Graphic Model Analysis of Frauds in Online Consumer Reviews

Chungsik Song*[1], Kunal Goswami*[2], Sang-Yoon Chang[§3], Euijin Choo[†4], Young Hee Park*[5]

[*]San Jose State University, San Jose, USA,
[§]Advanced Digital Science Center,
[†]Missouri University of Science and Technology
[1]chungsiksong@sjsu.edu, [2]kunal.goswami@sjsu.edu,
[3]sychg@adsc.com.sg, [4]chooe@mst.edu, [5]younghee.park@sjsu.edu

***Abstract*—We often rely on online consumer reviews and opinions posted on social media to make decisions in our daily lives. This article will address what are collectively referred to as opinion spam, which are opinions posted by fake reviewers who seek to promote or tear down target entities for financial gain. This has led industry and academic research to seek to develop an efficient and scalable framework to detect such opinion spam. In this paper, we propose a fraud detection framework by uncovering new features and network effects among reviewers and products. We study a Yelp data set for online reviews using graph-based methods that leverage the relational ties among reviewers, reviews, and businesses. We utilize clues from the study of structural properties of user graphs. We consider user networks in which reviewer nodes are connected to each other as friends. We investigate structural properties of user networks for recommended (non-spam) and fake (spam) reviewer groups. Through intensive computations involving these groups, we demonstrate that graphs for groups of recommended reviewers show characteristics of a small-world network. However, graphs for groups of fake reviewers reveal properties closer to those of a random network.**

Keywords—Opinion Spam; Graph Theory, Small World Network

I. INTRODUCTION

Online consumer reviews of products and businesses have become increasingly popular and have become valuable sources for consumer decision-making. As the Internet and Web are ubiquitous and have become an integral part of our daily lives, these reviews are now part of everyday decision-making and have direct influence on product and business sales [3, 14]. Yelp is one of the leading consumer review web sites, and alone contains more than 70 million reviews of businesses, with a market capitalization of roughly four billion dollars. However, the credibility of these reviews is fundamentally undermined when businesses commit review fraud, creating fake reviews for themselves or their competitors. Due to the financial gains associated with the positive reviews, opinion spam has made it difficult if not impossible to validate the authenticity of reviews. The financial benefits reaped from such reviews have even created a market of users paid to fabricate fake reviews either to fabricate hype to promote business or to tear down competitive products or businesses. Recent research has reported that one-third of all consumer reviews on the Internet are estimated to be fake [23].

The problem of opinion spam has been addressed by focusing on extracting and summarizing opinions from reviews using natural language processing and data mining techniques. This linguistic approach analyzed the language patterns of filtered versus non-spam users. In the end, classification using linguistic features has not proven to be efficient [17]. Jindal et al [8] proposed behavioral approaches that utilize the behavior of fake users, using several heuristics, such as duplicated reviews or acquiring bogus reviews from non-experts to generate pseudo-ground truth, or reference datasets. This data is then used for learning classification models together with carefully engineered features. However, the features might not be consistent even for datasets within the same domain, depending on the dataset source.

An unsupervised, general, and network-based framework to detect fraudsters and fake reviews was proposed to rely upon relationships among various entities in consumer reviews, including reviewers, businesses and reviews [26]. This presents a significant challenge, however, in analyzing big and complex relational datasets and requires innovative methodological approaches beyond basic data mining. There has been increasing interest in the study of large, real world, complex networks, in which graph theory is used to model the relationship between the entities [18]. Graph analysis yields a theoretical mathematical description of a network that is composed of numerous interrelated nodes [1, 7, 12, 16, 22, 26, 30] . Unlike most other ways of looking at data, graphs are designed to express relatedness. Graph networks can uncover patterns that are difficult to detect using traditional representations. Graph-based representations offer new methods of uncovering fraud rings and other sophisticated scams with a high-level of accuracy, and are capable of detecting advanced fraud scenarios in real-time.

In this paper, we propose a framework to identify fraud reviews by utilizing the friend relationship among reviewers, reviews, and businesses. Based on small-world networks [28], we analyze the structural properties of networks of reviewer groups with new meta data. We analyze the Yelp online review dataset using graph-based methods and approach the problem of fraud review detection as a network classification task involving the review networks. We also discover new features

using graph-based methods and identify suspicious users and reviews. The uncovered structural properties of a fake review group are used for the detection of fraudulent reviews. For analysis, we add new characteristics into the Yelp data by collecting new additional meta data: number of friends, number of reviews, and number of votes.

To do so, we first consider a user network that leverages the relational ties among reviewers. We generate two types of graphs for two user groups: one for a reviewer group whose reviews have been filtered as reviews not thought to be genuine reviews (spam), which we call the fake review group; another for the group whose reviews we believe to have been used to genuinely rate the value of business and products (non-spam), which we call the recommended review group. We calculate structural properties of these user graphs and compare them. In particular, we are interested in two distinguishing characteristics of small-world networks, the local clustering coefficient and global characteristic path length. We find that the structural properties of the graph for the "genuine reviewer group" are those of a small-world network. On the other hand, characteristics of the "fake review group" are closer to those of a random network. Second, we exploit structural properties of the fake review group in our extension of the framework to detect fraudulent reviews. The extended framework utilizes clues from meta data as well as from relational data, and exploits them collectively to spot suspicious users and reviews. It is demonstrated that our extended framework is more effective on Yelp review datasets than previous work.

We proceed in section 2 of this paper to summarize other work related to opinion fraud. We then discuss our methodology and motivations in section 3. We analyze the Yelp review dataset in section 4. Results are presented in section 5. Section 6 is our conclusion.

II. Related Works

Much of the previous work in opinion fraud has focused on review text content, behavioral analysis in supervised methods, and relational analysis of network effects using graph-based methods.

Characteristics of language that the opinion spammers use and how it differs from the language used in genuine reviews were investigated. Natural language processing and data mining techniques were used for opinion extraction and sentimental classification of reviews. Ott et al. [20] used supervised learning models to detect deceptive reviews based on linguistic features of reviews as well as features borrowed from studies in psychology. Feng et al. [5] investigated syntactic stylometry for deception detection, and show that features derived from context-free-grammar parse trees improve performance over shallow lexico-syntactic features. However, Mukherjee et al. [17] have shown that classifications using linguistic features were not efficient. The authors analyzed the effectiveness of linguistic and behavioral clues on a Yelp dataset with reviews that they either filtered out or recommended, and found that linguistic features are not as effective as behavioral clues.

Behavioral patterns in review data have been analyzed by data mining technology to detect fake reviews from genuine ones. Approaches in this category often leverage features indicative of spam extracted from metadata associated with user behavior (e.g., rating distribution), review content (e.g., number of capital letters), and product profile (e.g., brand and price). Jindal et al. [8] identified opinion spam by detecting duplicate reviews and using supervised learning with manually labeled training examples. Li et al. [12] used sentiment scores, product brands, and reviewer profile attributes to train classifiers, and used the two views from reviews and users under a co-training framework to spot fake reviews. Jindal et al. [9] proposed rule-based discovery of unusual patterns in review data associated with the rating and brand distribution of user reviews and found unexpected rules to highlight anomalies. Mukherjee et al. [15] utilized reviewing behaviors of users in an unsupervised Bayesian inference framework to detect opinion spammers and used frequent item set mining to find fraudulent reviewer groups. Xie et al. [29] monitored temporal behavior of products by tracking their average rating, review count, and ratio of one-time reviewers, to spot suspicious single-time reviewers. Unfortunately, these methods utilizing behavioral approaches are not generalizable: the models need re-training to account for differences between problem domains, such as book reviews versus movie reviews. Moreover, the features might not be consistent even for datasets within the same domain, depending on the dataset source. Consequently, feature extraction becomes a time-consuming yet pivotal sub-problem with attributes varying across domains.

Graph-based approaches that account for the network of reviewers, reviews, and products can more elegantly encapsulate structural signals that go beyond review content, simple heuristics, and behavioral analysis, and thus can be generalized across domains. Wang et al. [26] proposed the concept of a heterogeneous review graph to capture the relationships among reviewers, reviews, and businesses. To show how interactions between nodes in this graph can reveal the cause of spam, these authors proposed an iterative model to identify suspicious reviewers. Akoglu et al. [1] proposed a fast and efficient framework that detects fraudsters and fake reviews based on the rigorous theoretical foundations of belief propagation, and is linearly scalable. The Markov Random Field (MRF) has been utilized to model signed bipartite network of users and products that are connected through positive or negative review relations (signed edges). S. Rayana and L. Akoglu [22] developed a new holistic approach called SpEagle that utilizes clues from all metadata (text, timestamp, rating) as well as relational data (network) considered in [1]. Li et al. [12] constructed a user-IP-review graph to relate reviews that are written by the same users and from the same IPs.

Besides detecting individual spammers, there has also been work on identifying spammer groups through group level behavioral indicators of spam [16] and graph-based methods [7, 30]. Overall performance of these unsupervised approaches has been disappointing.

III. METHODOLOGY

A. Graph Theory and Small World Network

A set of concepts along and the relationship between them can lead to significant knowledge of a domain. A graph, consisting of nodes and edges, is then used to describe interactions between these concepts, where the node represents concepts and the edges represent their relationships. A graph is complete if every vertex is connected to every other vertex by an edge. In a given graph, there are many subsets of vertices that are complete, and we call these complete subgraphs cliques. The presence of many large cliques in a graph indicates that the graph has the kind of locally dense structure that we see in real small-world network [28]. Small-world networks display two distinguishing characteristics: they have a high clustering coefficient while still retaining a small characteristic path length.

1) Local clustering coefficient

The local clustering coefficient $C(p)$ measures the local density (cliquishness) of the neighborhood of each node in the network. It is given by the ratio of the actual triangle count at a vertex to the number of possible triangles at that vertex based on how many neighbors it has. A triangle is a complete graph on three vertices, and the triangle count at a vertex V is simply the number of triangles that contain V. Roughly speaking, the local clustering coefficient tells how well connected the neighborhood of the node is. If the neighborhood is fully connected, the clustering coefficient is 1. On the other hand, a value close to 0 means that there are hardly any connections in the neighborhood.

For an undirected graph, the local clustering coefficient C for a vertex that has k neighbors and t triangles is calculated as

$$\mathrm{C} = 2t/k(k-1) \tag{1}$$

On a random graph, the probabilities of vertex pairs being connected by edges are by definition independent, so there is no greater probability of two vertices being connected if they have a mutual neighbor than if they do not. This means that the clustering coefficient for a random graph is simply the same as the probability, or

$$\mathrm{X} \cong \frac{\mathcal{Z}}{\mathcal{N}_{\mathcal{V}}} \tag{2}$$

where $\mathcal{N}_{\mathcal{V}}$ is the number of vertices and $\mathcal{Z}$ is the mean degree of the dataset

2) Characteristic Path Length

The second property of a small-world network is that the length of the shortest path between two randomly chosen nodes tends to be small. Characteristic path length $L(p)$ is defined as the median of the means of the shortest path lengths connecting each vertex to all other vertices; it measures the typical separation between two vertices in the graph (global properties). Let $d_G(x,y)$ be the distance between the vertices $\mathrm{x},\mathrm{y} \in \mathcal{V}(\mathrm{G})$ for a connected graph G with vertices $\mathcal{V}(\mathrm{G})$. The characteristic path length is then defined as

$$L \equiv \frac{\sum_{x,y\in\mathcal{V}(G)} d_G(x,y)}{n(n-1)} \tag{3}$$

where n is the number of vertices in G and the sum is in the range of all pairs of vertices of G.

B. User Network with Friend Relationship

Small-world network properties can be found in the World Wide Web, social networks, the global economy system, and even in the human nervous systems. A recent study indicates that the network of neurons in the brain exhibits a small-world structure, and that deviance from this structure can be indicative of the potential for functional problems [21]. Researchers have used resting-state functional MR imaging and theoretical graph approaches and have shown abnormalities in intrinsic brain networks in patients with different abnormal conditions, including Alzheimer disease (AD), schizophrenia, attention deficit hyperactivity disorder, epilepsy, and traumatic brain injury [27]. For example, in patients with AD, Supekar et al [24] found a significant decrease in the clustering coefficient and small-world properties in patients with AD compared with control subjects, consistent with lower regional connectivity and disruption of global organization of brain networks. In general, real-world graphs should exhibit the small-world property. If they do not, that may be evidence of a problem, such as fraudulent activity in a small-world graph of transactions or trust relationships between businesses.

We analyze the small world properties of the large connected components of a Yelp reviewer graph and compute the local clustering coefficient and average path length for nodes contained in the graph. Yelp allows users to invite their friends to join Yelp or to make new friends on Yelp. friendship is a mutual relationship, which means that when a user adds another user as a friend, the first user will automatically be added as a friend of the second user.

We consider a user network graph, in which reviewer nodes are connected to each other by the friend relationship. We are given a user network $G = (U, E)$, in which reviewer nodes within the set $U = \{u_1, u_2, \ldots, u_n\}$ are connected with each other by the links $e = (u_i, u_j) \in E$. As previously indicated, we generate two types of graphs for two user groups: one for a reviewer group whose reviews have been filtered as reviews not thought to be legitimate reviews, which we call the fake review group; another for the group whose reviews we believe to have been used to legitimately rate the value of business and products, which we call the recommended review group. We calculate the local clustering coefficient and global characteristic path length for each graph. To gain additional insight into how the graph is structured, we look at the degree of each vertex, which is simply the number of edges that a particular vertex belongs to.

C. Detecting Fraudulent Reviewers and Fake Reviews

The online review dataset consists of a set of reviewers (users), a set of businesses, and reviews. Each review is written by a particular reviewer to rate a particular business on a scale from 1 to 5. This dataset can be represented by a reviewer-business bipartite network with review edges, in which reviewer nodes are connected to business nodes and the links represent the "reviewed" relationship associated with a rate which is ranged from1 to 5. We are given a review network $G = (V, E)$, in which a set of reviewer nodes $U = \{u_1, u_2, \dots, u_n\}$ and a set of business $P = \{p_1, p_2, \dots, p_n\}$ are connected with each other by the "review" links $e = (u_i, p_j, r) \in E,\ r \in \{1, 2, \dots, 5\}$ and $V = U \cap V$.

To spot fraudulent reviewers and fake reviews in online consumer reviews, we build our model on the framework proposed in [1], called FraudEagle, which exploited network effects among users and products. In this framework, the opinion fraud detection problem is formulated as a network classification task on signed networks that operate in a completely unsupervised fashion requiring no labeled data. The review network successfully captures the correlations of labels among users and products. The network edges are signed by ratings. An iterative, propagation-based algorithm exploits the network structure and the long-range correlations to infer the class labels of users, products, and reviews.

Besides the relational information between users and products, there exist a variety of metadata in review datasets. Those include the text content of reviews, timestamps, and star ratings. Earlier work [22] has used metadata to design features that are indicative of spam. We exploit findings from the analysis of user networks and add new features such as number of friends, number of reviews made and number of votes received into the framework. In particular, these metadata were used to estimate initial class probabilities for users, products, and reviews, which are incorporated as prior potentials of the nodes under a new MRF model.

IV. Evaluation

A. Data Description

In our experiments, we used a new dataset collected from Yelp.com using the crawler based on Scrapy open source library. We have extracted details for reviewers, businesses and reviews from particular websites. The next step is to patch them all together as a single system and run it; this crawler will simultaneously update the data base along with creating the positive and negative relationships between users and businesses; this crawler will also create a friend relationship node by node as it comes across the data. Table I shows the summary of data collected for the experiment.

TABLE I. Summary of Data

Reviews by the recommended review group	8,000
Reviews by the fake review group	3,000
Reviewers in the recommended review group	4,892
Reviewers in the fake review group	2,904
Businesses in recommended reviews	25
Businesses in fake reviews	384
Number of user-friends for recommended reviews	110,481
Number of user-friends for fake reviews	41,725

B. Data Analysis

Yelp developed a filtering algorithm that predicts whether a review is to be published or not. The algorithm is used to flag suspicious reviews and to filter those from the main Yelp page. Roughly 16% of restaurant reviews are filtered by Yelp. Recently, M. Luca and G. Zervas estimated the evidence of review fraud and the conditions under which it is most prevalent. They assembled two complementary datasets from Yelp and provided empirical support for using filtered reviews as a proxy for review fraud [13].

We have grouped collected review datasets into reviews by the recommended review group (non-spam) and reviews by the fake review group (filtered). In addition to grouping reviews, we also grouped reviewers in the Yelp dataset into the same two groups according to whether the reviews were considered to be recommended or fake. Statistical properties of these two reviewer groups have been calculated and compared.

1) Rate Distribution

Figure 1 shows the rating distribution in the two reviewer groups. For the recommended review group, the distribution has a characteristic "J" shape and shows that the reviews are skewed towards higher ratings. 70% of users in this group give reviews with ratings of 4 or 5, which are regarded as positive reviews.

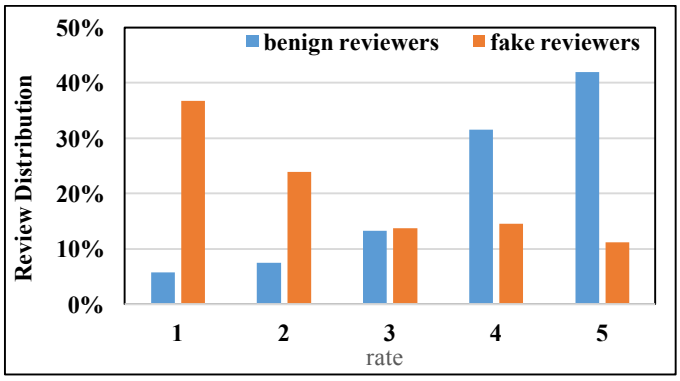


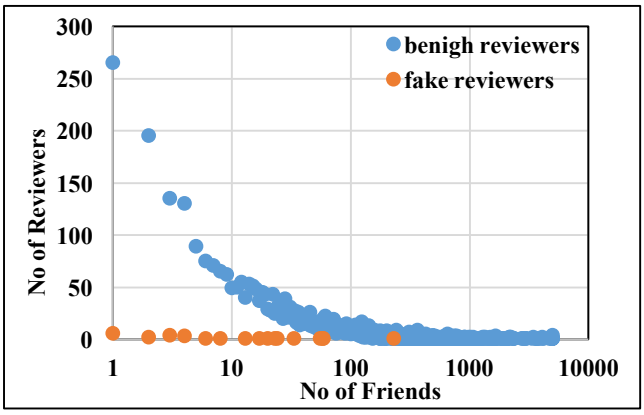
Fig. 1. Review Rate Distirbution

By contrast, the distribution of ratings for the fake review group shows the opposite behavior and data is skewed towards lower ratings. 60% of users in this fake review group give ratings of 1 or 2, which are regarded as negative reviews for a business. Users in the fake reviewer group give negative reviews, which is opposite to the behavior of regular users in online reviews, who instead give mostly positive reviews.

2) Number of Friends

We calculated and compared the number of friends of users in the recommended reviewer group and fake reviewer

group. Figure 2 shows the distribution of the number of friends in these two groups. Most users belonging to the fake reviewer group have few or no friends. However, users in the recommended review group have more friends on average. Users in the fake review group show the characteristics of one-time users [10] in a social network who have never made social connection with other users in the network. This feature is also manifested in the structural properties of the user graph



for the fake review group.

Fig. 2. Distribution of Number of Friends

B. Analysis of User Networks

We detail our study on reviewer networks and compute graphs of the properties of reviewers in the Yelp dataset. The graphs can be structured in many different ways. To gain additional insight into how the graph is structured, it is helpful to look at the degree of each vertex, which is simply the number of edges to which a particular vertex belongs. The local clustering coefficient and characteristic path length are computed to analyze the global and local structure of the graph. Results of these analyses are utilized as clues for metadata in the fraud detection framework.

1) Degree Distiution

We compute and show the degree distribution of the graph for the reviewer groups in Figure 3. The degree distributions of both reviewer groups appear as straight lines on a log-log graph, which is the typical power-law distribution observed in many real-world networks [4].

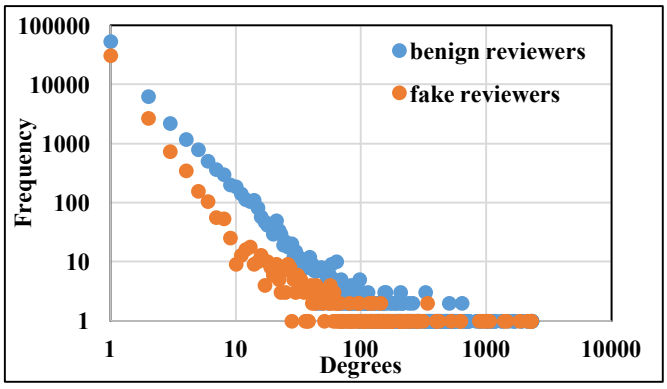


Fig. 3. Degree distribution of recommendedrecommended and fake reviewers

The mean degrees are calculated to be 3.07 and 2.26 for the recommended and fake review group, respectively. The graph of the recommended review group shows one degree more than that of the fake review group on average and indicates more connections among reviewers.

2) Local Clustering Coefficient

The local clustering coefficient is computed for the genuine reviewer group and the result is 0.038. This value is three orders of magnitude greater than that of the random network with the same vertex and edges. The large clustering coefficient indicates strong local clustering properties, which is a salient feature of the small network seen in many real-world networks. Table II shows the values of these structural parameters for other small-world networks as well as for random graphs that were generated on the same number of vertices and edges. On the other hand, the clustering coefficient of the fake review group is 0.005, which is relatively close to the value of a random network. By definition, the clustering coefficient is the ratio of action triangle count to possible triangle count. The calculated value of the clustering coefficient in the fake review group indicates that there are fewer actual triangles in the network and the graph is less connected than that of the recommended review group.

TABLE II. SMALE WORLD NETWORKS

Network	Number of vertices	Mean degree	Local Clustering Coefficient	
			Measured	Random graph
Internet (autonomous system)*	6,374	3.80	0.240	0.00060
World Wide Web (sites)*	153,127	35.20	0.110	0.00023
Power Grid*	4,941	2.70	0.080	0.00054
Biology Collaborations*	1,520,251	15.50	0.081	0.00001
Film Actor Collaborations*	449,913	113.40	0.200	0.00003
Yelp Users (recommended review group)	**37,022**	**3.07**	**0.038**	**0.00008**
Yelp Users (fake review group)	**22,403**	**2.26**	**0.005**	**0.00010**

(Note: The results with * come from [18]. The last two rows showing Yelp User data results from our work.)

3) Characteristic Path Length

The average path length of user networks for the recommended review group is calculated and the result is 4.138, which is in the range of the typical average path length in social networks [2, 25]. With the large clustering coefficient, our results indicate that the user graph of the recommended review group fits into the same range of average path length and large clustering coefficient values that we can see in other well-known small world networks. The average path length of user networks for the fake review group is even shorter and is 3.972. This result with a small

value of clustering coefficient indicates that the graph for the fake review group shows properties closer to those of random networks.

C. Fraudulent Reviewers and Fake Reviews

We exploit the clues from our structural analysis of the Yelp review dataset in the previous section and have added new features such as the number of friends, number of reviews made, and number of useful, funny, and cool votes received into the FraudEagle framework [1]. A low value for any of these features raises suspicion about the user. For computing beliefs about reviewer groups using belief propagation, we set the threshold value to 10^{-19}, having previously used 10^{-6}, with little difference in the results. The maximum number of iterations used to compute the beliefs was also increased, to 1000 iterations.

Results are summarized and compared to those calculated using FraudEagle (FE) [1] and SpEagle (SE) [22] framework in Table III and Table IV. To avoid a data imbalance issue between fake reviews and genuine reviews, we then randomly select 8,000 reviews from genuine reviews and 3,000 reviews from fake reviews. First, we include the number of friends as a feature and obtain results (friend counts), and then include the number of review counts of the reviewer (friend & review counts) and the number of review votes (friend, review & votes count) as a feature. We calculate average precision (AP) and area under the curve (AUC) as performance measures for both user and review rankings.

TABLE III. IMPROVEMENTS IN PERFORMANCE MATRIX FOR USER RANKINGS

	FE	SE	+Friend	+Friend +Review	+Friend +Review +Votes
Average Precision	0.4802	0.4370	0.5217	0.5654	0.6356
Area Under Curve	0.6226	0.5516	0.6634	0.7033	0.7659

TABLE IV. IMPROVEMENTS IN PERFORMANCE MATRIX FOR REVIEW RANKINGS

	FE	SE	+Friend	+Friend +Review	+Friend +Review +Votes
Average Precision	0.4798	0.4355	0.5201	0.5631	0.6316
Area Under Curve	0.6154	0.5474	0.6544	0.6925	0.7523

We can see the 20 ~ 30 % improvement in performance for average precision and area under the curve of user and review rankings. Our results show that the new features of number of friends, number of reviews made, and number of votes are important in detecting fraudulent reviewers and fake reviews.

V. CONCLUSIONS

Users in the fake review group, whose reviews were filtered as reviews not thought to be genuine, show different statistical characteristics from those of the recommended review group whose reviews were thought to be genuine. Users in the fake review group have fewer friends and give more negative ratings with rating 1 or 2. Graph theory based calculation shows that the user graph of the recommended review group has a small average path length and large clustering coefficient, which are characteristics of small-world networks. On the other hand, the average path length is shorter and local clustering coefficient is close to that of random networks in the fake reviewer graph. In future work, we will investigate whether this structural difference between the recommended and the fake review group gives us any clues to detecting fraud in online consumer reviews.

Our model for detecting fraudsters and fake reviews implements findings from the analysis of user networks and adds new features such as the number of friends, number of reviews made, and number of votes received. In particular, these metadata are used to estimate initial class probabilities for users, products, and reviews, which are incorporated as prior potentials of the nodes. The results show that number of friends, number of reviews made, and number of votes received are important features in detecting fraudulent reviewers and fake reviews.

REFERENCES

[1] L. Akoglu, R. Chandy, C. Faloutis, Opinion fraud detection in online reviews by network effects. In ICWSM, 2013

[2] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, "Four Degrees of Separation." arxiv.org/abs/1111.4570.

[3] J. Chevalier and D. Mayzlin, The Effect of Word of Mouth on Sales: Online Book Reviews, Journal of Marketing Research, 2006

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In SIGCOMM, pages 251–262, 1999.

[5] S. Feng, R. Banerjee, and Y. Choi. Syntactic stylometry for deception detection. In ACL, 2012.

[6] J. He and W.W. Chu, A Social Network-Based Recommender System, Data Mining for Social Network Data , 2010

[7] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang. Catchsync: catching synchronized behavior in large directed graphs. In KDD, pages 941–950, 2014.

[8] N. Jindal and B. Liu, Opinion spam and analysis. In WSDM, pages 219-230, 2008

[9] N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. In CIKM, 2010.

[10] R. Kumar, J. Novak, A. Tomkins, Structure and Evolution of Online Social Networks, KDD'06, August 20–23, 2006

[11] F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. In IJCAI, 2011.

[12] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective PU learning. In ICDM, 2014.

[13] M. Luca and G. Zervas, Fake it till you make it: Reputation, Competition, and yelp review fraud, 2015.

[14] M. Luca, Reviews, Reputations, and Revenue: The Case of Yelp.com, 2011.

[15] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In KDD, 2013.

[16] A. Mukherjee, B. Liu, and N. S. Glance. Spotting fake reviewer groups in consumer reviews. In WWW, 2012.

[17] A. Mukherjee, V. Venkataraman, B. Liu, and N. S. Glance. What Yelp fake review filter might be doing? In ICWSM, 2013.

[18] M. E. Newman, Networks: An introduction. New York, NY: Oxford University Press, 2010.

[19] M. E. Newman, Random graphs as models of networks, SFI working paper 2002.

[20] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock. Finding deceptive opinion spam by any stretch of the imagination. In ACL, pages 309–319, 2011

[21] J. Petrella, Use of graph theory to evaluate brain networks: A clinical tool for a small world, pages 317-320, Radiology 2011 259(2)

[22] S. Rayana and L. Akoglu, Collective Opinion Spam Detection: Bringing Review Networks and Metadata, KDD'15, 2015

[23] D. Streitfeld, Best Book Reviews Money Can Buy, 2012. http://nyti.ms/1cvg5b1.

[24] K. Supekar, Menon V , Rubin D , Musen M , Greicius MD . Network analysis of intrinsic functional brain connectivity in Alzheimer's disease . PLOS Comput Biol 2008, 4 (6)

[25] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The Anatomy of the Facebook Social Graph." arxiv.org/abs/1111.4503.

[26] G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. In ICDM, 2011.

[27] J. Wang, Zuo X , He Y . Graph-based network analysis of resting-state functional MRI . Front Syst Neurosci 2010 ; 4 : 16.

[28] D. Watts and S. Strogatz, Collective dynamics of 'small-world' networks. pages 440-442, Nature 1998 393(6684).

[29] S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In KDD, 2012.

[30] J. Ye and L. Akoglu. Discovering opinion spammer groups by network footprints. In ECML/PKDD, 2015.