

[25] Z. Wan, C. Zhang, P. Zhu, and Q. Hu, "Multi-view information-bottleneck representation learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 11, 2021, pp. 10 085–10 092.

[26] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, 2015, pp. 985–994.

[27] J. J. McAuley and J. Leskovec, "From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 897–908.

[28] M. Shi, Y. Tang, X. Zhu, D. Wilson, and J. Liu, "Multi-class imbalanced graph convolutional network learning," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, 2020.