Elevating GraphSAGE for Covertness: A Strategic Approach to Unmasking Fake Reviews in E-Commerce

Abhay Narayan[1], **Dameera Tharun**[1], **Madhu Kumar S D**[1], **Anu Mary Chacko**[1]

[1] National Institute of Technology Calicut, India
abhaynarayan09@gmail.com, {dameera_m210365cs,madhu,anu.chacko}@nitc.ac.in

Abstract

Fake reviews deliberately created and propagated as part of disinformation campaigns pose a significant threat to consumers, as they can lead them astray and result in financial losses and reputational damage. It is crucial to detect and mitigate these deceptive practices to maintain trust and integrity on online platforms. In this study, we propose a novel approach to enhance the effectiveness of fake review detection using GraphSAGE(Graph Sample and Aggregate), a graph-based technique, with a Covertness model. The innovative integration of GraphSAGE with a covertness measure aims to capture the intricate interactions and heterogeneity present in user reviews, by considering both user attributes and textual content. Our performance evaluations on publicly available real-world Amazon datasets demonstrate that our proposed model consistently achieves competitive Recall and AUC values across various training data percentages. Additionally, our comparative analysis against other graph-based models, including GCN, GAT, GeniePath, and GraphSAGE, highlights the superior performance of our proposed model. Our findings emphasize the robustness and potential of our model to accurately detect fake reviews in real-world scenarios. This study significantly contributes to advancing fake review detection methodologies by offering a promising approach to combat disinformation and safeguard consumer trust on online platforms.

Introduction

The ease and practicality of e-commerce has led to a growing prevalence of online purchasing. Due to the asymmetry of information and the absence of quality control centers in e-commerce, consumers typically rely on browsing product or service reviews prior to making online purchases. Online reviews are crucial sources of information for consumers to make informed purchasing decisions. Based on research conducted by Harvard University, each additional star in a product's rating on Yelp is associated with a 5-9% boost in revenue for that product(Luca 2016). Writing fake reviews is considered a form of disinformation in which actors orchestrate campaigns to manipulate public opinion, influence consumer behavior, and achieve economic objectives. These campaigns can distort the integrity of online reviews, making it difficult for consumers to make informed purchasing decisions based on authentic information. Driven by the prospect of monetary benefits, certain unscrupulous vendors often engage in collusion with spammers, with the intention of either undermining competitors or boosting their own enterprises through the dissemination of numerous fraudulent evaluations. The rampant dissemination of fraudulent evaluations renders it unfeasible for consumers to assess the genuine caliber of products just on review data. This significantly impacts consumers' buying experience and undermines the equitable competitive landscape among businesses. Furthermore, it significantly hampers the progress of the e-commerce sector. The study (Vainilavičius 2023) revealed a shocking discovery: almost half of the customer reviews on Amazon were deceptive. This result underscores the extensive prevalence of false reviews on e-commerce platforms.

Although there have been numerous attempts to develop automatic spam review detection systems, the majority of them heavily depend on learning from pre-determined characteristics and lack the ability to apply their knowledge to new situations.

Traditional statistical learning methods often employ supervised classifiers, such as support vector machines (SVMs)(Ott, Cardie, and Hancock 2013), logistic regression (LR)(Jindal and Liu 2008), and naïve Bayes(Li et al. 2011), to identify unexpected patterns by collecting review-specific semantic information(Yuan et al. 2019). These feature-focused approaches typically disregard relationships between reviews, users, and things. (Narayan, Madhu Kumar, and Chacko 2022; Jabeur et al. 2023) reviews earlier research employing machine learning techniques to detect financial fraud in e-commerce, encompassing fraudulent activities, such as deceptive product reviews.

Nevertheless, past experience cautions us against making rash decisions based just on the review; instead, we should confirm the accuracy of the material provided, including reviewers' credentials, preferences, biases, and views. In a similar vein, relying solely on review text for features might provide challenges because it's not always clear what a review text means and cannot always be trusted.

Our hypothesis is that modeling the data collected from reviews, users, and product could help significantly enhance the performance and generalizability of online spam review detection systems, thereby addressing the shortcomings of

the current approaches. Therefore, we develop a GraphSAGE network model incorporating a covertness measure. This model is designed to effectively capture the diversity and intricate relationships present in various features derived from user interactions and their corresponding reviews on items.The paper unfolds as follows:

Building on the foundation established in the "Related Work" section, the subsequent section, "Fundamentals," elucidates the critical concepts and methodologies that form the basis of our proposed approach. It delineates the fundamental elements necessary to comprehend the complexities involved in detecting fake reviews.

The "Proposed Model" section constitutes the core of our contribution, where we introduce the GraphSAGE network model enhanced with a covertness measure. This innovative approach endeavors to capture the heterogeneity and intricate interactions inherent in user reviews, thereby providing a robust framework for detecting deceptive practices.

Subsequently, the "Experiments and Results" section presents extensive experiments conducted to assess the efficacy of our proposed model. The results are meticulously presented and analyzed, shedding light on the model's performance across a range of scenarios and datasets.

Following the "Experiments and Results" section, the subsequent part delves into "Synthesizing Approaches: From Fake Reviews to SSIO Detection." In this section, we explore the connections and implications between fake reviews and state-sponsored information operations (SSIOs). We discuss how our proposed model, originally developed for fake review detection, can be adapted and applied to identify and respond to SSIOs, thereby broadening the scope of our research.

Finally, the "Conclusion" section synthesizes our findings and contributions, offering a conclusive summary and suggesting potential avenues for future research.

This paper not only contributes a novel GraphSAGE with Covertness model for fake review detection but also provides a structured exploration of the existing landscape, fundamental concepts, and experimental evaluations, culminating in a comprehensive understanding of the proposed approach and its implications.

Related Work

Characteristic-Focused Techniques

Conventional statistical methods rely on the extraction of various features from textual reviews, followed by the training of a language model. Jindal and Liu (Jindal and Liu 2008) were the first to identify three categories of spam reviews—namely, deceitful opinions, reviews solely focused on brands, and non-reviews. They conducted an analysis using real-world datasets from Amazon, extracting features centered around reviews, reviewers, and products. These features were then utilized as input for a logistic regression (LR) model.

In a recent study, Weng et al.(Weng et al. 2019) compiled 11 platform-independent features from word-level, semantic-level, and structural-level analyses to differentiate between fraudulent and normal items. They employed Xgboost as a binary classifier, and the evaluation outcomes demonstrated that their approach, CATS, achieved both high precision and recall. Ott et al. (Ott, Cardie, and Hancock 2013) tackled the problem by employing three strategies as features in naïve Bayes and SVM classifiers.

Another approach by Wang et al. (Wang et al. 2018) involved the use of a long short-term memory (LSTM) framework for spam review detection. They established three layers within the framework: the input layer for receiving data, the hidden layer of LSTM, and the output layer, each contributing to the prediction of spam reviews. (Hajek, Hikkerova, and Sahut 2023) suggest a novel approach to tackle this issue by employing aspect-based sentiment analysis (ABSA). Unlike traditional methods that solely evaluate the sentiment of the entire review, this approach assesses the sentiment of the individual aspects of the product under consideration.

Characteristic-Focused Techniques examine individual reviews, but they can be easily deceived by imposters. In contrast, graph methods take a broader perspective by connecting reviews, reviewers, and products, revealing fake review networks. This Adaptability advantage allows them to counter new deception methods.

Graph-Oriented Approaches

Graph-oriented techniques have gained widespread use in capturing textual characteristics across various entities. The initial application of a graph neural network (GNN) for spam review detection was introduced by Wang et al. (Wang et al. 2011). They constructed a heterogeneous "review graph" to illustrate the connections among reviewers, reviews, and online sellers. Shehnepoor et al. (Shehnepoor et al. 2017) employed spam features to model review datasets as heterogeneous information networks, transforming the spam review detection process into a classification challenge within such networks. In the classification phase, they introduced metapath concepts to identify feature importance and determine weights.

Liu et al.(Liu et al. 2018) presented a graph model based on neural networks, named Graph Embeddings for Malicious Accounts (GEM). This model took into account both "device aggregation" and "activity aggregation" in heterogeneous graphs. Up to now, these approaches have primarily concentrated on shallow encoders, specifically matrix factorization. In these methods, there is no parameter sharing, each node possesses a distinct embedding vector, and the inherent "transductive" features are incapable of generating embeddings for unseen nodes during training, thus neglecting the incorporation of node features.

In recent years, there has been a growing interest in employing the "message-passing" approach within graphs (Zhou et al. 2020). This methodology involves learning how to gather information from different types of neighbors through the use of Markov random field (MRF) techniques implicitly. Hamilton et al. (Hamilton, Ying, and Leskovec 2017) introduced the GraphSAGE model, demonstrating significant advancements compared to earlier methods like DeepWalk (Perozzi, Al-Rfou, and Skiena 2014) and SemiGCN (Kipf and Welling 2016). GraphSAGE over-

comes the constraint associated with applying Graph Convolutional Networks (GCN) in transductive settings with a specified Laplacian matrix.

In the realm of spam-bot detection, a model-based Graph Convolutional Neural Network (GCNN) is proposed in (Ali Alhosseini et al. 2019). This model suggests an inductive representation learning approach for spam review detection, leveraging reviewer profile information and the social network graph from Twitter datasets. The inductive representation learning method employed in this approach bears similarity to that of GraphSAGE. Various graph-oriented approaches have been proposed in the existing literature on spam review detection. However, there is a notable gap concerning the integration of comprehensive feature representations with the ability to handle covert or hidden patterns in user-generated content. Current methodologies often rely on shallow encoders and lack parameter sharing, thereby limiting their capacity to effectively capture nuanced relationships and transductive features. This motivates the need for an advanced approach that not only incorporates node features but also addresses the challenge of covert behavior within network data.

In this paper, we present a GraphSAGE network model that includes a covertness measure for classifying fake reviews. This model captures information from review text, user attributes, item attributes, and structured data. The concept of a covertness measure is introduced to address instances of a user engaging in illegal actions while attempting to conceal themselves within a crowd.

Fundamentals

In this section, we present the key concepts which are essential for comprehending our model.

Graph-Based Fraud Detection

Fraud detection on an attributed graph $G = (V, E, X)$ is formally defined as a machine learning problem within a graph-based framework, where:

- V denotes the set of nodes representing accounts or entities potentially involved in fraudulent activities.
- E represents the set of edges in the graph, signifying connections or interactions between nodes, such as relationships or transactions in the context of fraud detection.
- X encompasses the set of node properties, reflecting the traits or qualities possessed by each node in the network. These properties may include account-specific details, transaction histories, and other relevant information.

Let Y be the set of binary labels assigned to each node in the graph, indicating whether it is fraudulent ($Y(v) = 1$) or not ($Y(v) = 0$). The function $Y : V \to \{0, 1\}$ describes Y, where $Y(v) = 1$ signifies that node v is engaged in fraudulent activity, while $Y(v) = 0$ indicates that node v is not.

The primary goal of graph fraud detection is to develop a predictive model capable of accurately classifying nodes as fraudulent or non-fraudulent based on their attributes and interactions in the graph. This involves employing graph-based machine learning methods, such as graph neural networks, to effectively identify fraud by capturing the structural information and patterns inherent in the graph, along with the associated node attributes.

GraphSAGE

GraphSAGE, or Graph Sample and Aggregated, is a graph neural network model developed for large-scale graph learning. it can be described as follows:

- **Graph Representation:** Consider a graph $G = (V, E)$ with nodes V and edges E. Each node v_i is associated with a feature vector $x_i \in \mathbb{R}^d$, where d is the feature dimension.
- **Neighborhood Sampling:** GraphSAGE utilizes a sampling approach to acquire a fixed-size neighborhood around each node. Let $N(v_i)$ be the node v_i's neighborhood, and $S_i \subseteq N(v_i)$ represent the sampled neighborhood.
- **Aggregation Function:** Information from the sampled neighborhood S_i is aggregated for each node v_i using an aggregation function f_{agg}. The function combines neighboring node information to create a representative embedding for the central node:

$$h_i = f_{\text{agg}}(\{x_j | j \in S_i\})$$

- **Parameterized Aggregator:** The aggregation function f_{agg} is parameterized with learnable parameters, typically implemented as neural network layers. These parameters are trained to capture complex graph relationships.
- **Learnable Embeddings:** The resulting embeddings h_i for each node serve as representations for downstream tasks. These embeddings can be input into additional layers for specific tasks such as classification or regression.

GraphSAGE's training objective involves optimizing a task-specific loss function, with backpropagation used to update the model parameters.

In essence, GraphSAGE defines a mathematical framework for sampling and aggregating information from node neighborhoods in a graph, enabling the learning of embeddings that capture the graph's structural information for various applications.

Proposed Model

In this section, we present our innovative model for graph-based fake review detection, integrating a Covertness Measure inspired by Ovelgönne et al.(Ovelgönne et al. 2012) and leveraging the GraphSAGE framework. The aim is to enhance the accuracy of identifying deceptive reviews by capturing nuanced interactions and characteristics within the review graph.

Covertness Measure

Motivated by the work of Ovelgönne et al. (Ovelgönne et al. 2012), which introduces the concept of covertness centrality (CC) to address a user's illicit behavior concealed within a crowd, we incorporate a vertex's covertness measure. This measure is determined by two factors: its "commonness"

concerning a set C of centrality measures and its effective "communication" with a user-specified set of vertices.

The mathematical formulation of CC is as follows:

$$CC(v, \tau, \alpha) = \begin{cases} 0 & \text{if } CM(v) < \tau \\ \alpha CM(v) + (1-\alpha)CP(v) & \text{if } CM(v) \geq \tau \end{cases}$$

where $CM(v)$ is the commonness measure of vertex v, and the parameters τ and α control the threshold and the balance between commonness and communication potential, respectively.

The commonness measure (CM) is defined based on the notion of how frequently an actor's properties occur within the network. In the context of Ovelgönne et al. (Ovelgönne et al. 2012), it quantifies the prevalence of an actor's properties across the network:

$$CM(p) = \frac{|\text{Prop}(p)|}{|V|}$$

Here, $|\text{Prop}(p)|$ is the number of actors sharing the same properties as actor p, and $|V|$ is the total number of actors in the network. We use degree and closeness centrality as base properties for actors.

Additionally, the Communication Potential (CP) evaluates an actor's capacity to interact with a user-specified set of network nodes. Ovelgönne et al. (Ovelgönne et al. 2012), assesses how well an actor can connect with other actors based on their structural placements. We measure communication potential using Betweenness centrality:

$$\sum_{s \neq v \neq t} \frac{\sigma(s,t)}{d(s,t)}$$

where v is the vertex of interest, s and t are vertices in the network, $\sigma(s,t)$ is the number of shortest paths between s and t, and $d(s,t)$ is the number of shortest paths between s and t that do not pass through v.

GraphSAGE Model

We adopt the GraphSAGE (Graph Sample and Aggregation) model for learning node representations in the attributed graph. GraphSAGE enables us to leverage both local and global information by sampling and aggregating features from neighboring nodes. The model is trained to predict binary labels indicating whether a review is genuine or fake.

The GraphSAGE model is defined as follows:

$$h_v^{(l+1)} = \sigma\left(W^{(l)} \cdot \text{Aggregate}\left(\{h_u^{(l)}, \forall u \in \mathcal{N}(v)\}\right)\right)$$

where $h_v^{(l)}$ is the representation of node v at layer l, $W^{(l)}$ is the weight matrix at layer l, σ is the activation function, and $\mathcal{N}(v)$ represents the neighbors of node v.

Objective Function

The objective is to minimize the loss function L through training the GraphSAGE model on the labeled dataset, incorporating the Covertness Measure:

$$L = -\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \left[Y(v) \cdot \log(\hat{Y}(v)) + (1 - Y(v)) \cdot \log(1 - \hat{Y}(v)) \right]$$

where $|\mathcal{V}|$ is the total number of nodes, $Y(v)$ is the ground truth label, and $\hat{Y}(v)$ is the predicted probability of node v being a fake review.

The core of our investigation is encapsulated in Figure 1, which visually represents our proposed model. This graphical representation serves as a comprehensive depiction of the GraphSAGE network model enriched with a covertness measure.

Experiments and Results

Dataset

- **Amazon Review Dataset:** The dataset from Amazon comprises reviews in the Musical Instruments category (McAuley and Leskovec 2013). The dataset exhibits the following characteristics:
- **Labels:** The dataset includes a binary array denoting labels, where 1 indicates spam, and 0 signifies a benign review.
- **Features:** Features are represented by a sparse matrix of 25-dimensional handcrafted features.
- **Records:** The dataset contains a substantial number of reviews, totaling 11,944.

Utilizing the dataset provided in (McAuley and Leskovec 2013), the authors in (Dou et al. 2020) established three relationships by treating users as nodes in a graph:

1. **U-P-U:** Users who review at least one common product.
2. **U-S-U:** Users providing at least one star rating within a week.
3. **U-V-U:** Users exhibiting 5% textual similarity.

Table I displays the extensive statistics of the dataset, including all its relationships.

Relation	#Nodes	#Edges	Fraud%
U-P-U	11,944	175,608	9.5%
U-S-U	11,944	3,566,479	
U-V-U	11,944	1,036,737	
ALL	11,944	4,398,392	

Table 1: Dataset Statistics

Compared Methods:

We compare our proposed method with several state-of-the-art graph neural network techniques and their enhancements applied to homogeneous graphs to validate its effectiveness.

- **GCN:**Spectral graph convolutions were employed to generate a localized first-order approximation of a graph convolution network(Kipf and Welling 2016).
- **GAT**: The graph attention network employs an attention mechanism for aggregating information from neighboring nodes(Velickovic et al. 2017).

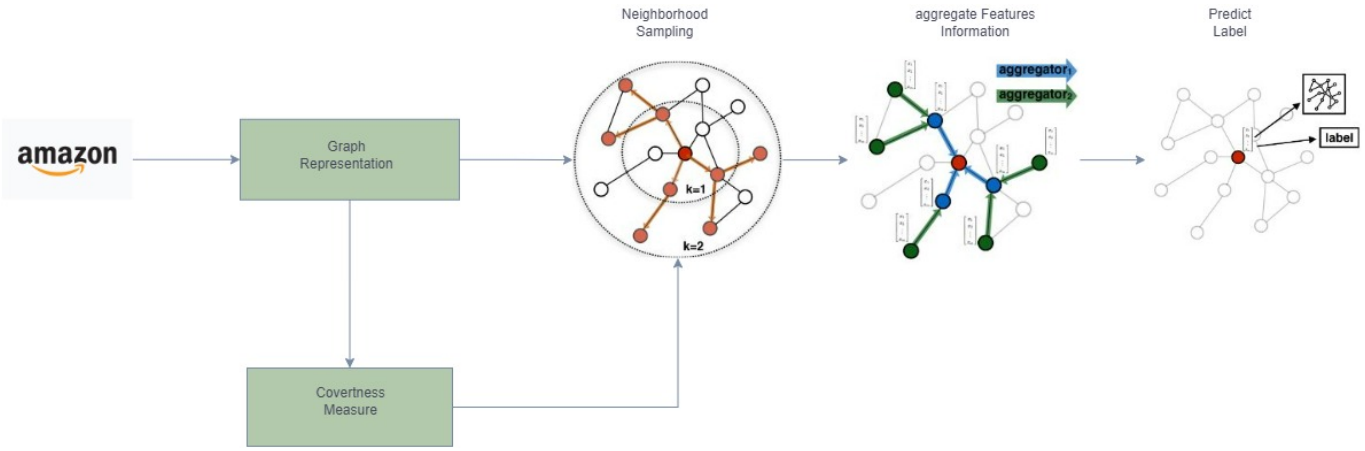

Figure 1: Proposed Model- GraphSAGE Enhanced with Covertness for Fake Review Detection

- **GraphSAGE:** is an inductive Graph Neural Network (GNN) model that employs a consistent sample of neighbors(Hamilton, Ying, and Leskovec 2017).
- **GeniePath:** Assigns a weight to each learned path in the graph, reflecting its importance(Liu et al. 2019).

All the methods mentioned above operate on a single relation, i.e., a homogeneous graph where all relations are merged together (as indicated by "ALL" in the table).

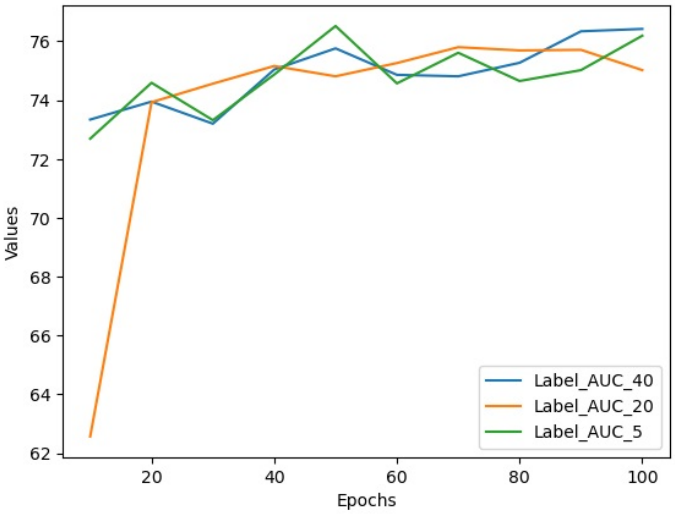

Figure 2: AUC With different training percentage

Evaluation Metric:

To evaluate the overall effectiveness of all classifiers, we utilize ROC-AUC (AUC) and Recall. AUC is computed based on the relative ordering of prediction probabilities for all instances, helping to mitigate the impact of imbalanced classes.

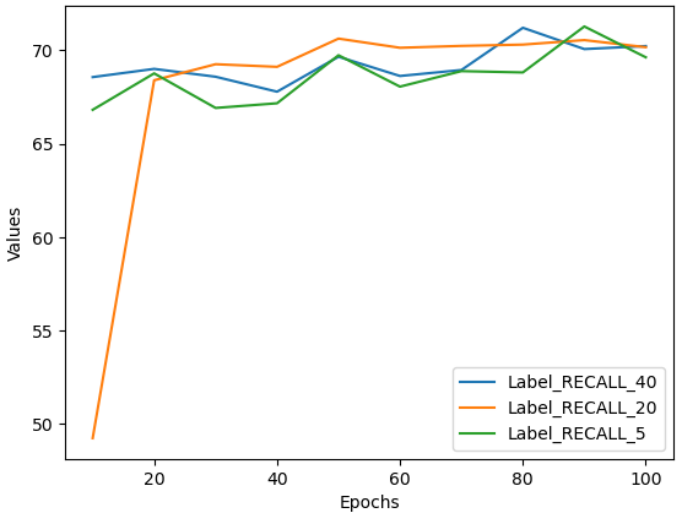

Figure 3: Recall With different training percentage

Performance Comparison:

Table II and Figures 1 and 2, offers a comprehensive view of the performance evaluation of various graph-based models on Amazon datasets, considering different training data percentages. A closer examination, with a specific focus on the GraphSAGE with Covertness model in comparison with others, reveals insightful patterns.

In terms of Recall, the GraphSAGE with Covertness model consistently exhibits competitive performance. At 5% training, it achieves a Recall of 69.71%, maintaining this performance at 20% training and demonstrating further improvement to 70.92% with 40% training. This trend suggests the model's ability to effectively capture subtle patterns and enhance its performance as the amount of training data increases.

Analyzing the AUC (Area Under the Curve), the GraphSAGE with Covertness model again stands out. It achieves an AUC of 75.10% at 5% training, maintains a high AUC

Table 2: Performance evaluation on Amazon datasets with various training data percentages.

		Methods				
Metric	Training%	GCN	GAT	GeniePath	Graph-SAGE	Graph-SAGE+ Covertness
Recall	5%	65.54	63.22	65.56	69.09	69.71
	20%	66.15	67.13	65.08	70.30	69.91
	40%	67.45	65.51	65.41	70.16	70.92
AUC	5%	74.44	73.89	71.56	70.71	75.10
	20%	75.13	72.10	71.89	73.97	74.52
	40%	74.34	75.16	72.65	75.27	76.79

of 74.52% at 20% training, and experiences a significant improvement to 76.79% with 40% training. This indicates the model's discriminative ability in distinguishing between positive and negative instances, particularly with a larger training dataset.

Comparing the GraphSAGE with Covertness model with other models, such as GCN, GAT, GeniePath, and GraphSAGE, reveals distinct characteristics. The GraphSAGE with Covertness model consistently outperforms or competitively matches these models in terms of Recall and AUC across various training percentages. It demonstrates a notable advantage in capturing nuanced patterns, especially with an increased amount of training data.

In conclusion, the GraphSAGE with Covertness model emerges as a robust performer in fake review detection on Amazon datasets. Its ability to maintain competitive Recall and AUC, particularly with larger training datasets, positions it as a promising approach for effectively discerning fraudulent activities in online reviews.

Synthesizing Approaches: From Fake Reviews to SSIO Detection

State-sponsored information operations(SSIO) are defined as anything that are done to accomplish both domestic and international geopolitical aspirations, whether human-initiated or not, through the employment of online social tactics using computerized and genuine end users(Bradshaw and Howard 2019). SSIOs primarily deploy the method of weaponizing three distinct forms of information: propaganda, misinformation (the unintended transmission of erroneous information), and disinformation (the intentional distribution of misleading information). Authoritarian regimes mostly utilize SSIOs to exert control over internal information, but only a limited number of countries have the necessary means and expertise to engage in operations that aim to impact global audiences(Mozur 2018).

Although this definition may seem disconnected from fake reviews, there are still several connections and implications that need to be explored. In general, the main purpose of fake reviews is to alter clients' purchasing choices and undermine opponents' influence in e-commerce.

Integration of fake review detection techniques to locate and respond to misinformation campaigns conducted by state actors is an interesting topic. Thus, for instance, the definition of covertness, which is used to identify covert nodes inside the network, could help reveal the covert nature of the activities of the state-supported entities responsible for conducting disinformation campaigns. Furthermore, the methodologies of fake review detection, such as graph-based models and neural networks, could be adjusted and tuned to support the detection of patterns that might indicate an SSIO. These models, which use semantic information acquired through sophisticated interactions and knowledge of the discovered outcomes of hidden behavior, may assist in identifying and preventing disinformation synthesized by state actors.

A possible approach on how graph-based models such as GraphSAGE with a covertness measure could be adapted to detect covert behavior in the context of SSIOs is as follows:

- **Leverage Encrypted Network Structure:** GraphSAGE typically utilizes both node attributes and network structure to explore information in data. In SSIOs, communication typically takes place via secure social media, making it challenging to understand the content posted there.
- **Exploit Network Dynamics:** Creating fake accounts may involve significant effort for SSIOs over a period of time. The adaptability of GraphSAGE to incorporate temporal dynamics using covertness measures is noteworthy. Identifying and analyzing how new nodes (fake accounts) integrate into the network is crucial. Such analysis could serve as an indicator of potential danger, particularly during product launch periods, when there is a sudden increase in activity or connections.
- **Integrate External Knowledge:**
  Including outside information, like identifying known SSIO actors or recognizing propaganda techniques, helps the proposed model get better at spotting suspicious behaviors or language patterns.
- **Consider Adversarial Techniques:** SSIO actors may alter their strategies to evade detection. Adversarial techniques could train GraphSAGE to solve this problem. This includes running simulations of SSIO campaigns so that the consequently trained model can accurately detect these changing tactics if encountered during simulation, as well as any other similar setting in the future, which could help it tackle this better.

Combining these strategies with the proposed model, we can move beyond content analysis and focus on the network dynamics and communication patterns within encrypted social networks used for SSIOs. This can help identify covert SSIO campaigns aimed at manipulating public perception or influencing consumer behavior.

Conclusion

In conclusion, this research introduces a novel approach for fake review detection, leveraging a GraphSAGE with Covertness model to effectively capture the intricate interactions and heterogeneity present in user reviews within the Amazon dataset. The model demonstrates consistent and competitive performance in terms of Recall and AUC across various training data percentages, outperforming or matching other graph-based models. This signifies the robustness of the proposed approach in distinguishing between genuine and fraudulent reviews. The model's effectiveness is further highlighted by its ability to improve performance with an increased amount of training data.

Moving forward, future research directions should explore the scalability and adaptability of the GraphSAGE with Covertness model to diverse datasets and online platforms. Additionally, investigations into optimizing the model's parameters and exploring potential extensions to handle evolving strategies employed by malicious entities in generating fake reviews would enhance its real-world applicability. The integration of advanced natural language processing techniques and the exploration of multi-modal data sources may provide additional layers of information for more accurate detection.

Furthermore, implementing the proposed model on an online platform and assessing its efficacy in real-world conditions would provide valuable insights into its practical effectiveness in combating propagation of disinformation in e-commerce platforms, including State-sponsored information operations(SSIO).

References

Ali Alhosseini, S.; Bin Tareaf, R.; Najafi, P.; and Meinel, C. 2019. Detect me if you can: Spam bot detection using inductive representation learning. In *Companion proceedings of the 2019 world wide web conference*, 148–153.

Bradshaw, S.; and Howard, P. N. 2019. The global disinformation order: 2019 global inventory of organised social media manipulation.

Dou, Y.; Liu, Z.; Sun, L.; Deng, Y.; Peng, H.; and Yu, P. S. 2020. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 315–324.

Hajek, P.; Hikkerova, L.; and Sahut, J.-M. 2023. Fake review detection in e-Commerce platforms using aspect-based sentiment analysis. *Journal of Business Research*, 167: 114143.

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

Jabeur, S. B.; Ballouk, H.; Arfi, W. B.; and Sahut, J.-M. 2023. Artificial intelligence applications in fake review detection: Bibliometric analysis and future avenues for research. *Journal of Business Research*, 158: 113631.

Jindal, N.; and Liu, B. 2008. Opinion spam and analysis. In *Proceedings of the 2008 international conference on web search and data mining*, 219–230.

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

Li, F. H.; Huang, M.; Yang, Y.; and Zhu, X. 2011. Learning to identify review spam. In *Twenty-second international joint conference on artificial intelligence*.

Liu, Z.; Chen, C.; Li, L.; Zhou, J.; Li, X.; Song, L.; and Qi, Y. 2019. Geniepath: Graph neural networks with adaptive receptive paths. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4424–4431.

Liu, Z.; Chen, C.; Yang, X.; Zhou, J.; Li, X.; and Song, L. 2018. Heterogeneous graph neural networks for malicious account detection. In *Proceedings of the 27th ACM international conference on information and knowledge management*, 2077–2085.

Luca, M. 2016. Reviews, reputation, and revenue: The case of Yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016).

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*, 897–908.

Mozur, P. 2018. A Genocide Incited on Facebook, With Posts From Myanmar's Military (Published 2018) — nytimes.com. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html. [Accessed 03-05-2024].

Narayan, A.; Madhu Kumar, S.; and Chacko, A. M. 2022. A review of financial fraud detection in e-commerce using machine learning. In *International Conference on Frontiers of Intelligent Computing: Theory and Applications*, 237–248. Springer.

Ott, M.; Cardie, C.; and Hancock, J. T. 2013. Negative deceptive opinion spam. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, 497–501.

Ovelgönne, M.; Kang, C.; Sawant, A.; and Subrahmanian, V. 2012. Covertness centrality in networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 863–870. IEEE.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710.

Shehnepoor, S.; Salehi, M.; Farahbakhsh, R.; and Crespi, N. 2017. NetSpam: A network-based spam detection framework for reviews in online social media. *IEEE Transactions on Information Forensics and Security*, 12(7): 1585–1595.

Vainilavičius, J. 2023. Millions of Amazon reviews fake, study finds. https://cybernews.com/security/millions-amazon-reviews-fake/. [Accessed 02-04-2024].

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *stat*, 1050(20): 10–48550.

Wang, C.-C.; Day, M.-Y.; Chen, C.-C.; and Liou, J.-W. 2018. Detecting spamming reviews using long short-term memory recurrent neural network framework. In *Proceedings of the 2nd International Conference on E-commerce, E-Business and E-Government*, 16–20.

Wang, G.; Xie, S.; Liu, B.; and Philip, S. Y. 2011. Review graph based online store review spammer detection. In *2011 IEEE 11th international conference on data mining*, 1242–1247. IEEE.

Weng, H.; Ji, S.; Duan, F.; Li, Z.; Chen, J.; He, Q.; and Wang, T. 2019. Cats: cross-platform e-commerce fraud detection. In *2019 ieee 35th international conference on data engineering (icde)*, 1874–1885. IEEE.

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)*, 1444–1449. IEEE.

Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2020. Graph neural networks: A review of methods and applications. *AI open*, 1: 57–81.