frontiers | Frontiers in Artificial Intelligence

REVIEW
published: 20 June 2022
doi: 10.3389/frai.2022.922589

Graph Learning for Fake Review Detection

Shuo Yu[1], *Jing Ren*[2], *Shihao Li*[1], *Mehdi Naseriparsa*[3] *and Feng Xia*[2]*

[1] School of Software, Dalian University of Technology, Dalian, China, [2] Institute of Innovation, Science and Sustainability, Federation University Australia, Ballarat, VIC, Australia, [3] Global Professional School, Federation University Australia, Ballarat, VIC, Australia

Fake reviews have become prevalent on various social networks such as e-commerce and social media platforms. As fake reviews cause a heavily negative influence on the public, timely detection and response are of great significance. To this end, effective fake review detection has become an emerging research area that attracts increasing attention from various disciplines like network science, computational social science, and data science. An important line of research in fake review detection is to utilize graph learning methods, which incorporate both the attribute features of reviews and their relationships into the detection process. To further compare these graph learning methods in this paper, we conduct a detailed survey on fake review detection. The survey presents a comprehensive taxonomy and covers advancements in three high-level categories, including fake review detection, fake reviewer detection, and fake review analysis. Different kinds of fake reviews and their corresponding examples are also summarized. Furthermore, we discuss the graph learning methods, including supervised and unsupervised learning approaches for fake review detection. Specifically, we outline the unsupervised learning approach that includes generation-based and contrast-based methods, respectively. In view of the existing problems in the current methods and data, we further discuss some challenges and open issues in this field, including the imperfect data, explainability, model efficiency, and lightweight models.

Keywords: graph learning, fake review detection, anomaly detection, social computing, data science

OPEN ACCESS

Edited by:
Tyler Derr,
Vanderbilt University, United States

Reviewed by:
Xinyi Zheng,
Carnegie Mellon University, United States
Kamal Berahmand,
Queensland University of Technology, Australia

***Correspondence:**
Feng Xia
f.xia@ieee.org

Specialty section:
This article was submitted to Frontiers in Machine Learning and Artificial Intelligence, a section of the journal Frontiers in Artificial Intelligence

Received: *18 April 2022*
Accepted: *23 May 2022*
Published: *20 June 2022*

Citation:
Yu S, Ren J, Li S, Naseriparsa M and Xia F (2022) Graph Learning for Fake Review Detection. Front. Artif. Intell. 5:922589. doi: 10.3389/frai.2022.922589

1. INTRODUCTION

With the prosperity of web services and social networks, e-commerce nowadays has gained tremendous popularity among a wide range of people. When this prosperity significantly boosts the sales of online businesses, it also leads to a considerable number of malicious sellers and spam activities on business websites. Driven by the monetary incentives, they utilize the vulnerability of online websites to make more profits, such as using fake identities in online review systems; thus, manipulating the reputation of the products and brands (Li et al., 2021a; Shehnepoor et al., 2021; Wang L. et al., 2021; Yang et al., 2021). Research on fake reviews have grown noticeably in the last 5 years, and **Figure 1** presents the research status by showing the number of papers with keyword "fake reviews" on SCI, EI, and DBLP, respectively. The anonymity characteristic of the Internet makes it challenging for the websites to handle fake reviews, and the high volume of freelancers and botnets worsen this situation (Wang et al., 2019; Wen et al., 2020; Hou et al., 2021; Li et al., 2021b). Moreover, fake reviewers adopt camouflage techniques to hide their identities (Hooi et al., 2016). These put forward the challenging problem of fake review detection.

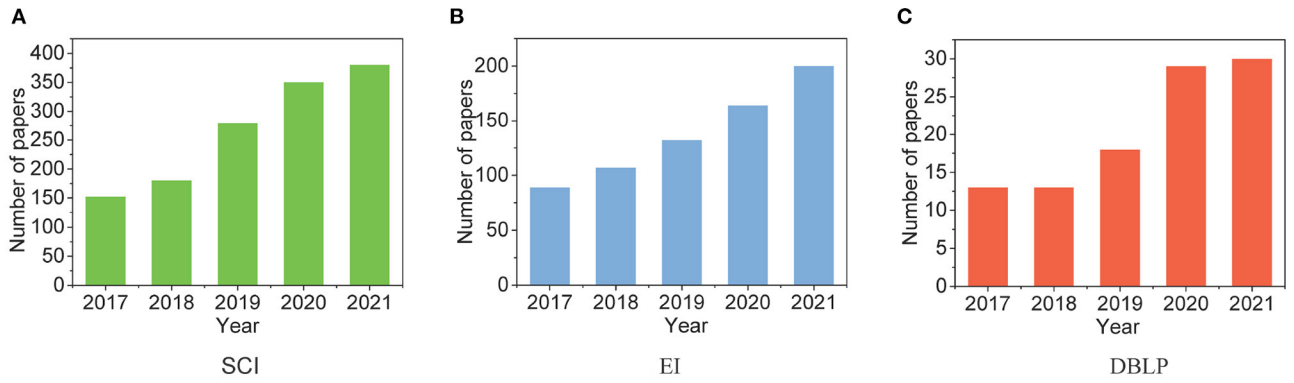

FIGURE 1 | Changes in the number of papers related to fake reviews in the past 5 years. **(A)** SCI. **(B)** EI. **(C)** DBLP.

Previous studies focus on distinguishing the content of news and learning text or image features to investigate fake review detection (Yuan et al., 2019; Branco et al., 2020). The rich information on the social context of reviews is then extracted and further analyzed (Yuan et al., 2017). Though context-based methods have made some progress, the explicit and implicit correlations among users and review contents are still unexplored. The effectiveness and accuracy of fake review detection are thus limited. As a result, graphs are built to represent interactive and complex relationships in fake review detection tasks. These methods utilize graph-structured data to formulate a binary relationship between reviews and reviewers, and they achieve significant performance in fake review detection. Despite the progress in the current studies, fake review detection still faces several challenges. Many models only focus on pairwise relationships and ignore the higher-order relationships. Moreover, though bipartite graphs are constructed, the representation of multi-source data is still unexplored.

Meanwhile, Graph Learning refers to the applications of machine learning models on graph data. With the rapid development in recent decades, graph learning has proven to be of great significance because of its wide applications. Graph learning-based fake review detection has thus been proposed and studied extensively in recent years. There are the following benefits of graph learning-based fake review detection. Firstly, most data in fake review detection contain rich relationships with each other; thus, the graphs effectively leverage the inter-connectivity in these real-world data (Yuan et al., 2017; Yu et al., 2017; Ma et al., 2022). Graphs powerfully capture the correlations among inter-dependent data objects. This nature is even more obvious in online review systems where users, items, attributes, and context are tightly associated with and impact each other by relations (Jerripothula et al., 2020; Rossi et al., 2020; Wang et al., 2020; Liu et al., 2021d). A variety of graphs are generated from data in review systems, and they significantly improve the performance of fake review detection. Second, graph learning effectively learns complicated relations and extracts knowledge from different kinds of graphs (Xu et al., 2020; Wang W. et al., 2021; Xia et al., 2021a). The objective of graph learning is to extract required features from graphs; then, the graph representation is applied for specific tasks (Guo et al., 2021; Xia and Ku, 2021; Xia et al., 2021b; Liu J. et al., 2022; Wang, 2022). In detail, many graph learning techniques, such as graph neural networks (GNNs), have been developed to learn the specific type of relations in the graph models and have been proven effective. Therefore, it is sensible to employ graph learning to model various relations in online review systems.

This paper presents the first literature survey of graph learning techniques for fake review detection. Though there have been surveys and reviews about graph-based anomaly detection (Akoglu et al., 2015; Pourhabibi et al., 2020) and deep learning-based graph anomaly detection (Ma et al., 2021), none of them focus on anomaly detection's down-stream application-fake review detection. While, Istanto et al. (2020) reviewed the fake review detection techniques published between 2015 and 2019, our survey focuses on graph learning's applications in one specific area of anomaly detection. Furthermore, our work focuses on techniques with graph learning and covers a wide range of time.

1.1. Contributions

The main contributions of this paper are summarized below:

- We provide a comprehensive analysis of the key challenges in graph learning-based fake review detection to assist readers with a better understanding of this downstream task.
- We summarize the current research progress in graph learning-based fake review detection, including supervised and unsupervised methods.
- We share and discuss significant future directions of graph learning-based fake review detection by summarizing open issues and challenges.

The rest of this paper is organized as follows: Section 2 explains the categories of fake reviews and presents a comprehensive study of the recent literature on the fake review issue. Section 3 reviews the graph learning methods for fake review detection. Section 4 summarizes the benchmark datasets used in fake review detection task. Section 5 analyzes the open issues and possible

TABLE 1 | Examples of fake reviews.

Types of fake review	Definition	Example
Untruthful opinions	These reviews intentionally misguide users of the review system by unjustly reviewing and rating target objects to manipulate the products' reputation.	(1) This little place in Soho is wonderful. World-class service. (2) Their artichoke chicken salad is the worst in NY.
Exclusive reviews	These reviews are given exclusively to specific brands, manufacturers, or sellers.	(1) The food is amazing! My friends and me are definitely coming back to this place. (2) Delicious, consistent, well-priced. Feels like its made with love.
Non-reviews	Non-reviews include two main sub-streams: (1) Advertisements and (2) Irrelevant content without opinions.	(1) Register to receive a gift. (2) akhdbfl (garbled)
Duplicates reviews	Different accounts post duplicate or near-duplicate reviews on products, either the same or different.	(1) Really charming. It is a great place to have a low-key lunch. (2) The food is simple and effective you should go. (3) It is a great place to have a low-key lunch. (4)The food is simple and effective-you should go.

future directions of fake review research. Finally, Section 6 concludes this paper.

2. THE STUDY OF FAKE REVIEWS

In this section, we first illustrate the typical examples of fake reviews as well as the existing categories for fake reviews. Then, we present a comprehensive study on the fake review issue and discuss our taxonomy for fake review detection approaches.

2.1. Categories of Fake Reviews

According to previous studies (Jindal and Liu, 2008; Li A. et al., 2019) and our summary, we categorize the fake reviews based on two factors: (a) their content and (b) their purposes. Definitions and examples of fake reviews are given in **Table 1**.

2.1.1. Untruthful Opinions

These reviews intentionally misguide the system users by unjustly reviewing and rating target objects to manipulate the products' reputation. The ratings are extremely high or low, and the review contents are either high praise or unfavorable comments.

2.1.2. Exclusive Reviews

These reviews target specific brands, manufacturers, or sellers. They are considered fake reviews whether they seem useful or not. That is because they are biased toward their targets and not objective enough.

2.1.3. Non-reviews

Non-reviews include two main sub-streams: (1) advertisements and (2) irrelevant contents without opinions (e.g., questions, answers, and random texts).

2.1.4. Duplicates Reviews

These are clearly spam. For instance, different accounts post duplicate or near-duplicate reviews on products, either the same or different.

2.2. A Taxonomy of Fake Review Detection Approaches

As illustrated in **Figure 2**, we summarize the fake reviews into three types, including fake review detection, fake reviewer detection, and fake review analysis. **Figure 2** summarizes comprehensive research around the taxonomy of fake review detection approaches. We will introduce these approaches in detail next.

2.2.1. Fake Review Detection

Online reviews play a pivotal role in consumers' decision-making. However, the existence of some fake reviews seriously misleads consumers' choices of products. Therefore, a lot of works are devoted to studying effective fake review detection. However, due to the high cost of manual data tagging, well-labeled reviews are very scarce. To address this problem, Rayana and Akoglu (2016) propose a collective opinion spam detection framework, which adopts active inference technology to select valuable nodes for labeling. They mainly design a label selection strategy based on three key characteristics of valuable nodes. They judge the node's value within a small budget so that the node is labeled and utilized for fake review detection. He et al. (2020) leverage Positive and Unlabeled (PU) learning to detect fake reviews, i.e., only a small number of positive samples and a large number of unlabeled samples are used to classify reviews. This scheme avoids the reliance on manually labeled data. Furthermore, they combine user behavior density to analyze fake reviews, which improves detection accuracy.

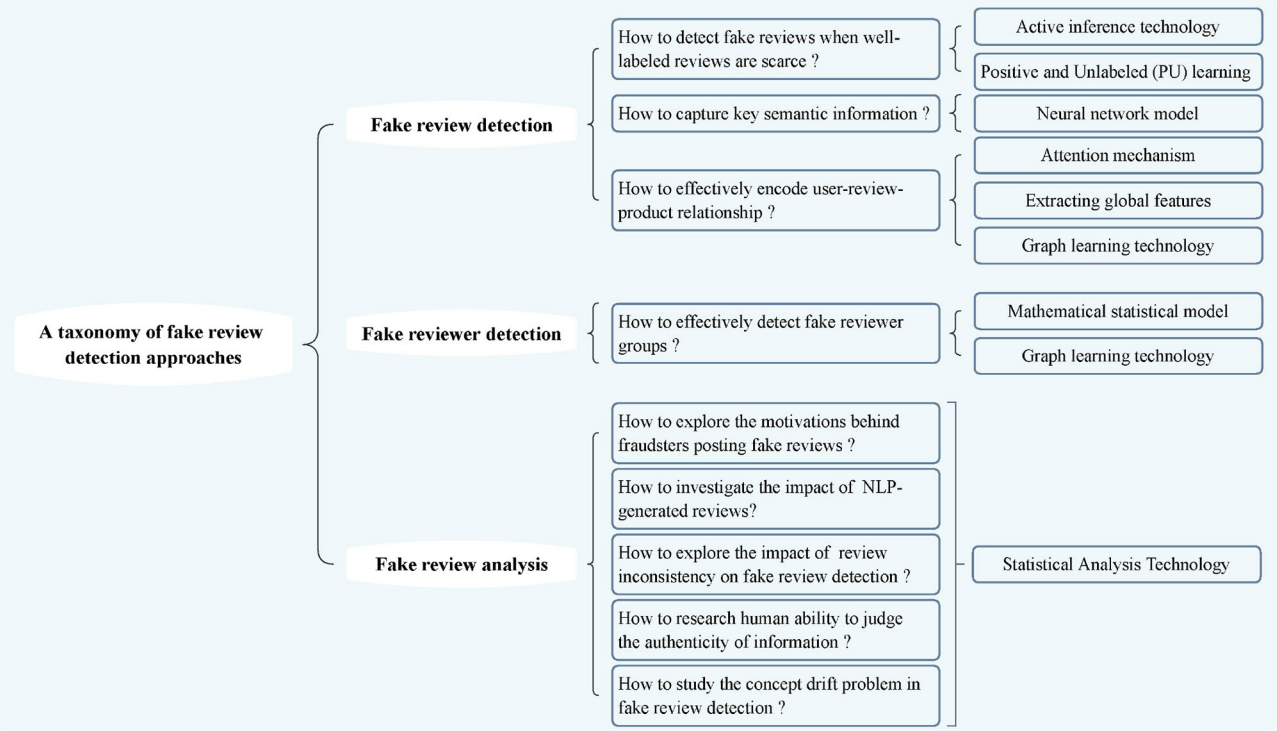

FIGURE 2 | A taxonomy of fake review detection approaches.

In addition, some studies are concerned about how to extract the useful features of reviews to detect fake reviews more accurately. Ren and Zhang (2016) design a neural network model which extracts document features to obtain corresponding representations. Compared with the manual discrete features model, the learn document vectors captured more critical semantic information, which leads to improving the performance of fake review detection. Fahfouh et al. (2020) exploit Paragraph Vector Distributed Bag of Words (PV-DBOW) and the Denoising Autoencoder (DAE) to obtain a global representation of a review. They focus on semantic information in the context of reviews to break the limitations of traditional classifiers. Hajek et al. (2020) explore the importance of hidden emotions contained in review texts. The proposed neural network model analyzes the semantic information in reviews. Also, the model paid attention to emotional features expressed by consumers in reviews when learning review embeddings. However, the above methods only consider the semantic features of the review and ignore its connection with the user and the product.

To address the problem mentioned above, Wang et al. (2016) learn the representation of reviews in a data-driven manner to avoid relying on experts' knowledge. Meanwhile, they combine the reviewer, product, and review text features to learn the representation of the review, which makes full use of global information and improves the performance of fake review detection. Yuan et al. (2019) design a hierarchical fusion attention network (HFAN) to learn the representation of reviews. It first extracts the semantic features of users and products. Then, it generates corresponding representations and encodes the user-review-product relationship to get the final review representation. This work highlights the importance of user and product information for learning review representation. Yu et al. (2019) mainly analyze the behavior of the stakeholders of the reviews and judge the falsity of the reviews. Specifically, they propose three indicators to calculate the fake degree of individuals, groups, and merchants. Then, they integrate them to detect fake reviews. Budhi et al. (2021) combine content and behavior features to detect fake reviews. To obtain global information, they summarize 133 features, including review text, user behavior, and product behavior features, respectively. Meanwhile, they also designed two sampling methods to solve the negative impact of imbalanced datasets on fake review detection.

The above methods have proved that the features of reviews, users, products, and their relationships all play a pivotal role in detecting fake reviews. The construction of the review graph effectively assists us to learn this information. Therefore some researchers applied graph learning for fake review detection, which leads to satisfying results. For instance, Li A. et al. (2019) construct a heterogeneous graph (i.e., Xianyu Graph) and a homogeneous graph (i.e., Comment Graph) to learn the local and global contexts of a review, respectively. Furthermore,

they utilized a GCN-based Anti-Spam (GAS) model to detect fake reviews in Xianyu. Sun and Loparo (2019) employ all heterogeneous data in social networks to detect fake reviews and convert them into classification tasks on heterogeneous information networks. Moreover, Noekhah et al. (2020) present a novel heterogeneous graph (MGSD) model to capture the relationships among entities, and their corresponding weight. They combine multiple features to obtain a new set of features for fake review detection. Various studies have confirmed the effectiveness of the graph learning method for fake review detection.

2.2.2. Fake Reviewer Detection

Driven by lots of profit, some businesses or users are devoted to publishing fake reviews to influence the consumption behavior of the consumers. Such behavior often leads to damaging the trust between businesses and consumers. Therefore, fake reviewer detection has attracted the increasing attention of researchers. For example, Li H. et al. (2017) analyze the number of reviews made by reviewers over a time period and found that they follow a fixed pattern. Specifically, multiple fake reviewers are likely to actively review the same set of products in a short time period (i.e., co-bursting). Thus, they design a two-mode Labeled Hidden Markov Model (LHMM) to detect fake reviewers. Kaghazgaran et al. (2018) utilize the Two-Face system to detect review manipulators. The system identifies the users with similarities to seed users through a so-called suspicious graph. They find that the difference in behavior features of manipulators and regular users is relatively small. Therefore, review manipulators are easier to identify by comparing their social features with regular users. Dhawan et al. (2019) believe that it would cause more dire consequences if fake reviewers collectively post fake reviews; thus, they propose a new framework to detect fake reviewer groups. Besides, Byun et al. (2021) construct a user similarity projection graph and divide the corresponding community. In the next step, they extract the abnormal feature to classify opinion spammers. Xu et al. (2021) firstly constructs the reviewer-projection graph; then, they adopt the Clique Percolation Method (CPM) to detect the opinion spammer group. The above methods demonstrate their superior performance in detecting fake reviewers.

2.2.3. Fake Review Analysis

There have been many studies on fake review analysis to deal with the exponential growth in the number of fake reviews. Luca and Zervas (2016) analyzes fake reviews on Yelp to explore the primary motivation behind the fake reviewers who post fake reviews. For example, they find that chain restaurants are less likely to receive fake reviews when compared to independent restaurants due to their established reputation. This finding helps people understand how a business's reputation affects its motivation for posting fake reviews. Hovy (2016) explores the impact of the application of natural language processing techniques in the generation of fake reviews. They utilize various language models to generate fake reviews based on meta-information and try to detect these fake reviews. Finally, they find that NLP-generated reviews are more difficult to detect; thus, they reflect the dual character of the application of NLP techniques.

In addition, Shan et al. (2021) explore the impact of review inconsistency on fake review detection. They present three types of review inconsistency, including rating-sentiment inconsistency, content inconsistency, and language inconsistency. Based on their findings, the review inconsistency of fake reviews is noticeably high. Therefore, review inconsistency is a fruitful measure to improve the accuracy of fake review detection. Banerjee and Chua (2021) are dedicated to researching users' perception of language nuances. Thus, they invite 380 participants to judge the authenticity of three hotel reviews. The results verify that linguistic cues assist the users in judging the authenticity of reviews to a certain extent. However, the human ability to judge the authenticity of information is almost equivalent to random guessing. Mohawesh et al. (2021) focus on the drift problem concept in fake review detection. They find that the drift problem concept is common in fake review detection and the classifier performance decreases over time which reminds us to update the classifier frequently. Multiple studies on fake review analysis from different perspectives bring us comprehensive thinking and provide new ideas for problem-solving.

3. GRAPH LEARNING FOR FAKE REVIEW DETECTION

Graph Neural Networks (GNNs) have accomplished decent success in many tasks (e.g., node classification, sub-graph classification, graph classification, link prediction) owing to their capability of capturing node attributes and graph structure information. Therefore, many fake review detection methods thoroughly investigate GNNs to utilize their powerful capacities. It should be noted that Section 2 mainly introduces the definitions of different kinds of fake reviews and relevant fake review detection tasks, while this section focuses on graph learning-based methods. Instead of solving one task with one approach, graph learning-based models have their advantages in coping with multiple tasks at the same time.

In this section, GNN-based methods for fake review detection are classified into two categories: (a) supervised and (b) unsupervised. The supervised methods consider fake review detection a binary classification problem, while the unsupervised methods define it as a cluster problem. Here, we firstly summarize the supervised methods; then, we follow our discussion by explaining the unsupervised methods. **Table 2** summarizes the important notations that are used in this paper. **Table 3** summarizes the main characteristics of the graph learning papers for fake review detection. We category the methods to three detection tasks as listed in Section 2.2, wherein, "FRD", "FRerD", and "FRA" represent "Fake Review Detection", "Fake Reviewer Detection", and "Fake review Analysis", respectively. In **Table 3**, we list several representative graph learning methods for fake review detection. These methods are specifically compared in several perspectives. GAS, PC-GNN, IHGAT, AO-GNN are supervised methods, while DeepFD, IN-GNN, and PAMFUL

TABLE 2 | Commonly used notations with explanations.

Notation	Explanation
$\mathcal{G}$	A graph.
$\mathcal{V}$	The set of nodes in a graph.
$\mathcal{E}$	The set of edges in a graph.
$\mathbf{X}$	Node feature matrix of a graph
v_i	A node in the node set V
$e_{i,j}$	An edge in the edge set E
$\mathbf{h}_i$	The node representation vector of node vi.
$\mathcal{C}$	Unlabeled node set.
$\mathbf{Z}$	Output representation of the encoder.
$\mathbf{A}$	The adjacency matrix of a graph.
$\hat{\mathbf{A}}$	The reconstruction adjacency matrix.
$\hat{\mathbf{Z}}$	The reconstruction feature matrix.
$\mathbf{D}$	The node degree matrix.
$\sigma(\cdot)$	Activation function

are unsupervised. Wherein, IHGAT focuses on link level and represents relation, others focus on node level. PAMFUL can only detect fake reviewer while other methods except PC-GNN can detect fake review. PC-GNN can recognize both fake review and fake reviewer at the same time. Among these methods, PC-GNN, AO-GNN, and DeepFD are with better generalization ability (Du et al., 2020; Betlei et al., 2021; Hibshman et al., 2021). Some detailed parameter settings are unspecified and some datasets are not publicly available such as Xianyu Graph, Alibaba Review Graph, and Alibaba Group, thus limiting the repetition of these methods to some extent.

3.1. Supervised Methods

Fake review detection is defined as a task that determines whether a review is fake or not. Therefore, some studies consider it as a binary classification task, which can be defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{C}\}$, where each node in $\mathcal{V}$ has been labeled either fake or not in $\mathcal{C}$. The supervised methods identify the fake nodes that significantly differ from the normal nodes in $\mathcal{G}$. **Figure 3** illustrates a general framework of the supervised methods with graph neural networks for fake review detection. GraphSAGE, proposed by Hamilton et al. (2017) and GCN proposed by Kipf and Welling (2017) are commonly used comment node embedding method. Li A. et al. (2019) adopt Graph Convolutional Networks (GCNs) to address the anti-spam problem at Xianyu (The largest second-hand goods app in China). Their method is called GCN-based Anti-Spam (GAS), which contains two primary inputs: (a) XianYu Graph (heterogeneous graph) and (b) Comment Graph (isomorphic graphs). The GCN-based methods mainly focus on isomorphic graphs. The graph which is utilized in the fake review detection process contains two types of nodes: (a) user and (b) item. Traditional GCN can not be processed directly; however, GAS extends GCNs for heterogeneous graphs to obtain local information and aggregate the information to the edges from the Xianyu Graph. The formula of each layer in GAS is presented as follows:

$$\mathbf{h}_e^l = \sigma(\mathbf{W}_E^l \cdot AGG_E^l(\mathbf{h}_e^{l-1}, \mathbf{h}_{U(e)}^{l-1}, \mathbf{h}_{I(e)}^{l-1})) \tag{1}$$

wherein, $\mathbf{h}_e^l$ is the representation of the edge and $\mathbf{W}_E^l$ is the learning parameter, AGG is the concatenation operation. $\mathbf{h}_e^{\mathbf{h}-1}$, $\mathbf{h}_{U(e)}^{l-1}$, and $\mathbf{h}_{I(e)}^{l-1}$ are the edge, user, and item embeddings from the $l-1$ layer. In addition, GAS considers global information through the Comment Graph. The node in the Comment Graph represents the comment, while the edge conveys the similarity between the comments. Moreover, GCN is utilized to obtain global information. Finally, local and global information are spliced and classified through labeled data sets.

Liu et al. (2021b) propose a method called PCGNN for imbalanced supervised learning. The imbalanced supervised learning proves to be a suitable approach for fake review detection. In this approach, they develop a label-balanced sampler to pick nodes and edges for sub-graph training. Then, they design a neighborhood sampler to choose neighbors for over-sampling the minority class and under-sampling the majority class neighborhoods, respectively. The sampler picks nodes and edges for the construction of the adjacent matrix. In the chosen step, the sampler generates samples for the minority class and under-samples the neighbors in the majority class. The overall loss function is formulated as follows:

$$\mathcal{L} = \mathcal{L}_{gnn} + \alpha \mathcal{L}_{dist}, \tag{2}$$

wherein, $\mathcal{L}_{gnn}$ is the cross-entropy loss of the graph neural network. $\mathcal{L}_{dist}$ denotes the loss for learning the parameter in the neighborhood sampler. α denotes the balanced parameter.

The majority of recent studies have put their efforts into the innovation of the new representation method; however, a few works focus on fake review detection. With the rapid emergence of fake reviews, we encourage more studies on fake review detection in a supervised way.

3.2. Unsupervised Methods

In some circumstances, fake review detection is defined as an unsupervised problem. For example, when background knowledge is not available to mark the data, unsupervised fake review detection methods effectively identify the fake reviews. This problem is defined as a graph $\mathcal{G} = \{\mathcal{A}, \mathcal{E}\}$, where the unsupervised methods learn a mapping function to embed the node features to a latent space. Berahmand et al. (2021) propose a new random walk model to integrate network structure and node attributes, based on the assumption that two nodes on the network will be linked since they are nearby in the network, or connected for the reason of similar attributes. The unsupervised methods detect all fake review nodes in the graph automatically. The detection of fake review nodes is based on the Poisson distribution or fake score of the nodes in a low-dimensional latent space. There are two groups of unsupervised fake review detection methods: (a) generation-based and (b) contrast-based methods, respectively. This categorization takes the different designs of pretext decoders and objective functions into account.

TABLE 3 | Comparative review of graph learning methods for fake review detection.

Methods	Detection task	Task level	Supervised/unsupervised	Scalability	Generalization ability	Datasets
GAS (Li A. et al., 2019)	FRD	Node	Supervised	✓	✗	Xianyu Graph
PC-GNN (Liu et al., 2021b)	FRD& FRerD	Node	Supervised	✓	✓	YelpChi, Amazon, and Alibaba Review Graph
IHGAT (Liu et al., 2021a)	FRD	Link	Supervised	✓	✗	Alibaba Group
AO-GNN (Huang et al., 2022)	FRD	Node	Supervised	✓	✓	YelpChi, Amazon, and Books
DeepFD (Ding et al., 2019)	FRD	Node	Unsupervised	✓	✓	Yelp, Amazon, and DDos
IN-GNN (Liu B. et al., 2022)	FRD	Node	Unsupervised	✓	✗	MisInfdect and Pheme
PAMFUL (Zhao et al., 2022)	FRerD	Node	Unsupervised	✗	✗	Bitcoin-Alpha, Weibo

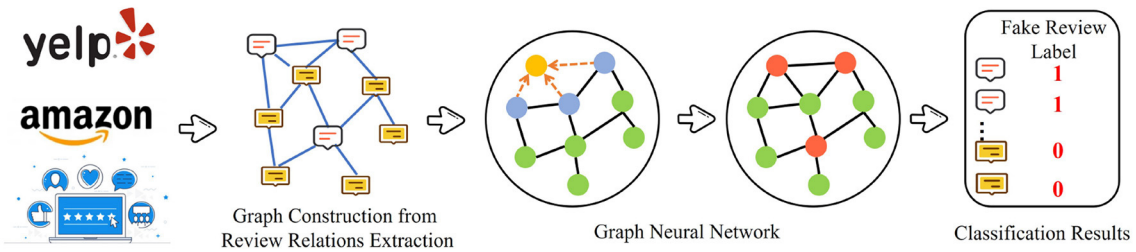

FIGURE 3 | Supervised methods with Graph Neural Network for fake review detection.

The generation-based methods focus on retaining more structural information by reconstructing the input graph, and they minimize the differences between the reconstructed and the input graphs, respectively. However, the contrast-based methods maximize the difference between the two corresponding views. As for the fake review detection, the generation-based methods pay more attention to the detection step after encoding, while the contrast-based methods pay more attention to the design of the discriminator; thus, they directly detect the fake review node.

3.2.1. Generation-Based Methods

The generation-based methods aim to reconstruct and employ the input graph to serve as the supervision labels. The generation-based methods include auto-regressive, flow-based, auto-encoder methods. The detection technique for fake detection always employs the auto-encoder to learn the representation of the node. Fake review detection is formulated as a task that performs the anomaly detection tasks in attributed networks. **Figure 4** illustrates a general framework of the generation-based methods with auto-encoder, which is designed for fake review detection. Wang et al. (2018) view the fake review detection problem as identifying the suspicious dense blocks in the attributed bipartite graph. They propose a deep learning model named *DeepFD* to differentiate between normal and suspicious users. DeepFD contains three primary components. The first component reconstructs the input graph by applying the encoder result of the node. The second component preserves different behaviors among diverse users. These two components preserve the structural information and behavioral characteristics. The last component detects the fake review. In the graph reconstruction component, the loss function is formulated as:

$$\mathcal{L}_{recon} = ||(\hat{S} - S) \odot H||_2^2 \tag{3}$$

wherein, $\odot$ denotes the Hadamard product, $\hat{S} = \{\hat{s}_1, \hat{s}_2, \ldots\ldots, \hat{s}_i\}$, $\hat{s}_i$ presents the learned graph structure of node i, and $S = \{s_1, s_2, \ldots\ldots, s_i\}$, s_i denotes the initial graph structure of node i. H denotes the weight vector. By minimizing this loss function, the node representation preserves global information. In the preservation component of the user behavior, for node i and j, the distance of their embedding and the similarity are defined as follows:

$$dis_{ij} = ||(\mathbf{h}_i^{(K)} - \mathbf{h}_j^{(K)})||_2^2 \tag{4}$$

$$sim_{ij} = exp(-\lambda \cdot dis_{ij}) \tag{5}$$

wherein, $\mathbf{h}_i^{(K)}$ denotes the vector representations of the user node i of layer K, $\mathbf{h}_j^{(K)}$ presents the vector representations of the user node j of layer K. The loss function of this component is formulated as follows:

$$\mathcal{L} = \mathcal{L}_{recon} + \alpha \mathcal{L}_{sim} + \gamma \mathcal{L}_{reg} \tag{6}$$

wherein, $\mathcal{L}_{recon}$ is defined in Equation 3, $\mathcal{L}_{sim}$ denotes the dis_{ij} between all the nodes, $\mathcal{L}_{reg}$ is L2-norm regularizer term. After

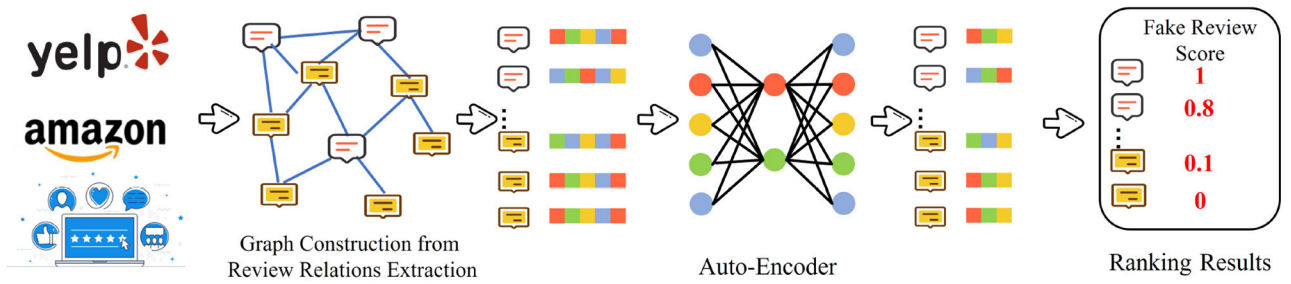

FIGURE 4 | Generation-based methods with Auto-Encoder for fake review detection.

obtaining embedding results, they adopt the DBSCAN algorithm, which is one of the most common density-based clustering algorithms to detect fake reviews.

Some studies adopt an auto-encoder to receive attribute and structure information for detecting anomalies. Ding et al. (2019) adopt two decoders (named structure reconstruction and attribute reconstruction decoders) to decode the result of the encoder. These encoders preserve the structure and attribute information simultaneously. According to the embedding result, this method effectively receives the anomaly score of the node. The method adopts GCN to encode the attribute network, the structure reconstruction decoder is trained by the output of the attributed network encoder Z, the structure reconstruction result is presented as follows:

$$\hat{\mathbf{A}} = sigmoid(\mathbf{Z}\mathbf{Z}^{\mathbf{T}}) \tag{7}$$

The attribute reconstruction decoder utilizes the graph convolutional layer to predict the original node attributes as follows:

$$\hat{\mathbf{X}} = f_{Relu}(\mathbf{Z}, \mathbf{A}|\mathbf{W}^{(3)}) \tag{8}$$

wherein $\mathbf{W}^{(3)}$ denotes the learning parameter. The objective function is formulated as follows:

$$\mathcal{L} = (1-\alpha)||\mathbf{A} - \hat{\mathbf{A}}||_F^2 + \alpha||\mathbf{X} - \hat{\mathbf{X}}||_F^2 \tag{9}$$

wherein, α denotes an important controlling parameter which balances the impacts of structure and attribute reconstructions, respectively. Finally, the anomaly score of each node i is calculated as follows:

$$score(v_i) = (1-\alpha)||\mathbf{h} - \hat{\mathbf{h}}_\mathbf{i}||^2 + \alpha||\mathbf{x}_\mathbf{i} - \hat{\mathbf{x}}_\mathbf{i}||^2 \tag{10}$$

By contrast, Li Y. et al. (2019) propose a spectral convolution and deconvolution-based framework, named *SpecAE*. SpecAE encodes node attributes and topological relations at the same time. This method sharpers the features with their neighbors in order to reconstruct the features. To magnify the difference between the current node and its neighbors, the result of the encoder is formulated as follows:

$$\mathcal{Y} = (1+\alpha)\mathbf{X} - \alpha\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X} \tag{11}$$

The propagation rule of the decoder layer is given as follows:

$$Deconv(\mathbf{Z}, \mathbf{A}) = \sigma((1+\alpha)\mathbf{Z} - \alpha\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{Z})\mathbf{W}_g \tag{12}$$

wherein, $\mathbf{W}_g$ denotes the trainable weight matrix in the deconvolution layer.

Generation-based methods aim to reconstruct the attribute or structure features. In these methods, the input data serve as the supervision signals. In the detection step, all methods utilize the representation of the encoder to calculate the anomaly score of the node and rank the anomalous degree.

3.2.2. Contrast-Based Methods

The contrast-based methods are built on the idea of mutual information maximization. These methods learn representations by contrasting positive instance pairs against negative instance pairs. Contrast-based methods are trained by a specific anomaly detection-aware target. **Figure 5** illustrates a general framework of the contrast-based methods which is designed for anomaly detection.

The success of the contrast-based methods largely relies on the definition of the contrastive instance pair. The contrast-based methods that are utilized for fake review detection mainly focus on the design of instance pairs and the discernibility of positive and negative instances, respectively. Liu et al. (2021c) propose a novel contrastive self-supervised learning framework for anomaly detection on attributed networks, called *CoLA*. The instance pair in CoLA can efficiently capture local information and node attribute. Specifically, they design "target node" vs. "local subgraph". For positive instance pairs, the initial node is set as the target node; then, the sampled subgraph is composed of the neighbor nodes of the target node. For negative instance pairs, the initial node is randomly selected from the list of nodes except the target node. Instance pairs are employed to train the GNN model for anomaly detection. The input of GNN consists of the target node, local subgraph, and label. In the GNN component, CoLA

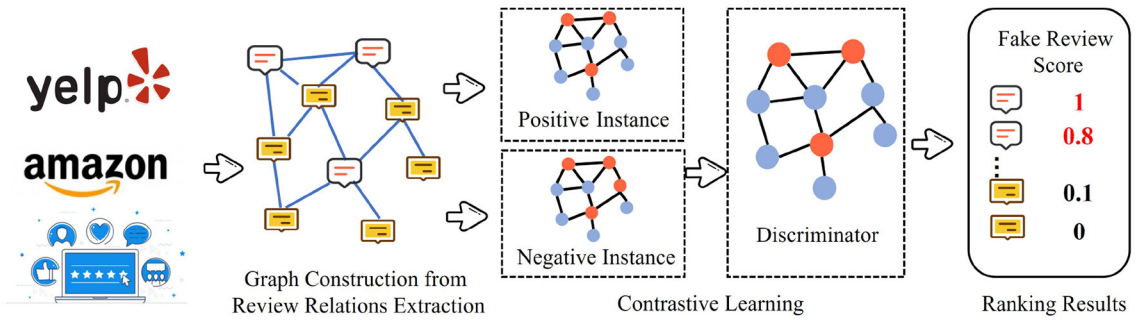

FIGURE 5 | Contrast-based methods for fake review detection.

adopts GCN due to its high efficiency. Moreover, the target node embedding is the output of the GCN model, which is denoted by z_i^{tn}. The local subgraph embedding is presented as follows:

$$e_i^{lg} = Readout(\mathbf{Z}_i) = \sum_{k=1}^{n_i} \frac{(\mathbf{Z}_i)_k}{n_i} \tag{13}$$

Secondly, in the discriminate part, CoLA applies the bilinear scoring function to produce the predicted score of a node which is calculated as follows:

$$s_i = Discriminator(z_i^{lg}, z_i^{tn}) = \sigma(\mathbf{z}_i^{lg}\mathbf{W}^{(d)}\mathbf{z}_i^{tn\,T}) \tag{14}$$

wherein, $\mathbf{W}^{(d)}$ denotes the weight matrix of discriminator, and $\sigma(\cdot)$ presents the logistic sigmoid function. Moreover, the final anomaly score of v_i is obtained by computing the average value of multi-round differences between the scores of negative and positive pairs:

$$f(v_i) = \frac{\sum_{r=1}^{R}(s_{i,r}^{(-)} - s_{i,r}^{(+)})}{R} \tag{15}$$

wherein, $f(\cdot)$ is the mapping function of the anomaly score, which is the goal of anomaly detection. R denotes the number of sampling round. $S^{(+)}$ and S^{-} are positive and negative predicted scores, respectively.

Ding et al. (2020) consider the difference between the nodes for anomaly detection. They propose adversarial graph differentiation networks (AEGIS), which learn anomaly-aware node representations to detect anomalies effectively. For graph differentiative layer l, the representation of the node is learned by the feature difference and node feature itself. The equation is expressed as follows:

$$h_i^{(l)} = \sigma(\mathbf{W}_1\mathbf{h}_i^{(l-1)} + \sum_{j\in\mathcal{N}_i} \alpha_{ij}\mathbf{W}_2\Delta_{i,j}^{(l-1)}) \tag{16}$$

wherein, h_i^{l-1} and h_i^{l-1} are representation of node i in layer i. $\Delta_{i,j}^{(l-1)}$ denotes the feature difference between node i and node j. After learning the anomaly-aware node representations, the second phase aims to train a generative adversarial network. This phase accurately models the distribution of normal data. The loss function, which is defined in generator G and discriminator D are as follows:

$$\mathcal{L}_{\mathbf{G}} = \mathbb{E}_{\tilde{z}\sim p(\tilde{z})}[\log(1 - D(\tilde{G(z)}))] \tag{17}$$

$$\mathcal{L}_{\mathbf{D}} = -\mathbb{E}_{z\sim Z}[\log D(z)] - \mathbb{E}_{\tilde{z}\sim p(\tilde{z})}[\log(1 - D(\tilde{G(z)}))] \tag{18}$$

wherein, z denotes the node representation of normal node and $\tilde{z}$ denotes the generated anomaly. G and D are the generator and discriminator functions, respectively. Finally, the anomaly score of node i is computed as follows:

$$score(x_i^{'}) = p(y_i^{'} = 0|z_i^{'}) = 1 - D(z_i^{'}) \tag{19}$$

4. DATASETS

The graph learning research on the fake review detection has not produced an abundant number of datastets. In this section, we provide a comprehensive review on the existing datasets which are utilised in previous studies. **Table 4** presents the detailed statistics of these datasets.

4.1. Yelp

Yelp's website publishes rich crowd-sourced reviews about businesses. The Yelp dataset captures the relevant data about the businesses, reviews, and users. Specifically, the reviews in the following Yelp datasets contain various items such as product and user information, timestamp, ratings, and a plain text review.

Yelp adopts a filtering algorithm that effectively identifies fake/suspicious reviews. After the identification step, the algorithm stores the identified fake reviews into a filtered list. The filtered reviews are also made public on a business Yelp page. While the Yelp page of a business displays the recommended reviews, it is also possible to view the filtered/unrecommended reviews through a link at the bottom of the page. The Yelp

TABLE 4 | The statistics of fake review datasets.

Datasets	#users	#products	#reviews	Labeled
YelpCHI	38,063	201	67,395	Yes
YelpNYC	160,225	923	359,052	Yes
YelpZip	260,277	5,044	608,598	Yes
Amazon Reviews	34,686,770	6,643,669	2,441,053	No
Amazon FineFoods	256,059	74,258	568,454	No
Amazon Movies	889,176	253,059	7,911,684	No
BeerAdvocate	33,387	66,051	1,586,259	No
RateBeer	40,213	110,419	2,924,127	No
CellarTracker	44,268	485,179	2,025,995	No
SWMReview	966,942	15,094	1,132,373	No
Epinions	49, 290	139, 738	664, 824	No

anti-fraud filter is not perfect (hence the "near" ground truth); however, it has been found to produce accurate results (Weise, 2011). The following Yelp datasets are all labeled datasets that contain both recommended and filtered reviews.

4.1.1. YelpCHI

YelpCHI (Mukherjee et al., 2013) is a labeled dataset that includes 67,395 reviews from 201 hotels and restaurants by 38,063 reviewers in the Chicago area.

4.1.2. YelpNYC

YelpNYC (Rayana and Akoglu, 2015) is a labeled dataset that includes 359,052 reviews from 923 restaurants by 160,225 reviewers in New York City.

4.1.3. YelpZip

YelpZip (Rayana and Akoglu, 2015) is a labeled dataset that includes 608,598 reviews for restaurants, starting with a zipcode number which is increased incrementally. Note that the zipcodes are organized by geography; thus, the reviews for restaurants are ordered in a continuous region of the U.S. map, including NJ, VT, CT, and PA.

4.2. Amazon

Amazon is a retail giant in e-commerce with billions of review data. Amazon dataset was first collected and utilized in McAuley J. and Leskovec (2013), McAuley J. J. and Leskovec (2013). To obtain this enormous data, they started with a list of 75 million asin-like strings (Amazon product identifiers) that they collected from the Internet Archive. Almost around 2.5 million of the strings had at least one review. They further divide this dataset into 26 parts based on the top-level category of each product (e.g., books, movies). The reviews in the Amazon dataset contain various items such as product and user information, ratings, and a plain text review.

4.2.1. Amazon Reviews

Amazon Reviews dataset (McAuley J. and Leskovec, 2013) consists of reviews from Amazon. The dataset includes 34,686,770 reviews from 6,643,669 users on 2,441,053 products, spanning a period of 18 years from June 1995 to March 2013. Note that this dataset contains potential duplicates.

4.2.2. Amazon FineFoods

Amazon FineFoods (McAuley J. J. and Leskovec, 2013) consists of reviews of fine foods from Amazon. The dataset includes 568,454 reviews from 256,059 users on 74,258 products, spanning from October 1999 to October 2012.

4.2.3. Amazon Movies

Amazon Movies dataset (McAuley J. J. and Leskovec, 2013) consists of movie reviews from Amazon. The dataset includes 7,911,684 reviews from 889,176 users on 253,059 products, spanning from August 1997 to October 2012.

4.3. Other Datasets

4.3.1. BeerAdvocate

This dataset consists of beer reviews from BeerAdvocate (McAuley J. J. et al., 2012). The data span a period of more than 10 years, from January 1998 to November 2011, including 1,586,259 reviews from 33,387 users on 66,051 beers. Each review includes ratings in terms of five aspects: appearance, aroma, palate, taste, and overall impression. Reviews include product and user information followed by these five ratings and a plain text review.

4.3.2. RateBeer

This dataset consists of beer reviews from RateBeer (McAuley J. J. et al., 2012). The data span a period of more than 10 years, from April 2000 to November 2011, including 2,924,127 reviews from 40,213 users on 110,419 beers. Each review includes ratings in terms of five aspects: appearance, aroma, palate, taste, and overall impression. Reviews include product and user information followed by these five ratings and a plain text review.

4.3.3. CellarTracker

This dataset consists of wine reviews from CellarTracker (McAuley J. J. and Leskovec, 2013). The data include 2,025,995 reviews from 44,268 users on 485,179 wines. Reviews include product and user information, ratings, and a plain text review.

4.3.4. SWMReview

The SoftWare Marketplace (SWM) Review dataset (Akoglu et al., 2013) was collected by crawling the software product (app) reviews under the entertainment category from a popular online software marketplace. The product apps consist of a diverse set of categories (e.g., games, movies, news, sports). The complete collection includes 1,132,373 reviews from 966,842 unique users for 15,094 apps and spans 198 weeks between July 2008 and April 2012. As part of a review, a user rates a product from 1 (worst) to 5 (best).

4.3.5. Epinions

Epinions (Kumar et al., 2018) is a consumers opinion site where users review items such as cars, books, movies, software, etc. In addition to the normal reviews, the consumers can assign the items numeric ratings between 1 (min) to 5 (max). Users also express their Web of Trust, i.e., a list of reviewers whose reviews

and ratings have been consistently valuable. Moreover, users define Block list, i.e., a list of authors whose reviews have been consistently offensive, inaccurate, or not valuable. The dataset consists of 664,824 reviews from 49,290 users rating 139,738 different items at least once. The total number of trust statements is 487,181.

5. OPEN ISSUES

This section shares key challenges and open issues with respect to the search for fake review detection, including imperfect data, explainability, and lightweight models.

5.1. Imperfect Data

The accuracy and integrity of data are the premise to ensure the effectiveness of graph learning methods. Therefore, in fake review detection, the effectiveness of graph learning highly depends on data quality and data usability. However, the graph learning methods are often negatively affected by imperfect data (e.g., missing data, noise data, imbalanced data, and limited data). Moreover, the extensive participation of users in the review process leads to the omission of information or automatic reviews. These shortcomings cause the review data to be inaccurate and of poor quality; thus, the imperfect data leads to insufficient feature learning. As a result, graph learning outcomes will be correspondingly biased. Therefore, how to develop data-efficient fake review detection methods remains an open issue.

5.2. Explainability

In the practical application of fake review detection, sufficient evidence and reasons are required to indicate that a review is fake. However, commonly such explainability is the missing part of graph learning methods, which have long been criticized for their black-box nature. The graph learning-based model obtains the vector representation of the review by building the relationship between the reviews and the model feature. The detection results are then achieved through the representation. Researchers have been trying to solve the explainability problem of graph learning. However, the existing methods focus on explaining the importance of nodes or relationships in graphs, ignoring the structure factor in graph learning methods, which is more intuitive and straightforward for a human to understand. Therefore, one of the future challenges is to explore the explainability of fake review detection based on graph learning.

5.3. Efficiency

In the fake review detection task, the enormous size of the review data has become a significant problem (Ying et al., 2018). In such big data, the number of nodes and relationships is huge, which increases the cost of training the model. This problem is especially important in fake review detection because the model should embed nodes into low-dimensional space, detect a large number of fake reviews, and detect fake reviewers. Therefore, it is imperative to study more efficient algorithms to speed up the training and detection of large-scale data.

5.4. Lightweight Models

Detecting fake reviews is a comprehensive task. The enterprise is capable of detecting fake reviews by constructing complex graph structures and building graph relationships. Therefore, the detection task often performs a huge number of comparisons through a large amount of data. Furthermore, studies generally focus on how to improve the accuracy of the model by adding more parameters and layers to the models. However, fake review detection should also be performed on ordinary users' hardware or embedded platforms in real life. Since the learning models have numerous parameters and layers, utilizing them on an embedded platform has become a key challenge. To this end, the detection models should be effectively streamlined and optimized; thus, they run smoothly on devices with limited computing and hardware powers.

6. CONCLUSION

We present a survey on fake review detection methods based on graph learning in this paper. Graph-based methods are more advantageous over others because they utilize graph-structured data to formulate a binary relationship between reviews and reviewers. As a result, graph learning achieve significant performance in fake review detection. In this survey, we firstly introduce the various types of fake reviews, including untruthful opinions, exclusive reviews, non-reviews, and duplicate reviews. Moreover, we clarify different types of fake reviews by providing relevant examples for each of them. We categorize the fake review issues into three types: fake review detection, fake reviewer detection, and fake review analysis. Thirdly, we discuss the supervised and unsupervised fake review detection, which utilizes graph learning. The paper discusses the graph learning features, including the representation learning methods, detection methods, and the loss function. We analyze the unsupervised mechanisms, including generation-based and contrast-based models, respectively. Also, this paper presents a summary of the data sets that are utilized for graph-based fake review detection. Finally, we discuss the challenges and open issues, including the imperfect data, explainability, efficiency of the model, and how to propose lightweight models. This survey could be a guide for both junior and senior scholars to study the fake review issue in-depth.

AUTHOR CONTRIBUTIONS

SY and FX contributed to conception and design of the study. All authors contributed to manuscript writing and revision, read, and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant 62102060.

REFERENCES

Akoglu, L., Chandy, R., and Faloutsos, C. (2013). "Opinion fraud detection in online reviews by network effects," in *Proceedings of the International AAAI Conference on Web and Social Media*, Cambridge, MA.

Akoglu, L., Tong, H., and Koutra, D. (2015). Graph based anomaly detection and description: a survey. *Data Mining Knowledge Discov.* 29, 626–688. doi: 10.1007/s10618-014-0365-y

Banerjee, S., and Chua, A. Y. (2021). Calling out fake online reviews through robust epistemic belief. *Inform. Manage.* 58, 103445. doi: 10.1016/j.im.2021.103445

Berahmand, K., Nasiri, E., Rostami, M., and Forouzandeh, S. (2021). A modified deepwalk method for link prediction in attributed social network. *Computing* 103, 2227–2249. doi: 10.1007/s00607-021-00982-2

Betlei, A., Diemert, E., and Amini, M.-R. (2021). "Uplift modeling with generalization guarantees," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21* (New York, NY: Association for Computing Machinery), 55–65. doi: 10.1145/3447548.3467395

Branco, B., Abreu, P., Gomes, A. S., Almeida, M. S. C., Ascensão, J. T., and Bizarro, P. (2020). "Interleaved sequence rnns for fraud detection," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20* (New York, NY: Association for Computing Machinery), 3101–3109. doi: 10.1145/3394486.3403361

Budhi, G. S., Chiong, R., Wang, Z., and Dhakal, S. (2021). Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. *Electron. Commerce Res. Appl.* 47, 101048. doi: 10.1016/j.elerap.2021.101048

Byun, H., Jeong, S., and Kim, C.-K. (2021). Sc-com: Spotting collusive community in opinion spam detection. *Inform. Process. Manage.* 58, 102593. doi: 10.1016/j.ipm.2021.102593

Dhawan, S., Gangireddy, S. C. R., Kumar, S., and Chakraborty, T. (2019). "Spotting collective behaviour of online frauds in customer reviews," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019* (Macao), 245–251. doi: 10.24963/ijcai.2019/35

Ding, K., Li, J., Agarwal, N., and Liu, H. (2020). "Inductive anomaly detection on attributed networks," in *29th International Joint Conference on Artificial Intelligence, IJCAI 2020*, 1288–1294. doi: 10.24963/ijcai.2020/179

Ding, K., Li, J., Bhanushali, R., and Liu, H. (2019). "Deep anomaly detection on attributed networks," in *Proceedings of the 2019 SIAM International Conference on Data Mining, ICDM '21*, Calgary, AL, 594–602. doi: 10.1137/1.9781611975673.67

Du, M., Pentyala, S., Li, Y., and Hu, X. (2020). "Towards generalizable deepfake detection with locality-aware autoencoder," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20* (New York, NY: Association for Computing Machinery), 325–334. doi: 10.1145/3340531.3411892

Fahfouh, A., Riffi, J., Mahraz, M. A., Yahyaouy, A., and Tairi, H. (2020). PV-DAE: a hybrid model for deceptive opinion spam based on neural network architectures. *Expert Syst. Appl.* 157, 113517. doi: 10.1016/j.eswa.2020.113517

Guo, T., Bai, X., Tian, X., Firmin, S., and Xia, F. (2021). Educational anomaly analytics: features, methods, and challenges. *Front. Big Data* 4, 811840. doi: 10.3389/fdata.2021.811840

Hajek, P., Barushka, A., and Munk, M. (2020). Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. *Neural Comput. Appl.* 32, 17259–17274. doi: 10.1007/s00521-020-04757-2

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). "Inductive representation learning on large graphs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS '17*, 1025–1035.

He, D., Pan, M., Hong, K., Cheng, Y., Chan, S., Liu, X., et al. (2020). Fake review detection based on pu learning and behavior density. *IEEE Network* 34, 298–303. doi: 10.1109/MNET.001.1900542

Hibshman, J. I., Gonzalez, D., Sikdar, S., and Weninger, T. (2021). "Joint subgraph-to-subgraph transitions: Generalizing triadic closure for powerful and interpretable graph modeling," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21* (New York, NY: Association for Computing Machinery), 815–823. doi: 10.1145/3437963.3441817

Hooi, B., Song, H. A., Beutel, A., Shah, N., Shin, K., and Faloutsos, C. (2016). "Fraudar: bounding graph fraud in the face of camouflage," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, San Francisco, CA, 895–904. doi: 10.1145/2939672.2939747

Hou, M., Ren, J., Febrinanto, F., Shehzad, A., and Xia, F. (2021). "Cross network representation matching with outliers," in *2021 International Conference on Data Mining Workshops (ICDMW)* (Auckland: IEEE), 951–958. doi: 10.1109/ICDMW53433.2021.00124

Hovy, D. (2016). "The enemy in your own camp: how well can we detect statistically-generated fake reviews -an adversarial study," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 351–356. doi: 10.18653/v1/P16-2057

Huang, M., Liu, Y., Ao, X., Li, K., Chi, J., Feng, J., et al. (2022). "AUC oriented graph neural network for fraud detection," in *Proceedings of the ACM Web Conference 2022, WWW '22* (New York, NY: Association for Computing Machinery), 1311–1321. doi: 10.1145/3485447.3512178

Istanto, R. S. H., Mahmudy, W. F., and Bachtiar, F. A. (2020). "Detection of online review spam: a literature review," in *Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology, SIET '20*, Malang, 57–63. doi: 10.1145/3427423.3427434

Jerripothula, K. R., Rai, A., Garg, K., and Rautela, Y. S. (2020). Feature-level rating system using customer reviews and review votes. *IEEE Trans. Comput. Soc. Syst.* 7,1210–1219. doi: 10.1109/TCSS.2020.3010807

Jindal, N., and Liu, B. (2008). "Opinion spam and analysis," in *Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08*, Palo Alto, CA, 219–230. doi: 10.1145/1341531.1341560

Kaghazgaran, P., Caverlee, J., and Squicciarini, A. (2018). "Combating crowdsourced review manipulators: a neighborhood-based approach," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18*, Marina Del Rey, CA, 306–314. doi: 10.1145/3159652.3159726

Kipf, T. N., and Welling, M. (2017). "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations (ICLR)*.

Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., and Subrahmanian, V. (2018). "REV2: fraudulent user prediction in rating platforms," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18*, 333–341. doi: 10.1145/3159652.3159729

Li, A., Qin, Z., Liu, R., Yang, Y., and Li, D. (2019). "Spam review detection with graph convolutional networks," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, 2703–2711. doi: 10.1145/3357384.3357820

Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., et al. (2017). "Bimodal distribution and co-bursting in review spam detection," in *Proceedings of the 26th International Conference on World Wide Web, WWW '17*, Beijing, 1063–1072. doi: 10.1145/3038912.3052582

Li, Y., Huang, X., Li, J., Du, M., and Zou, N. (2019). "Specae: spectral autoencoder for anomaly detection in attributed networks," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19*, 2233–2236. doi: 10.1145/3357384.3358074

Li, Z., Hui, P., Zhang, P., Huang, J., Wang, B., Tian, L., et al. (2021a). "What happens behind the scene? Towards fraud community detection in e-Commerce from online to offline," in *WWW '21: Companion Proceedings of the Web Conference 2021* (New York, NY: Association for Computing Machinery), 105–113. doi: 10.1145/3442442.3451147

Li, Z., Wang, H., Zhang, P., Hui, P., Huang, J., Liao, J., et al. (2021b). "Live-streaming fraud detection: a heterogeneous graph neural network approach," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21* (New York, NY: Association for Computing Machinery), 3670–3678. doi: 10.1145/3447548.3467065

Liu, B., Sun, X., Meng, Q., Yang, X., Lee, Y., Cao, J., et al. (2022). Nowhere to hide: Online rumor detection based on retweeting graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2022.3161697

Liu, C., Sun, L., Ao, X., Feng, J., He, Q., and Yang, H. (2021a). "Intention-aware heterogeneous graph attention networks for fraud transactions detection," in *KDD '21* (Association for Computing Machinery), 3280–3288. doi: 10.1145/3447548.3467142

Liu, J., Xia, F., Feng, X., Ren, J., and Liu, H. (2022). Deep graph learning for anomalous citation detection. *IEEE Trans. Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2022.3145092

Liu, Y., Ao, X., Qin, Z., Chi, J., Feng, J., Yang, H., et al. (2021b). "Pick and choose: a GNN-based imbalanced learning approach for fraud detection," in *Proceedings of the Web Conference 2021, WWW '21*, 3168–3177. doi: 10.1145/3442381.3449989

Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., and Karypis, G. (2021c). Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Trans. Neural Netw. Learn. Syst*. doi: 10.1109/TNNLS.2021.3068344

Liu, Y., Yang, S., Zhang, Y., Miao, C., Nie, Z., and Zhang, J. (2021d). Learning hierarchical review graph representations for recommendation. *IEEE Trans. Knowledge Data Eng*. doi: 10.1109/TKDE.2021.3075052

Luca, M., and Zervas, G. (2016). Fake it till you make it: reputation, competition, and yelp review fraud. *Manage. Sci*. 62, 3412–3427. doi: 10.1287/mnsc.2015.2304

Ma, R., Pang, G., Chen, L., and van den Hengel, A. (2022). "Deep graph-level anomaly detection by glocal knowledge distillation," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, WSDM '22* (New York, NY: Association for Computing Machinery), 704–714. doi: 10.1145/3488560.3498473

Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., et al. (2021). A comprehensive survey on graph anomaly detection with deep learning. *IEEE Trans. Knowledge Data Eng*. 704–714. doi: 10.1109/TKDE.2021.3118815

McAuley, J., and Leskovec, J. (2013). "Hidden factors and hidden topics: understanding rating dimensions with review text," in *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, Rio de Janeiro, 165–172. doi: 10.1145/2507157.2507163

McAuley, J. J., and Leskovec, J. (2013). "From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews," in *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, 897–908. doi: 10.1145/2488388.2488466

McAuley, J. J., Leskovec, J., and Jurafsky, D. (2012). "Learning attitudes and attributes from multi-aspect reviews," in *12th IEEE International Conference on Data Mining, ICDM '12*, Brussels, 1020–1025. doi: 10.1109/ICDM.2012.110

Mohawesh, R., Tran, S., Ollington, R., and Xu, S. (2021). Analysis of concept drift in fake reviews detection. *Expert Syst. Appl*. 169, 114318. doi: 10.1016/j.eswa.2020.114318

Mukherjee, A., Venkataraman, V., Liu, B., and Glance, N. (2013). "What yelp fake review filter might be doing?" in *ICWSM '13*, Cambridge, MA, 409–418.

Noekhah, S., binti Salim, N., and Zakaria, N. H. (2020). Opinion spam detection: using multi-iterative graph-based model. *Inform. Process. Manage*. 57, 102140. doi: 10.1016/j.ipm.2019.102140

Pourhabibi, T., Ong, K.-L., Kam, B. H., and Boo, Y. L. (2020). Fraud detection: a systematic literature review of graph-based anomaly detection approaches. *Decis. Support Syst*. 133, 113303. doi: 10.1016/j.dss.2020.113303

Rayana, S., and Akoglu, L. (2015). "Collective opinion spam detection: bridging review networks and metadata," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15*, Sydney, NSW, 985–994. doi: 10.1145/2783258.2783370

Rayana, S., and Akoglu, L. (2016). "Collective opinion spam detection using active inference," in *Proceedings of the 2016 SIAM International Conference on Data Mining, ICDM '16*, Miami, FL, 630–638. doi: 10.1137/1.9781611974348.71

Ren, Y., and Zhang, Y. (2016). "Deceptive opinion spam detection using neural network," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, 140–150.

Rossi, R. A., Ahmed, N. K., Koh, E., Kim, S., Rao, A., and Abbasi-Yadkori, Y. (2020). *A Structural Graph Representation Learning Framework*. New York, NY: Association for Computing Machinery. doi: 10.1145/3336191.3371843

Shan, G., Zhou, L., and Zhang, D. (2021). From conflicts and confusion to doubts: examining review inconsistency for fake review detection. *Decis. Support Syst*. 144, 113513. doi: 10.1016/j.dss.2021.113513

Shehnepoor, S., Togneri, R., Liu, W., and Bennamoun, M. (2021). HinRNN: a graph representation learning neural network for fraudster group detection with no handcrafted features. *IEEE Trans. Neural Netw. Learn. Syst*. doi: 10.1109/TNNLS.2021.3123876

Sun, Y., and Loparo, K. (2019). "Opinion spam detection based on heterogeneous information network," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, 1156–1163. doi: 10.1109/ICTAI.2019.00277

Wang, H., Zhou, C., Wu, J., Dang, W., Zhu, X., and Wang, J. (2018). "Deep structure learning for fraud detection," in *2018 IEEE International Conference on Data Mining (ICDM)*, Singapore, 567–576. doi: 10.1109/ICDM.2018.00072

Wang, J., Wen, R., Wu, C., Huang, Y., and Xion, J. (2019). "FDGARs: fraudster detection via graph convolutional networks in online app review system," in *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19* (New York, NY: Association for Computing Machinery), 310–316. doi: 10.1145/3308560.3316586

Wang, J., Wen, R., Wu, C., and Xiong, J. (2020). "Analyzing and detecting adversarial spam on a large-scale online app review system, in *WWW '20: Companion Proceedings of the Web Conference 2020* (New York, NY: Association for Computing Machinery). doi: 10.1145/3366424.3383756

Wang, L., Li, P., Xiong, K., Zhao, J., and Lin, R. (2021). "Modeling heterogeneous graph network on fraud detection: a community-based framework with attention mechanism," in *CIKM '21: Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (New York, NY: Association for Computing Machinery), 1959–1968. doi: 10.1145/3459637.3482277

Wang, W., Xia, F., Wu, J., Gong, Z., Tong, H., and Davison, B. D. (2021). Scholar2vec: vector representation of scholars for lifetime collaborator prediction. *ACM Trans. Knowledge Discov. Data* 15, 1–19. doi: 10.1145/3442199

Wang, X., Liu, K., He, S., and Zhao, J. (2016). "Learning to represent review with tensor decomposition for spam detection," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, TX, 866–875. doi: 10.18653/v1/D16-1083

Wang, Y. (2022). "Fair graph representation learning with imbalanced and biased data," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, WSDM '22* (New York, NY: Association for Computing Machinery), 1557–1558. doi: 10.1145/3488560.3502218

Weise, K. (2011). *A Lie Detector Test for Online Critics*. Bloomberg Businessweek.

Wen, R., Wang, J., Wu, C., and Xiong, J. (2020). "ASA: adversary situation awareness via heterogeneous graph convolutional networks," in *WWW '20: Companion Proceedings of the Web Conference 2020* (New York, NY: Association for Computing Machinery), 674–678. doi: 10.1145/3366424.3391266

Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., et al. (2021a). Graph learning: a survey. *IEEE Trans. Artif. Intell*. 2, 109–127. doi: 10.1109/TAI.2021.3076021

Xia, F., Yu, S., Liu, C., Li, J., and Lee, I. (2021b). Chief: clustering with higher-order motifs in big networks. *IEEE Trans. Netw. Sci. Eng*. 9:990–1005. doi: 10.1109/TNSE.2021.3108974

Xia, T., and Ku, W.-S. (2021). "Geometric graph representation learning on protein structure prediction," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21* (New York, NY: Association for Computing Machinery), 1873–1883. doi: 10.1145/3447548.3467323

Xu, G., Hu, M., and Ma, C. (2021). Secure and smart autonomous multi-robot systems for opinion spammer detection. *Inform. Sci*. 576, 681–693. doi: 10.1016/j.ins.2021.07.072

Xu, J., Yu, S., Sun, K., Ren, J., Lee, I., Pan, S., et al. (2020). "Multivariate relations aggregation learning in social networks," in *JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (New York, NY: Association for Computing Machinery), 77–86. doi: 10.1145/3383583.3398518

Yang, Y., Xu, Y., Sun, Y., Dong, Y., Wu, F., and Zhuang, Y. (2021). Mining fraudsters and fraudulent strategies in large-scale mobile social networks. *IEEE Trans. Knowledge Data Eng*. 33, 169–179. doi: 10.1109/TKDE.2019.2924431

Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., and Leskovec, J. (2018). "Graph convolutional neural networks for web-scale recommender systems," in *Proceedings of the 24th ACM SIGKDD international Conference on Knowledge Discovery & Data Mining*, London, 974–983. doi: 10.1145/3219819.3219890

Yu, C., Zuo, Y., Feng, B., An, L., and Chen, B. (2019). An individual-group-merchant relation model for identifying fake online reviews: an empirical study on a Chinese e-commerce platform. *Inform. Technol. Manage*. 20, 123–138. doi: 10.1007/s10799-018-0288-1

Yu, S., Xia, F., Zhang, K., Ning, Z., Zhong, J., and Liu, C. (2017). "Team recognition in big scholarly data: exploring collaboration intensity," in *3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress* (Orlando, FL: IEEE), 925–932. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.155

Yuan, C., Zhou, W., Ma, Q., Lv, S., Han, J., and Hu, S. (2019). "Learning review representations from user and product level information for spam detection," in *2019 IEEE International Conference on Data Mining (ICDM)*, Beijing, 1444–1449. doi: 10.1109/ICDM.2019.00188

Yuan, S., Wu, X., Li, J., and Lu, A. (2017). "Spectrum-based deep neural networks for fraud detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17* (New York, NY: Association for Computing Machinery), 2419–2422. doi: 10.1145/3132847.3133139

Zhao, T., Jiang, T., Shah, N., and Jiang, M. (2022). A synergistic approach for graph anomaly detection with pattern mining and feature learning. *IEEE Trans. Neural Netw. Learn. Syst.* 33: 2393–2405. doi: 10.1109/TNNLS.2021.3102609

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.