Fraudulent User Detection on Rating Networks Based on Expanded Balance Theory and GCNs

Wataru Kudo, Mao Nishiguchi, Fujio Toriumi

The University of Tokyo, Japan

{kudo, nishiguchi, tori}@crimson.q.t.u-tokyo.ac.jp

Abstract—Rating platforms provide users with useful information on products or other users. However, fake ratings are sometimes generated by fraudulent users. In this paper, we tackle the task of fraudulent user detection on rating platforms. We propose an end-to-end framework based on Graph Convolutional Networks (GCNs) and expanded balance theory, which properly incorporates both the signs and directions of edges. Experimental results on four real-world datasets show that the proposed framework performs better, or even best, in most settings. In particular, this framework shows remarkable stability in inductive settings, which is associated with the detection of new fraudulent users on rating platforms. Furthermore, using expanded balance theory, we provide new insight into the behavior of users in rating networks, that fraudulent users form a faction to deal with the negative ratings from other users. The owner of a rating platform can detect fraudulent users earlier and constantly provide users with more credible information by using the proposed framework.

Index Terms—graph convolutional networks, rating networks, fraud detection, balance theory

I. INTRODUCTION

With the recent spread of e-commerce platforms, user-generated content, such as ratings or reviews, is becoming more and more important in the decision-making process of consumers. Many online marketplaces or trading platforms (e.g., Amazon, Yelp, and Bitcoin Alpha) allow users to rate the quality of contents, or the trustworthiness of other users. The average ratings or posted comments help inform users before purchasing, visiting or trading. Because users attach great weight to such information in personal consumption, fraudulent users have great financial motivation to generate fake ratings [1]–[3]. Identifying and understanding fraudulent users may help platform providers maintain their credibility. Our goal is to detect fraudulent users on rating platforms based on their social rating behavior.

Previous studies [4], [5] have shown that the graph-based approaches are effective for the task of fraud detection. Rev2 [5], the state-of-the-art approach of this task appropriately models the interdependencies among users on rating platforms.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '19, August 27–30, 2019, Vancouver, Canada

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6868-1/19/08?/\$15.00

<http://dx.doi.org/10.1145/3341161.3342929>

Following the success of them, we also regard online rating data as a social graph. We call this graph a “rating network.” Rating networks have two characteristic properties: 1) signed edges and 2) directed edges. Reviews are associated with rating values in most cases and the meanings of edges greatly differ depending on the rating values. The edges of low ratings are associated with a negative relationship, while high ratings are associated with a positive one. Here, a rating network should be regarded as a kind of signed network, even if the rating values themselves are all positive. The direction of edges is also important, especially in user-to-user homogeneous rating networks, in which users can both rate and be rated. Rating others and being rated by others are different interactions; therefore, distinguishing between them could provide a better model of users behavior.

Recently, Graph Convolutional Networks (GCNs) have performed remarkably well on several graph-related tasks, including node classification [6], [7] and link prediction [8], [9]. One of the appealing characteristics of GCNs is their end-to-end learning manner, which incorporates both the graph structures and attributes of each node. In this respect, GCNs appear to be suitable for the task of detecting fraudulent users on online rating platforms. However, in applying existing GCNs to rating networks, it is necessary to handle both the signs and directions of the edges.

As for handling signs, the Signed Graph Convolutional Network (SGCN) [10] is a successful approach. In SGCN, they effectively incorporate 2-hop relationships between nodes by applying balance theory [11], [12], a social theory for signed networks. However, SGCN is designed for undirected networks and can not consider 2-hop relationships properly in directed networks. It is assumed that the appropriate expansion of balance theory is effective to settle this problem and capture the social behavior of users on rating networks.

The aim of this paper is to construct a detection model using expanded balance theory and GCNs. To the best of our knowledge, this work is the first to apply GCNs to the task of fraudulent user detection on online rating platforms. We conduct experiments on four real-world datasets to compare the existing state-of-the-art approach to three other GCN-based models considered to be effective, including the proposed framework.

Our main contribution in this paper is:

- We propose a novel GCN-based end-to-end framework using expanded balance theory for fraudulent user de-

tection that incorporates both the signs and directions of edges.

- From the experimental results, we show that the proposed framework performs well, or even best compared to the existing state-of-the-art algorithm and other GCN-based methods. It also shows remarkable stability in inductive settings, which is associated with the detection of new fraudulent users on a rating platform.
- We present an analysis of a rating platform based on expanded balance theory and provide new insight into the behavior of fraudulent users in rating networks.

It is worth noting that the proposed method maintains high performance even when few edges are observed. In practical cases, the proposed framework is useful when the owner of a rating platform must cope with the detection of new fraudulent users quickly, as well as when applied to general detection tasks.

The rest of this paper is organized as follows: In Section II, we review existing studies related to our work. We then describe our proposed framework in Section III, followed by Section IV, in which we discuss the experimental results and present a discussion. Finally, we conclude in Section V.

II. RELATED WORK

A. Fraudulent User Detection on Online Rating Platforms

Several efforts have been made in exploring the detection of fraudulent users. Existing research can be categorized into two major approaches: 1) Feature-based approaches and 2) Graph-based approaches.

1) *Feature-based Approaches*: Feature-based approaches attempt to create a feature that represents each user's behavior. Features from texts [13], [14] and timestamps [15], [16] have been frequently studied. More related to our work, another work [17] modeled consensus, or the wisdom of crowds, from rating values on reviews. Although those approaches show good interpretability, they are not easy to generalize to a different platform or domain.

2) *Graph-based*: Graph-based approaches use the relations among the entities in a rating network. Approaches based on belief propagation [18] or iterative algorithms [5], [19] have been common. The most successful approach is Rev2 [5], an iterative algorithm with theoretical guarantees that outperforms the existing methods, including feature-based approaches. It deals with a rating network as a signed network.

B. Graph Convolutional Networks

The application of neural networks on graph data has seen recent rapid development. Taking rating networks as edge-attributed networks, the Relational Graph Convolutional Network (R-GCN) [9] appears related to our task of detecting fraudulent users. R-GCN can incorporate different discrete types of edges in the graph.

As for the efforts on signed networks, the Signed Graph Convolutional Network (SGCN) [10] showed good performance in the link sign prediction task [20]. In SGCN [10], they claimed that special attention is needed to handle negative

edges correctly in GCNs and then designed convolutional operations based on balance theory [11], [12], a social theory for signed networks that has shown its effectiveness in related tasks such as signed network embeddings [21], [22].

Balance theory formulates the intuition that “the friends of my friends are friends, and the enemies of my friends are enemies”. In SGCN, they applied this theory to infer the relationship (“Friends” or “Enemies”) between entities without the direct edge. Their implementation of balance theory makes it possible to incorporate signs of edges. However, there remains the problem of how to handle the edge directionality properly. In SGCN, the final representations of each node are informed by limited types of neighbors.

III. PROPOSED FRAMEWORK

One of our key motivations is to build a classification model that properly incorporates the sign and direction of edges. To achieve that, we propose an expansion of balance theory, and design convolutional operations based on the expanded balance theory following the way in SGCN [10] (hereinafter, referred to as the “Original SGCN”). In this section, we provide the basic theory behind the proposed framework, followed by the formulas in which the theory is incorporated.

A. Expanded Balance Theory for Signed Directed Networks

We hypothesize that the basic idea of balance theory can be effectively applied to the task of node classification on online rating platforms. However, considering the properties of rating networks, where a different edge direction implies a different relation between nodes, it could be necessary to provide more detailed definitions of “Friends” and “Enemies” than that of the Original SGCN. Below, we introduce our expansion of balance theory that establishes eight distinct types of relationships between 2-hop neighbors.

An edge has two binary attributes (sign and direction), so there are four possible patterns in an edge. Then, taking 2-hop neighbors into account, there are 16 patterns of combinations. Consequently, as shown in Fig. 1, we can define eight different types of “Friends” and “Enemies.”

Fig. 1. Inference of relationships between 2-hop neighbors by proposed definition.

With this new definition, we can distinguish four different meanings each for “Friends” and “Enemies.” For example,

we can explain that “Out-Out Friends” (hereinafter, “O/O F”) is derived from the intuition “Someone liked by my favorite person is my friend” (as well as its opposite, replacing ‘liked’ and ‘favorite’ with their antonyms).

B. Flow of Model Construction

The proposed framework designs the convolutional operation based on the newly introduced definitions shown in Fig. 1, which enables the models to take both the signs and directions of the edges into account, and to detect various types of interactions.

In the rest of this section, we describe how we construct the initialized vectors of each node, followed by detailed explanations and the formulas of our framework. Main notations of this paper are shown in Table I.

TABLE I
NOTATIONS

Notation	Description
$\mathcal{U} = \{u_1, \dots, u_n\}$	set of nodes
$sign \in \{+, -\}$	sign of edge
$dir \in \{out, in\}$	edge direction
x_i	initialized vector of u_i
$N_i^{sign,dir}$	neighbors of u_i along with edges of sign and direction
$h_i^{sign,dir}$	1-st layer hidden representation of u_i
$agg \in AGG$	aggregator in the 2nd layer

1) *Initialized Vectors*: We construct the initialized vector of each node from its rating distribution. We count the rating values that the nodes gave to others and were given by others separately, and then normalize each, followed by concatenation.

2) *Details and Formulas*: Following the Original SGCN, our framework has two convolutional layers and one fully connected layer.

In the 1st convolutional layer, we construct four representation vectors for each node. Each vector is calculated by the aggregation of the neighbor nodes along with the four patterns of edges (derived from two binary attributes of edges: signs and directions). In each process, we first average the initialized vectors of the corresponding neighbors and then concatenate the averaged vector with the initialized vector of the node itself, followed by transformation by the corresponding weight matrix and non-linear activation function. The output of this 1st layer is four vector representations of u_i , which are denoted as $h_i^{+,out}$, $h_i^{-,out}$, $h_i^{+,in}$, $h_i^{-,in}$. With $h_i^{sign,dir}$ as the notation of the four vectors in one statement, we formally define the 1st layer as follows:

$$h_i^{sign,dir} = \tanh \left(W_1^{sign,dir} \left[\sum_{j \in N_i^{sign,dir}} \frac{x_j}{|N_i^{sign,dir}|}, x_i \right] \right) \quad (1)$$

where $\tanh()$ is hyperbolic tangent function and $N_i^{sign,dir}$ denotes the set of u_i itself and the neighbors connected with

u_i by the corresponding edges. Each $W_1^{sign,dir} \in \mathbb{R}^{d_1 \times 2d_0}$ is a linear transformation matrix to be learned, corresponding to the sign and direction of edges. Note that d_1 is the dimension of $h_i^{sign,dir}$.

In the 2nd convolutional layer, with the input of hidden representations after the first layer $\{h_i^{sign,dir} | i = 1, 2, \dots, n\}$, we get eight types of vectors for each node. Here, the eight types correspond to the all eight kinds of relations from a node to a 2-hop neighbor (shown in Fig. 1). In this layer, we have eight independent aggregators that work differently. Each is responsible for aggregating certain types of neighbors' output of the 1st layer, along with certain types of edges. Here, we define $AGG = \{agg\} = \{O/O F, O/I F, I/O F, I/I F, O/O E, O/I E, I/O E, I/I E\}$ as a set of eight aggregators. In this layer, the objectives of aggregation (i.e., along with what kind of edge and which representation of $h_i^{sign,dir}$ to select) differ depending on the aggregators. Furthermore, each aggregator has two different objectives to aggregate that are associated with the signs. Here, we define two map functions denoted by $\mathcal{N}_i$ and h'_i , both of which take agg and $sign$ as arguments, and return the type of neighbors and the representation to aggregate, respectively. The two functions are defined in Table II.

TABLE II
DEFINITION OF TWO MAP FUNCTIONS. $\overline{sign}$ MEANS THE COMPLEMENT OF THE SIGN.

$agg \in \mathbf{AGG}$	Map Functions	
	$\mathcal{N}_i(agg, sign)$	$h'_i(agg, sign)$
O/O F	$N_i^{sign,out}$	$h_i^{sign,out}$
O/I F	$N_i^{sign,out}$	$h_i^{sign,in}$
I/O F	$N_i^{sign,in}$	$h_i^{sign,out}$
I/I F	$N_i^{sign,in}$	$h_i^{sign,in}$
O/O E	$N_i^{sign,out}$	$h_i^{\overline{sign},out}$
O/I E	$N_i^{sign,out}$	$h_i^{\overline{sign},in}$
I/O E	$N_i^{sign,in}$	$h_i^{\overline{sign},out}$
I/I E	$N_i^{sign,in}$	$h_i^{\overline{sign},in}$

Note that we obtain the output representations of this layer z_i as the concatenation of vectors from the eight aggregators. Now we can formally define the 2nd layer as follows:

$$z_i^{(agg,sign)} = \sum_{j \in \mathcal{N}_i(agg,sign)} \frac{h'_j(agg,sign)}{|\mathcal{N}_i(agg,sign)|} \quad (2)$$

$$z_i^{agg} = \tanh(W_2^{agg} [z_i^{(agg,+)}, z_i^{(agg,-)}, h'_i(agg,+), h'_i(agg,-)]) \quad (3)$$

$$z_i = \parallel_{agg \in AGG} z_i^{agg} \quad (4)$$

where $W_2^{agg} \in \mathbb{R}^{d_2 \times 4d_1}$ is the transformation matrix to be learned in this layer.

Finally, at each node, we feed z_i to the fully connected layer, getting the probability that u_i is a fraudulent user. We then obtain the final output as follows:

$$\hat{y} = \text{sigmoid}(W_3 Z) \quad (5)$$

where $W_3 \in \mathbb{R}^{1 \times 8d_2}$ is the weight matrix to be learned and Z is the final embedding matrix of nodes, composed of $\{z_i | i = 1, 2, \dots, n\}$. The cross-entropy between $\hat{y}$ and the ground truth label is minimized.

In comparing the proposed framework and the Original SGCN, we can briefly summarize the difference: the final representation of a node can be informed by all eight types of 2-hop relationships in the proposed framework, while only two (“O/O F” and “O/O E”) were used in the Original SGCN.

IV. EXPERIMENTS

A. Datasets

We adopted the four real-world rating networks used in the experiments of Rev2 [5]: Bitcoin OTC, Bitcoin Alpha, Amazon, and Epinions. The definition of ground truth (i.e., whether some users are fraudulent or benign) also follows their experiments. Note that we preprocess the Epinions dataset in our experiment by extracting a subgraph of random sampling nodes in terms of computational cost.

B. Settings

1) *Proposed Framework*: After calculating the initialized vector of each node, we need to convert raw rating networks (i.e., edges have raw rating values) to signed networks, since the proposed framework requires signed networks as input. We simply divide all edges of rating networks into two classes (positive and negative) using as a threshold the center of possible rating values in each dataset. In Amazon the set of possible ratings is $\{1, 2, 3, 4, 5\}$, so we truncated the edges with a rating value of 3.

For model selection, we train using two-thirds of the given training data, and then select the best model with the remaining one-third. The performances reported below are for the test data held out from the training data. All other GCN-based models follow this procedure. As for hyperparameters, dimension of representation vectors for a relationship (denoted as d_1, d_2) is set to 32, following the Original SGCN.

2) *Baselines*: We compare performances between the proposed framework and the following three methods in our experiments.

- Rev2 [5] is the state-of-the-art method for fraudulent user detection in all four datasets in this experiment.
- R-GCN [9] is a Graph Convolutional Network model for edge-attributed networks. Here we conducted the experiments in three variants of R-GCN. In two of them, we use raw rating values as the discrete edge types, and the base is set 3 or the same as unique count of rating values. The other takes rating networks as signed network. We report the results from the best variants of each dataset.

- Original SGCN [10] is a GCN-based model for signed networks. As described in Section III-A, limited types of relationships between nodes are considered unlike the proposed framework. Although it’s originally designed for link sign prediction, we use the final embedding to the node classification task where parameters are learned jointly.

C. Results

1) *Transductive Settings*: The task is to classify the users on rating networks into two classes (fraudulent/benign) in transductive settings, where the nodes used to classify are involved in the graph of the learning phase. Following the settings of [5], we adopt the metrics called Area Under the Curve (AUC) and calculate the average AUC score of 10-fold cross validation. Table III shows the results of our models and baselines described above. We observe that GCN-based models (R-GCN, Original SGCN, and ours) performed better than the existing state-of-the-art baseline (Rev2) in most cases. We also observe that models that utilize balance theory to convolutional operations (Original SGCN and ours) achieved better performance than R-GCN.

TABLE III
AVERAGE AUC VALUES OF 10-FOLD CROSS VALIDATION. IN EPINIONS DATASETS, BIRDNEST [23], WHICH WAS USED IN REV2 [5], IS NOT AVAILABLE IN OUR EXPERIMENT.

Methods	OTC	Alpha	Amazon	Epinions
Rev2	0.893	0.840	0.857	0.854
R-GCN	0.960	0.926	0.818	0.767
Original SGCN	0.994	0.959	0.871	0.959
Proposed	0.996	0.970	0.875	0.973

2) *Inductive Settings*: We also conducted experiments in the inductive setting, which can be regarded as fraud detection towards newcomers to the rating platforms. In each dataset, we first extract a small percentages of reviews generated early and used them as training data. Then, in the test phase, the evaluation was conducted using all reviews. Note that GCN-based models do not require any retraining because the learned convolution kernel can be directly utilized for the inference task.

We compared the proposed framework and the Original SGCN, which performed very well in the transductive setting. Fig. 2 shows the results for the classification of the nodes that did not exist in the training phase. Since we randomly divide the nodes into the training set and the validation set in the training phase, we repeat the training and the evaluation 30 times and record the averaged AUC values in order to reduce the influence of randomness. We observe that the proposed framework performed better in most cases. In particular, there were significant improvements in the case with an extremely low ratio of training data.

Fig. 2. AUC for the inductive settings.

D. Comparison of Models

In this section, we interpret the above results by comparing models. We obtained results for three criteria: 1) effectiveness of GCN-based end-to-end approaches, 2) effectiveness of utilizing balance theory, and 3) effectiveness of incorporating directions of edges.

1) Effectiveness of GCN-based end-to-end approaches:

Results in the transductive settings indicate the effectiveness of GCN-based models in the semi-supervised fraudulent user detection task. Even simple implementation of GCN (R-GCN) achieves comparable performance to Rev2. Although Rev2 can capture the interaction among entities on rating network, extracted representations are not necessarily suitable for distinguishing fraudulent and benign users. This suggests that GCN-based models succeed in extracting distinguishing features from interactions among entities.

2) Effectiveness of utilizing balance theory:

In transductive settings, SGCN-based models (Original SGCN and ours) outperformed R-GCN in all cases. Significantly, this implies that considering the sign of edges by applying balance theory greatly improves the performance of fraudulent user detection on rating networks.

3) Effectiveness of incorporating directions of edges:

Although our expansion provides no significant improvement in transductive settings, the proposed framework does outperform the Original SGCN in most cases of inductive settings. To uncover the reason for this, we focus on the Bitcoin OTC dataset, in which the difference of AUC in inductive settings was the greatest between the Original SGCN and the proposed framework.

We show how fraudulent and benign users have relationships shown in Fig. 1 with their 2-hop neighbors in 1) the entire Bitcoin OTC network (Fig. 3) and 2) the earliest 5% edges of Bitcoin OTC network (Fig. 4). Note that legends indicate the permutation of the first edges' sign (from source to 1-hop) and the second edges' sign (from 1-hop to 2-hop). For example, "P_N" indicates that the first edge is positive and the second edge is negative.

We observe that edges in the early days (Fig. 4) are quite limited in variety compared to the complete network shown in Fig. 3. In particular, it is significant that relationships containing negative edges (such as "Enemies" relationships or "N_N" relationships in "Friends") do not appear. In terms of model comparison, the Original SGCN considers only "O/O F" and "O/O E", while the proposed framework can incorporate all types. In transductive settings, where all nodes and edges

Fig. 3. Number of 2-hop relationships in Bitcoin OTC. These are separately counted depending on whether the source user is fraudulent or benign.

Fig. 4. Number of 2-hop relationship in the earliest 5% edges of Bitcoin OTC. These are separately counted depending on whether the source user is fraudulent or benign.

can be used in the training phase, interactions with "O/O F" and "O/O E" already appear characteristic enough, as shown in Fig. 3. We can assume that is the reason why there is no significant difference between the Original SGCN and the proposed framework in transductive settings.

In inductive settings with a low training edge ratio, where observable relationships in training data can be quite limited as shown in Fig. 4, it may be necessary to incorporate interactions exhaustively, including those that cannot be captured by the Original SGCN. The results shown in Fig. 2 suggest that fraudulent users have distinctive features, even in the early stage (i.e., before increasing negative edges), and that the proposed framework succeeds in extracting them while the Original SGCN fails.

In general, the proposed framework can be effectively used when observable relationships are limited and distinctive relationships among users are unknown. In a practical case, the owner of a rating platform can detect new fraudulent users quickly. Since the proposed framework uses the learned parameter in classifying and requires only a rating network as input, there is no need for retraining of the detection task, even for the newcomers. That is a promising advantage compared to the existing methods such as Rev2, which require an additional iterative learning process to evaluate newcomers.

E. Insight into the behavior of fraudulent users

Analysis based on 16 types of 2-hop relationships derived from our expanded balance theory reveals that fraudulent users early on (Fig. 4) have few negative edges from their 1-hop neighbors, while they have a lot in the grown network (Fig. 3). Furthermore, it is counterintuitive that they have many “I/I F” relationships, which correspond to the case when a fraudster is trusted by someone who is trusted by another, even in the grown network. In this section, we aim to uncover how and when those relationships increase as the network grows.

Following the previous section, we focus on the Bitcoin OTC network and analyze how fraudulent users are rated by others. In Fig. 5, we show transitions in the number of 1) positive edges from fraudulent users to fraudulent users and 2) negative edges from benign users to fraudulent users, as the network grows. It is indicated that the positive edges between fraudulent users increase in a short time corresponding to the increase of negative edges from benign users to fraudulent ones. This implies a behavior of fraudulent users that they form a faction to deal with the negative ratings from other users. This results imply that incorporating time series with interactions among users is likely to improve performance.

Fig. 5. Transition graph of the positive edges from fraudulent to fraudulent and negative edges from benign to fraudulent

As indicated above, expanded balance theory proposed in this paper is helpful as an analytical framework and can lead to a useful insight. Applying the theory to provide more detailed analyses on rating networks is one of our future directions.

V. CONCLUSION

In this paper, we investigated the task of fraudulent user detection on rating networks. We proposed an end-to-end GCN-based framework using expanded balance theory, which effectively incorporates both the signs and directions of edges. Experimental results show that the proposed framework performs better, or even best, in most settings. In particular, this framework shows remarkable stability in inductive settings, which is associated with the detection of new fraudulent users on rating platforms. In practical cases, our framework helps the owner of a rating platform detect fraudulent users earlier and constantly provide users with more credible information. We also analyzed a rating network using our proposed theory and provided a new insight into fraudulent users on rating platforms, that fraudulent users form a faction to deal with the

negative ratings from other users. For our future work, we will explore the way to incorporate time series with interactions among users in a GCN-based framework. Furthermore, we will tackle more detailed analyses using expanded balance theory for a better understanding of rating networks.

REFERENCES

- [1] Lappas, Theodoros, Gaurav Sabnis, and Georgios Valkanias. "The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry." *Information Systems Research* 27.4 (2016): 940-961.
- [2] Luca, Michael, and Georgios Zervas. "Fake it till you make it: Reputation, competition, and Yelp review fraud." *Management Science* 62.12 (2016): 3412-3427.
- [3] Kumar, Srijan, and Neil Shah. "False information on web and social media: A survey." *arXiv preprint arXiv:1804.08559* (2018).
- [4] Akoglu, Leman, Hanghang Tong, and Danai Koutra. "Graph based anomaly detection and description: a survey." *Data mining and knowledge discovery* 29.3 (2015): 626-688.
- [5] Kumar, Srijan, et al. "Rev2: Fraudulent user prediction in rating platforms." *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 2018.
- [6] Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." *arXiv preprint arXiv:1609.02907* (2016).
- [7] Velickovi, Petar, et al. "Graph attention networks." *arXiv preprint arXiv:1710.10903* (2017).
- [8] Kipf, Thomas N., and Max Welling. "Variational graph auto-encoders." *arXiv preprint arXiv:1611.07308* (2016).
- [9] Schlichtkrull, Michael, et al. "Modeling relational data with graph convolutional networks." *European Semantic Web Conference*. Springer, Cham, 2018.
- [10] Derr, Tyler, Yao Ma, and Jiliang Tang. "Signed graph convolutional networks." *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018.
- [11] Heider, Fritz. "Attitudes and cognitive organization." *The Journal of psychology* 21.1 (1946): 107-112.
- [12] Cartwright, Dorwin, and Frank Harary. "Structural balance: a generalization of Heider's theory." *Psychological review* 63.5 (1956): 277.
- [13] Fayazi, Amir, et al. "Uncovering crowdsourced manipulation of online reviews." *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. ACM, 2015.
- [14] Sandulescu, Vlad, and Martin Ester. "Detecting singleton review spammers using semantic similarity." *Proceedings of the 24th international conference on World Wide Web*. ACM, 2015.
- [15] Xie, Sihong, et al. "Review spam detection via temporal pattern discovery." *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012.
- [16] Minnich, Amanda J., et al. "Trueview: Harnessing the power of multiple review sites." *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015.
- [17] Lim, Ee-Peng, et al. "Detecting product review spammers using rating behaviors." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.
- [18] Akoglu, Leman, Rishi Chandu, and Christos Faloutsos. "Opinion fraud detection in online reviews by network effects." *Seventh international AAAI conference on weblogs and social media*. 2013.
- [19] Wang, Guan, et al. "Identify online store review spammers via social review graph." *ACM Transactions on Intelligent Systems and Technology (TIST)* 3.4 (2012): 61.
- [20] Leskovec, Jure, Daniel Huttenlocher, and Jon Kleinberg. "Predicting positive and negative links in online social networks." *Proceedings of the 19th international conference on World wide web*. ACM, 2010.
- [21] Leskovec, Jure, Daniel Huttenlocher, and Jon Kleinberg. "Signed networks in social media." *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2010.
- [22] Kim, Junghwan, et al. "Side: Representation learning in signed directed networks." *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2018.
- [23] Hooi, Bryan, et al. "Birdnest: Bayesian inference for ratings-fraud detection." *Proceedings of the 2016 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2016.