# Detecting spammer groups using centrality measure

Miss Natthanicha Suriyamongkol



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2022

# การตรวจจับกลุ่มผู้สร้างสแปมด้วยค่าความเป็นศูนย์กลาง

น.ส.ณัฏฐณิชา สุริยะมงคล



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

| | |
|---|---|
| Thesis Title | Detecting spammer groups using centrality measure |
| By | Miss Natthanicha Suriyamongkol |
| Field of Study | Computer Science |
| Thesis Advisor | Asst. Prof. SUKREE SINTHUPINYO, Ph.D. |

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the FACULTY OF ENGINEERING
(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

..... Chairman
(Assistant Professor NATTEE NIPARNAN, Ph.D.)
..... Thesis Advisor
(Asst. Prof. SUKREE SINTHUPINYO, Ph.D.)
..... Examiner
(JESSADA THUTKAWKORAPIN, Ph.D.)
..... External Examiner
(Assistant Professor Denduang Pradubsuwan, Ph.D.)

ณัฏฐณิชา สุริยะมงคล : การตรวจจับกลุ่มผู้สร้างสแปมด้วยค่าความเป็นศูนย์กลาง. ( Detecting spammer groups using centrality measure) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุกรี สินธุภิญโญ

*การจ้างผู้ชำนาญเขียนรีวิวปลอมบนอีคอมเมิร์ช (e-Commerce) แพลตฟอร์มเพื่อดึงดูดลูกค้าเป็นที่พบเจอมากขึ้นในยุคที่อีคอมเมิร์ชแพลตฟอร์มนั้นถูกใช้เป็นช่องทางหลักในการซื้อสินค้าและบริการ ด้วยปริมาณรีวิวปลอมที่มากขึ้นส่งผลให้ประสิทธิภาพในการตรวจจับของผู้เชี่ยวชาญ (fraud/spam specialist) ลดน้อยลง การวิจัยนี้จึงได้นำเสนอวิธีการตรวจจับกลุ่มผู้สร้างสแปม เพื่อช่วยลดประมาณผู้ต้องสงสัยและเพิ่มประสิทธิภาพในการตรวจสอบให้แก่ผู้เชี่ยวชาญ โดยการวิจัยนี้ได้นำ XGBoost สำหรับจัดหมวดหมู่ (classification) มาใช้ในการตรวจจับการสร้างสแปมแบบบุคคล และใช้วิธีการหาค่าความเป็นศูนย์กลาง (centrality measure) และ Structural Clustering Algorithm for Network (SCAN) มาช่วยในการตรวจจับกลุ่มผู้สร้างสแปม จากผลการวิจัยแสดงให้เห็นว่าวิธีที่นำเสนอนี้สามารถลดจำนวนผู้ต้องสงสัยที่สร้างสแปมในการส่งตรวจ อีกทั้งยังมีค่าความถูกต้อง (accuracy) และค่าความแม่นยำ (precision) ที่มากกว่างานวิจัยก่อนหน้าเมื่อพิจารณากลุ่มที่มีความน่าสงสัยในการเป็นสแปมสูงสุดหนึ่งร้อยกลุ่มแรก*



| | | |
|---|---|---|
| สาขาวิชา | วิทยาศาสตร์คอมพิวเตอร์ | ลายมือชื่อนิสิต ..... |
| ปีการศึกษา | 2565 | ลายมือชื่อ อ.ที่ปรึกษาหลัก ..... |

## 6370087621 : MAJOR COMPUTER SCIENCE
KEYWORD: Spam review, Fraud review, Spam group, Spammer group detection, Spammer behavioral

Natthanicha Suriyamongkol : Detecting spammer groups using centrality measure. Advisor: Assistant Professor. SUKREE SINTHUPINYO, Ph.D.

*The practice of hiring professional spammers to create fake reviews on e-commerce platforms in order to attract customers can easily be found on the internet. As a result, the number of suspicious spammers has rapidly increased, hindering the performance of fraud/spam specialists' investigations. This study proposes a solution to mitigate the overwhelming number of investigative cases and improve investigation performance. The study utilizes XGBoost for classification to detect individual spammers; and centrality measure and the Structural Clustering Algorithm for Networks (SCAN) to identify spammer groups. Experimental evaluations show that this approach effectively reduces the number of spammers and increases accuracy and precision compared to previous studies, particularly concerning the hundred highest suspicious groups.*



| | | |
|---|---|---|
| Field of Study: | Computer Science | Student's Signature ..... |
| Academic Year: | 2022 | Advisor's Signature ..... |

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and appreciation to Asst. Prof. Sukree Sinthupiyo, Ph.D., for his invaluable guidance, unwavering support, and exceptional expertise throughout my research journey. His mentorship and dedication have been instrumental in shaping this thesis. I am grateful for his insightful feedback, constructive criticism, and patience in refining my ideas and methodologies.

I would also like to thank the members of my thesis committee, Asst sincerely. Prof. Nattee Niparnan, Ph.D., Jessada Thutkawkorapin, Ph.D., and Asst. Prof. Denduang Pradubsuwan, Ph.D., for their valuable time, expertise, and thoughtful input. Their constructive feedback and suggestions have significantly enriched my work and helped me to develop a more comprehensive understanding of the subject matter.

I am indebted to my family and friends for their unwavering support, love, and encouragement throughout this challenging academic journey.



In conclusion, the successful completion of this thesis would not have been possible without the invaluable support and contributions of numerous individuals and organizations. While I have mentioned some names here, I would like to acknowledge and express my gratitude to all those who have played a role in this journey, whether mentioned explicitly or not. Thank you all for being integral to this transformative experience.

Natthanicha Suriyamongkol

# TABLE OF CONTENTS

**Page**

# LIST OF TABLES

**Page**

# LIST OF FIGURES

**Page**



# 1. INTRODUCTION

## 1.1. Motivation

The eCommerce industry was developed in the early 1990s and has been continuously growing in the digital age, in which the Internet and mobile banking have become the backbone of financial payment. When the pandemic hit, the enforcement of lockdowns, social distancing, and others were applied. Many retailers inevitably had to create online channels to survive their businesses. The eCommerce has been rapidly growing worldwide throughout this pandemic (Brewster, 2022; OECD, 2020; UNCTAD, 2021); most people ineluctably buy products, services, foods, and beverages via eCommerce. The number of products on eCommerce has increased across the number of customers (Coppola, 2022; Law, 2021), so reviews have become a critical factor in customers' purchasing decisions. Since the reviews play a role in deciding for the customers to buy the products/services, the more positive reviews the products/services get, the more customers the retailers get (Tang et al., 2020). Statistically, Luca and Mintel (Luca, 2016; Press, 2015) showed that 90 percent of customers read the reviews to support the making financial decision, and 70 percent of the customers are inclined to trust the reviewers of other customers. However, some retailers cunningly do the dirty on the customers by hiring an individual review spammer or a group of review spammers to do positive reviews or even defame competitors to maximize profit gains (Xu & Zhang, 2015).



Spammer groups are groups of reviewers who purposely work together to produce spam/fraud reviews for promoting or demoting targeted products (Mukherjee et al., 2011). They usually write spam reviews for products/services to earn more profit (Hussain et al., 2021). They camouflage their identity with different

reviewer IDs; for example, one spammer can have multiple IDs, and various persons share a single ID (Vidanagama et al., 2020). Writing spam reviews service gets more attention from retailers; it quickly finds the writing spam review service on the Internet because it is not illegal. Many spam reviews from the spammer groups are enough to control the product's sentiment or perspective (Mukherjee et al., 2012; Wang et al., 2018); grouped spam is more urgent to detect than individual spam (Li et al., 2021).

Spam or spammer investigation is required considerable resources in terms of spam/fraud specialists or ground-truth labels. Moreover, the spammer with skills in the spam review business can easily handcraft a fake review that is just like a genuine review (Mukherjee et al., 2012; Ott et al., 2011; Ye & Akoglu, 2015). Therefore, manually labeling reviews is an unpleasant process (Mukherjee et al., 2011). To minimize the resources, pointing out grouped review spam is affordable compared to the individual spam review (Mukherjee et al., 2012; Zhang et al., 2018).

Several studies and solutions have been explored and proposed. Mukherjee et al. (Mukherjee et al., 2012) initiated the area of review spammer group detection using frequent item set mining (FIM) which the customer has reviewed in common to identify candidate spammer groups. Group indicator set, for instance, review time, content, and size, and individual indicator set, such as review time, content, and rating. Ye and Akoglu (Ye & Akoglu, 2015) first applied two network-based markers to cluster reviewers. Zhang, Wu & Cao (Zhang et al., 2018) introduced a spammer group detection using FIM with a supervised learning method to find the spammer group. Li et al. (Li et al., 2017) proposed the Labelled Hidden Markov Model (LHMM) to spot individual and group spammers. De Meo et al. (Meo et al., 2017) proposed the centrality measure from the trusted network and then predicted helpfulness-based

reputation; as a result, the trust relationship produced the most helpful reviews on the community platforms. Hussain et al. (Hussain et al., 2021) presented a spammer group detection that utilizes the Structural Clustering Algorithm for Networks (SCAN) algorithm to highlight the candidate spam groups and use individual and group indicators to calculate spamming scores for each group.

Although many studies have introduced the group of spam detection solutions, there are still unexplored areas to unveil. We have borrowed the centrality measures, individual, group behavioral indicators, linguistic indicators, and the SCAN algorithm. These new combinations improve the precision and recall compared to the previous study with the same dataset. Finally, this study could be an assistance to the investigation process.

## 1.2. Objective and Contribution

This study proposed a novel approach to spot the spammer group using centrality measures, co-reviewer graph, and Structural Clustering Algorithm for Networks (SCAN). The proposed framework helps to optimize the investigation process by scoping down the number of suspicious spammers and minimizing the full-time equivalent (FTE) of spammer specialists, in other words reducing the workload or investigation cases of the spammer specialists.



First, we find the individual spammer using XGBoost–the supervised learning method, with two individual indicator sets; 1) linguistic set {Linguistic Inquiry and Word Count (LIWC) (Pennebaker & Lay, 2002), readability (DuBay, 2004), credibility, Part of Speech (POS) tagging, and evidentiality of content (Su et al., 2010)}; 2) reviewer behavior set {review gap, life tenure, rating entropy, and rating deviation}.

Next, we apply Eigenvector Centrality to calculate a product vector of disease scores and apply the scores to create co-reviewer graphs, in other words, paring reviewers who show the similarity in review post time, review rating, and vector of disease scores. The co-reviewer graphs are fed into the SCAN algorithm to identify candidate spam groups. Then, using group indicator set {review tightness and rating variance} to calculate the spam scores of each candidate spammer group. Finally, we get the suspicious spammer groups attached with support reasons and the number of suspicious individual spammers in the groups.

To the best of our knowledge, the contribution of this paper is:

1. This study is the first study that applies centrality measures of spammers into the products/services dimension to identify the relationship of suspicious spammers and use it to form candidate spam groups.
2. This study utilizes linguistic review, reviewer behavior, and characteristics of the candidate group indicator set to identify the spammer group.
3. We achieve higher precision and recall performance in the co-reviews graph with the SCAN algorithm compared to Hussain et al. (Hussain et al., 2021) with the same dataset, which is the YelpNYC dataset. It contains reviews for restaurants in New York City collected from Yelp.com from 2004 to 2015. The dataset collects the information of user ID, product ID, review date, review rating, and label of spam/suspicious reviews. It has 160,225 reviewers (17.29% spammers) and 359,052 reviews (10.27% spammer reviews).

# 2. BACKGROUND

## 2.1. eCommerce

eCommerce is known as electronic commerce or Internet commerce. eCommerce also refers to the physical store on the Internet. Global data reports, that in 2020, eCommerce has been rapidly grown by 27.6% and expanded worldwide (Shepherd, 2022). Most things could be able to order from the eCommerce – products, services, foods, and beverages. eCommerce could be recognized as a community since customers can share their experience on the ordered products/services, moreover, the experience in the form of reviews is of value to other customers to make the financial decision.

## 2.2. YelpNYC Dataset

Yelp publishes crowd-sourced reviews about business platforms that develop the Yelp.com website and the Yelp mobile app (Yelp, 2022). YelpNYC dataset contains reviews for restaurants in New York City collected from Yelp.com. The dataset collects the information of user ID, product ID, review date, review rating, and label of spam/suspicious reviews. Even though the label YelpNYC dataset produced from Yelp anti-fraud filter had not reached the perfect, Weise has generated the accurate results (Weise, 2021). Consequently, Rayana and Akoglu (Rayana & Akoglu, 2015) were the first to use the YelpNYC dataset to propose their study. This dataset helps conduct the fundamental of spammer group detection due to its rich information.

## 2.3. XGBoost for Classification

XGBoost is the short-term Extra Extreme Gradient Boosting; it is also designed to be used with large, complicated datasets and support regression and classification problems (Chen & Guestrin, 2016). In the classification part, the probability of each class goes to fifty percent if it tries to predict two classes. It then uses the Residual to calculate the similarity score for the leaf.

$$Similarity\ Score = \frac{(\sum Residual)^2}{\sum[PreviousProbability_i \times (1 - PreviousProbability_i)] + \lambda} \quad 1)$$

Since the XGBoost is one of the decision tree algorithms, the samples are split and repeatedly calculated the similarity score. $\lambda$ (lambda) is added to prune the leaves easier. Then calculate the Gain of the roots, then use the Gain to select the branch's threshold. It is continuously building the tree until it reaches the number of levels.



$$Gain = left_{similarity} + right_{similarity} - Root_{similarity} \quad 2)$$

In addition, Cover is the value to limit adding the leaf in XGBoost. Cover could be interpreted as the Minimum Child Weight.

$$\text{Cover} = \sum [PreviousProbability_i \times (1 - PreviousProbability_i)] \quad 3)$$

To prune the tree, it calculates the difference between the Gain associated with the lowest branch and the $\gamma$ (gamma) we have picked. The tree will be pruned if the difference is a negative number. Next, determine the Output Value for each leaf, and the computation of the tree is complete.

$$\text{Output Value} = \frac{\sum Residual}{\sum[PreviousProbability_i \times (1 - PreviousProbability_i)] + \lambda} \quad (4)$$

Now, it can make a new prediction; likewise, the other boosting method, XGBoost for Classification, also makes new predictions by starting with the initial prediction by converting this probability to a log(odds) value and then adding the log odds of the initial prediction to the output of the tree, which is scaled by a learning rate $\varepsilon$ (eta). Finally, using the Logistic Function to convert the log(odds) value into a probability, continuously building another tree with the new residual and keep building trees until the residuals are minimal or reading the maximum number of trees.

$$Probability = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} \quad 5)$$



## 2.4. Eigenvector Centrality

Eigenvector Centrality EC (v) is one of the five popular centrality measures – Degree Centrality, Closeness Centrality, Betweenness Centrality, PageRank, and Eigenvector Centrality (Meo et al., 2017). EC (v) was built based on the ideas of Eigenvector and Eigenvalue to form the relationship among the network (EigenvectorCentrality, 2022). Eigenvector represents the relationship of all vertices in

the network as an adjacency matrix and then transfers the important/influence factor to the successive vertices via Eigenvalue. The Eigenvalue returns each vertex's significant/influence value in the network. The idea of an eigenvector shows in the following equation, where A is the adjacency matrix of the network, $\vec{v}$ is the eigenvector, and $\lambda$ is the Eigenvalue:

$$A\vec{v} = \lambda\vec{v} \quad (6)$$

The equation is read as A times $\vec{v}$ gives the same result as scaling the eigenvector $\vec{v}$ by eigenvalue $\lambda$. In order to get the values of $\lambda$, the eigenvector and their eigenvalues of a matrix A need to be found in the first place. The common way to rewrite the equation is to factor the $\lambda$ out and write it as $\lambda$ times $I$ where $I$ is the identity matrix with 1's down the diagonal (3Blue1Brown, 2016).

$$A\vec{v} = (\lambda \mathrm{I})\vec{v} \quad (7)$$

On both sides is matrix-vector multiplication, subtracting $(\lambda \mathrm{I})\vec{v}$ off both sides and factoring out the $\vec{v}$ on the left-hand side.

$$A\vec{v} - (\lambda \mathrm{I})\vec{v} = \vec{0} \quad (8)$$

$$(A - (\lambda \mathrm{I}))\vec{v} = \vec{0} \quad (9)$$

Now, we get a new matrix that is always true if $\vec{v}$ is the zero vector. However, we want a non-zero eigenvector. The only way to make the product of a

matrix with a non-zero vector become zero is if the transformation associated with that matrix squishes space into a lower dimension, that is squishification function corresponds to a zero determinant for the matrix.

$$det(A - (\lambda \mathrm{I})) = \vec{0} \quad 10)$$

We can tweak $\lambda$, as that value of $\lambda$ changes, the matrix itself changes, and so the determined of the matrix changes. The goal is to find a value of $\lambda$ that will turn this determinant to zero, which means the tweaked transformation of squished space into a lower dimension. For example, $\lambda$ can only be an eigenvalue if this determinant happens to be zero, we can conclude that the only possible eigenvalues are $\lambda$ equals 2 and $\lambda$ equals 3.

$$\det\left(\begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}\right) = (3-\lambda)(2-\lambda) = 0 \quad 11)$$

$$\lambda = 2 \; or \; \lambda = 3 \quad 12)$$



Finally, when we replace all the values back into equation 6, it can be interpreted that each node $(v_1, v_2)$ will transfer the important/influence factor via the eigenvalues $\lambda$. This is the concept of Eigenvector Centrality with Eigenvectors and Eigenvalues.

$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 2 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \; ; \; When \; \lambda = 2 \quad 13)$$

## 2.5. Structural Clustering Algorithm for Networks (SCAN)

SCAN is the algorithm to cluster vertex in the network, which stands out in resistance to noise and handling clusters of different shapes and sizes (Xu et al., 2007). The idea of SCAN is to use the information of who associates with whom to identify clusters of individuals with common interests or unique relationships. For computing the SCAN, first, the immediate neighborhood of a vertex needs to be defined (i.e., the set of people that an individual knows).

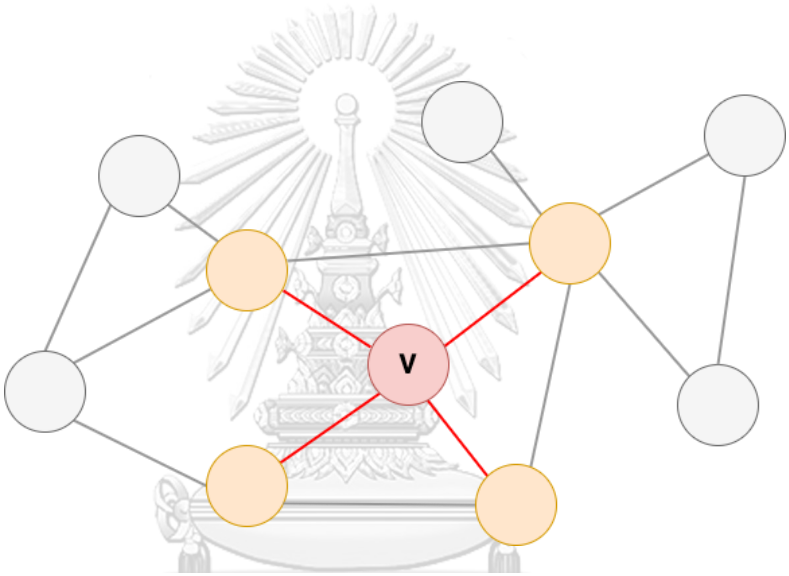


*Figure 1: The neighborhood of a vertex*



Secondly, the structural similarity is computed to measure the similarity of the two vertices; a significant similarity value means this pair is a clique member and a small value for hub and outliers.

$$Structural\ Similarity\ (v,w) = \frac{\text{Structural Similarity (v, w)}}{\sqrt{(no.v's\ neighbors) \times (no.\ w's\ neighbors)}} \quad (14)$$

It is considered the core whether the vector has several neighborhoods with a structural similarity score greater than the threshold. Direct Structure Reachable of v

and w is defined when v is the core and w, whose has structural similarity higher than the threshold, is a neighbor of v. Finally, all vertices, including cores and the Direct Structure Reachable, are added to the SCAN algorithm to get Structure-Connected Cluster as shown in Figure 2.

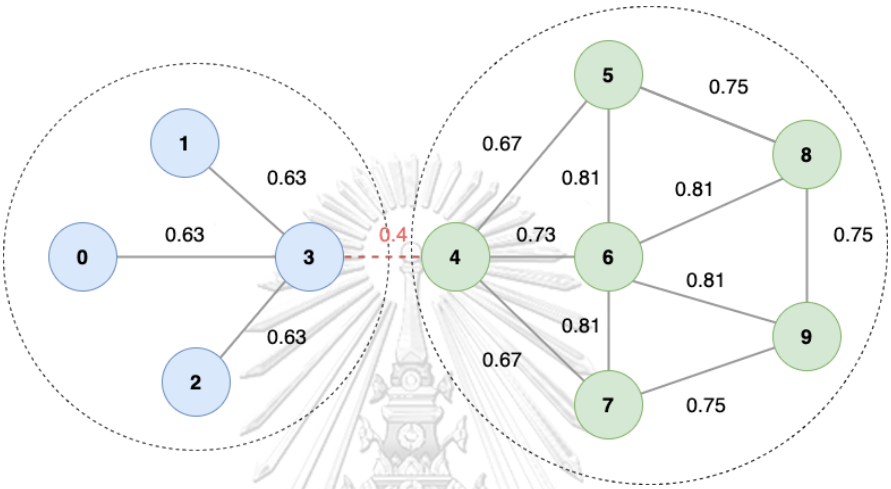


*Figure 2: The Structure-Connected clusters*

# 3. REVIEW OF LITERATURE

The following section summarizes various approaches we use to build spammer group detections.

## 3.1. Spammer Group Detection and Diversification of Customers' Reviews

This study proposed a novel method to detect the suspicious spammer groups on the eCommerce platform (Hussain et al., 2021). Co-reviewer graphs hold the reviewers who share similarities in review post time and rating indicators. Next, the co-reviewer was fed into the Structural Clustering Algorithm for Networks (SCAN) algorithm to spot the candidate spam groups. Individual indicator set {time burstiness, maximum number of reviews, and average rating deviation}; and group indicator set {review tightness, product tightness, rating variance, group size, and reviewer ratio} are used to calculate the spam score for each candidate spam group. The candidate spam group is identified as a spammer if the score exceeds the predefined threshold value. This study's outcome helped identify the group of spammers and classify the non-spam reviews. However, it still has an area to improve precision and recall in identifying the spammer group method, for instance, by adding the linguistic indicators.

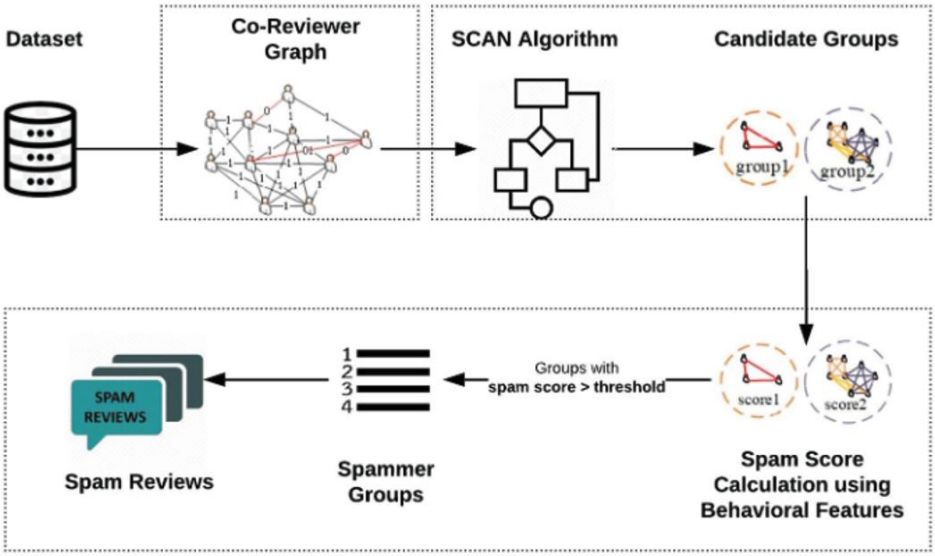


*Figure 3: The framework of the Spammer Group Detection (SGD) method*

## 3.2. Using Centrality Measures to Predict Helpfulness-Based Reputation in Trust Networks

Centrality measures are essential to identify the most influential or central position in the trusted network (Meo et al., 2017). This study applied five centrality measures – Degree Centrality, Closeness Centrality, Betweenness Centrality, PageRank, and Eigenvector Centrality– to calculate the centrality-based reputation (CBR) scores. The CBR scores were investigated to determine whether they could be used as a helpfulness-based reputation (HBR) reviewer. As a result, it showed that CBR scores could predict the HBR ones, and the Eigenvector Centrality performed the best among the other measures. To prove the result, they applied Gradient Boosting Regression to highlight the relationship between CBR and HBR scores, and the experiments confirmed that CBR scores are good predictors of HBR ones.

## 3.3. Do Reviewers' Words and Behaviors Help Detect Fake Online Reviews and Spammers? Evidence from a Hierarchical Model

While most studies in spam detection on eCommerce platforms focused on reviewers' behavior and review metadata such as time and rating, this study was interested in the fact that the same reviewer who writes reviews often shows a unique style or pattern because each reviewer has their distinct style (Le et al., 2022; Pennebaker & Lay, 2002). Hierarchical Logistic Regression (HLR)-base model proposed in detecting the fake reviews based on both linguistic {LIWC factors (Pennebaker & Lay, 2002), readability, credibility, evidentiality, and POS}; and behavioral characteristics {rating entropy, rating deviation, review count, lift tenure, and review gap}. The outcome showed that HLR could identify fake reviews and review spammers more accurately than the standard machine-learning algorithms – Support Vector Machine (SVM), K-Nearest Neighbors algorithm (KNN), Naive Bayes classifier (NB), and Random Forest (RF). This study evaluated the proposed model by recency and duration analyses as shown in Figure 4.

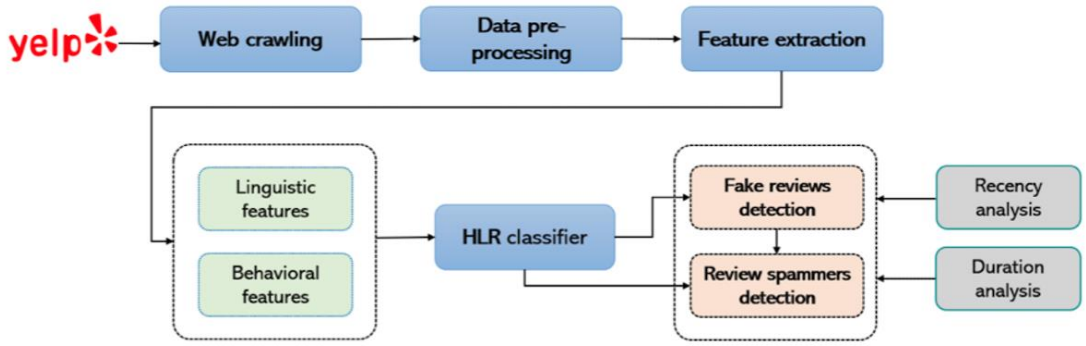


*Figure 4: The system framework of the HLR model.*

## 3.4. Collective Opinion Spam Detection: Bridging Review Networks and Metadata

This study proposed SPEAGLE, which utilizes the metadata (text, timestamp, and rating) and the relational data (review network) from the three real-world review datasets from Yelp.com (Rayana & Akoglu, 2015). SPEAGLE could detect fake reviews and spammers. Moreover, this study designed a light version of SPEAGLE called SPLITE, and it performed more efficiently than the SPEAGLE due to using a subset of features to reduce computational overhead. Finally, SPEAGLE was the most significant scale qualitative evaluation performed to date for the opinion spam problem, outperforming several baselines and state-of-the-art methods.

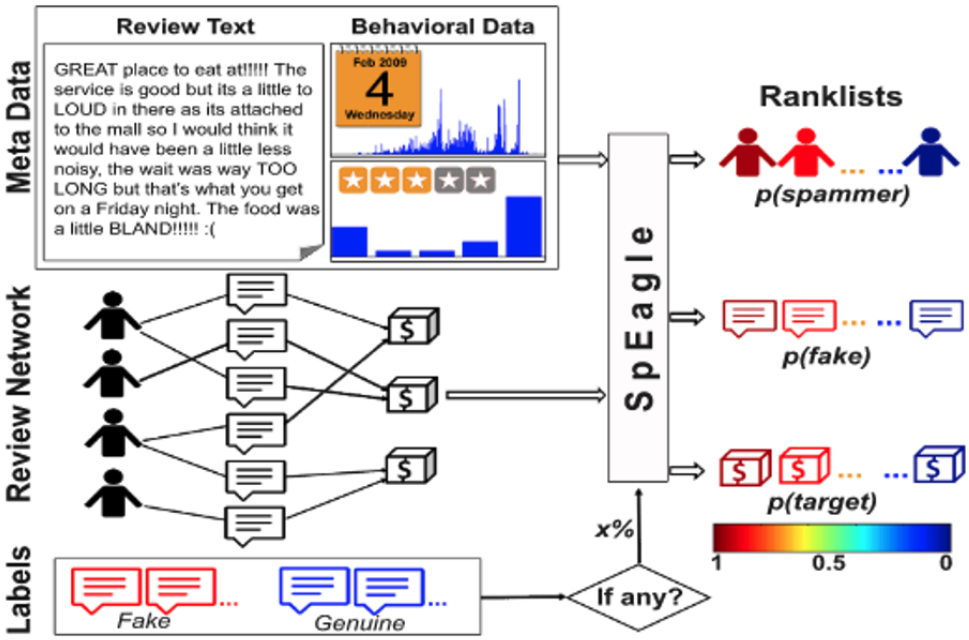


*Figure 5: The system framework of the SPEAGLE*

# 4. PROPOSED GROUP OF THE SPAMMER DETECTION

This chapter introduces the methodology of the proposed spammer detection group and the dataset we use for our spammer detection.

Streitfeld, Sun and Loparo (Sun & Loparo, 2019) revealed one-third of customer reviews were a fraud. Indeed, D'Onfro (D'Onfro, 2013) reported that the number of fake reviews on Yelp rose to 20% in 2013 from only 5% in 2006. Furthermore, in many previous studies, the Yelp dataset is the most frequently used in review spammer detections. Thankfully, we received the Yelp dataset from Shebuti Rayana. Rayana and Akoglu (Rayana & Akoglu, 2015) proposed the first spammer detection study that used the Yelp dataset. Moreover, the YelpNYC dataset is the same dataset our main study applied to detecting group review spam for training their model (Hussain et al., 2021). The YelpNYC is a review text file containing reviews from New York hotels and restaurants from 2004 to 2015. There are 359,052 reviews written by 160,225 reviewers, 10.27% spam reviews, and 17.79% spammers.



Since our main contribution is constructing a centrality measure of spammers into the products/services dimension to identify the relationship of suspicious spammers, the spammer is seen as an analogy for a vector of disease, and products/services are the disease container. Hence the reviewers who have reviewed the disease container (products/services) will have more chance of getting the disease or being one of the spammer groups. However, this process requires the reviewers' health records (whether they are spammers). We first utilize the XGBoost for classification to predict the potential spammers based on the linguistics review and reviewers' behavior. We give attention to the linguistic aspects because the writing style is unique and challenging to develop in a short period. Therefore,

distinguishing spam reviews using linguistics is crucial (Le et al., 2022; Ott et al., 2011). Behaviors of reviewers are no less important, according to many previous studies that have been proposed. (Choo et al., 2015; D'Onfro, 2013; Hussain et al., 2021; Wang et al., 2018; Zhang et al., 2018). The definition of each indicator is clarified as follows:

1. LIWC approach was developed by Pennebaker and Lay (Pennebaker & Lay, 2002). It separates the review content into the 11 linguistic dimensions– effective processes, cognitive processes, negations, pronouns, quantifiers, social words, tentative words, word count, family-related words, leisure-related words, and words longer than six letters.
2. Readability was introduced over the past 80 years (DuBay, 2004). The higher score the review gets, the more genuine score gets.
3. POS tagging considers the particular part of speech into four main categories– verbs, adverbs, adjectives, and superlatives.
4. Evidentiality ranks the reviewer's confidence in model adverbs, lexical, auxiliary verbs, and epistemic adjectives (Su et al., 2010).
5. Time burstiness with three days threshold is defined based on the study that the spammer accounts activated in a short time to achieve their goal.
6. A maximum number of reviews in a single day could spot the spammer because spammers usually post many reviews in a single day.
7. Average rating deviation is another indicator that can identify the spammer. Logically, a spammer gives the same product ratings to promote or defame the product.



We build a co-reviewer graph (Hussain et al., 2021; Kaghazgaran et al., 2019). The co-reviewer graph links two reviewers who show similarities in reviewing products and time. The co-reviewer pairs the similarity between reviewer V and W,

represented by CRS(V, W, p), where p is a product V and W both reviewed in common. t is the time when the reviewers reviewed product p, and R represents the rating score of the product p.

$$CRS(V,W,p) = \begin{cases} 0, & \left(\left|t_V^p - t_W^p\right| > \alpha\right) OR \left(\left|R_V^p - R_W^p\right| > \beta\right) \\ 1, & otherwise \end{cases} \quad (15)$$

Then, we use co-reviewer similarity (CRS) to compute the weight edge between reviewer V and reviewer W, represented by $\lambda(V,W)$; the weight edge equals one means both V and W show similarity; in other words, V and W are co-reviewers.

$$\lambda(V,W) = \begin{cases} 0, & \forall_p \in P_A \cap P_B, CRS(V,W,p) = 0 \\ 1, & otherwise \end{cases} \quad (16)$$

Next, we utilize eigenvector and eigenvalue to rank a product's disease/spam score in the network using the individual suspicious spammer from XGBoots. Finding the network's centrality is our inspiration in this step. As aforementioned, we name the product reviewed by a spammer as a product review vector of disease to infer that once the reviewers review this product, they have more chance of being a spammer. The disease/spam will spread to any reviewer who comes to review the product vector of the disease. For example, the reviewer, who often reviews the product vector of the disease, will show a high risk get the disease; in other words, this reviewer has a high chance of being a spammer.

Last but not least, to form the group of suspicious reviewers, we feed the co-reviewer graphs into the SCAN algorithm. As a result, we get a group of strong relationships among reviewers, and the group member shares common interests. As mentioned in Review of Literature chapter, the SCAN algorithm is driven by three main fragments: structural similarity score, identifying core node process, and Direct Structure Reachable (DSR).

Once we get the suspicious group of spammers from the SCAN algorithm, we apply the group indicator as the following to increase the precision of our detection.

1. Review tightness shows the ratio of the number of reviews in a group and the cardinality of the Cartesian product–the reviewer set and the product set in the group.
2. Rating variance identifies the rating tone of the group, which the spammer group mostly reviews in the same direction.
3. The reviewer ratio informs whether the suspicious group has ever dominated the product reviews.



Finally, in each group, we sorted the suspicious group by the percentage of individual spammers (suspicious individual spammers from the XGboost for Classification). For instance, the first rank of the suspicious group is the group with the highest percentage of individual spammers; let us say it has five individual spammers out of six members in the group, or 83%.

To achieve our objective, optimize the investigation process and minimize the full-time equivalent (FTE). We provide a report of suspicious spammer groups with

supporting reasons using 95% confidential intervals, the interquartile range (IQR), and the number of suspicious individual spammers in each group.

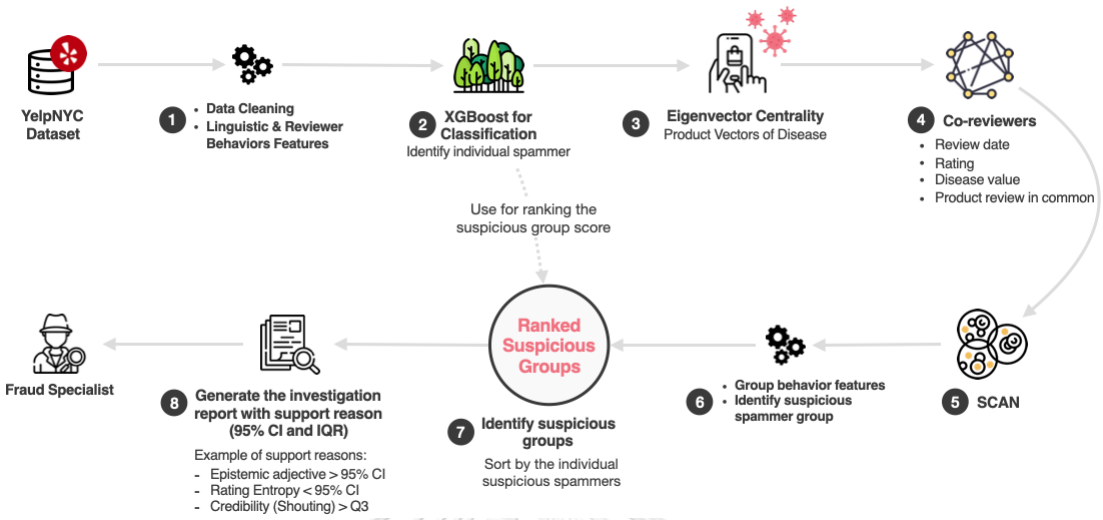


*Figure 6: The framework of the proposed group of the spammer detection method*

**Algorithm 1: SCAN Algorithm**

**input**: $V = v_{n-1}, v_{n-2}, v_{n-3} \dots v_0$
**output**: G = $g_{n-1}, g_{n-2}, g_{n-3} \cdots g_0$
:
Prepare an empty dict for suspicious groups $G$
$grp\_id \leftarrow 0$
**for** *each v in V* **do**
  **if** $grp\_id = 0$ *or* $num\_member(grp) > 0$ **then**
    $grp\_id \leftarrow grp\_id + 1$
    G $\leftarrow Group(grp\_id)$
  **if** *v is a core* **then**
    **for** *each y in neighbors of v* **do**
      **for** *each x in vertex of y* **do**
        **if** $x \notin \forall G$ *and* $similarity\_scores(x) > threshold$ **then**
          G $\leftarrow x$
    **if** $num\_member(G) = 1$ **do**
      G $\leftarrow \emptyset$

# 5. RESULTS

In this chapter, we will summarize our proposed group of spammer detection to Hussain's detection (Hussain et al., 2021). The results we compare are not from their original framework of spammer group detection. We aim to develop a framework that first applies the centrality measures of influencer/spammer into the products/services dimension. Second, we combine three characteristic indicators: linguistic review, reviewer behavior, and group behavior. Last, we propose a different strategy for identifying the spammer group.

We use the same YelpNYC dataset in the implementation process with Hussain's study. On the one hand, Hussain et al. used YelpNYC only for the training model part and another dataset for the testing part. On the other hand, we use YelpNYC in both the training and testing parts. YelpNYC dataset was collected from 2004 to 2015. It has 160,225 reviewers (17.29% spammers) and 359,052 reviews (10.27% spammer reviews).



For this study, we focus on the high recall and spammer rate of each group of spammers because the purpose of the study is to decrease the burden of investigation resources, such as the workload of spammer specialists and time-consuming. Since we have a ground truth of an individual spammer, we propose that the suspicious spammer groups will be predicted correctly if the groups have more than 50% of the individual spammers. Then, we compare the detection performance by selecting the most suspicious spammer group from 30 to 179 groups. As a result, our result shows higher precision and recall, with fewer suspicious reviewers in the group that led to our objective.

| No. Group | Fraud Precision (%) | | Fraud Recall (%) | | Accuracy (%) | |
|---|---|---|---|---|---|---|
| | Previous Study | Proposed Study | Previous Study | Proposed Study | Previous Study | Proposed Study |
| 30 | 7 | 29 | 100 | 75 | 7 | 43 |
| 40 | 7 | 28 | 100 | 80 | 7 | 42 |
| 50 | 6 | 25 | 100 | 80 | 6 | 48 |
| 60 | 7 | 22 | 100 | 80 | 7 | 50 |
| 70 | 6 | 19 | 100 | 80 | 6 | 47 |
| 80 | 6 | 17 | 100 | 80 | 6 | 49 |
| 90 | 7 | 15 | 100 | 80 | 7 | 49 |
| 100 | 7 | 15 | 100 | 82 | 11 | 48 |
| 110 | 7 | 15 | 100 | 82 | 19 | 52 |
| 120 | 7 | 15 | 100 | 82 | 26 | 56 |
| 130 | 7 | 15 | 88 | 82 | 31 | 59 |
| 140 | 7 | 15 | 88 | 82 | 36 | 62 |
| 150 | 7 | 15 | 78 | 82 | 39 | 65 |
| 160 | 7 | 15 | 70 | 82 | 42 | 67 |
| 170 | 7 | 15 | 64 | 82 | 45 | 69 |
| 179 | 7 | 15 | 64 | 82 | 48 | 70 |

*Table 1: Performance comparison of fraud group class of spammer groups detection using YelpNYC dataset.*

| No. Group | No. Reviewers | | No. Suspicious | |
|---|---|---|---|---|
| | Previous Study | Proposed Study | Previous Study | Proposed Study |
| 30 | 239 | 227 | 239 | 115 |
| 40 | 318 | 301 | 318 | 162 |
| 50 | 390 | 389 | 390 | 178 |
| 60 | 506 | 490 | 506 | 208 |
| 70 | 614 | 574 | 614 | 243 |
| 80 | 674 | 663 | 674 | 263 |
| 90 | 799 | 758 | 799 | 290 |
| 100 | 858 | 827 | 835 | 329 |
| 110 | 943 | 927 | 835 | 335 |
| 120 | 1036 | 992 | 835 | 335 |
| 130 | 1107 | 1121 | 835 | 335 |
| 140 | 1181 | 1171 | 835 | 335 |
| 150 | 1236 | 1231 | 835 | 335 |
| 160 | 1300 | 1327 | 835 | 335 |
| 170 | 1396 | 1380 | 835 | 335 |
| 179 | 1538 | 1429 | 835 | 335 |

*Table 2: Comparison of the number of investigations where the number of reviewers are in normal and spammer groups, and the number of suspicious are in the spammer group.*

# 6. DISCUSSION

## 6.1. Summary of Findings

This study contributes a framework of group spammer detection to facilitate the investigation process and reduce the workload of spammer specialists. We perform several experiments and angles of comparison with the previous study (Hussain et al., 2021), and the results show that our framework performs better in optimizing the investigation workload. Moreover, the framework can minimize the previous study's workload and reveals higher precision and recall.

This study has aggregated numerous ideas and inspirations from many superb studies, including Hussain et al. In the development phase, we develop the spammer group detection following Hussain et al. step by step and test the performance by the same dataset as them; we find that the precision from their detection can be improved based on our measurement method. The suspicious groups from the mocked detection contain many non-spammer or normal reviewers, and the size of the suspicious groups need to be smaller to reach this proposal's objective. To improve the precision and reduce the number of suspicious reviewers at once, we try to spot the suspicious spammers and track them by the product they reviewed. Since the spammer groups aimed to control the product's perspective (Mukherjee et al., 2012), they were inclined to write many reviews on the same product. We take how spammer groups work together as an advantage by adopting Network Centrality to capture the product many suspicious spammers reviewed. Hence, we can create co-reviewers with higher confidence that they might work together as a spammer group. The result confirms that we get higher precision and a smaller group as expected which is a significant point of our objective.

Another interesting finding from the result is that the linguistic spammer indicator can guide us to find some suspicious spammers. Still, some real spammers were very good at camouflaging themselves since their writing styles and linguistics showed an unsuspicious sign. However, three indicators, linguistic, reviewer behaviors, and group behaviors, perform much better than only one or two indicators; they can tell us more than just the writing style but demonstrate the behaviors that help to create rule-based to prevent or catch the spammer group better.

## 6.2. Future Work

Even though our framework performs better than the previous study, it still has several areas that can be improved to increase the precision and find new contributions of the spammer group detection. For instance, using up-to-date datasets because customer behaviors have been changing during the pandemic hit (Brewster, 2022; OECD, 2020; UNCTAD, 2021). The e-commerce platforms promoted new business strategies to get customers into their platforms, such as giving a free shipping campaign or a flash sale period. Therefore, spammers and their groups must update their tricks to avoid being caught and make their spam reviews look authentic. That means we must update or add new behavior indicators up-to-date and retain the model.

Another crucial point to determine the performance matrix in the future is the ground-truth labels of the spammer group, which is the most challenging work because it requires the spammer specialists to make the ground-truth labels, and we need to collaborate and work closely with the e-commerce platform.

# REFERENCES


3Blue1Brown. (2016). *Eigenvectors and eigenvalues*. Retrieved 2022-06-12 from https://www.3blue1brown.com/

Brewster, M. (2022). *Annual Retail Trade Survey Shows Impact of Online Shopping on Retail Sales During COVID-19 Pandemic*. Retrieved 2022-09-17 from https://www.census.gov/library/stories/2022/04/ecommerce-sales-surged-during-pandemic.html

Chen, T., & Guestrin, C. (2016, 8). XGBoost: A Scalable Tree Boosting System.

Choo, E., Yu, T., & Chi, M. (2015). Detecting Opinion Spammer Groups Through Community Discovery and Sentiment Analysis. In (pp. 170-187). https://doi.org/10.1007/978-3-319-20810-7_11

Coppola, D. (2022). *E-commerce worldwide - statistics & facts*. Retrieved 2022-09-17 from https://www.statista.com/topics/871/online-shopping/#topicHeader__wrapper

D'Onfro, J. (2013). *No Title*. Insider. Retrieved 2022-07-02 from https://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9

DuBay, W. (2004). *The Principles of Readability*. Retrieved 2022-05-07 from https://www.researchgate.net/publication/228965813_The_Principles_of_Readability

EigenvectorCentrality. (2022). *Eigenvector centrality - Wikipedia*. Retrieved 2022-05-10 from https://en.wikipedia.org/wiki/Eigenvector_centrality.

Hussain, N., Mirza, H. T., Ali, A., Iqbal, F., Hussain, I., & Kaleem, M. (2021). Spammer group detection and diversification of customers' reviews. *7*, e472. https://doi.org/10.7717/peerj-cs.472

Kaghazgaran, P., Alfifi, M., & Caverlee, J. (2019, 11). Wide-Ranging Review Manipulation Attacks.

Law, T. (2021). 19 Ecommerce Statistics that will Guide Your Strategy in 2021. Retrieved 2022-09-17 from

Le, T.-K.-H., Li, Y.-Z., & Li, S.-T. (2022). Do Reviewers' Words and Behaviors Help Detect Fake Online Reviews and Spammers? Evidence From a Hierarchical Model. *10*, 42181-42197. https://doi.org/10.1109/ACCESS.2022.3167511

Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., & Shao, J. (2017, 4). Bimodal Distribution and Co-Bursting in Review Spam Detection.

Li, J., Lv, P., Xiao, W., Yang, L., & Zhang, P. (2021). Exploring groups of opinion spam using sentiment analysis guided by nominated topics. *171*, 114585. https://doi.org/10.1016/j.eswa.2021.114585

Luca, M. (2016). *Reviews, Reputation, and Revenue: The Case of Yelp.com*. Retrieved 2022-06-17 from https://www.hbs.edu/ris/Publication Files/12-016_a7e4a5a2-03f9-490d-b093-8f951238dba2.pdf.

Meo, P. D., Musial-Gabrys, K., Rosaci, D., Sarnè, G. M. L., & Aroyo, L. (2017). Using Centrality Measures to Predict Helpfulness-Based Reputation in Trust Networks. *17*, 1-20. https://doi.org/10.1145/2981545

Mukherjee, A., Liu, B., & Glance, N. (2012, 4). Spotting fake reviewer groups in consumer reviews.

Mukherjee, A., Liu, B., Wang, J., Glance, N., & Jindal, N. (2011). Detecting group review spam.

OECD. (2020). *E-commerce in the time of COVID-19*. Retrieved 2022-09-17 from https://www.oecd.org/coronavirus/policy-responses/e-commerce-in-the-time-of-covid-19-3a2b78e8/

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. 309--319. https://aclanthology.org/P11-1032

Pennebaker, J. W., & Lay, T. C. (2002). Language Use and Personality during Crises: Analyses of Mayor Rudolph Giuliani's Press Conferences. *36*, 271-282. https://doi.org/10.1006/jrpe.2002.2349

Press, M. (2015). *Seven In 10 Americans Seek Out Opinions Before Making Purchases*. Retrieved 2022-06-17 from https://www.mintel.com/press-centre/social-and-lifestyle/seven-in-10-americans-seek-out-opinions-before-making-purchases

Rayana, S., & Akoglu, L. (2015, 8). Collective Opinion Spam Detection.

Shepherd, J. (2022). *The Social Agency For Growth-Focused Brands. The Social Shepherd*. Retrieved 2022-06-25 from https://thesocialshepherd.com/

Su, Q., Huang, C.-R., & Chen, H. K.-y. (2010). Evidentiality for Text Trustworthiness Detection.

Sun, Y., & Loparo, K. (2019, 11). Opinion Spam Detection Based on Heterogeneous Information Network.

Tang, X., Qian, T., & You, Z. (2020). Generating behavior features for cold-start spam review detection with adversarial learning. *526*, 274-288. https://doi.org/https://doi.org/10.1016/j.ins.2020.03.063

UNCTAD. (2021). *How COVID-19 triggered the digital and e-commerce turning point*. Retrieved 2022-01-26 from https://unctad.org/news/how-covid-19-triggered-digital-and-e-commerce-turning-point

Vidanagama, D. U., Silva, T. P., & Karunananda, A. S. (2020). Deceptive consumer review detection: a survey. In (Vol. 53, pp. 1323-1352).

Wang, Z., Gu, S., Zhao, X., & Xu, X. (2018). Graph-based review spammer group detection. *55*, 571-597. https://doi.org/10.1007/s10115-017-1068-7

Weise, K. (2021). *A Lie Detector Test for Online Reviewers*. bloomberg.com. Retrieved 2021-10-12 from https://www.bloomberg.com/news/articles/2011-09-29/a-lie-detector-test-for-online-reviewers.

Xu, C., & Zhang, J. (2015, 11). Towards Collusive Fraud Detection in Online Reviews.

Xu, X., Yuruk, N., Feng, Z., & Schweiger, T. A. J. (2007). SCAN.

Ye, J., & Akoglu, L. (2015). Discovering Opinion Spammer Groups by Network Footprints. In (pp. 267-282). https://doi.org/10.1007/978-3-319-23528-8_17

Yelp, I. (2022). *Yelp - Wikipedia*. Retrieved 2022-06-29 from https://en.wikipedia.org/wiki/Yelp.

Zhang, L., Wu, Z., & Cao, J. (2018). Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model. *6*, 2559-2568. https://doi.org/10.1109/ACCESS.2017.2784370

# VITA

| | |
|---|---|
| **NAME** | Natthanicha Suriyamongkol |
| **DATE OF BIRTH** | 26 July 1992 |
| **PLACE OF BIRTH** | Bangkok Thailand |
| **INSTITUTIONS ATTENDED** | Thammasat University |
| **HOME ADDRESS** | 199/1 Mu 2.<br>Lamsai Lamlukka<br>Pathumtani<br>12150 |
| **PUBLICATION** | Suriyamongkol N, Banani S, Thiemjarus S, Noothong W. Preventative Care with Disease Risk Prediction on Health Tracking Application. International Convention on Rehabilitation Engineering and Assistive Technology (i-CREATe), Canberra Australia, August 26th to 29th, 2019. |
| **AWARD RECEIVED** | - Selected to receive Samart Innovation Funds for the graduation research project, 2015<br>- Highest GPA graduate, Computer Engineering, Thammasat University, 2015<br>- Highest GPA award, Computer Engineering, Thammasat University, 2013 |