Identify Online Store Review Spammers via Social Review Graph

GUAN WANG, SIHONG XIE, BING LIU, and PHILIP S. YU, University of Illinois at Chicago

Online shopping reviews provide valuable information for customers to compare the quality of products, store services, and many other aspects of future purchases. However, spammers are joining this community trying to mislead consumers by writing fake or unfair reviews to confuse the consumers. Previous attempts have used reviewers' behaviors such as text similarity and rating patterns, to detect spammers. These studies are able to identify certain types of spammers, for instance, those who post many similar reviews about one target. However, in reality, there are other kinds of spammers who can manipulate their behaviors to act just like normal reviewers, and thus cannot be detected by the available techniques.

In this article, we propose a novel concept of review graph to capture the relationships among all reviewers, reviews and stores that the reviewers have reviewed as a heterogeneous graph. We explore how interactions between nodes in this graph could reveal the cause of spam and propose an iterative computation model to identify suspicious reviewers. In the review graph, we have three kinds of nodes, namely, reviewer, review, and store. We capture their relationships by introducing three fundamental concepts, the trustiness of reviewers, the honesty of reviews, and the reliability of stores, and identifying their interrelationships: a reviewer is more trustworthy if the person has written more honesty reviews; a store is more reliable if it has more positive reviews from trustworthy reviewers; and a review is more honest if many other honest reviews support it. This is the first time such intricate relationships have been identified for spam detection and captured in a graph model. We further develop an effective computation method based on the proposed graph model. Different from any existing approaches, we do not use an review text information. Our model is thus complementary to existing approaches and able to find more difficult and subtle spamming activities, which are agreed upon by human judges after they evaluate our results.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications

General Terms: Algorithms

Additional Key Words and Phrases: Spammer detection, review graph

ACM Reference Format:
Wang, G., Xie, S., Liu, B., and Yu, P. S. 2012. Identify online store review spammers via social review graph. ACM Trans. Intell. Syst. Technol. 3, 4, Article 61 (September 2012), 21 pages.
DOI = 10.1145/2337542.2337546 http://doi.acm.org/10.1145/2337542.2337546

1. INTRODUCTION

Online social intelligence is becoming ubiquitous recently. Users of similar interests form communities with user-generated contents to help each other in searching, advertising, decision making, etc. Online shopping review communities are such examples, where users share their experiences about products and stores by posting reviews. Such reviews are an important resource to help people make wise choices

This work is supported by the National Science Foundation, under grant IIS-0905215, OISE-0968341, DBI-0960443, and IIS-1111092.
Authors' address: G. Wang, S. Xie, B. Liu, and P. S. Yu, Computer Science Department, University of Illinois at Chicago; email: gwang26@uic.edu.

DOI 10.1145/2337542.2337546 http://doi.acm.org/10.1145/2337542.2337546

for their purchases. Due to this reason, the review system has become a target of spammers who are usually hired or enticed by companies [Consumerist] to write fake reviews to promote their own products or services, and/or to distract customers from their competitors. Driven by profits, there are more and more spam reviews in major review Web sites, such as PriceGrabber.com, Shopzilla.com, or Resellerratings.com [Consumerist]. Spammers are starting to corrupt the online review system and confuse the consumers.

1.1. Challenges

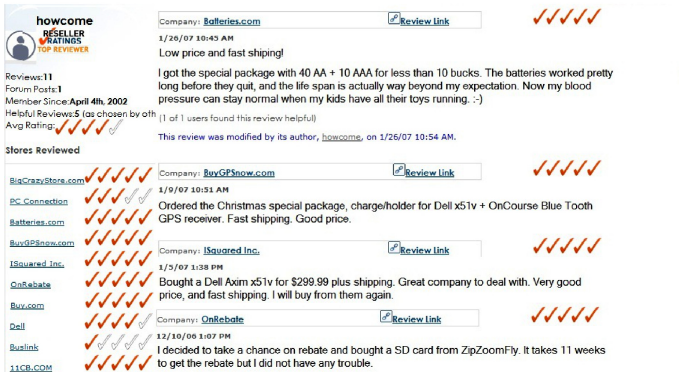
Automatically identifying spammers for e-commerce is an urgent yet under exploration task. Unlike other kinds of better studied spam detection, for instance, search engine spam [Gyngyi and Garcia-Molina 2005] or email spam [Carreras et al. 2001], review spam is much harder to detect. The main reason is that spammers can easily disguise themselves as genuine reviewers, which makes it very difficult for a human user to recognize them, let alone by computers; while in other forms of spam detection tasks, one can tell spam from nonspam without much difficulty.

Previous review spam detection algorithms used behaviors of reviewers to catch product review spammers, for instance, review text similarities, rating similarities and deviations, the number of spammed products, etc. According to the existing studies, these behaviors are useful clues of certain types of spamming activities. For instance, if a reviewer uses a significant amount of similar text in multiple reviews about the same product, or a reviewer rates different products constantly high or low within a short period of time, (s)he may be a spammer [Jindal and Liu 2008; Lim et al. 2010], since these activities can imply one type of spamming, that is, trying to make the most impact out of the least effort.

Unfortunately, since we are interested in *store* reviews in this work, even if we can borrow experiences from previous studies, these clues may not be discriminative or sufficient to tell store review spammers from benign reviewers. For example, although it looks suspicious for a person posting multiple reviews to the same product, it can be quite normal for a person posting more than one review to the same store due to multiple purchasing experiences. More specifically, as one person has the same writing style on review writing, it may be normal to have similar reviews from one reviewer across multiple stores because unlike different products, different stores basically provide the same services. Moreover, many ordinary users write reviews only sporadically. It is reasonable for them to write multiple reviews within a short period of time for different shopping experiences. Therefore, reviewer behaviors proposed in the existing approaches for product reviews may not be sufficient for catching spammers of store reviews. Besides, there are other types of spammers who carefully design their fake reviews in a very deceptive way that would slip through existing spammer detection techniques.

Figure 1 shows the profiles of two reviewers of stores from Resellerratings.com that demonstrate such inaccuracy and insufficiency. The first case (Figure 1(a)) is that a reviewer kept posting similar reviews that seem unrelated to the reviewed stores in a short period of time. These duplications and short-time rush phenomena could be a sign of spamming and they have been used as important clues to catch spammers [Jindal and Liu 2008; Jindal et al. 2010; Lim et al. 2010]. However, after manually checking the content of the links plus other evidences gathered from the Internet, we found that these duplications were posted by a genuine warm-hearted reviewer who tried to warn other consumers about a cluster of bad stores that have many complaints.

(a) seems suspicious, but benign reviewer



(b) seems normal, but really suspicious reviewer

Fig. 1. Reviewer examples.

The second case (Figure 1(b)) is even more subtle. The person's behaviors may seem normal at a first glance. However, we can be almost sure that (s)he is a spammer if we go beyond behaviors of this single reviewer and consider the stores this reviewer commented on, and other people's reviews of the same stores, and their reviews about other stores. (more discussions in Case Study, Section 4).

These examples reveal the incompleteness of the existing spammer detection methods and the need to look for a more sophisticated and complementary framework. However, the following challenges are major obstacles towards such a framework.

(1) There is no ground truth of whether a review is written by a spammer or not. By reading the review text alone, we usually do not have enough clues to tell spam from nonspam.
(2) Spammers' behaviors may be hard to capture. For example, in order to successfully mislead customers, spammers can make their writing styles and review habits look very similar to those of ordinary reviewers as shown in the second case in Figure 1.
(3) A spammer can also write good and honest reviews, because they could be real customers of some online stores themselves sometimes. Even more complicated, a current good reviewer could be a spammer before, and we do not know when a reviewer would write a spam review.
(4) For a given product or online store, there could be more spam reviews than normal ones. This could happen if they hired spammers to register multiple users and write good things about them, while at the same time, there are fewer customers who actually purchased stuffs from them.

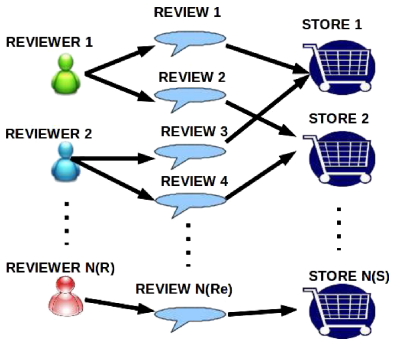


Fig. 2. Review graph.

These obstacles are the core challenges that make simple behavioral heuristics insufficient. To capture sophisticated spammers, we need to consider more clues than just simple reviewer behaviors.

1.2. Contributions

Our first contribution is to propose a heterogeneous graph model with three types of nodes to capture spamming clues. In our opinion, clues of telling if a reviewer is innocent or not include the reviewer's reviews, all the stores (s)he commented on, and reviews from other reviewers who have shopping experiences on the same stores. Therefore, we propose a novel heterogeneous graph, referred to as the *review graph*, to capture relations among reviewers, reviews, and stores. These are three different kinds of nodes in the review graph. A reviewer node has a link with a review if (s)he wrote it. A review node has an edge to a store node if it is about that store. A store is connected to a reviewer via this reviewer's review on the store. Each node is also attached with a set of features. For example, a store node has features about its average rating, its number of reviews, etc. Figure 2 illustrates such a review graph.

Our second contribution is the introduction of three fundamental concepts that can be computed from the review graph, that is, the *trustiness* of reviewers, the *honesty* of reviews, and the *reliability* of stores, and the identification of their interrelationship: a reviewer is more trustworthy if the person has written more honesty reviews, a store is more reliable if it has more positive reviews from trustworthy reviewers, and a review is more honest if it is supported by many other honest reviews. Furthermore, if the honesty of a review goes down, it affects the reviewer's trustiness, which has an impact on the store (s)he reviews. And depending on how this reviewer's opinion varies with other reviewers' opinions about the same store, other reviewers' trustiness may change. These intertwined relations are revealed from the review graph. Comparing with the existing methods [Jindal and Liu 2008; Jindal et al. 2010; Lim et al. 2010], which only consider review relations, mainly duplications, and use them to depict a reviewer's behavior, our model handles more subtle clues from different sources, for instance, reviewers, reviews, and stores. The three concepts are more complete representations of spamming clues. This is the first time such intricate relationships have been identified for spam detection and captured in a graph model.

Our third contribution is the development of an effective iterative computation method for the three concepts, based on the proposed graph model. Starting from the common senses that are uniquely related to store review system and its spamming scenarios, we derive how one concept impacts another. The mathematical forms of these relations are also significantly different from the classic graph reinforcement formulations such as those in HITS [Kleinberg 1999].

Modeling review spam analysis problem with a review graph to include more hidden spamming clues is a brand new angle. To the best of our knowledge, we are the first to take this angle and the first to apply a node reinforcement method to analyze the nontrivial relations of a reviewer's trustiness, a store's reliability, and a review's honesty.

This article is organized as follows: In Section 2, we discuss previous work on spam detection and their limitations. In Section 3, we formally define the spam detection problem and propose our solution in detail. In Section 4, we show our experimental results based on all the reviews and reviewers of a popular store review Web site.

2. RELATED WORK

Detecting opinion spam is a less-studied and more difficult area than detecting other forms of spam, such as Web search engine spam [Gyngyi and Garcia-Molina 2005] and email spam [Carreras et al. 2001]. Search engine spam refers to spam on Web pages, which has two main types, that is, content spam and link spam. Link spam is spam on hyperlinks, which almost does not exist in online reviews. Content spam adds irrelevant but popular words in target Web pages in order to fool search engines to make the pages relevant to a large number of search keywords, which again hardly occurs in reviews. Email spamming refers to broadcasting irrelevant emails from a spam source to a large number of email addresses that it gathers. It is also different line of research. Furthermore, there is ground truth for both these forms of spam. Simply by looking at the Web pages or emails, one knows whether they are spam or not. However, for opinion spam detection, we have no such ground truth. As we have demonstrated in Figure1, even human readers cannot decide whether a store review is a spam or not.

Jindal and Liu [2008] proposed the problem of *product* review spammer detection [Jindal and Liu 2008]. They identified 3 categories of spams: fake reviews (also called untruthful opinions), reviews on brand only, and nonreviews. Their spam detection method uses supervised learning. They first engineered a set of features about reviews, reviewers, and products. They then semiautomatically labeled spam reviews using mainly text similarity and some manual effort. Those labeled spam reviews are duplicate or near duplicate reviews, and are treated as positive training examples and the rest are used as negative ones. Based on the features and the training data, a classifier is constructed to detect spam reviews. Their approach relies heavily on text similarity, therefore it is only good for certain types of spamming activities, that is, duplicated review spamming.

Another spam review detection algorithm using *unexpectedness* rules mining is proposed in Jindal et al. [2010]. They treated each review as a record associated with a rating class (positive, negative and neutral). A rule mining algorithm is used to produce a list of unexpectedness rules. Their study found that the top unexpected rules indeed imply abnormal reviews and reviewers. However, their method does not identify true spammers, but only finds some strange behaviors as unexpected rules.

Lim et al. [2010] studied the behaviors of reviewers for spammer detection. They also find several features to capture of spamming behaviors, for instance., multiple ratings/reviews on a single product or a group of products, and rating deviations. Each reviewer has some scores on these features. And a linear combination of the scores indicates the suspicious degree of a reviewer. The method is unsupervised, which saves a lot of human labeling efforts. However, based on their experiments, this work essentially also relies on duplications, that is, multiple reviews from the same reviewer targeting the same item or item group. Thus, it is only suited for a special kind of spamming. Although our work also exams reviewers' behaviors, they are only a small part of our solution. As discussed in the introduction section, our review graph can

capture much more information and also a rich set of interrelationships of reviewers, reviews, and stores for spam detection, rather than just studying the individual behavior of each reviewer as in Lim et al. [2010].

Wu et al. developed a spam review detection algorithm based on the ranking of products (in their case, hotels) [Wu et al. 2010]. Their observation is that spam reviews may distort product ranking more than random chosen reviews. Their methods is thus restricted to the situation where the ranking of products is available, which cannot be applied to general review datasets.

Another work [Mukherjee et al. 2011] studied a special spamming activity, that is, group spamming, where a group of spammers are assumed to write spam reviews together on a few target stores. Their method, however, focuses on catching spammer groups and is not applicable to the general case of discovering individual spammers.

More recently, researchers applied psychological and linguistic clues to identify review spam [Ott et al. 2011]. They first use Mechanical Turk to obtain suspicious reviews in hotel rating domain. After that, they identify genres that frequently appear in those suspicious reviews as features to classify other reviews. Our work has two major differences from their work. First, our model is completely unsupervised, while their work requires Mechanical Turk to first identifies suspicious reviews as training data. Second, it is still an open question how linguistic clues will uncover spams. For example, after knowing that certain words are possibly indicators to spams, spammers could simply change their writing styles to avoid them. Moreover, the words they identified as spam indicators are quite normal, for instance, "family," "Chicago," "vacation," which may appear in any truthful reviews.

The general opinion mining research does not perform spam detection, although it studies many aspects of opinion and also the quality of reviews [Liu 2010; Pang and Lee 2008]. The quality of reviews is related to our work but low quality reviews do not mean spam reviews.

Comparing with all the previous work, ours is the first to explore how a reviewers relations with other reviewers and stores reveal clues of spamming. Complementing to existing work, our method is able to find more subtle and sophisticated spamming activities, using reinforcements of reviewers trustiness, store's reliability, and review's honesty.

The general idea of reinforcement based on the graph link information has been applied in many different scenarios. PageRank [Page et al. 1998] and HITS [Kleinberg 1999] are successful examples in link-based ranking using reinforcement. But they are not applicable to our spam detection case. PageRank discovers prestige nodes in a graph. But prestige is not related to spamming activities. We cannot say a prestige node has a larger chance to be a spammer or a benign reviewer. HITS has authority and hub nodes influencing each other, which are quite different from our three types of nodes interactions. In our framework, we do not have the concepts of authority or hub. Specially, HITS computes authority weight of a node by the summation of the weights of hubs linked to it. In our case, simply using summation of one kind of scores of trustiness, reliability, or honesty to infer another would not work. We derived the interaction functions from unique characteristics of the review spam detection domain (see Section 3.1 for detailed discussions). They turned out to be nonlinear functions rather than linear functions as in HITS. Moreover, we have three different types of nodes and each node also has attached features, which dont exist in both PageRank or HITS.

Reinforcement based on the graph link information has been applied in many different scenarios including bias opinion holder and controversy entity identification [Lauw et al. 2006], authority discovery and truth identification from multiple conflicting sources [Yin et al. 2008], and in identifying fair research paper reviews in an

evaluation system [Lauw et al. 2008]. Their problem settings and detailed techniques are different from review spam detection, therefore they are not applicable to our work.

Other kinds of link based analysis, for instance, belief propagation, have also been used for detection of other forms of untruthful information, such as online auction fraud [Pandit et al. 2007] and accounting risk [McGlohon et al. 2009]. However, they are not comparable with our work. First, in Pandit et al. [2007] and McGlohon et al. [2009], the graphs are homogeneous, generated based on transactions, for instance, auctions or credited/debited relations. While in our model, it is hard to use a homogeneous graph to represent three different entities and their relations. Second, even if they demonstrated that belief propagation is well suited in their situations, it is not clear how the algorithm can work for our heterogeneous graph. For example, in their cases, a malicious node has larger probability to link to another malicious node than a benign node, but in our case, an unreliable store may links to many trustworthy reviewers or many untrustable reviewers, depending on their opinions, etc. This is only one example. The situations that our model deals with are more complex, because review spam has little clear-cut logic that can be easily captured as in auction fraud or accounting risk.

3. SPAM DETECTION MODEL

In this section, we first introduce the assumptions based on which we build the proposed model. Then we define three factors in the model, namely, reviewers' trustiness, reviews' honesty, and stores' reliability. At the end of the section, we discuss the iterative computation framework, which computes the three factors.

3.1. Intuitive Assumptions and Observations

We start by exploring possible causes of spamming activities. Grounded in common sense, we first make the following assumptions.

— Spammers are usually for profits, so they have connections to stores that would benefit spammers to promote their prominence or defame other stores [Consumerist].
— There are several categories of stores in terms of their quality. High quality stores are excellent in terms of their product qualities and customer services. Low quality stores are either junk ones that sell poor quality products or fake ones that never deliver products.
— Spammers are usually hired by low quality stores. We make this assumption based on the following reasons.
  (i) Low quality stores are the ones that suffer if customers know the truth about them.
  (ii) They are the ones in desperate need to be promoted by customers, which may be the only way to get attention of more customers.
  (iii) They are the ones that are less competitive, but want to defame their competitors in order to get themselves more profits.

  Such stores have a stronger motivation to hire spammers to write dishonest reviews. On the other hand, stores with better reputation, stable consumer population, and good revenue may not hire spammers at all, since they lose much more if they are caught doing so. Even if a good store really entices spammers to say good things about them, it may not be very harmful. Therefore, we assume that less reliable stores are more likely to be involved in review spamming.
  However, good stores may also hire people to damage the reputation of their competitors.

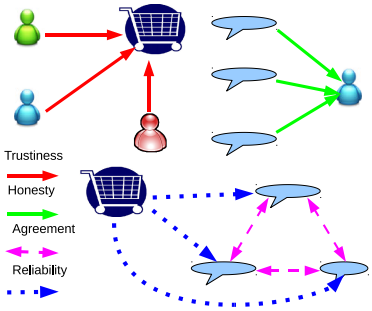


Fig. 3. Influences among different types of nodes.

— Harmful spam reviews always try to deviate from the truth. Therefore, they can be either positive reviews about lousy stores, or negative reviews about good stores.
— Not all reviews deviating from mainstream are spam. People may feel differently or have different experiences about the same service.

Based on these assumptions, we can infer that enthusiastic praises about low quality (therefore less reliable) stores and unreasonable complains about high quality (therefore more reliable) stores are more suspicious. Although there may be exceptions in practice, these types of spammers are our main concerns, since we believe that they cover a significant volume of harmful spamming activities. Therefore, we have the following observations.

(1) We can judge the honesty of a review given the reliability of the store it was posted to, plus the agreement (to be defined later) that the review with its surrounding reviews about the same store. Other surrounding reviews' agreement (to be defined later) with the same store.
(2) If we have the honesty scores of all the reviews of a reviewer, we can infer his/her trustiness, because one is certainly more trustworthy if one wrote more reviews with high honesty scores.
(3) Now we go back to see how to depict a store's reliability. From common sense again, a store is more reliable if it is reviewed by more trustworthy reviewers with positive reviews, and less reliable if it is reviewed by more trustworthy reviewers with negative reviews.

Figure 3 shows these influences among a store's reliability, a review's honesty, and a reviewer's trustiness. They are intertwined and affecting each other. These influences have some resemblance to the well-known authority and hub relations [Kleinberg 1999], but the relations in our case are very different from those in authority-hub analysis. First, a reviewer cannot gain more trust by writing more reviews, nor can we use the mean of review honesty to capture a reviewer's trustiness. For example, reviewer A has only one review with confidence 1. Reviewer B has ten reviews and average confidence is 0.9. In reality, reviewer B should be more trustworthy than A. Second, reviews have impact on each other. If a review deviates from most of the others, it could be a clue to spamming.

3.2. Basic Definitions

From these observations, we define variables that quantify the qualities of reviewers, reviews, and stores.

Table I. Features Associated with Nodes and Their Notations

Notation	Definition
r	a reviewer
v	a review
s	a store
α_r^i	reviewer r's i^{th} review
β_r^s	reviewer r's review on store s
κ_v	review v's author id
n_r	the number of r's reviews
H_r	the honesty summation of r's reviews
Ψ_v	review v's rating
Γ_v	store id of v's
t_v	review v's posting time
U_s	the set of reviews on store s

Definition 3.1 (*Trustiness of reviewers*). The trustiness of a reviewer r (denoted by $T(r)$) is a score of how much we can trust r. For ease of understanding and computation, we limit the range of $T(r)$ to $(-1, 1)$.

Definition 3.2 (*Honesty of reviews*). The honesty of a review v (denoted by $H(v)$) is a score representing how honest the review is, $H(v) \in (-1, 1)$.

Definition 3.3 (*Reliability of stores*). The reliability of a store s (denoted by $R(s)$) is a score representing the quality of store s. $R(s) \in (-1, 1)$.

For ease of presentation, we give the notations of node features that are related to the computation of the definitions in Table I.

3.3. Reviewers Trustiness

To model the trustiness, let's see how we can tell if a reviewer is trustworthy or not, based on which we can devise several computational ideas.

(1) A reviewer's trustiness does not depends on the number of his/her reviews, but on the summation of their honesty scores.
For example, a reviewer with ten reviews could be less trustworthy than a reviewer with only one review, if those ten are fake and the one is honest. But if two reviewers have similar average review honesty scores, one is more trustworthy than the other if (s)he has more honest reviews.
(2) A reviewer's trustiness score should be positive/negative if he/she writes more reviews with positive/negative honesty scores. While the score should be negative if he/she has many reviews with negative honesty scores.
(3) For a given reviewer, his/her trustiness does not grow/drop linearly with the number of high/low honesty reviews that (s)he wrote. It grows/drops faster when the number of such reviews is smaller, and slows down when the number is larger.
For example, when a reviewer has already written 100 honest reviews hence becoming highly trustworthy, his/her trustiness does not improve that much by writing one more honest review. However, if we have a reviewer with only 2 reviews, his/her trustiness will increase significantly as the third high honesty review appears.

Here we define a reviewer r's trustiness to be dependent on the summation of the honesty scores of all r's reviews,

$$H_r = \sum_{i=1}^{n(r)} H\left(\alpha_r^i\right), \tag{1}$$

where $n(r)$ is the number of reviews from r.

The proposed trustiness score of r should be a function T satisfying the given intuitions, which are formally expressed with the following relations.

$$T(i) < T(j),\ if\ H_i < H_j \tag{2}$$

$$T(r) < 0,\ if\ H_r < 0,\ T(r > 0)\ if\ H_r > 0 \tag{3}$$

$$\frac{dT(r)}{dH_r} = T(r)(K - T(r)). \tag{4}$$

Relation (2) represents the first idea: A reviewer's trustiness depends on the summation of the honesty scores of all r's reviews. One reviewer is more trustworthy than another if one has a larger honesty score. Relation (3) depicts the second idea: We don't trust a reviewer whose reviews tend to be dishonest, and we trust a reviewer whose reviews tend to be honest. Relation (4) represents the third idea that is similar to the population growth model [Pearl and Reed 1920]: the growing (or dropping) speed of trustiness is the product of current trustiness level and the room of improvement. Suppose K is the upper bound of the trustiness score, $K - T(r)$ is how much more trustiness one can gain by writing more honest reviews. It becomes smaller as $T(r)$ grows. Solving these relations gives us the general form of trustiness function

$$T(r) = \frac{K}{1 + e^{-KH_r}}. \tag{5}$$

The function is the well-known logistic function and is shown in Figure 4.

A reviewer's trustiness is a continuous value over its confidence. It increases as the reviewer has more and more high honest reviews, and decrease with more and more low honest reviews. It grows (or drops) faster at the beginning, when the number of reviews is relatively small, and approaches stable situations when the number of reviews gets larger.

Notice this general form is in the range $(0, K)$. As we mentioned before, the upper bound of trustiness score is +1 in our case, and its range should be $(-1, 1)$ to have more practical meanings.

Therefore, we rescale the trustiness score T of a reviewer r as

$$T(r) = \frac{2}{1 + e^{-H_r}} - 1. \tag{6}$$

The general shape of the trustiness function is the well-known logistic curve and is shown in Figure 4.

In reality, reviews from the same reviewer can have different degrees of honesty, due to many reasons, for instance, subjective bias. A normal reviewer p may post an unreasonable review i with $H(p(i)) < 0$, while a spammer q may have a honest review j with $H(q(j)) > 0$. However, it is still possible to distinguish malicious reviewers from benign ones, since the trustiness score is controlled by the summation of a reviewer's review honesty, that is, the general trend of a user's reviews. To calculate $T(r)$, we need to know the honesty values of r's reviews, which we define next.

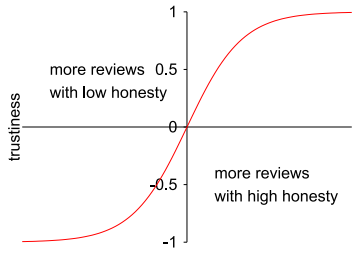


Fig. 4. Relation between *trustiness* and *honesty*.

3.4. Review Honesty

How do we interpret a review? When we shop online and read a review, we usually keep two factors in mind. The first one is the store we are looking at. If the store is a good one such as Apple.com, we tend to trust the positive reviews on it. If we are reading reviews about a store we have never heard of, or we knew it was a bad one, we tend to doubt the high rating reviews posted on it. The second factor is the surrounding reviews, which are other reviews about the same store within a certain time window Δt, for instance, 3 months before and after the posting time of the target review. We are likely to trust the mainstream opinions held by most surrounding reviews, rather than outliers opinions.

Based on these observations, we model a review's honesty with two factors.

(1) The *reliability* of the store it reviews.
(2) The *agreement* between this review and other reviews about the same store within a given time window.

We will discuss reliability in the next section. We study agreement first by introducing the surrounding set.

Definition 3.4 *Surrounding Set*. The surrounding set of review v is the set of v's surrounding reviews.

$$S_v = \{i \,|\Gamma_i = \Gamma_v,\ |t_i - t_v| \leq \Delta t\}.$$

$\forall i, j \in S_v$ agree with each other when their opinions about the same aspects of the store are close. However, opinion mining is too costly to let us assess opinion of a reviewer [Hu and Liu 2004; Popescu and Etzioni 2005]. Fortunately, in addition to the review text, reviews usually have rating information about a store. Even if two 5-star ratings may mean different aspects of a store, such as customer service and delivery, they are correlated with each other. Therefore, we make some assumptions here.

— A review's rating about a store reflect its opinion.
— Two reviews with similar rating scores about the same store have similar opinions about the store.

From the assumptions, $\forall i, j \in S_v$ agree with each other if their ratings are similar, for instance, they are 5-star and 4-star, or 1-star and 2-star.

$$|\Psi_i - \Psi_j| < \delta, \tag{7}$$

where δ is a given bound (we use 1 in a 5-star rating system in this article).

Thus, we can partition the surrounding set S_v as

$$S_v = S_{v,a} \cup S_{v,d} \tag{8}$$

$$S_{v,a} = \{i \mid \ |\Psi_i - \Psi_v| < \delta\} \tag{9}$$

$$S_{v,d} = S_v \setminus S_{v,a}. \tag{10}$$

We also take the reviewers' trustiness scores into consideration. One review should be good even if it does not agree with any surrounding reviews when it is written by a trustworthy reviewer while the surrounding reviews are posted by untrustworthy reviewers. Similarly, one review may be bad even if it agrees with the surrounding reviews, since they may all come from spammers. Therefore, we define the agreement score of review v within time window Δt as

$$A(v, \Delta t) = \sum_{i \in S_{v,a}} T(\kappa_i) - \sum_{j \in S_{v,d}} T(\kappa_j). \tag{11}$$

Notice that trustiness score T could be either positive or negative. This equation means that if one review agrees with other reviews by trustworthy reviewers, its agreement score increases. On the other hand, if it agrees with untrustworthy reviewers, its score decreases. This equation also promotes a benign user's review agreement score when it is surrounded by spam reviews that it does not agree with. Because if spammers' trustiness scores are negative, a benign review's agreement score gets promoted by subtracting the negative numbers.

$A(v, \Delta t)$ can be positive or negative. Here we normalize it to $(-1, 1)$ to make later computation easier.

$$A_n(v, \Delta t) = \frac{2}{1 + e^{-A(v,\Delta t)}} - 1. \tag{12}$$

Review v's *honesty* $H(v)$ is defined as follows.

$$H(v) = |R(\Gamma_v)| A_n(v, \Delta t), \tag{13}$$

where $R(\Gamma_v)$ is the reliability of store Γ_v, which we will define later. Here we just want to note that, by definition, $R(\Gamma_v)$ can be positive (for reliable stores) or negative (for poor quality stores). We take its absolute value as an amplifier of $A_n(v, \Delta t)$. This is consistent with previous discussions. If a store's $|R(\Gamma_v)|$ is large, it is either quite good or quite bad. A review on this store should have a high honesty score if it agrees with many other honest reviews. If $|R(\Gamma_v)|$ is small, the store's reliability is difficult to tell, and a review's honesty score gets diminished a little by that fact.

3.5. Store Reliability

To define reliability, we have similar intuitions to trustiness. A store is more reliable if it has more trustworthy reviewers saying good things about it, while it is more unreliable if more trustworthy reviewers complains about it. The increasing/decreasing trend of reliability should also be a logistic curve so as to be consistent with our common sense. Therefore, we define the reliability R of store s as

$$R(s) = \frac{2}{1 + e^{-\zeta}} - 1, \tag{14}$$

where $\zeta = \sum_{v \in U_s, T(\kappa_v) > 0} T(\kappa_v)(\Psi_v - \mu)$ and μ is the median value of the entire rating system, for instance, 3-star in a 5-star rating range.

ALGORITHM 1: Iterative Computation Framework

Input: The set of store $\mathscr{S}$, review $\mathscr{V}$, and reviewer $\mathscr{R}$
The agreement time window size Δt, review similarity threshold δ, and the *round* of iterations
Output: The set of *reliability* R, *honesty* H, and *trustiness* T
// Initialization step Initialize store's reliability and reviewer's trustiness to 1, compute review's agreement using these initial values and Δt, δ according to (11) and (12)
roundCounter = 0 **repeat**
  for $v \in \mathscr{V}$ **do**
    compute $H(v)$ using (13);
  end
  for $r \in \mathscr{R}$ **do**
    compute $T(r)$ using (6);
  end
  for $s \in \mathscr{S}$ **do**
    compute $R(s)$ using (14);
  end
  for $v \in \mathscr{V}$ **do**
    update v's agreement using new $\mathscr{T}$ based on (11) and (12);
  end
  roundCounter++;
until *roundCounter* < *round*;

Therefore, a store's reliability depends on all trustable reviewers who post reviews on it, and their ratings. When considering reliability, we only consider reviewers with positive trustiness score because their ratings really reflect the store's quality. In contrast, whatever a less trustworthy reviewer says about a store, it is less trustable. For example, we do not know the real intention to rate a store as a good one or bad one for a review, if it comes from a potential spammer.

3.6. Iterative Computation Framework

Integrating the pieces of information of the review graph together, we have an iterative computation framework to compute reliability, trustiness, and honesty, by exploring the inner dependencies among them. The algorithm is given in the figure below, which is self-explanatory.

The overall time complexity of our framework is

$$O(k(N_r + N_v)),$$

where k is the number of iterations. k is small (usually 4 or 5) as the algorithm converges quite fast in practice.

In every round, to compute trustiness T for every reviewer, we only need to access each review once and get its honesty value. Therefore, this step is linear to the number of reviews N_v. For a similar reason, the reliability of R computation is linear to the total number of reviewers N_r. We will show in Section 4.1 that in our case, the computation of honesty H for all reviews is also linear to the number of reviews N_v, from a probabilist point of view. Therefore, the whole running time is linear to the number of reviews and reviewers.

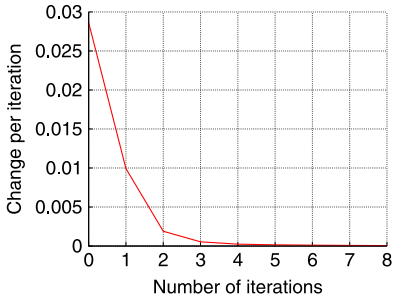


Fig. 5. Convergence.

4. EVALUATION

4.1. Dataset and Data Features

We use the store review data from www.resellerratings.com, one of the largest hosts of store reviews, for our experiments. The Web site provides a unique url for every reviewer's profile, containing metadata such as reviewer' id, all his/her reviews and ratings with posting times about stores, and links to those stores. At every store's page, there is information about its average rating score, all reviewers with their reviews.

The data we crawled is a snapshot of complete information from the Web site on October 6, 2010. We clean the data by removing users and stores with no review. After that, we have 343,603 reviewers who wrote 408,470 reviews on 14,561 stores in total.

4.2. Computational Performance

Figure 5 shows the convergence of the iterative computation. Let $\mathscr{T}i$ and $\mathscr{T}i+1$ be the vectors of trustiness scores of iteration i and $i+1$. The change

$$c_i = 1 - cos(\mathscr{T}i, \mathscr{T}i+1)$$

converges quickly after a few iterations.

Now we analyze the complexity when computing review honesty. As we defined in the previous section, the honesty of a review is determined by the reliability of the store it comments on and the agreement between itself and other reviews of its surrounding set. To access the reliability core of a given store is only an $O(1)$ operation. However, it seems to cost $O(m^2)$ to compute the agreement of a review v, where m is the number of reviews of store Γ_v. Fortunately, the review number distribution follows *power-law* [Barabasi and Albert 1999] in the review graph. It means that only a few stores have many reviews, while most stores have only a few reviews.

Figure 6 illustrates the review number k's probability distribution over stores. It is a double-log plot of review distributions. Fitting it with a straight line, we have

$$P(k = m) = \eta m^{-1.43}.$$

Therefore, total operations on computing agreements G in one iteration is

$$\begin{aligned} Op(G) &= \sum_m P(k = m)m^2 = \sum_m \eta m^{-1.43} m^2 \\ &= \sum_m \eta m^{0.57} < \sum_m \eta m = O(N_v) \end{aligned}$$

that is linear in the total number of reviews.

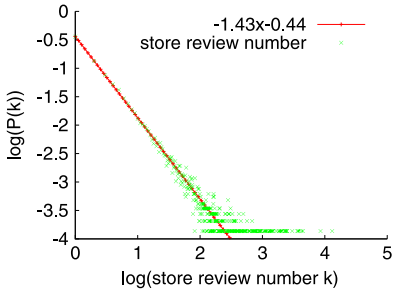


Fig. 6. Review distribution.

4.3. Spammer Detection Results Evaluation

4.3.1. Evaluation Criteria. Review spam detection's unique challenge lies in its lack of ground truth, that is, the label indicating which review is a spam or which reviewer is suspicious does not exist naturally in the review data. To obtain such labels, human judges are needed to read the reviews, search the Internet regarding the truth of the reviews, and use their intuitions to make the judgment, which requires significant amount of effort. Thus, it is too labor-intensive and time-consuming for human judges to go over all reviews and label them. Instead, we let our algorithms identify highly suspicious spammer candidates first, and then recruit human judges to examine them to decide how many are indeed suspicious. Through this, we evaluate the effectiveness of our algorithms.

We use IR-based evaluation strategy. First we let our algorithm identify highly suspicious spammers candidates. Then we recruit human judges to make the judgments on the candidates about whether they seem to be real spammers. Therefore, we have precision as our performance measure. Similar evaluation approaches have been used in previous review spam detection research [Jindal and Liu 2008; Lim et al. 2010], and in auction fraud detection [Pandit et al. 2007], therefore this is a well-established way of performance evaluation. The details of our evaluation strategy is as follows. Spammer detection result evaluation involves a great deal of human judgment, so it is different from ordinary numerical analysis and curve comparison. To suit this special evaluation task, we adopt the following two strategies. They are also used by previous research [Jindal and Liu 2008; Lim et al. 2010]. We follow the same line in our experiments.

IR-Based Evaluation Strategy. The goal of spammer detection method is to identify suspicious reviewer candidates for further investigation. Thus, it is similar to an information retrieval task, which tries to present the users the most relevant items first. Therefore we borrow the evaluation measures from information retrieval to show the effectiveness of our spammer detection algorithm. There are two such standard measures: precision and recall. For our task, precision is defined as the fraction of spammers from all candidates retrieved by the algorithm; recall is the percentage of retrieved spammers among all *real* spammers. Since we have no ground truth about *real* spammers, we simply use the precision as the evaluation measure, which is also commonly used in relevance judgment of information retrieval.

Human Judgments Consistency Criterion. Since there is no "spammer" label in any kind of review data, human evaluation is necessary to judge if a target is trustworthy. A spammer detection algorithm is effective, if different human evaluators agree with each other about their judgments and concur with the system on the same

set of results. Therefore, we use human judgment consistency as another important evaluation criterion.

Our human evaluators are 3 computer science major graduate students who also have extensive online shopping experiences. They work independently on spammer identification.

4.3.2. Human Evaluation Process. Judging suspicious spammers is a complicated task for human and often involves intuition and searching for additional information, especially when we target at more subtle spamming activities. To decide if a candidate is a spammer requires human judges not only to read his/her reviews and ratings, but also to collect evidences from relations with other reviewers, stores, and even the Internet.

To standardize this complex judgment process, our human evaluators agree upon three conditions and put them together as our evidence to claim that a reviewer is a potential spammer.

— A reviewer is suspicious if (s)he has a significant number of reviews giving opposite opinions to others' reviews about the same stores.
  For example, if a reviewer gives high ratings to all the stores he has reviewed, while other reviewers usually rate these stores low, the reviewer is problematic.
— A reviewer is suspicious if (s)he has a significant number of reviews giving opposite opinions about some stores as compared to the ratings from the Better Business Bureaus (BBB)[1]
  For example, if a reviewer gives high ratings to all the stores (s)he reviewed, but BBB gives them *F*s (the rating range is from *A* to *F*), the reviewer is clearly suspicious.
— A reviewer is suspicious if (s)he has a significant number of reviews saying opposite opinions about some stores as compared to evidences presented by general Web search results.
  For example, if a reviewer gives high ratings to all stores he/she reviewed, but Google search results about these stores often contain information of them having fake reviews, this reviewer is problematic.

Note that each of these steps involves human labor effort, intuition, and background knowledge, so the entire evaluation is very hard to be computerized. For example, for the third step, our human judges actually need to go through several search result pages and understand the content, in order to make a judgment.

Although every single condition may not be convincing enough to prove spamming activities, all of them together can be a confident claim of spamming. We invite our human judges to evaluate the top 100 suspicious reviewers identified by our model. Note that no existing work focused on such a large scale and subtle individual cases.[2] Human evaluators gave their independent judgments based on various information from resellerratings.com, business honesty information from BBB, and search results from Google, and by reading reviews, according to the given conditions.

Since our human resource limitation, it is inconvenient to let every evaluator go through all 100 reviewers and look for evidences themselves. Therefore, the authors spent significant amount of time collecting evidences to show if a reviewer is likely to be a spammer or not.

---

[1]Better Business Bureau is a well-known corporation that endeavors to a fair and effective marketplace. It gathers reports on business reliability, alert the public to business or consumer scams, and enforce the mutual trustiness between consumers and companies

[2]In a closely related work [Lim et al. 2010], they only assign 50 reviews to each human judge. Moreover, their judging criteria are not as complicated as ours.

Table II. Human Evaluation Result

	Evaluator 1	Evaluator 2	Evaluator 3
Evaluator 1	**49**	33	37
Evaluator 2	-	**35**	23
Evaluator 3	-	-	**40**

4.3.3. Precision and Consistency. In our evaluation, if more than one evaluators regard a reviewer as a spammer, we label it as a suspicious spammer. Our evaluators identified 49 out of 100 suspicious candidates to be spammers. The precision is 49%. Although our precision is not very high compared to the previous work, we are dealing with much more subtle and complex cases (not simple duplications), which existing studies could not handle. Besides, our precision is meaningful, since human evaluators agree with each other on their judgments.

Table II shows the agreement of human judges. For example, Evaluator 1 identified 49 suspicious reviewers, out of which 33 were recognized by Evaluator 2 and 37 were caught by Evaluator 3. To explore their agreement, we used *Fleiss's kappa* [Fleiss and Cohen 1973], which is an interevaluator agreement measure for any number of evaluators. The kappa among 3 evaluators is 60.3%, which represents almost substantial agreement [Landis and Koch 1977]. It is also important to note that those reviewers who were not identified as spammers are not necessarily innocent. It is simply because the human judges have not found enough strong evidences to conclude that they are spammers.

4.3.4. Compare with Baseline. Since our work is the first one utilizing a review graph and targeting at subtle spamming activities, there is no existing work that is comparable. Besides, previous studies are mainly focusing on different forms of duplicate reviews, in order to catch spammers. Given the differences between our work and previous studies, we want to demonstrate that suspicious reviewers found by our method can hardly be identified by existing techniques.

We choose to compare with the approach in Lim et al. [2010], because it is the state of the art of behavior-based spammer detection techniques. They use some types of duplicate reviews as a strong evidences of spamming. They first look for candidates who have multiple reviews about one target (in our case store), and then compute spamming scores to capture spammers from the candidates.

In our top 100 suspicious candidate list, only 3 candidates can be found based on the duplication criterion. Moreover, only 1 out of these 3 candidates is finally labeled as a spammer by our human evaluators. This result means that our work detects different types of spamming activities from existing researches. They can seldom find the spammer types that we are able to find. By no means do we claim that the existing methods are not useful. Instead, our method aims to find those that cannot be found by previous methods. there are only 59 candidates that they can find in our dataset. Only 3 of these 59 candidates are in our top 100 suspicious list.

4.3.5. Suspicious Spammer Case Study. Here we take a close look at reviewers ranked highly suspicious by our algorithm. We pick two candidates: one often promotes stores while the other demotes stores.

The first case is reviewer *howcome*,[3] whose reviews are mostly positive. When we examined these reviews, we found many of them are problematic, such as those highly rated reviews for UBid, OnRebate, ISquared Inc, Batteries.com, and

[3]http://www.resellerratings.com/profile.pl?user=57665

Table III. Top 10 Reliable Stores

Store Name	Reselleratings Rating	BBB Rating
TigerDirect	7.44	A
SuperMediaStore	9.27	A^+
OneCall	9.33	A^+
Newegg	9.77	A^+
Mwave	9.18	B^-
LA Police Gear	9.11	A^-
iBuyPower	8.33	B^-
FrozenCPU	9.44	A^+
eWiz	9.08	C
eForcity	8.55	A^-

BigCrazyStore.com. For example, UBid is widely complained at different review Web sites like ConsumerAffairs, epinions, CrimesOfPersuasion, and ResellerRatings. On-Rebate is *sued* for failing to pay rebates to customers. ISquared Inc is rated as D by BBB. Batteries.com is generally lowly rated at ResellerRatings. BigCrazyStore has very few records about its quality on the Internet. And this reviewer's review is the only one about BigCrazyStore on ResellerRatings. All these evidences lead us to conclude that this reviewer is suspicious.

The second case is *shibbyjk*,[4] All reviews of this reviewer are complains. And all stores being complained are high quality stores according to BBB, for instance, 1SaleADay(A), StarMicro(B^+), 3B Tech(B), and Accstation(A^-). From these unfair ratings, one may argue that this reviewer may be still normal but just picky. However, after reading all his/her negative reviews, we found that all complains are about transaction problems. It is fishy because the chance of all these high-standard companies having transaction problems and credit card frauds with this particular customer is very low. Besides, transaction problem is a good excuse to make false complains, since the truth is hard to be verify.

4.4. Store Reliability Evaluation

Our algorithm can also give every store a reliability score. In this section, we pick the top 10 and the bottom 10 stores to show their ratings on ResellerRatings.com and BBB. In this way, we demonstrate the accuracy of reliability scores that our model can offer. Note that we are only showing that our method is consistent with BBB score, but not suggesting the true quality of the listed stores. This consistency, however, does give us some evidence that our technique is effective.

Tables III and IV list top and bottom stores identified by our algorithm. (We randomly choose a store if multiple stores have the same reliability score.) As we can see, our algorithm generally gives high reliability scores to quality stores, and low scores to lousy ones. However, there are a few exceptions, where the BBB and our technique are not in agreement: one good and one bad. For example, CCI Camera City is an interesting case. Its BBB rating is quite high. However, when we put this company's name into Google and looked into the result pages (at the time when this article was written), we found 6 out of 10 results in the first page of Google search are customer complaints about it. This phenomenon shows that our algorithm works and no source could provide 100% accurate information about a company's quality, for instance, BBB

[4]http://www.resellerratings.com/profile.pl?user=565595

Table IV. Bottom 10 Reliable Stores

Store Name	Reselleratings Rating	BBB Rating
86^{th} Street Photo	0.30	F
Best Price Cameras	1.43	F
Dealer Cost Car Audio	1.23	F
USA Photo Nation	0.20	F
Camera Addict	0.59	F
CCI Camera City	0.44	A^+
OC System	3.00	F
Shop Digital Direct	0.35	F
Camera Giant	0.21	F
Infiniti Photo	0.28	F

also has blind spots. Another disagreement case is eWiz, but we do not have enough evidence to tell which rating is close to the truth.

4.4.1. Controversial Store Case Study. Despite the challenging nature, it will be interesting to see whether there are some stores that receive good/bad ratings from Resellerratings.com, but bad/good ratings from our social review graph model. We call such stores *controversial stores*, because we cannot claim which source provides the truth for sure. However, the results may raise the alert for customers to consider more carefully about the controversial stores and enable the hosting site to investigate the stores.

(1) Highly rated by our review graph model but lowly rated by Resellerratings.
   — *Just Deals*. Its rating is 0.95 by our model but 3.95/10 by Resellerratings. We searched its name in Google and found that it is an active store with advertisement on Google, accounts on Facebook and Twitter and many fans.
   — *Walmart*. Its rating is 0.96 by our model but 5.03/10 by Resellerratings. We know that it is a great company.
   — *Wirefly.com*. Its rating is 1 by our model but 4.83/10 by Resellerratings. This store also actively advertises on Google, Facebook and Twitter with many fans. There are some search results in Google contain the word "scam," but it turns out to be their advertising strategy. For example, the first result containing the word "scam" is on their Web site explaining why it is possible for them to offer products at very low price.
(2) Lowly rated by review graph model but highly rated by Resellerratings.
   — *Chumbo Corp*. Its rating is -1 by our model but 7.09/10 by Resellerratings. We found it is not in business anymore.
   — *Bunta Tech*. Its rating is -1 by our model but 5.65/10 by Resellerratings. We found little relevant information about this particular store, for instance, the first 5 pages in Google search results did not contain much information on this store except for reviews from Resellerratings.
   — *Compubuzz.com*. Its rating is -1 by our model but 5.13/10 by Resellerratings. Google returns many cheating complains upon searching the name of this store, for instance, 5 out of 10 results in the first result page.

Controversial store results could be used to further calibrate spamming detection results and provide more perspectives when raising spamming alerts. Such result is another unique outcomes of our social review graph model.

4.5. Review Honesty

As the third type of nodes in the review graph, every review also has a honesty score. However, unlike a reviewer or a store, it is harder and not informative to evaluate each single review. First, it is very hard, if not impossible, to judge one or two sentences are true or not. Second, it is not very useful to evaluate every single review when we already have reviewer and store trusty scores.

5. CONCLUSION

To fight spammers, we introduced a novel review graph model and an iterative reinforcement method that utilizes influences among reviewers, reviews, and stores that reviewers have reviewed. Our work is the first to consider clues that are out of the box of single reviewer's behaviors. Our method demonstrated how the review graph information reflects causes of spamming and reveals important clues of different types of spammers. We proposed a novel way to compute trustiness, honesty, and reliability scores, and demonstrated the effectiveness of interpreting reviewer's veracity and store's quality. Comparing with existing work, our method identified more subtle spamming activities with good precision and human evaluator agreement.

6. FUTURE WORK

Review spam detection is a challenging and under-exploration area, where only a few attempts have been made. In our opinion, there are three potential approaches that could help to produce more precise detection results. The first one is to apply ensemble methods to lift the performance. In our approach, we do not include text information. It is promising to ensemble our work with existing ones as classifiers with text features. The second potential approach is to apply different link based ranking algorithms in the heterogeneous review graph. We have three types of nodes in the graph. However, we have not explored link features and more node features. Links between different types of nodes could be treated differently to improve spam discovery performance. The third direction could be further exploring finer-grained forms of spams. For example, many spammers only write one reviews with one user name and they register many different user names. Our current approach does not intentionally handle such cases. Moreover, there are more and more ways to generate review spams. Automatically identifying such spams is still of high demand.

REFERENCES

BARABASI, A. AND ALBERT, R. 1999. Emergence of scaling in random networks. *Science*.

CARRERAS, X., MARQUEZ, L. S., AND SALGADO, J. G. 2001. Boosting trees for anti-spam email filtering. In *Proceedings of the 4th International Conference on Recent Advances in Natural Language Processing*. 58–64.

CONSUMERIST. Resellerratings cracks down on thecellshop.net's review bribing. http://consumerist.com/2008/05/reselleratings-cracks-down-on-thecellshopnets-review-bribing.html.

FLEISS, J. AND COHEN, J. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. In *Educational and Psychological Measurement*.

GYNGYI, Z. AND GARCIA-MOLINA, H. 2005. Web spam taxonomy. In *Proceedings of the Workshop on Adversarial IR on the Web*.

HU, M. AND LIU, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

JINDAL, N. AND LIU, B. 2008. Opinion spam and analysis. In *Proceedings of the International Conference on Web Search and Web Data Mining (WSDM'08)*. ACM, New York, NY, 219–230.

JINDAL, N., LIU, B., AND LIM, E. 2010. Finding unusual review patterns using unexpected rules. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*. ACM, New York, NY, 1549–1552.

KLEINBERG, J. 1999. Authoritative sources in a hyperlinked environment. In *J. ACM 46*, 5, 604–632.

LANDIS, J. AND KOCH, G. 1977. The measurement of observer agreement for categorical data. *Biometrics*.

LAUW, H., LIM, E. P., AND WANG, K. 2006. Bias and controversy: Beyond the statistical deviation. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*. ACM, 625–630.

LAUW, H. W., LIM, E., AND WANG, K. 2008. Bias and controversy in evaluation system. *IEEE Trans. Knowl. Data Engin. 20*, 11, 1490–1504.

LIM, E., NGUYEN, V., JINDAL, N., LIU, B., AND LAUW, H. 2010. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*. ACM, New York, NY, 939–948.

LIU, B. 2010. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*.

MCGLOHON, M., BAY, S., ANDERLE, M., STEIER, D., AND FALOUTSOS, C. 2009. Snare: A link analytic system for graph labeling and risk detection. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

MUKHERJEE, A., LIU, B., WANG, J., GLANCE, N., AND JINDAL, N. 2011. Detecting group review spam. In *Proceedings of the 20th International Conference Companion on World Wide Web*. 93–94

OTT, M., CHOI, Y., CARDIE, C., AND HANCOCK, J. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

PAGE, L., BRIN, S., MOTWANI, R., AND WINOGRAD, T. 1998. The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford Digital Library Technologies Project.

PANDIT, S., CHAU, D., WANG, S., AND FALOUTSOS, C. 2007. Netprobe: A fast and scalable system for fraud detection in online auction networks. In *Proceedings of the 16th International Conference Companion on World Wide Web*. 201–210.

PANG, B. AND LEE, L. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retrieval 2*, 1–2.

PEARL, R. AND REED, L. 1920. On the rate of growth of the population of the united states since 1790 and its mathematical representation. *Proc. Nat. Acad. Sci*.

POPESCU, A. AND ETZIONI, O. 2005. Extracting product features and opinions from reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*.

WU, G., GREENE, D., SMYTH, B., AND CUNNINGHAM, P. 2010. Distortion as a validation criterion in the identification of suspicious reviews. In *Proceedings of the 1st Workshop on Social Media Analytics*.

YIN, X., HAN, J., AND YU, P. S. 2008. Truth discovery with multiple conflicting information providers on the web. *IEEE Trans. Knowl. Data Engin. 20*, 6.

Received December 2010; revised March 2011; accepted May 2011