Identifying Singleton Spammers via Spammer Group Detection

Dheeraj Kumar[1], Yassien Shaalan[2], Xiuzhen Zhang[2(✉)], and Jeffrey Chan[2]

[1] Lyles School of Civil Engineering, Purdue University, West Lafayette, USA
kumar299@purdue.edu
[2] School of Science (Computer Science), RMIT University, Melbourne, Australia
{yassien.shaalan,xiuzhen.zhang,jeffrey.chan}@rmit.edu.au

Abstract. Opinion spam is a well-recognized threat to the credibility of online reviews. Existing approaches to detecting spam reviews or spammers examine review content, reviewer behavior and reviewer-product network, and often operate on the assumption that spammers write at least several if not many fake reviews. On the other hand, spammers setup multiple sockpuppet IDs and write one-time, singleton spam reviews to avoid detection. It is reported that for most review sites, a large portion, sometimes over 90%, of reviewers are singletons (identified by the reviewer ID). Singleton spammers are difficult to catch due to the scarcity of behavioral clues. In this paper, we argue that the key to detect singleton spammers (and their fake reviews) is to detect group spam attacks by inferring the hidden collusiveness among them. To address the challenge of lack of explicit behavioral signals for singleton reviewers, we propose to infer the hidden reviewer-product associations by completing the review-product matrix by leveraging the product and review metadata and text. Experiments on three real-life Yelp datasets established that our approach can effectively detect singleton spammers via group detection, which are often missed by existing approaches.

Keywords: Opinion spam · Singleton spammers · Sockpuppet IDs
Inductive matrix completion

1 Introduction

Online shoppers are ever increasing and the product reviews influence their buying decisions to a great extent. According to the *Local Consumer Review Survey 2016* by BrightLocal[1], 84% of the online shoppers trust product reviews as much as personal recommendations. Positive reviews and higher star ratings results in substantial financial gains for the businesses, while negative reviews can cause reputation damage and financial losses. As a result of such financial incentives,

D. Kumar—performed this research while working as a Research Officer at RMIT University, Melbourne, Australia under the supervision of Prof. Xiuzhen Zhang.

[1] https://www.brightlocal.com/learn/local-consumer-review-survey/.

© Springer International Publishing AG, part of Springer Nature 2018
D. Phung et al. (Eds.): PAKDD 2018, LNAI 10937, pp. 656–667, 2018.
https://doi.org/10.1007/978-3-319-93034-3_52

opinion spam (fake reviews which deliberately mislead readers) is prevalent. There is an estimate from 2012 that one-third of consumer reviews on the Internet were fake[2]. Similarly, the number of fake reviews on yelp.com rose from 5% in 2006 to 20% in 2013 [8].

Opinion spam detection has attracted significant research [1,5–7,11,13–15,17]. Existing approaches to detecting spam reviews and spammers focus on extracting spam signals from review texts [5,11,14], reviewer behaviour [7,15], or the reviewer-product networks [1,13]. Many approaches assume spammers write at least several if not many fake reviews for multiple products. It is reported, however, that the majority of reviewers in most of the opinion websites are singletons, i.e. writes only one review. According to [17], over 90% of the reviewers of resellerratings.com write only one review. Indeed, our analysis showed that a majority, in the range of 65% to 70%, of the reviewers of Yelp datasets [13] are singletons (See Table 1 for details). Detecting singleton spam reviews and singleton spammers is challenging due to the lack of obvious spamming signals. Existing approaches based on reviewer behavior and reviewer-product networks are not effective for detecting singleton spammer reviewers [14,17]. Sandulescu and Ester [14] argued that the key to catch singleton spammers can only be found in the review texts. A drawback of their text-based approach is that the collusiveness among reviewers is overlooked and therefore may not be effective.

To address the challenge of scarcity of spam signals for singleton reviewers, we argue that the key to effective detection of singleton spammers is via identifying spammer groups. A spammer group is *"a group of reviewers writing fake reviews together to promote or to demote some target products"* [9]. Here reviewers are defined by the reviewer ID. The actual person behind different IDs could be a single person, multiple persons, or a combination of both. There is a great incentive for opinion spammers to create multiple sockpuppet IDs to write singleton reviews since it helps them avoid detection and provide higher revenues by writing many fake reviews. However, existing approaches to detecting spammer groups [9,18–20] are not directly applicable for singleton spammers since they assume that participants of spammer groups frequently write reviews for multiple products together (hence, non-singleton).

In this paper, we propose SSGD (*singleton spammer group detection*), a novel approach to detecting singleton spammers via spammer group detection. The intuition behind SSGD is that given the purpose of group spamming is to promote or demote the reputation of products within a short time window, a burst of changes in signals such as rating and number of reviews can indicate the occurrence of spam attacks. From the likely spam attacks, the candidate singleton spammers and their targeted products are identified. We further examine the review textual content, rating and time to infer latent reviewer-product associations and uncover the collusiveness among singleton reviewers. We formulate the problem of inferring (hidden) reviewer-product association as a review-product matrix completion problem. The sparse review-product associa-

[2] http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html.

tion matrix is *completed* using the additional information such as review text and metadata (star rating and date), the product description text and the product "also bought" and "also viewed" network. Lastly, the inferred reviewer-product associations are clustered to detect spammer groups consisting of (mostly) singleton spammers.

We conducted experiments on three real-world opinion spam datasets on yelp.com (YelpChi, YelpNYC, and YelpZIP) [13] to evaluate the effectiveness of SSGD for detecting singleton spammers by detecting them in groups. We benchmarked SSGD against five approaches in the literature for detecting spammer groups and individual spammers based on reviewer behavior, reviewer-product network and review text. SSGD outperformed all these approaches in terms of both recall and precision for singleton spammer detection. To the best of our knowledge, this paper is the first attempt at identifying singleton spammers via detecting hidden collusiveness among them for group spam attacks.

2 Related Work

Existing studies in the literature on detecting spam reviews and spam reviewers can be broadly classified as reviewer behaviour based [7,15], reviewer-product network based [1,13], and review text based [5,11,14]. A detailed survey of these techniques can be found in [4]. Reviewer behavior-based and reviewer-product network-based approaches are not effective for detecting singleton spammers, as they focus on reviewers with multiple reviews. Text-based approaches [11, 14] can address the challenge of lacking behavioral clues for one-time singleton reviewers by examining the psycholinguistic features in review contents [11] or by examining the pairwise content similarity between reviews [14]. However, the text-based approaches totally ignore the collusiveness among reviewer IDs and other review metadata, and our experiments show that they are not very effective in detecting singleton reviewers.

There are some recent studies on detecting singleton spam reviews [14,17]. Xie et al. [17] constructed a multidimensional time series consisting of the average rating, number of all and singleton reviews for time windows of fixed duration. It then detects spam attacks by finding abnormal sections in each time series. Though not specifically for singleton reviews, [21] monitors a list of carefully selected indicative signals of opinion spam over time and design efficient techniques to both detect and characterize abnormal events in real-time. The indicative features used for temporal spam detection in [21] are a superset of the time series used in [17]. One thing to note is that the approaches presented in [17,21] does not label individual singleton reviews as spam or genuine, but predict the time when a product is most likely to be a victim of a spam attack.

Our research is also related to studies concerning spammer group detection [9, 18–20]. Mukherjee et al. [9] first proposed an approach to detect spammer groups using *frequent itemset mining* (FIM) to find a set of candidate groups. It then uses several behavioral models derived from the collusion phenomenon among suspected fake reviewers to detect fake reviewer groups. The approach proposed

in [20] uses only the network footprint information (user-product graph) to detect spammer groups. The approaches proposed in [18,19] uses pairwise features of reviewers to detect spammer groups which are defined for only those reviewer pairs who have adequate reviewing histories. Recently, Li et al. [6] proposed an HMM-based approach to detect spammers' co-bursting behavior to detect spammer groups. For all these approaches, the assumption about the spammer groups is that reviewers in a group write fake reviews for multiple products together and detection is based on the explicit reviewer-product associations, hence all of these approaches would miss the singleton spammers.

More generally, our research is related to spam detection on social media [16], where, most studies focus on finding clusters of linked nodes and the singleton spammers are likely ignored. In addition, our approach of inferring hidden associations is related to inferring network structures from data [2], which is an important data mining task in many domains.

3 The Proposed Approach: SSGD

We propose a novel approach to catch singleton spammers by discovering hidden collusiveness among them and detecting spammer groups. Detecting singleton spammer group is challenging as singleton reviewers write only one review and there is little information available about each individual singleton spammer. However, there are still (hidden) signals available for detection of singleton spammer groups. Such spammer groups aim to influence average ratings and impressions of target products by using following two tactics [14,15]:

- Inject enough fake reviews to affect the average ratings;
- Flood the most recent review pages with fake reviews as most buyers read only top several reviews before forming an opinion about a business.

Both these approaches require a (singleton) spammer group to generate a relatively abnormal number of positive or negative reviews over a short period of time, depending on what influences they wish to exert. Figure 1 shows the framework of our proposed approach SSGD. We first apply the spam attack detection approach [21] to identify the target products and the attack time in the multivariate time series for a set of indicative signals: average rating, number of positive/negative reviews, rating entropy, the ratio of singletons and first-timers, youth score, and temporal gap entropy to detect abnormal changes/bursts as potential spam attacks. The magnitude of abrupt change is used to assign an anomaly score for each detected attack. Further details about this step are well documented in [21] and are not reproduced here for brevity.

The spam attacks produce a review-product subgraph of target products and (potentially) spam reviews. The matrix representing the review-product network is very sparse without meaningful review-product associations. This data sparsity problem is similar to the one encountered in the recommendation systems, where matrix completion has shown great success in dealing with such sparsity. A detailed survey of the novel techniques to infer hidden network structure from

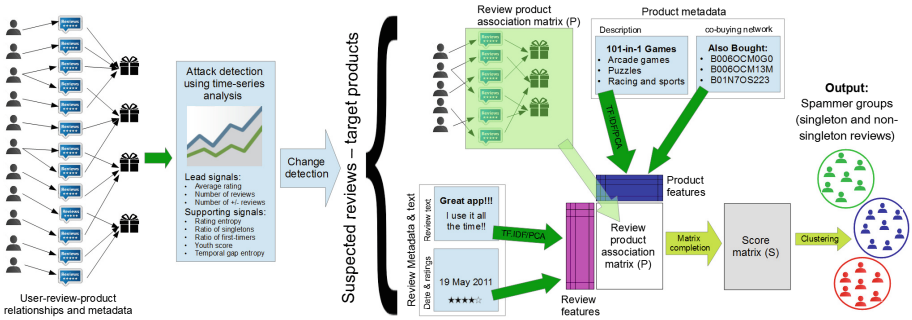

Fig. 1. Schematics of the proposed approach (SSGD)

the data for matrix completion is given in [2]. For the problem of inferring hidden collusiveness among reviewers, we use the *inductive matrix completion* (IMC) algorithm [10], which uses additional information such as review text and metadata (star rating and date) and product metadata such as its description text. Once the "completed", enriched review-product matrix is obtained from IMC, we cluster the reviews (using the inferred associations among reviews as the feature vector) to discover groups of similar reviews (reviewers) targeting the attacked products. We next describe these steps in details.

3.1 Inferring Hidden Reviewer-Product Associations

The spam attacks define a review-product bipartite graph of N_r suspicious reviews and N_p target products as nodes. An edge is present between review $i, 1 \leq i \leq N_r$ and product $j, 1 \leq j \leq N_p$ if review i belongs to the product j. This bipartite review-product graph is represented as a review-product associations matrix $P \in \mathbb{R}^{N_r \times N_p}$, where $P_{ij} = 1$ if review i belongs to product j, otherwise, $P_{ij} = 0$. The sparse review-product association matrix P is then "completed" using the IMC algorithm (described next) to learn hidden reviewer-product associations.

The IMC algorithm was used in [10] to predict *gene-disease associations by combining multiple types of evidence (features) for diseases and genes to learn latent factors that explain the observed gene-disease associations*. The spammer group detection is similar to this problem since we have a (sparse) review-product association matrix and based on the features of reviews and products, the aim is to discover groups of users who are most likely to write fake reviews for a group of products. IMC can be interpreted as a generalization of the transductive multi-label learning formulation: *low rank empirical risk minimization for multi-label learning* (LMEL) [22] and assumes that the associations matrix is generated by applying feature vectors associated with its rows as well as columns to a low-rank matrix Z (representing actual person behind various sockpuppet IDs). The goal is to recover Z using observations from P. Let $x_i \in \mathbb{R}^{f_r}$, and $y_j \in \mathbb{R}^{f_p}$ denotes the feature vector for review i, and product j respectively. Let $X \in \mathbb{R}^{N_r \times f_r}$

denote the training feature matrix of N_r reviews, where the i^{th} row is the review feature vector x_i, and let $Y \in \mathbb{R}^{N_p \times f_p}$ denote the training feature matrix of N_p products, where the j^{th} row is the product feature vector y_j. The inductive matrix completion problem is to recover a low-rank matrix $Z \in \mathbb{R}^{f_r \times f_p}$ using the observed entries from P. Denote the set of observed entries (i.e., training review-product associations) by Ω. The entry P_{ij} of the matrix is modeled as $P_{ij} = x_i^T Z y_j$ and the goal is to learn Z using the observed entries Ω. Z is of the form $Z = WH^T$, where $W \in \mathbb{R}^{f_r \times k}$ and $H \in \mathbb{R}^{f_p \times k}$, and k is small. The low-rank constraint on Z is NP-hard to solve. The standard relaxation of the rank constraint is the trace norm, i.e., sum of singular values. Minimizing the trace-norm of $Z = WH^T$ is equivalent to minimizing $\frac{1}{2}(||W||_F^2 + ||H||_F^2)$. The factors W and H are obtained as solutions to the following optimization problem:

$$\min_{\substack{W \in \mathbb{R}^{f_r \times k}; \\ H \in \mathbb{R}^{f_p \times k}}} \sum_{(i,j) \in \Omega} \ell(P_{ij}, x_i^T W H^T y_j) + \frac{1}{2}\lambda(||W||_F^2 + ||H||_F^2), \tag{1}$$

The loss function ℓ penalizes the deviation of estimated entries from the observations. A common choice for loss function is the squared loss function ($\ell_{sq}(a, b) = (a - b)^2$). The regularization parameter λ trades off accrued losses on observed entries and the trace-norm constraint. IMC adapt the LEML solver [22] for solving (1). The solver uses alternating minimization (fix W and solve for H and vice versa) to optimize (1). The resulting optimization problem in one variable (W or H) is solved using the conjugate gradient iterative procedure. The features used to learn hidden reviewer-product association using IMC are described next.

We use the *term frequency - inverse document frequency* (TF-IDF) as the text feature for review/product description. TF-IDF formally measures how concentrated the occurrences of a given word is into relatively few documents [12]. The terms with the highest TF-IDF scores are often the terms that best characterize the topic of the document. Before extracting the TF-IDF feature for the review texts and product descriptions, we filter out the stop words as they are extremely common words which would appear to be of little value in deciding review text similarity. We project the review and product TF-IDF features to a lower dimensional space. In particular, we use *principal component analysis* (PCA) that performs a linear mapping of the data onto the lower dimensional space by maximizing the variance of the data in the new representation. We choose the leading 200 eigenvectors of the covariance matrix as the text features for reviews and product description. Another set of features for reviews consists of the date and the star rating associated with each review. For product feature, apart from the TF-IDF features obtained from the review content and product description, the product "also bought" and "also viewed" network is useful in identifying similar products which could be the common target of a group of spammers. In case this information is not available, identity matrix (I) is used as the product features representing each product as independent of others.

3.2 Finding and Ranking Spammer Groups

The output of the IMC algorithm is a completed review-product association matrix called as *score matrix* ($S \in \mathbb{R}^{N_r \times N_p}$). Higher values of S_{ij} indicate a greater likelihood of review i being written for product j. The learned product association for each review is used as the review feature vector to cluster them to detect spammer groups. We consider the N_r rows of the score matrix S as the N_p- dimensional feature vector for each review. To detect spammer groups, we need to find the set of reviews which are most likely being written by a group of a few spammers (sockpuppet IDs). These reviews would form a dense cluster in the N_p- dimensional feature representation given by score matrix S. The genuine reviews are expected to be at a large distance from other genuine reviews and spammer groups.

We choose the popular density-based clustering algorithm DBSCAN [3] to cluster spammer groups since it does not require the number of clusters to seek as an input, which is not known for our problem. Also, the reachability distance (ϵ) parameter of DBSCAN provides an easily tunable parameter for detecting a dense cluster of spammers, leaving out noise points (genuine reviews which may not belong to any spammer group). Most of the reviews in the detected groups are singleton reviews, hence review group is similar to reviewer group, however, in some cases, multiple reviews belonging to the same user can be a classified in different spammer groups (some spammers write a few genuine looking reviews to camouflage their campaign or may be part of multiple spam campaigns). The candidate spammer groups are then ranked based on the average intra-cluster distance between the reviews belonging to a group in the feature vector space representation. The groups whose members are close to each other are given high ranking as compared with the groups whose members are sparse.

4 Experiment Setup

We performed experiments on three publicly available online review datasets that are widely used in the opinion spam literature: YelpChi, YelpNYC, and YelpZIP [13]. These datasets contain reviews for restaurants in Chicago and NYC and in areas defined by a zip code in the NY state. All datasets contain review metadata such as star rating and date as well as review text. Their basic statistics is given in Table 1. It can be seen that all datasets contain a high percentage of singleton reviews, ranging from 65.35% to 70.55%.

The Yelp datasets have "near" ground truth labels for spam reviews based on the fake/suspicious filtering algorithm used at Yelp.com. The author of a spam review is labeled a spammer. Although the Yelp anti-fraud filter is not perfect, it was found to produce accurate results, and the spam reviews and spammers thus labeled were used as ground truth for evaluating opinion spam detection algorithms [13]. Table 2 also shows the portion of singleton spammers in the ground truth spammers in the three Yelp datasets. It can be seen that a significant portion of spammers on the Yelp datasets are singleton spammers.

Table 1. Basic statistics of the three datasets (In the parlance of customer reviews, restaurants are the products.)

	# Reviewers	# Products	#Reviews per reviewer	#Reviews per product	#Singleton reviewers (%)
YelpChi	38,063	201	1.77	335.30	70.55%
YelpNYC	160,225	923	2.25	389.00	66.15%
YelpZIP	260,277	5,044	2.34	120.66	65.35%

Table 2. Singleton spammers in the Yelp datasets

	# Spammers	Singleton spammers (% of # Spammers)
YelpChi	8,919	76.03%
YelpNYC	36,885	63.91%
YelpZiP	80,466	62.97%

We compare SSGD against five state-of-the-art baseline approaches for detecting spammer groups and individual spammers – two are spammer group detection approaches, one is a spam review detection algorithm utilizing network as well as metadata and two are text-based approaches, one being specialized for detecting singleton spammers, as described below:

1. Spammer group detection baselines:
   (a) FIM (Frequent Itemset Mining) [9]: This approach assumes that spammer groups are groups where reviewers (identified by reviewer IDs) frequently write reviews together. The candidate groups are then ranked based on the group spam features described in [9], which are found effective in distinguishing spammer and non-spammer groups.
   (b) NFS-GroupStrainer [20]: This approach detects spammer groups based on the footprint of reviewers on the reviewer-product network. It first finds targeted products using a graph-based measure *Network Footprint Score* (NFS) which quantifies the statistical distortion caused by spamming activities in the reviewer-product graph. A hierarchical clustering algorithm called GroupStrainer is then applied on the two-hop subnetwork of the targeted products to find spammer groups.
2. Spam review detection utilizing network as well as metadata baseline:
   (a) SpEagle: The SpEagle algorithm [13] utilizes clues from all review metadata (text, timestamp, rating) as well as the reviewer-review-product network to find suspicious users and reviews and the targeted products.
3. Review text-based baselines: The text-based approaches use only the textual review contents to detect spam reviews and accordingly singleton reviewers.
   (a) Ott: Ott et al. [11] built a supervised classification model based on a comprehensive set of psycholinguistic features extracted from review text.

(b) DSR [14]: This approach detects singleton spam reviews by computing the semantic similarity among pairs of reviews.

5 Results and Discussion

We next report the results of SSGD for detecting spammers and spammer groups on the three Yelp datasets.

5.1 Recall and Precision for Singleton Spammer Detection

We evaluated the average precision and recall of the singleton spammers (a majority among all the spammers in all of the three datasets) detected by each approach as a measure of its effectiveness. The results are given in Table 3 where the maximum value of achieved average precision/recall is shown in bold.

Table 3. Average precision and recall (%) of singleton spammers detected by each approach

		SSGD	FIM [9]	NFS - GroupStrainer [20]	SpEagle [13]	Ott [11]	DSR [14]
YelpChi	Precision	**23.50**	7.23	18.24	20.14	16.20	1.03
	Recall	**87.57**	22.21	63.60	50.00	49.80	10.20
YelpNYC	Precision	**21.15**	5.21	16.87	18.28	16.24	5.12
	Recall	**74.25**	21.17	60.42	55.89	72.81	28.15
YelpZip	Precision	**24.39**	6.25	18.69	17.57	22.63	5.08
	Recall	**88.79**	25.78	66.15	59.37	71.61	13.24

SSGD achieves the highest average precision and recall compared to other approaches across all datasets. The results in Table 3 show that the groups detected by SSGD contain more ground-truth singleton spammers. Spammer group detection approaches FIM and NFS-GroupStrainer obtain poor precision and recall as they cannot capture singleton spammers effectively. SpEagle, like SSGD, which is also based on the review-product network and review and product metadata, is not that effective in detecting singleton spammers as indicated by its low average precision/recall scores. A possible reason may be that SpEagle infers the spam probability for reviewers based on the possible targeted restaurants, rather than on the spam attacks as in SSGD, due to which many genuine reviews are also labeled as spammers resulting in higher false negatives and hence lower value of recall. The text-based approaches: Ott [11] and DSR [14] show very different performance for detecting spammers. The Ott [11] approach shows much better performance than DSR, which confirms that supervised learning based on psycholinguistic features from review contents is effective for detecting spam reviews. Still, the Ott approach does not perform as well as SSGD

in terms of recall or precision for detecting singleton spammers. This indicates that the collusiveness among singleton reviewers during spam attacks embed strong signals complementary to the linguistic features for detecting singleton spammers.

Table 4. Singleton reviewers in spam attacks

	#products	#reviewers	#Singleton reviewers (%)
YelpChi	48	5,026	93.12%
YelpNYC	45	18,243	91.84%
YelpZIP	48	26,926	92.57%

We next investigate the fraction of singleton spammers in the spammer groups detected by SSGD. Table 4 lists number of restaurants (products), number of unique reviewers and percentage of singleton reviewers in the top 50 spam attacks detected by SSGD. Comparing the percentage of singleton reviewers in Tables 1 and 4, it clearly shows the sharp increase in the number of singleton reviewers during spam attacks. This result reaffirms that spammers tend to write singleton reviews from multiple IDs (sockpuppets) to avoid being caught. This also shows the effectiveness of the first step of SSGD for identifying spam attacks. Our approach to identifying candidate spam activities via examining the temporal dynamics of multiple signals at the review level as well as the review meta-data is effective for detecting spam activities from singleton reviewers.

5.2 Qualitative Analysis of Detected Spammer Groups

The spammer groups detected by SSGD (under default settings of DBSCAN of reachability distance parameter $\epsilon = 0.01$ and a minimum of 3 reviewers in a cluster) mostly consist of singleton reviewers who either gave all high (4–5) or all low (1–2) star ratings and wrote nearly identical reviews for a set of restaurants within a short time duration. Table 5 lists statistics for the top 5 groups detected by SSGD on the Yelp datasets. Among the top 5 spammer groups, the group consists of 20–90 reviewers (many are singletons) targeting 4–9 products. The timestamps and rating distribution of most groups are concentrated for maximum impact. The group spam targeted restaurants exhibit some common characteristics such as of the same cuisine, located in the same locality, etc.

6 Conclusions

Opinion spam is a prevalent problem hampering the credibility of online reviews. Existing methods often focus on reviewers who have written multiple reviews and spammer are detected by their abnormal behaviors. However, majority of

Table 5. Summary of the statistics of the top 5 SSGD detected groups in the Yelp datasets (#P: number of products, #U: number of users, Time & Rating distribution (s: scattered, c: concentrated))

	ID	#P	#U	Time	Rating	(near) Duplicate	Restaurant Description
YelpChi	1	4	24	s	c	10/30	#2 (same cuisine)
	2	5	25	c	c	12/29	In same area
	3	4	21	c	s	7/25	Hot dog
	4	6	55	s	c	32/64	Two attacks 2 month apart
	5	5	40	c	c	21/48	#3 (same cuisine)
YelpNYC	1	6	48	c	c	28/53	5 in same area
	2	5	39	s	c	22/43	#3 (same cuisine)
	3	6	37	s	c	28/47	breakfast
	4	7	52	c	c	45/59	-
	5	8	60	s	s	38/71	3 in same area
YelpZIP	1	8	63	c	s	42/74	4 in same area
	2	7	54	c	c	36/67	Fine dining
	3	9	87	s	s	47/109	Pizza
	4	7	41	c	c	13/52	#4 (same cuisine)
	5	6	42	c	c	28/49	-

reviewers are singleton reviewers, and are often overlooked by existing opinion spam detection approaches. In this paper we proposed a novel approach to detecting spammer groups consisting of singleton reviewers. Our approach comprises several strategies to address the challenge of scarcity of explicit signals for singleton reviewers. Especially we focus on identifying the collusiveness among singleton reviewers via detecting coordinated spam attacks. We experimented on three real-life Yelp datasets to evaluate our approach. Our results showed that the problem of singleton spam is widespread – many online review sites have mostly singleton spammers, and many group spam attacks involve singleton reviewers. Experiments show that our approach can more accurately capture singleton spammers than existing approaches and can detect spammer groups of singleton spammers overlooked by existing approaches. For future work, we will investigate approaches that make use of more hidden signals for accurate singleton spam detection.

Acknowledgments. This work was supported by the Australian Research Council (ARC) linkage project grant LP120200128.

References

1. Akoglu, L., Chandy, R., Faloutsos, C.: Opinion fraud detection in online reviews by network effects. In: ICWSM 2013 (2013)
2. Brugere, I., Gallagher, B., Berger-Wolf, T.Y.: Network structure inference, a survey: motivations, methods, and applications. arXiv preprint arXiv:1610.00782 (2016)

3. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD 1996 (1996)
4. Heydari, A., Tavakoli, M., Salim, N., Heydari, Z.: Detection of review spam: a survey. Expert Syst. Appl. **42**(7), 3634–3642 (2015)
5. Jindal, N., Liu, B.: Opinion spam and analysis. In: WSDM 2008 (2008)
6. Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., Shao, J.: Bimodal distribution and co-bursting in review spam detection. In: WWW 2017 (2017)
7. Lim, E.P., Nguyen, V.A., Jindal, N., Liu, B., Lauw, H.W.: Detecting product review spammers using rating behaviors. In: CIKM 2010 (2010)
8. Luca, M., Zervas, G.: Fake it till you make it: reputation, competition, and yelp review fraud. Manag. Sci. **62**(12), 3412–3427 (2016)
9. Mukherjee, A., Liu, B., Glance, N.: Spotting fake reviewer groups in consumer reviews. In: WWW 2012 (2012)
10. Natarajan, N., Dhillon, I.S.: Inductive matrix completion for predicting gene-disease associations. Bioinformatics **30**(12), i60–i68 (2014)
11. Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding deceptive opinion spam by any stretch of the imagination. In: ACL 2011 (2011)
12. Rajaraman, A., Ullman, J.D.: Mining of Massive Datasets. Cambridge University Press, Cambridge (2011)
13. Rayana, S., Akoglu, L.: Collective opinion spam detection: Bridging review networks and metadata. In: KDD 2015 (2015)
14. Sandulescu, V., Ester, M.: Detecting singleton review spammers using semantic similarity. In: WWW 2015 Companion (2015)
15. Savage, D., Zhang, X., Yu, X., Chou, P., Wang, Q.: Detection of opinion spam based on anomalous rating deviation. Expert Syst. Appl. **42**(22), 8650–8657 (2015)
16. Wu, L., Hu, X., Morstatter, F., Liu, H.: Adaptive spammer detection with sparse group modeling. In: ICWSM, pp. 319–326 (2017)
17. Xie, S., Wang, G., Lin, S., Yu, P.S.: Review spam detection via time series pattern discovery. In: WWW 2012 Companion (2012)
18. Xu, C., Zhang, J.: Combating product review spam campaigns via multiple heterogeneous pairwise features. In: SIAM International Conference on Data Mining, pp. 172–180 (2015)
19. Xu, C., Zhang, J., Chang, K., Long, C.: Uncovering collusive spammers in Chinese review websites. In: CIKM 2013 (2013)
20. Ye, J., Akoglu, L.: Discovering opinion spammer groups by network footprints. In: ECML PKDD 2015 (2015)
21. Ye, J., Kumar, S., Akoglu, L.: Temporal opinion spam detection by multivariate indicative signals. In: WSDM 2016 (2016)
22. Yu, H.F., Jain, P., Kar, P., Dhillon, I.S.: Large-scale multi-label learning with missing labels. In: ICML 2014 (2014)