Loan Default Analysis with Multiplex Graph Learning

Binbin Hu, Zhiqiang Zhang, Jun Zhou*, Jingli Fang, Quanhui Jia, Yanming Fang, Quan Yu, Yuan Qi
Ant Financial Services Group, Hangzhou , China
{bin.hbb,lingyao.zzq,sophone.fjl,jun.zhoujun,yuan.qi}@antfin.com,{quanhui.jia,yanming.fym,jingmin.yq}@mybank.cn

ABSTRACT

Aiming to effectively distinguish loan default in the Mobile Credit Payment Service, industrial efforts mainly attempt to employ conventional classifier with complicated feature engineer for prediction. However, these solutions fail to exploit multiplex relations existed in the financial scenarios and ignore the key intrinsic properties of the loan default detection, *i.e.,* communicability, complementation and induction. To address these issues, we develop a novel attributed multiplex graph based loan default detection approach for effectively integrating multiplex relations in financial scenarios. Considering the complexity of financial scenario, an Attributed Multiplex Graph (AMG) is proposed to jointly model various relations and objects as well as the rich attributes on nodes and edges. We elaborately design relation-specific receptive layers equipped with adaptive breadth function to incorporate important information derived from local structure in each aspect of AMG and stack multiple propagation layer to explore the high-order connectivity information. Furthermore, a relation-specific attention mechanism is adopted to emphasize relevant information during end-to-end training. Extensive experiments conducted on the large-scale real-world dataset verify the effectiveness of the proposed model compared with state of arts. Moreover, AMG-DP has also achieved a performance improvement of 9.37% on KS metric in recent months after successful deployment in the Alipay APP.

KEYWORDS

Loan Default Analysis, Multiplex Graph, Graph Neural Network

ACM Reference Format:

Binbin Hu, Zhiqiang Zhang, Jun Zhou, Jingli Fang, Quanhui Jia, Yanming Fang, Quan Yu, Yuan Qi. 2020. Loan Default Analysis with Multiplex Graph Learning. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October19–23, 2020, Virtual Event, Ireland.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3340531.3412724

* The corresponding author.

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412724

1 INTRODUCTION

In recent years, the *Mobile Credit Payment Service* (*e.g.,* Ant Credit Pay of Ant Financial [1]) has gradually became indispensable in our daily life. Analogous to the credit card services in commercial banks, the loan default behaviours [9, 22] have seriously impaired the operations of mobile credit payment services. To avoid the economic loss caused by the frauds, *Loan Default Prediction*, which aims to predict whether a user will fail to make repayments in the future, is at the heart of the risk management system in financial institutions [11, 25].

Since the loan default prediction problem can be intuitively formulated as a binary classification problem, the industrial solutions generally employ tree-based (or neural network-based) classifier with user-related features generated by subtle feature engineering. Conventional methods make the prediction mostly relying on the statistical features extracted from various aspects, *e.g.,* user profile, credit history and transaction behaviors. However, in real-world financial institutions, the underlying reasons driving a user to fail to make required repayments is quite complicated, and such multiplex information is hard to be expressed by these feature-based approaches. Recently, a few research efforts attempt to analyze anomalies in financial scenarios via graph learning. Nevertheless, these methods ignore the plentiful attributes on the edges and are only designed for specific scenarios (*e.g.,* cash-out [9], fraud [22, 29] and malicious account [11]), resulting in unsatisfactory performance on our complex financial scenarios.

In our study, we mainly investigate into the following three key intrinsic properties of the loan default prediction problem in mobile credit payment services, which are still underexplored in the existing industrial approaches. **(P1) Communicability**: financial default of a certain user is not completely motivated by his/her inherent factors. In addition, we have reason to believe a user maybe reluctant to make repayments due to frequent fund transfer behaviors, which are characterized by the local structure in the network. **(P2) Complementation**: the portrait of financial defaulter is complex, which is hard to fully expressed through single source of information, especially for the new users with few behaviors. Hence, capturing fine-grained interactions in the multiplex scenarios while incorporating the reciprocity between complementary relations is in urgent demand. **(P3) Induction**: the new users appear every day in the practical financial scenarios, which prompts models to keep the ability for making prediction on such a group of users with only limited profile information.

To address these issues, we aim to leverage rich multiplex interactions in financial scenarios in a principled way, and propose a novel **A**ttributed **M**ultiplex **G**raph based Loan **D**efault **P**rediction approach, called AMG-DP for short. First of all, for better integrating rich information derived from various relations and abundant

[1] https://www.antfin.com/

attributes in objects and links, we propose to model the scenario of loan default prediction as an **A**ttributed **M**ultiplex **G**raph (AMG for short). Fig. 2(a) exhibits an well-established AMG in our scenario, which covers different kinds of relations as well as various kinds of objects [12, 15]. Meanwhile, both objects and links are associated with plentiful attributes. Inspired by the analysis of real data, we propose to exploit the rich local structure information under our AMG framework through recently emerging graph neural networks [24, 26, 28, 30]. As shown in Fig. 2(b), our proposed model is composed of three parts, namely input layer, relation-specific receptive layer and output layer. Specifically, we pre-process the raw feature in the input layer. And in the relation-specific receptive layer, an adaptive breadth function is designed to explore important information derived from local structure in each aspect of AMG. Then, multiple embedding propagation layers are stacked to exploit high-order information (**P1**). At last, an relation-specific attention mechanism is adopted to emphasize relevant views during training since different views might have distinct weight to decide personalized credit exposure (**P2**). Moreover, our proposed model is naturally potential to make predictions on new-coming users, benefitting from the ability of inferring embedding via local structure in AMG (**P3**). To evaluate the performance of AMG-DP, we collect a large-scale real-world dataset from a mobile credit payment platform. Extensive experiments conducted on this dataset verify the effectiveness of AMG-DP. Further studies demonstrate the rationality of each component designed and the high quality of learned representations. To sum up, our work has the following contributions:

- To our knowledge, it is the first time that the loan default prediction problem has been generally modeled as attributed multiplex graph for comprehensively exploring various relations in financial scenarios.
- With the analysis of real-world data, we investigate into the three key intrinsic properties of the loan default prediction problem: communicability, complementation and induction, which are not fully explored in previous studies.
- To integrated above three properties into a unified model, we develop a novel AMG-DP model equipped with relation-specific receptive layer and attention mechanism, sufficiently leveraging the local structure and multiplex relations to characterize users' credit risk.
- We perform extensive experiments on a large-scale real-world dataset and verify the effectiveness, stability and interpretation of our proposed model. These merits of the proposed model are extremely important to the financial scenarios. Moreover, AMG-DP also gains a performance improvement of 9.37% on KS metric in recent months when deployed in the Alipay APP.

2 PROBLEM FORMULATION

We represent rich information derived from the scenario of mobile credit payment service as attributed multiplex graph (AMG). Formally, given a relation set $\mathcal{R}$, we define the AMG as follows.

Definition 2.1. **Attributed Multiplex Graph** (AMG). An AMG is denoted as $\mathcal{G} = \{\mathcal{G}^r | r \in \mathcal{R}\} = \{\mathcal{V}, \mathcal{E}, \mathcal{X}_{\mathcal{V}}, \mathcal{X}_{\mathcal{E}}\}$. Here, $\mathcal{G}^r = \{\mathcal{V}^r, \mathcal{E}^r, \mathcal{X}^r_{\mathcal{V}}, \mathcal{X}^r_{\mathcal{E}}\}$ is a relation-specific graph with relation-specific nodes (*i.e.,* $\mathcal{V}^r$) and edges (*i.e.,* $\mathcal{E}^r$) *w.r.t.* $r \in \mathcal{R}$. $\mathcal{V} = \{\mathcal{V}^r | r \in \mathcal{R}\}$ and $\mathcal{E} = \{\mathcal{E}^r | r \in \mathcal{R}\}$ represent all nodes and edges in the AMG, respectively. $\mathcal{X}_{\mathcal{V}} = \{x_v | v \in \mathcal{V}\}$ ($\mathcal{X}_{\mathcal{E}} = \{x_e | e \in \mathcal{E}\}$) is the set of attributes on nodes (edges) denoting node v (edge e) has feature x_v (x_e). Meanwhile, we use the superscript r to denote the set of attributes on nodes or edges *w.r.t.* relation r(*i.e.,* $\mathcal{X}^r_{\mathcal{V}}$ or $\mathcal{X}^r_{\mathcal{E}}$). Note that the AMG degrades into a single graph when $|\mathcal{R}| = 1$.

Compared to previous work [2, 8, 9, 17, 22, 23], our defined AMG is a more general framework to model real-world unstructured network data, which consists of hundreds of millions of nodes and edges of multiple types, and each node and edge is associated with plentiful attributes. Specially, multi-view [16, 18] and multi-relational [5] networks can be easily modeled by AMG as a special case, which ignores node types and attributes on the nodes and edges. Several efforts have been made to complex network analysis with representation learning. While, most of these works are not suitable to directly applied in our scenario for promising performance due to their intrinsic limitations [2, 15].

In our study, we naturally abstract the scenario of mobile credit payment service as an AMG, which jointly integrates abundant attributes and interaction information between users and various objects. Here, we further formalize the loan default prediction problem based on AMG as follows:

Definition 2.2. **AMG based Loan Default Prediction**. AMG based loan default prediction aims to infer the binary label (whether he/she is a defaulter) of user $u \in \mathcal{U}$ based on his/her intrinsic attributes and the large scale interactive information involved in the AMG $\mathcal{G}$. Note that the user set $\mathcal{U}$ is a subset of the node set of the AMG $\mathcal{G}$ (*i.e.,* $\mathcal{U} \subseteq \mathcal{V}$).

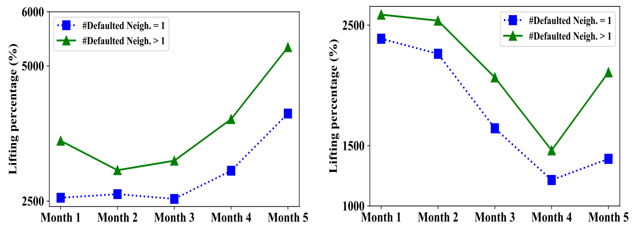
It is worthwhile to note that loan default prediction aims to predict whether a user will fail to make repayments in the future, which is inherently different from fraud detection or anomaly detection [14, 19]. Moreover, we explore and exploit several key intrinsic properties for loan defaulters (Section 3.1), which may be not existed in fraud detection and anomaly detection. On the other hand, deep graph learning under AMG framework provides a new perspective for loan default analysis, which naturally captures various relations and abundant attributes to characterize loan defaulters.

3 METHODOLOGY

In this section, we firstly analyze how the multiplex relations existed in financial scenarios influence the loan default prediction in real-world datasets. Next, we present the proposed **A**ttributed **M**ultiplex **G**raph based Loan **D**efault **P**rediction (callded AMG-DP for short) in detail.

3.1 Multiplex Relations in Financial Scenarios

In this section, we investigate into multiplex relations in financial scenarios intuitively, which inspire us to characterize the **communicability** and **complementation** for loan default prediction. With the real-world datasets (See the descriptions in Sec. 4) in Ant Credit Pay, we count the number of neighbors who are defaulter (called defaulted neighbor) for each user *w.r.t.* each relation (We take *Transfer* and *Social* relations as examples in our analysis). Next, we divide all users into three distinct groups *w.r.t.* the number of



(a) Transaction (b) Social

Figure 1: The lifting percentages of defaulter rate in users with different amount of defaulted neighbors against users without any cash-out neighbor *w.r.t.* transaction and social relation.

their defaulted neighbors (*i.e.,* #Defaulted Neigh. = 0/ = 1/ >1 in Fig. 1). The fraction of defaulter is calculated in each group and we present the statistical results on five months in Fig. 1. From the results, we have the following observations.

- From the macro-level (**Complementation**), different relations offer different perspectives to characterize defaulters, which yields different probabilities of defaulters within the same group on each month in Fig. 1. It derives to capture fine-grained interactions in financial scenarios and incorporate complementary relations for performance improvements.
- From the micro-level (**Communicability**), users with more defaulted neighbors are more likely to be defaulters. It implies that the loan default of a certain user is likely to be influenced by the interactions with people around(*e.g.,* frequent transfer). With graph model, we naturally integrate such a pattern as local structure into our proposed model.

Motivated by the above observations from real-world data, we develop the novel AMG-DP to exploit the rich information derived from local structure and complementary resources implicated in multiplex relations simultaneously. As shown in Fig. 2(b), the architecture of the proposed AMG-DP is comprised of three layers: (1) Input layer, which pre-process the raw attributes to eliminate the effect of missing values and the scale of attribute values. (2) Relation-specific receptive layer, which comprehensively explores rich semantics derived from multiplex relation through projection, aggregation and propagation. (3) Output layer, which emphasize relevant complementary information for final prediction with an adaptive fusion function. In the ensuing discussion, we will zoom into each well-designed part of AMG-DP.

3.2 Input Layer

In our framework, the input is the information of nodes and edges derived from the AMG. Considering the scale of attribute values and the existence of missing values, the original attributes of nodes and edges are discretized to a new feature space with a fixed dimension, denoted as $\mathbf{X}_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}| \times F_{\mathcal{V}}}$ and $\mathbf{X}_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}| \times F_{\mathcal{E}}}$. Due to multiplex relations in AMG, we rewrite $\mathbf{X}_{\mathcal{V}}$ and $\mathbf{X}_{\mathcal{E}}$ as $\{\mathbf{X}^r_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}| \times F^r_{\mathcal{V}}}\}_{r \in \mathcal{R}}$ and $\{\mathbf{X}^r_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}| \times F^r_{\mathcal{E}}}\}_{r \in \mathcal{R}}$, where $F_{\mathcal{V}} = \sum_{r \in \mathcal{R}} F^r_{\mathcal{V}}$ and $F_{\mathcal{E}} = \sum_{r \in \mathcal{R}} F^r_{\mathcal{E}}$.

Remark: For efficient storage and computation, all the matrices are stored as sparse form in our implementation.

3.3 Relation-specific Receptive Layer

As mentioned above, we model our financial scenario as AMG, involving multiplex relations. Hence, we introduce relation-specific receptive layer to incorporate rich semantics derived from multiplex relations.

3.3.1 Projection. Firstly, we assume that the users associated with different relations have different feature spaces, which can be captured through relation-specific matrices (*i.e.,* $\mathbf{M}^r_{\mathcal{V}} \in \mathbb{R}^{F^r_{\mathcal{V}} \times d}$ and $\mathbf{M}^r_{\mathcal{E}} \in \mathbb{R}^{F^r_{\mathcal{E}} \times d}$). Here d is the dimension size of node and edge embeddings. The projection operation is implemented as follows,

$$\mathbf{H}^r = \mathbf{X}^r_{\mathcal{V}} \cdot \mathbf{M}^r_{\mathcal{V}}, \quad \mathbf{E}^r = \mathbf{X}^r_{\mathcal{E}} \cdot \mathbf{M}^r_{\mathcal{E}}, \tag{1}$$

where $\mathbf{H}^r$ and $\mathbf{E}^r$ are the projected embeddings for nodes and edges *w.r.t.* relation r, respectively.

3.3.2 Aggregation. To comprehensively explore and extensively exploit the local structure in the relation-specific receptive layer, an adaptive breadth function $\phi(\mathbf{h}^r_u, \mathcal{E}^r, \mathbf{E}^r; \Theta^r_\phi)$ for user u is introduced to adaptively capture important information implied in local structure, and generate the propagated information of user u *w.r.t.* relation r, denoted as $\mathbf{h'}^r_u$. Meanwhile, for better representation learning, the function $\phi(\cdot)$ will take the rich information on the edges (*i.e.,* $\mathcal{E}^r$ and $\mathbf{E}^r$) into account. Here, Θ^r_ϕ denotes the parameter set of $\phi(\cdot)$ *w.r.t.* relation r. Inspired by the recent emerging attention mechanism [20], given a user u, we formally specify the adaptive breadth function $\phi(\cdot)$ for neighbor aggregation as follows,

$$\begin{aligned} \alpha_{u,i} &= \frac{\exp(\mathbf{v}^{r\mathrm{T}} f(\mathbf{W}^r_{\phi 1}[\mathbf{h}^r_u || \mathbf{h}^r_i || \mathbf{e}^r_{u,i}]))}{\sum_{(u,j) \in \mathcal{E}^r} \exp(\mathbf{v}^{r\mathrm{T}} f(\mathbf{W}^r_{\phi 1}[\mathbf{h}^r_u || \mathbf{h}^r_j || \mathbf{e}^r_{u,j}]))}, \\ \mathbf{h'}^r_u &= f(\mathbf{W}^r_{\phi 2} \sum_{(u,i) \in \mathcal{E}^r} \alpha_{u,i}(\mathbf{h}^r_i + \mathbf{e}^r_{u,i})). \end{aligned} \tag{2}$$

Here, the weight matrices $\mathbf{W}^r_{\phi 1} \in \mathbb{R}^{d \times 3d}$, $\mathbf{W}^r_{\phi 2} \in \mathbb{R}^{d \times d}$ and the context vector $\mathbf{v}^r \in \mathbb{R}^d$ are parameters (*i.e.,* Θ^r_ϕ). $||$ denotes the concatenate operation and $f(\cdot)$ is the activation function, which is defined as $\tanh(\cdot)$ in our experiments.

3.3.3 Propagation. After aggregating the neighbor information, we formulate the message propagation function towards user u as follows,

$$\mathrm{ReLU}(\mathbf{W}^r_p[\mathbf{h}^r_u || \mathbf{h'}^r_u]), \tag{3}$$

where $\mathbf{W}^r_p \in \mathbb{R}^{d \times 2d}$ is the trainable transformation matrix.

A single aggregation may be inadequate in capturing complex interactions in AMG. To enhance the expressiveness, we stack multiple embedding propagation layers to explore the high-order connectivity information. Formally, given a user u, we recursively obtain its embedding as:

$$\mathbf{h}^{r(l)}_u = \mathrm{ReLU}(\mathbf{W}^{r(l)}_p[\mathbf{h}^{r(l-1)}_u || \mathbf{h}'^{r(l)}_u]), \tag{4}$$

wherein the information propagated in l^{th} order network for user u (*i.e.,* $\mathbf{h}'^{r(l)}_u$) is calculated by $\phi(\mathbf{h}^{r(l-1)}_u, \mathcal{E}^r, \mathbf{E}^r; \Theta^{r(l)}_\phi)$, parameterized

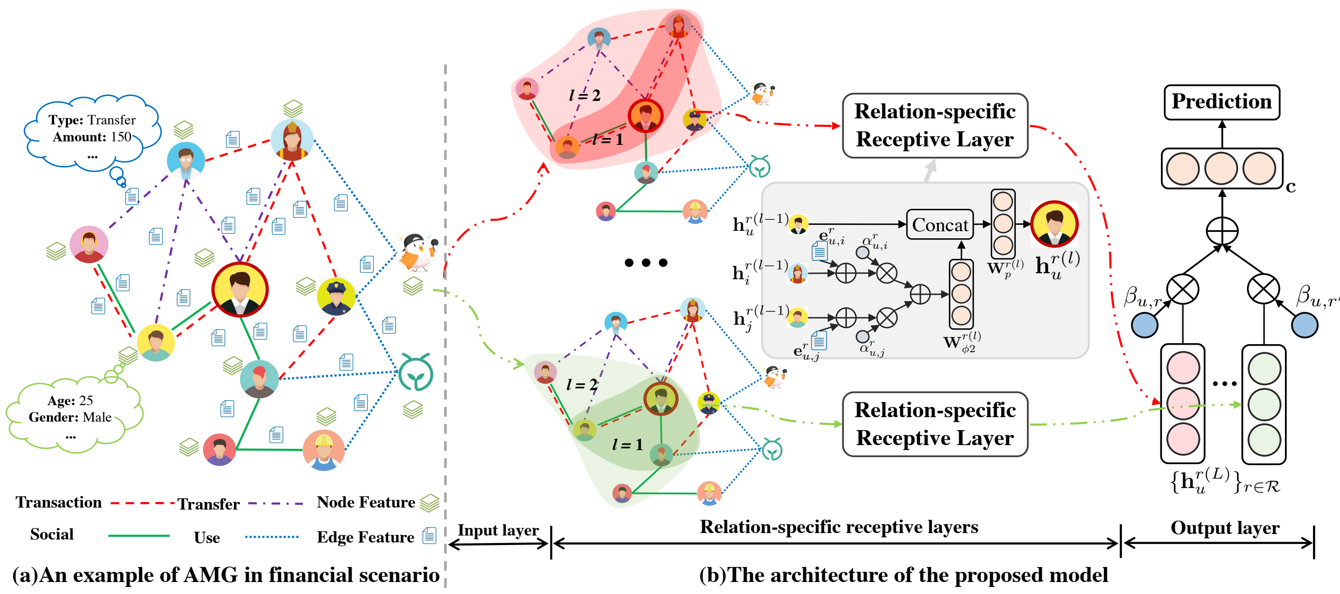


Figure 2: Overview of our work. (a) A toy exampled of AMG in financial scenario. (b) The framework of our proposed model AMG-DP with two embedding propagation layers. Intuitively, different relation-specific receptive layers (*e.g.,* transaction and social) are capable of capturing disparate topological structures *w.r.t.* relations in AMG.

by $\Theta_{\phi}^{r(l)}$. Moreover, $\mathbf{h}_u^{r(0)}$ is set as $\mathbf{h}_u^r$ at the initial information propagation iteration.

3.4 Output Layer

Note that the classification target in loan default prediction is the set of user nodes in the AMG, hence we will focus on the user nodes $u \in \mathcal{U}$ in the output layer. Intuitively, different relations may have disparate topological structures, as well as the different semantic meanings for prediction. Hence, these representations(*i.e.,* $\{\mathbf{h}_u^{r(L)}\}_{r\in\mathcal{R}}$ after L^{th} propagation) learned based on the local structure via each relation-specific receptive layer offer strengths to complement each other for promising performance improvement. To characterize the fine-grained differentiation of importance or relevance in the multiplex graph, an adaptive fusion function $\varphi(\{\mathbf{h}_u^{r(L)}\}_{r\in\mathcal{R}};\Theta_\varphi)$, parameterized by Θ_φ is learned to weigh multiplex relations and integrate multiple representations for loan default prediction. Specifically, we implement the adaptive fusion function with relation-specific attention mechanism, which is formulated as follows:

$$\beta_{u,r} = \frac{\exp\left(\boldsymbol{v}_\varphi^T f(\mathbf{W}_\varphi \mathbf{h}_u^{r(L)} + \mathbf{b}_\varphi)\right)}{\sum_{r'\in\mathcal{R}} \exp\left(\boldsymbol{v}_\varphi^T f(\mathbf{W}_\varphi \mathbf{h}_u^{r'(L)} + \mathbf{b}_\varphi)\right)}, \qquad \mathbf{z}_u = \sum_{r\in\mathcal{R}} \beta_{u,r}\cdot \mathbf{h}_u^{r(L)}, \tag{5}$$

where $\{\mathbf{W}_\varphi, \mathbf{b}_\varphi, \boldsymbol{v}_\varphi\}$ form the parameter set Θ_φ of the adaptive fusion function. Similarly, we set $f(\cdot)$ as $\tanh(\cdot)$.

3.5 Model Learning

To learn the parameters of our method, cross-entropy is adopted as the loss function to train the model in an end-to-end mode. Formally, we define the loss function over all labeled users (*i.e.,* $\mathcal{U}_L$) as:

$$\mathcal{L} = -\sum_{u\in\mathcal{U}_L} y_u \log(\sigma(\mathbf{c}^{\mathrm{T}}\mathbf{z}_u)) + (1-y_u)\log(1-\sigma(\mathbf{c}^{\mathrm{T}}\mathbf{z}_u)), \tag{6}$$

where y_u is the ground truth and $\mathbf{c}$ is the parameter vector of the final classifier.

3.6 Discussion

AMG-DP is a flexible approach to leverage multiplex relations and plentiful attributes in well-established AMG. In the practical financial application, we are more concerned about the newcome users with limited information. Hence, the proposed approach has the following advantages in this case: (1) Our model has the inductive ability to make prediction for these newcome users by aggregating their neighbors with multiplex relations. (2) Fine-grained complementary information carried by multiplex relations can be effectively integrated into our model for performance improvement. For efficient numerical computation, we follow the trick in [10] to perform attention calculation in adaptive breadth function. Therefore, the complexity and storage is in linear with $O(|\mathcal{E}^r|)$ for the relation-specific receptive layer *w.r.t.* relation r. Overall, the upper bound of model complexity is $O(|R| * \max_{r\in R}(\mathcal{E}^r))$. In practice , the number of relations (*i.e.,* $|R|$) is small and the interactions of newcome users are sparse, which makes our model suitable for large-scale data in real applications.

Table 1: Statistics of the AMG.

Relation	#S	#T	#S-T
User $\xrightarrow{transfer}$ User	9.00×10^7	9.00×10^7	4.32×10^9
User $\xrightarrow{transaction}$ User	1.10×10^8	1.10×10^8	5.23×10^9
User $\xrightarrow{social}$ User	1.20×10^8	1.20×10^8	3.14×10^9
User $\xrightarrow{use}$ Applet	1.10×10^8	1.60×10^6	3.86×10^7

4 EXPERIMENTS

In this section, we evaluate the performance of our model on a real-world dataset and present the detailed result analysis, with the aim of answering the following research questions:

- **RQ1**: Does our model outperform the state-of-the-art methods on the loan default prediction task with the real-world dataset.
- **RQ2**: How do the key components (*e.g.,* attention mechanism and propagation layer) of our model benefit the prediction.
- **RQ3**: How about the quality of embeddings generated by the proposed model.
- **RQ4**: How about the deployment and performance of the proposed model in the real-world system.

4.1 Experiment Stetup

4.1.1 Dataset Description. We utilize the dataset from Ant Credit Pay, which is an online credit payment service provided by Ant Financial Services Group. We extract ten sub-datasets for the evaluation namely *Month 1* to *Month 10.* Note that we keep the loan default rate at approximately 0.05 through random sampling in each month and consequently each sub-dataset contains about 1.5 million users for classification. To evaluate the performance of each method, we utilize the first five month (*i.e.,* Month 1 $\sim$ 5) for training [2] and predict the default probability of users in the remaining five month (*i.e.,* Month 6 $\sim$ 10), respectively. Based on the real-world dataset, we construct an attributed multiplex graph, consisting of various relations between users, as well as the interactions between users and applets of Alipay The detailed descriptions of the AMG are shown in Table 1. To enrich user information, we extract 108 attributes related to credit exposure for each user. Note that the volume of the experimental dataset and the corresponding AMG is extremely large, which brings more challenges.

4.1.2 Baselines and Evaluation Metrics. We compare our model with several state-of-the-art methods, which fall into three main groups: feature based methods (*i.e.,* **MLP**, **DeepForest** [32] and **Xgboost** [3]), unsupervised graph embedding methods (*i.e.,* **Node2vec** [6] and **Metapath2vec** [4]), GNN methods with single graph (*i.e.,* **GraphSAGE** [7] and **GAT** [21]), heterogeneous graph (*i.e.,* **HAN** [23]) and multi-view graph (*i.e.,* **SemiGNN** [22]). Note that several multi-view based methods (*e.g.,* MVE [16], MNE [27] and GATNE [2]) are not selected as baselines, since these methods **cannot scale up to such a large-scale dataset**. Nevertheless, our selected baselines have a comprehensive coverage of the existing graph based prediction methods.

We evaluate the performance of the loan default prediction via **AUC** (*i.e.,* Area Under the ROC Curve) and **KS** (*i.e.,* Kolmogorov Smirnov) , which are widely adopted in financial scenarios. A larger AUC or KS indicates a better performance.

4.1.3 Implementation. For scaling up to large-scale datasets adopted in the paper, we implement all the models in our experiments on parameter server based distributed learning systems [31]. For fair comparison, we set learning rate = 1e-4, regularizer = 1e-5, batch size = 256, embedding size = 64 and select ADMA as optimizer for all models. Concretely, we set the architecture of MLP as [256, 128, 64]. And we set the number of trees = 100, depth of each tree = 5 for Xgboost and DeepForest. For random walk based methods (*i.e.,* Node2vec and Metapath2vec), we set the number of walks per node as 10, the walk length as 100 and the window size as 5. We select metapath "User $\xrightarrow{transaction}$ User $\xrightarrow{social}$ User" for Metapath2vec, which performs best on validation set. For GNN based methods (*i.e.,* GraphSAGE, GAT, HAN, SemiGNN and AMG-DP), we set the number of sampled neighbor as 100 and the number of propagation layer as 2. Specially, we set the pool strategy in GraphSAGE as "max pool" and set the number of attention heads in GAT as 8.

4.2 Performance Comparison (RQ1)

We present the performance comparison of AMG-DP and the baseline methods in Table 2. The major findings can be summarized as follows.

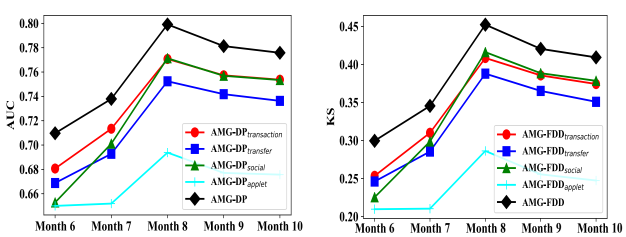
- Our model consistently outperforms all the baselines on all the months by a large margin. In particular, AMG-DP achieves remarkable improvement over the strongest baseline *w.r.t.* AUC by 2.88 $\sim$ 3.63% and KS by 9.05 $\sim$ 14.69%, respectively. The results show the effectiveness and stability of AMG-DP for loan default prediction, which has a more principled mechanism to jointly leverage local structure and complementary information in AMG.
- Feature based methods (*i.e.,* MLP, DeepForest and Xgboot) achieve relatively pool performance on all months, indicating that handcrafted feature engineering is insufficient to capture the portrait of loan defaulter in our complex financial scenario, further limiting performance. With the help of the powerful ability of deep neural networks for modeling non-linear functions, these methods still yield competitive performance.
- Generally, unsupervised graph representation methods (*i.e.,* Node2vec and Metapath2vec) perform unsatisfactorily on all months. It implies that pure graph structure contains limited information and these methods fail to make predictions for unseen target users. As a comparison, the performance of Node2vec$_f$ and Metapath2vec$_f$ verifies the that incorporating profile information is beneficial to characterize user's credit risk.
- We can observe that GNN-based methods (*i.e.,* GraphSAGE, GAT, SemiGNN and HAN) work well among these baselines, which benefit from the integration of local structure information. It is noteworthy that SemiGNN consistently underperforms other GNN-based methods. An intuitive explanation is that SemiGNN only explores first-order structure

[2]We hold out the last month in training data as the validation set for parameter setting

Table 2: Results of effectiveness experiments. We underline the best performance from the baselines for each comparison. We use "*" to indicate the improvement of AMG-DP over the best performance from the baselines is significant based on paired t-test at the significance level of 0.01. †: We adopt Xgboost as the final classification model for these unsupervised graph embedding methods. ‡: We utilize learned embedding as well as intrinsic attributes for prediction. The last row indicates the percentage of improvements gained by the proposed method compared to the best baseline.

Methods	Month 6		Month 7		Month 8		Month 9		Month 10	
	AUC	K-S	AUC	K-S	AUC	K-S	AUC	K-S	AUC	K-S
MLP	0.6247	0.2111	0.6449	0.24	0.7004	0.3253	0.6832	0.2938	0.6841	0.2890
DeepForest	0.6614	0.2359	0.6846	0.2709	0.7422	0.3672	0.7293	0.3442	0.7281	0.3372
Xgboost	0.6625	0.2342	0.6845	0.2701	0.7432	0.3676	0.7295	0.3430	0.7272	0.3355
Node2vec†	0.5652	0.0980	0.5667	0.1018	0.5728	0.1105	0.5629	0.094	0.5613	0.0903
$\text{Node2vec}_f^{\dagger\ddagger}$	0.6728	0.2406	0.6934	0.2765	0.7557	0.3847	0.7378	0.3545	0.7290	0.3310
Metapath2vec†	0.5710	0.1001	0.5783	0.1121	0.5801	0.1198	0.5732	0.1132	0.5701	0.1096
$\text{Metapath2vec}_f^{\dagger\ddagger}$	0.6801	0.2511	0.7023	0.2801	0.7637	0.3909	0.7462	0.3618	0.7371	0.3489
GraphSAGE	0.6807	0.2536	0.7134	0.3102	0.7707	0.4086	0.7573	0.3857	0.7537	0.3744
GAT	0.6890	0.2613	0.7140	0.3116	0.7711	0.4078	0.7573	0.3833	0.7541	0.3747
HAN	0.6823	0.2600	0.7103	0.3110	0.7701	0.4034	0.7512	0.3801	0.7512	0.3712
SemiGNN	0.6727	0.2465	0.6999	0.2874	0.7634	0.3910	0.7504	0.3694	0.7445	0.3616
AMG-DP	**0.7097***	**0.2997***	**0.7379***	**0.3456***	**0.7991***	**0.4524***	**0.7813***	**0.4206***	**0.7758***	**0.4094***
Improvement	3.00%	14.69%	3.35%	10.91%	3.63%	10.71%	3.17%	9.05%	2.88%	9.26%

information and ignores the rich attributes implied in neighbors. Meanwhile, HAN achieves comparable performance to GraphSAGE and GAT due to the integration of heterogeneous information. However, it overlooks rich attributes on edges and depends heavily on extensive domain knowledge (*e.g.,* selection of meta-paths).



(a) (b)

Figure 3: Performance change of AMG-DP when using different relations of data.

4.3 In-depth Analysis of AMG-DP (RQ2)

In this section, we perform a series of in-depth analysis to better understand the traits of AMG-DP.

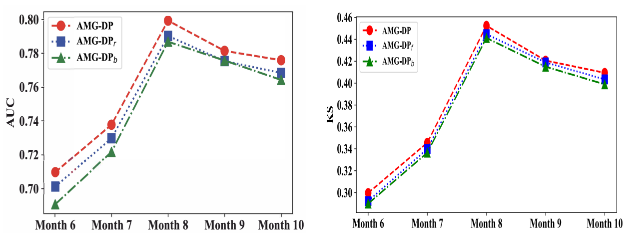
4.3.1 Impact of Different Relations. As mentioned above, our model characterizes defaulters in different aspects to enhance performance, benefiting from the utilization of multiplex relations. Thus, we aim to analyze the impact of different relations in this subsection. Note that we use AMG-DP$_*$ to denote variants of AMG-DP only using one specific relation (*e.g.,* AMG-DP$_{transaction}$ means AMG-DP only adopts the transaction relation for prediction). As shown in Fig. 3, we can observe that the performance will drop a lot if only one specific relation is utilized in AMG-DP. It indicates the effectiveness of local structure and complementary information embodied in multiplex relations. It is not surprising that both social and transaction relations achieve good performance since people with similar credit risk tend to gather together and such a pattern can be naturally modeled via graph model.

4.3.2 Effect of Attention Mechanism. In order to verify the effectiveness of attention mechanism in AMG-DP, we respectively remove the attention mechanism in the breadth function $\phi(\cdot)$ (*i.e.,* AMG-DP$_b$) and fusion function $\varphi(\cdot)$ (*i.e.,* AMG-DP$_f$). As shown in Fig 4, AMG-DP consistently outperforms AMG-DP$_b$ and AMG-DP$_f$, which demonstrates that the proposed attention mechanism is essential and helps better utilize local structure and multiplex relations to improve the model's performance in two aspects. First, AMG-DP provides a more reasonable way to incorporate the reciprocity between complementary relations in AMG, which is potential to fully characterize the portrait of financial defaulters. Second, the importance of each neighbor should be adaptively captured for a specific target user instead of being treated equally.

Besides the performance effectiveness, we further present the macro-level analysis of the attention distributions on each month in Fig 5 (a). We can observe that the distribution of attention values are indeed very skew, implicating some relations (*e.g., Transaction*) are more important to consider than the others (*e.g., Applet*). These findings are consistent with previous observations for impact of different relations, where these relations are likely to achieve better performance.

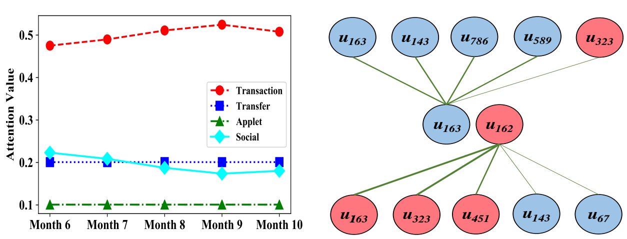
From the micro-level view, we select a defaulter u_{162} and a normal user u_{163} in the test set as an illustrative example. As shown in Fig. 5 (b), we can observe that u_{163}, u_{143}, u_{786} and u_{589} obtain higher attention value for u_{163} while u_{163}, u_{323} and u_{451} obtain higher attention value for u_{162}, which is intuitive since they belong to the same label with the target user. This finding provides explicit evidence to demonstrate that our proposed attention mechanism is capable of emphasizing important neighbors derived from local structure in AMG, which are useful for loan default prediction.

4.3.3 Effect of Propagation Layer. At last, we show the results of performance comparison *w.r.t.* the number of propagation layers in Fig. 6. In this experiment, we only stack two propagation layers at most, since our dataset is extremely large. It is clearly that the performance increases as we stack more propagation layers, indicating the usefulness of high-order local structure information [1].



(a) (b)

Figure 4: Performance comparison of attention mechanism.



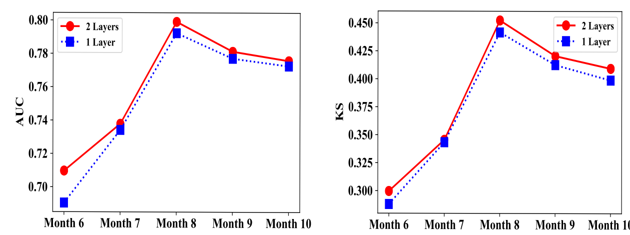
(a) (b)

Figure 5: The detailed analysis of attention mechanism: (a) The distribution of attention values of AMG-DP *w.r.t.* relations. (b) An illustrative example of the interpretability of attention distributions under *Transaction* relation. Red means defaulter and blue means normal user. The edge thickness denotes the attention value between nodes (A thicker edge means a larger attention value). Best viewed in color.

4.4 Visualization (RQ3)

To examine the quality of user representations learned by AMG-DP, we visualize the embeddings of users on month 9 using t-SNE [13] algorithm in Fig. 7 (For comparison, we visualize the embeddings learned by several competitive baselines (*i.e.,* GraphSAGE, GAT and HAN) in table 2.).

From the plots, we observe that GraphSAGE cannot effectively identify defaulters due to its limited ability of expression for multiplex graph. On the other hand, GAT and HAN can reasonably



(a) (b)

Figure 6: Performance comparison *w.r.t.* the number of propagation layers.

separate the defaulters from normal users, benefiting from the integration of attention mechanism and heterogeneous information, respectively. Nevertheless, AMG-DP has a more crisp boundary and denser cluster, which indicates the high quality of user representations learned by the proposed AMG-DP.

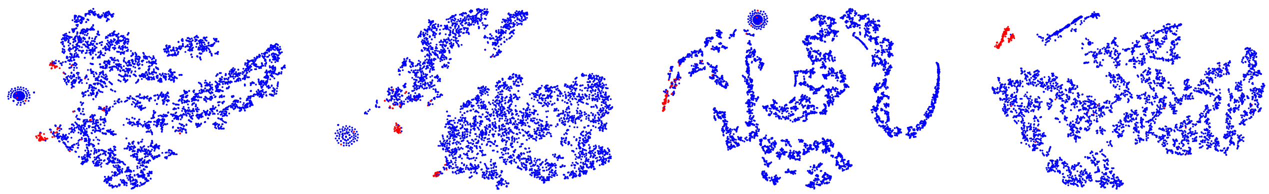
4.5 Deployment and Performance (RQ4)

To further demonstrate the effectiveness of the proposed AMG-DP, we deploy it in the Alipay App for loan default prediction. As shown in Fig. 8, the deployment pipeline consists of two parts: offline training and online prediction. For offline training, aiming to construct AMG, feature extractor and relation extractor are two important component to explore rich attributes (*e.g.,* user profile and credit history) and various relations (*e.g.,* transaction, transfer and social) from original database, respectively. Meanwhile, a series of training instances are also generated from database through negative sampling strategy, in order to keep the loan default rate at approximately 0.05. Then, we feed these training instances as well as AMG into AMG-DP for training and upload the optimal model to model server. For online prediction, given a fresh batch of users, relevant neighbors are collected *w.r.t.* different relations from AMG and the model server is able to make predictions based on these information. It is worthwhile to note that the AMG is updated monthly to keep the user's attributes and local structure to stay up-to-date.

Compared to the existed deployed baseline model SMART (A distributed implementation of Xgboost), We can observe that the proposed AMG-DP achieve **9.37%** improvement on the main metric (*i.e.,* KS) in recent months. The promotion of loan default prediction performance furture verifies the effectiveness of our proposed AMG-DP. For online prediction, our deployed model takes approximately **1 hour** to score **200 million** customers. And the inference time scales linearly *w.r.t.* the number of clients.

5 CONCLUSION

In the paper, we study the defaulter detection problem with attribute multiplex graph for comprehensively exploring and exploiting various relations in financial scenarios. For this purpose, we propose a novel AMG-DP model under the AMG framework. Specifically, we pre-process the raw feature in the input layer and incorporate rich semantics derived from multiplex relations in the relation-specific receptive layer. In the output layer, an adaptive fusion function is



(a) Graphsage (b) GAT (c) HAN (d) AMG-DP

Figure 7: The result of Visualization (Red: Defaulters, Blue: Normal users). Best viewed in color.

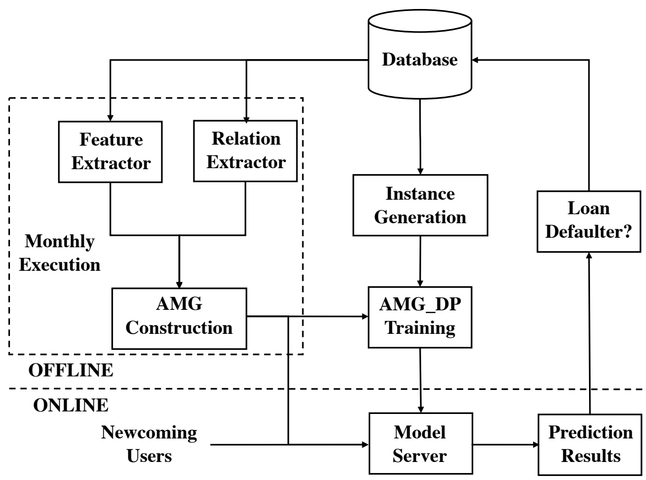


Figure 8: The deployment of AMG-DP.

designed to emphasize relevant complementary information during training. Extensive experiments on a large-scale real-world dataset verify the effectiveness, stability and interpretation of our proposed model. Meanwhile, we uncover the deployment details of AMG-DP on real-world system (*i.e.,* Alipay APP) and the proposed AMG-DP achieves 9.37% improvement on KS metric in recent months.

REFERENCES

[1] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. Grarep: Learning graph representations with global structural information. In *CIKM*. 891–900.
[2] Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Representation Learning for Attributed Multiplex Heterogeneous Network. In *SIGKDD*. 1358–1368.
[3] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *SIGKDD*. 785–794.
[4] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *SIGKDD*. 135–144.
[5] Ming-Han Feng, Chin-Chi Hsu, Cheng-Te Li, Mi-Yen Yeh, and Shou-De Lin. 2019. Marine: Multi-relational network embeddings with relational proximity and node attributes. In *WWW*. 470–479.
[6] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *SIGKDD*. 855–864.
[7] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NIPS*. 1024–1034.
[8] Binbin Hu, Yuan Fang, and Chuan Shi. 2019. Adversarial Learning on Heterogeneous Information Networks. In *SIGKDD*. 120–129.
[9] Binbin Hu, Zhiqiang Zhang, Chuan Shi, Jun Zhou, Xiaolong Li, and Yuan Qi. 2019. Cash-out User Detection based on Attributed Heterogeneous Information Network with a Hierarchical Attention Mechanism. In *AAAI*. 946–953.
[10] Ziqi Liu, Chaochao Chen, Longfei Li, Jun Zhou, Xiaolong Li, Le Song, and Yuan Qi. 2019. Geniepath: Graph neural networks with adaptive receptive paths. In *AAAI*. 4424–4431.
[11] Ziqi Liu, Chaochao Chen, Xinxing Yang, Jun Zhou, Xiaolong Li, and Le Song. 2018. Heterogeneous graph neural networks for malicious account detection. In *CIKM*. 2077–2085.
[12] Yao Ma, Suhang Wang, Chara C Aggarwal, Dawei Yin, and Jiliang Tang. 2019. Multi-dimensional Graph Convolutional Networks. In *SDM*. 657–665.
[13] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008), 2579–2605.
[14] Guansong Pang, Longbing Cao, and Ling Chen. 2016. Outlier detection in complex categorical data by modelling the feature value couplings. In *IJCAI*.
[15] Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. 2020. Unsupervised Attributed Multiplex Network Embedding. In *AAAI*.
[16] Meng Qu, Jian Tang, Jingbo Shang, Xiang Ren, Ming Zhang, and Jiawei Han. 2017. An attention-based collaboration framework for multi-view network representation learning. In *CIKM*. 1767–1776.
[17] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and S Yu Philip. 2018. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering* 31, 2 (2018), 357–370.
[18] Yiwei Sun, Suhang Wang, Tsung-Yu Hsieh, Xianfeng Tang, and Vasant Honavar. 2019. Megan: A generative adversarial network for multi-view network embedding. *arXiv preprint arXiv:1909.01084* (2019).
[19] Acar Tamersoy, Kevin Roundy, and Duen Horng Chau. 2014. Guilt by association: large scale malware detection by mining file-relation graphs. In *SIGKDD*. 1524–1533.
[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*. 5998–6008.
[21] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR*.
[22] Daixin Wang, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Zhou Jun, Yang Shuang, and Qi Yuan. 2019. A Semi-supervised Graph Attentive Network for Financial Fraud Detection. In *ICDM*.
[23] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous Graph Attention Network. In *WWW*. 2022–2032.
[24] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. 2019. A comprehensive survey on graph neural networks. *arXiv preprint arXiv:1901.00596* (2019).
[25] Dianmin Yue, Xiaodan Wu, Yunfeng Wang, Yue Li, and Chao-Hsien Chu. 2007. A review of data mining-based financial fraud detection research. In *WiCOM*. 5519–5522.
[26] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *SIGKDD*. 793–803.
[27] Hongming Zhang, Liwei Qiu, Lingling Yi, and Yangqiu Song. 2018. Scalable Multiplex Network Embedding.. In *IJCAI*. 3082–3088.
[28] Ziwei Zhang, Peng Cui, and Wenwu Zhu. 2018. Deep learning on graphs: A survey. *arXiv preprint arXiv:1812.04202* (2018).
[29] Qiwei Zhong, Yang Liu, Xiang Ao, Binbin Hu, Jinghua Feng, Jiayu Tang, and Qing He. 2020. Financial Defaulter Detection on Online Credit Payment via Multi-view Attributed Heterogeneous Information Network. In *WWW*. 785–795.
[30] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, and Maosong Sun. 2018. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434* (2018).
[31] Jun Zhou, Xiaolong Li, Peilin Zhao, Chaochao Chen, Longfei Li, Xinxing Yang, Qing Cui, Jin Yu, Xu Chen, Yi Ding, et al. 2017. Kunpeng: Parameter server based distributed learning systems and its applications in alibaba and ant financial. In *SIGKDD*. 1693–1702.
[32] Zhihua Zhou and Ji Feng. 2017. Deep Forest: Towards An Alternative to Deep Neural Networks.. In *IJCAI*. 3533–3539.