entropy

MDPI

Article

Online Auction Fraud Detection in Privacy-Aware Reputation Systems

Jun-Lin Lin [1,2,*] and Laksamee Khomnotai [3]

1 Department of Information Management, Yuan Ze University, Taoyuan 32003, Taiwan
2 Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan 32003, Taiwan
3 Faculty of Management Science, Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima 30000, Thailand; s1009212@mail.yzu.edu.tw
* Correspondence: jun@saturn.yzu.edu.tw; Tel.: +886-3-463-8800 (ext. 2611)

Received: 19 April 2017; Accepted: 2 July 2017; Published: 5 July 2017

Abstract: With a privacy-aware reputation system, an auction website allows the buyer in a transaction to hide his/her identity from the public for privacy protection. However, fraudsters can also take advantage of this buyer-anonymized function to hide the connections between themselves and their accomplices. Traditional fraudster detection methods become useless for detecting such fraudsters because these methods rely on accessing these connections to work effectively. To resolve this problem, we introduce two attributes to quantify the buyer-anonymized activities associated with each user and use them to reinforce the traditional methods. Experimental results on a dataset crawled from an auction website show that the proposed attributes effectively enhance the prediction accuracy for detecting fraudsters, particularly when the proportion of the buyer-anonymized activities in the dataset is large. Because many auction websites have adopted privacy-aware reputation systems, the two proposed attributes should be incorporated into their fraudster detection schemes to combat these fraudulent activities.

Keywords: online auction; privacy; anonymity; fraudster detection

1. Introduction

Rapid progress in Internet technology and electronic payment has made online auctions more prevalent and convenient [1]. In online auctions, merchandise is often purchased from a complete stranger. Therefore, building trust between potential buyers and sellers is important to ensure the success of auction websites. Most auction websites are equipped with a reputation system to evaluate the credibility of each auction account. The reputation system uses a simple scheme to compute and publish a reputation score for each auction account; this scheme is based on a collection of opinions that other auction accounts hold about the account. For example, on eBay, the seller and buyer in a transaction can give each other a positive, negative, or neutral rating. Intuitively, sellers with more positive ratings and fewer negative ratings are more reputable and are likely to draw more sales.

The lucrative opportunity associated with a favorable online reputation attracts both honest and fraudulent sellers to pursue high reputation scores. Honest sellers achieve higher reputation scores by providing improved services (e.g., higher quality products, lower prices, and faster response) to their buyers. However, fraudulent sellers use a deceitful scheme, known as inflated reputation fraud [2], to boost their reputation scores. In this scheme, fraudulent sellers perform many transactions for low-priced merchandise within a group of collusive accounts to boost the positive ratings of the group's members [3]. Because the cost of conducting the scheme is low, inflated reputation fraud is prevalent in online auctions. In this paper, we focus on detecting inflated reputation fraud. Notably, inflated reputation fraud is often the first step toward other fraudulent activities, such as selling counterfeit products or failing to deliver products.

Most recent approaches for detecting inflated reputation fraud are based on Social Network Analysis (SNA) [1–13]. These SNA-based approaches construct a social network of buyers and sellers based on their past transactions, and then detect fraudsters by finding cohesive groups in the network. However, some auction websites adopt a privacy-aware reputation system that enables buyers to hide their links to sellers. Fraudsters can also use this function to hide the links within their collusive group, making them hard to detect with traditional SNA-based approaches.

This paper presents a solution for detecting inflated reputation fraud in auction websites that use a privacy-aware reputation system. To the best of our knowledge, all SNA-based approaches in the literature use either synthetic datasets [6] or real datasets crawled from auction websites. Thus, these approaches have no access to the hidden links between buyers and sellers. We propose two privacy-related attributes to quantify the proportion of hidden links associated with each account, and show that the addition of these two attributes enhances the prediction accuracy for detecting fraudsters.

The remainder of this paper is organized as follows. The second section reviews previous work on the reputation systems in online auctions and the existing methods for detecting inflated reputation fraud. The third section describes the privacy-aware reputation systems on auction websites and proposes two privacy-related attributes associated with each user. The fourth section describes the dataset used in this study. The fifth section presents a performance study to evaluate the effectiveness of using the proposed privacy-related attributes to detect fraudsters. Finally, a discussion and concluding remarks are given in the sixth and seventh sections, respectively.

2. Related Work

2.1. Reputation Systems in Online Auction

Two factors are crucial to the success of an online auction website [14]. The first is how easily buyers can find sellers. The second is the trust that the website facilitates through its reputation system. The reputation systems in online auctions are essentially recommendation systems. Both parties in a transaction can give each other a positive, negative, or neutral rating, and the reputation system calculates a reputation score for each user based on all the ratings that the user has received from his/her past transactions, and the reputation score is available to the public. A third party can also access detailed information about each rating that a user has received so far. Detailed information is provided for the following aspects of a transaction:

- Date and time of the transaction.
- Seller and buyer of the transaction. This information can be used to construct a social network of users (see Section 2.2).
- Merchandise description.
- The rating (positive, negative or neutral) that the user received from his/her counterpart in the transaction.
- Textual feedback comment.

Such a reputation system builds trust in online auctions that lack typical human interaction [15], forming a large-scale, word-of-mouth network among users [16]. Reputable sellers can not only gain trust but also generate price premiums from potential buyers [17–19]. By contrast, a high proportion of negative ratings reduces the sales price [20]. A high proportion of neutral ratings impairs sales for sellers with high proportions of positive ratings, but facilitates sales for sellers with high proportions of negative ratings [21]. In case a negative rating is received, textual feedback comments and reactions are important for rebuilding trust [22]. The information that sellers provide to buyers can also affect the sellers' reputation [23].

2.2. Constructing Social Networks from Reputation System

Based on the transaction history, a transaction network can be constructed, in which each node indicates an auction account and each link depicts a transaction between two auction accounts. Although the transaction network provides a complete view of the social interactions among auction accounts, a single factor prevents the use of the transaction network: on auction websites, the complete transaction history is not available to the public. Notably, only after at least one party in a transaction rates his/her counterpart does the transaction appear in the reputation system, which is open to the public. However, providing a rating after each transaction is not mandatory.

By contrast, a rating network is constructed based on the rating history of auction accounts. On many auction websites (e.g., eBay, Taobao, Ruten, and Yahoo! Kimo), the rating history is accessible to the public. Similar to a transaction network, each node in a rating network indicates an auction account, but each link depicts a rating relationship between two connected nodes. Because providing a rating after each transaction is not mandatory, each link in the rating network corresponds to a link in the transaction network, but not vice versa. Because the transaction history is not available to the public, and inflated reputation fraud requires the accumulation of positive ratings, most previous studies have adopted the rating network as a suitable surrogate for the transaction network [4,8,10,11].

2.3. Methods for Detecting Fraudsters in Online Auction

Previous techniques for detecting inflated reputation fraud used mostly user-level features such as the median, mean, sum, or standard deviation of the merchandise prices that a user sold or bought over a period [5,24]. The reputation systems on most auction websites also play a significant role in fraudster detection. Studies have shown that recent negative ratings are useful for predicting future fraud, and that experienced buyers can use the reputation system to avoid potential fraudulent auctions [25]. However, this approach does not fully utilize the information provided by the reputation system to uncover the interaction among users, who may still be deceived by fraudsters [26].

More recent approaches incorporate network-level features to combat inflated reputation fraud. Because inflated reputation fraud requires a collusive group of users to give each other positive ratings, a cohesive relation occurs within the collusive group. Many SNA-based approaches can identify cohesive subgroups in a network (see Table 1), and some of these approaches (e.g., *k*-core and *k*-plex) have been applied to detect collusive groups of fraudsters in a rating network [2,4,11]. In addition to basic features, such as degree and betweenness [11], more sophisticated features (e.g., neighbor diversity, neighbor driven attributes, credibility, and density) have been proposed for fraudster detection (see Table 2).

Table 1. Subgroups in Social Network Analysis (SNA).

Subgroups	Description
clique	A maximal fully connected subnetwork of a network *G*
n-clique	A maximal subnetwork of a network *G* in which every pair of nodes is connected by a path in *G* of length *n* or less
n-clan	An *n*-clique which has a diameter less than or equal to *n*
k-core	A maximal connected subnetwork of a network *G* in which each node is connected to at least *k* other nodes in the subnetwork
k-plex	A maximal subnetwork of a network *G* in which each node is connected to at least *n*-*k* other nodes in the subnetwork, where *n* is the number of nodes in the subnetwork.

Table 2. SNA-based network-level features for detecting fraudsters in online auctions.

Attribute	References
k-core	[2,4,11]
core/periphery ratio	[2]
center-weight	[4]
credibility	[3]
density	[3,7]
degree (in-degree, out-degree)	[11]
normalized betweenness	[11]
k-plex =2 and (size = 5 or 6 or 7)	[11]
n-clique = 1 and (size = 3 or 4 or 5)	[11]
neighbor diversity	[13]
neighbor driven attributes	[12]

3. Privacy-Aware Reputation System and Privacy-Related Attributes

3.1. Privacy-Aware Reputation System

Many auction websites have adopted privacy-aware reputation systems, where the buyer in a transaction can decide whether to hide his/her identity from the public. Ruten, Yahoo! Kimo Auction, and eBay adopted privacy-aware reputation systems in 2008, 2009, and 2013, respectively.

On Yahoo! Kimo Auction, after winning the bid of the merchandise, within 60 days, the buyer has the option of hiding from the public the information about both the seller and the merchandise in the rating that the buyer receives from the seller. By doing so, in the rating that the seller receives from the buyer, the information about the buyer is also hidden from the public. A similar buyer-anonymized function is also available in the reputation system of Ruten, except that transactions can be set to the hidden mode within 6 months, instead of 60 days.

If a buyer chooses to hide his/her identity in a transaction, the transaction is referred to as an anonymous transaction. Since Ruten adopted the privacy-aware reputation system in 2008, a substantial proportion of the transactions on Ruten have been anonymous. A random sample of 190,782 transactions on Ruten between 2008 and 2011 across 24 categories of merchandise showed that 11.38% of the transactions were anonymized [27]. The proportions of anonymous transactions across different categories of merchandise varied from 0.87% in the Books and Stationery category to 27.45% in the Women's Intimates and Sleepwear category and 28.57% in the Real Estate and Specialty Services category. The results reflect that buyers often demand privacy when purchasing personal or intimate products. Although the original intention of anonymous transactions is to protect customers' privacy, fraudsters can abuse anonymous transactions to hide their criminal activities from the public (see Section 5.2.1).

Once a transaction is anonymized, the following information that originally appears in the ratings of the buyer and the seller is no longer available to the public:

- In the rating that the buyer receives from the seller, the seller ID and information about the merchandise are hidden.
- In the rating that the seller receives from the buyer, the buyer ID is hidden.

Although the rating (i.e., positive, neutral, or negative) and textual comments of an anonymous transaction are still public, third parties do not know who gave the rating. Consequently, a third party cannot construct the link between the buyer and the seller of an anonymous transaction. Thus, the privacy of the buyer is protected.

3.2. Privacy-Related Attributes

Although privacy-aware reputation systems provide more shopping privacy to buyers, fraudsters can also exploit anonymous transactions to hide the links to collusive auction accounts. Consequently, if a third party crawls an auction website to build the rating network of auction accounts (see Section 2.2), the transactions within a collusive group of auction accounts may be hidden and

thus cannot be reconstructed. Because many fraud detection approaches [1–13] employ the rating network to detect fraudsters, anonymous transactions in the reputation system may render these approaches unfeasible to detect fraudsters who take advantage of anonymous transactions. To the best of our knowledge, most fraud detection approaches are based on datasets crawled from auction websites instead of datasets directly provided by the auction websites. Therefore, fraudsters exploiting anonymous transactions are likely to be overlooked.

For example, previous studies have shown that a fraudster requires accomplices to provide positive ratings; thus, intensive transactions must occur between them. Consequently, they are likely to appear in the 2-core subgraphs of the rating network [2,11]. However, if the fraudster and accomplices use anonymous transactions to hide the buyers' identities in their transactions, then the links among them may not appear in rating networks constructed by third parties. Consequently, the members of the collusive group may not belong to the same 2-core subgraphs.

To overcome this problem, we propose two privacy-related attributes to capture the proportion of anonymous activities associated with each auction account. Let n denote the number of positive ratings that a user has received. Because a transaction can be either anonymous or non-anonymous, we can decompose n into n_a and n_n, where n_a is the number of positive ratings resulting from anonymous transactions, and n_n is the number of positive ratings resulting from non-anonymous transactions. Depending on the user's role in a transaction (buyer or seller), a rating can be given as a seller or as a buyer. We can further decompose n_a into two parts: the number of anonymous positive ratings given by sellers (denoted as n_{as}) and the number of anonymous positive ratings given by buyers (denoted as n_{ab}). That is,

$$n = n_a + n_n = (n_{ab} + n_{as}) + n_n. \quad (1)$$

In this paper, we use the number of anonymous positive ratings that an account has received from its buyers (i.e., n_{ab}) as the first privacy-related attribute. Notably, because only the buyer in a transaction has the right to decide whether to anonymize the transaction, an account with a high n_{ab} is likely to belong to a fraudster who uses a large number of anonymized accomplices to boost its rating.

The second attribute is the *anonymous ratio* (denoted by R_a), which is defined as the number of anonymous positive ratings divided by the number of all positive ratings that an account has received or given to other accounts. Intuitively, an account with a high R_a is likely to belong to a fraudster. Section 5 describes an experiment that applies both R_a and n_{ab} to detect fraudsters in a real world dataset (see Section 4) by using decision trees and artificial neural networks.

4. Data Collection and Dataset Preparation

A dataset collected from Ruten [28] was used in this study. A subset of this dataset was also used in our previous work [12,13]. The data collection process proceeded in a level-by-level manner [4,8,10,11,29] and is explained as follows:

Step 1. Collecting accounts (first level). Ruten regularly releases a list of recently suspended accounts, together with the reasons for the suspension. Our data collection process began with the collection of all 9168 accounts suspended by Ruten in July 2013. Because some of these accounts were not fraud-related (e.g., selling alcohol or prescribed medicine), we manually checked the 9168 accounts and retained only the 3101 whose suspension reasons were fraud-related, such as evaluation hype, selling counterfeit products, fake bidding, and failure to deliver products. Furthermore, because inflated reputation fraud works by accumulating positive ratings from accomplices, we removed the accounts that had not yet received any ratings. The remaining 1064 accounts were denoted as L_1 accounts. Ruten altered the status of one L_1 account to normal in October 2013. Therefore, the L_1 accounts included 1063 fraudster accounts and 1 non-fraudster account. Notably, for 132 of the 1063 L_1 fraudster accounts, all of the positive ratings they received were anonymous. Furthermore, 121 of these 132 accounts had an anonymous ratio R_a of 1, indicating that all of the positive ratings they had received and given to other accounts were anonymous.

Step 2. Collecting accounts (second level). We then collected all of the non-anonymous accounts that had received ratings from or given ratings to any L_1 account. Consequently, 3475 new accounts were discovered and were denoted as L_2 accounts. Because each L_2 account was linked to at least one L_1 account and all L_1 accounts were not anonymous, each L_2 account had an anonymous ratio of <1. Among the 3475 L_2 accounts, 149 of them were suspended by Ruten due to fraudulent activities and were treated as fraudster accounts in this experiment. Table 3 shows the numbers of fraudster and non-fraudster accounts in the L_1 and L_2 accounts.

Table 3. Numbers of fraudster and non-fraudster accounts in the L_1 and L_2 accounts.

Level	Fraudsters	Non-Fraudsters	Total
L_1	1063	1	1064
L_2	149	3326	3475
Total	1212 (26.7%)	3327 (73.3%)	4539

Step 3. Collecting accounts (third level). To reveal the accounts that were involved in transactions with these L_2 accounts, we further collected all non-anonymous accounts that had received ratings from or given ratings to any of the 3475 L_2 accounts. In this step, 233,169 new accounts were discovered and were denoted as L_3 accounts. On average, each L_2 account transacted with 233169/3475 = 67 L_3 accounts. By contrast, on average, each L_1 account transacted with only 3475/1064 = 3.2 L_2 accounts. Notably, non-fraudster accounts received positive ratings from many accounts, whereas fraudster accounts received positive ratings mostly from their accomplices. In Table 1, the proportion of fraudster accounts was much higher in L_1 accounts (1063/1064) than in L_2 accounts (149/3475). Therefore, the ratio between the numbers of L_2 accounts and L_1 accounts (approximately 3.2) was much lower than the ratio between the numbers of L_3 accounts and L_2 accounts (approximately 67).

Step 4. Constructing the social network. We constructed a social network comprising all of the L_1, L_2, and L_3 accounts, where each node in the network represents an account. If an account had given at least one positive rating to another account before 31 July 2013, then the nodes representing the two accounts were connected through a link in the social network. The resulting network contained 237,708 (= 1064 + 3475 + 233,169) nodes and 348,259 links. Notably, 121 nodes in the social network were not connected to any other node. They represented the 121 L_1 accounts with an anonymous ratio of 1, as described in Step 1. Notably, a user can be a buyer, a seller, or both in the network. Among the L_1 accounts, 96 were buyers, 884 were sellers, and 84 were both. Among the L_2 accounts, 2561 were buyers, 58 were sellers, and 856 were both.

Step 5. Calculating SNA-related attributes. Based on the social network created in the previous step, we calculated several SNA-related attributes (shown in Table 4) for the nodes representing the L_1 or L_2 accounts to build a dataset for this performance study. We did not include the L_3 accounts in the dataset, because the social network did not include all of the accounts that had received ratings from or given ratings to the L_3 accounts. Therefore, the resulting dataset contained 4539 (=1064 + 3475) records (Table 3). The dataset is available at the supplementary of this paper.

Table 4. SNA-related attributes.

Notation	Definition
R_a	Anonymous ratio. See Section 3.2.
n_{ab}	Anonymous count. See Section 3.2.
k-core	The largest *k* value of all *k*-core components that the node resides. See [2,4].
CW	Center weight. See [4].
nBetweenness	Normalized betweenness. See [11].
binary_*k*-core	Binary attributes indicating whether the node is in a *k*-core component. In this study, k = 2 to 6 are used, as in [11].
2-plex_and_size = *s*	Binary attributes indicating whether the node is in a 2-plex component with size = *s*. In this study, s = 5 to 7 are used, as in [11].
ND_r	Neighbor diversity on the number of received ratings [13].
NDA_{mean}	Mean of the numbers of received ratings of the node's neighbors [12].

5. Performance Study

5.1. Experimental Design

The experiment was designed from two perspectives: attributes and datasets. Concerning the attributes used to build a classifier for detecting fraudsters, our goal was to evaluate whether the addition of the two privacy-related attributes (i.e., R_a and n_{ab}) can improve the performance of the existing sets of attributes used in previous work. Five sets of attributes were considered in this study. The first set contained only one attribute, *k*-core [2], and the second set contained two attributes, *k*-core and center weight (CW) [4]. The third set (denoted as S_9) contained eight binary attributes (binary_*k*-core for $k = 2$ to 6, and 2-plex_and_size = s for $s = 5$ to 7, shown in Table 4) and one numeric attribute (normalized betweenness) [11]. The fourth set contained only one attribute, ND_r (neighbor diversity on the number of received ratings [13]). The fifth set contained only one attribute, NDA_{mean} (the mean of the numbers of received ratings of the node's neighbors [12]). In this performance study, we tested these five sets of attributes and then evaluated whether their performance can be improved by adding R_a and n_{ab}.

Regarding the datasets, our goal was to evaluate whether a given approach can detect fraudsters effectively among users with various proportions of anonymous transactions. Let D_{100} denote the dataset collected as described in Section 4; we generated three subsets of D_{100} (D_0, D_{0+}, and D_{15} shown in Table 5) based on the anonymous ratio R_a. Dataset D_0 contained the accounts that have never engaged in any anonymous transactions ($R_a = 0$), and D_{0+} contained accounts that have engaged in at least one anonymous transaction ($R_a > 0$). Thus, $D_0 \cap D_{0+} = \varphi$ and $D_0 \cup D_{0+} = D_{100}$. Dataset D_{15} contained the top 15% of accounts based on R_a. Thus, $D_{15} \subset D_{0+} \subset D_{100}$. Ordering these datasets by the proportion of anonymous transactions gives $D_{15} > D_{0+} > D_{100} > D_0$, and by testing these datasets, we could verify whether a given approach can still perform effectively if anonymous transactions become prevalent. Notably, the last column of Table 5, baseline accuracy, represents the prediction accuracy of always predicting that an account is a non-fraudster (or fraudster) account if non-fraudster (or fraudster) accounts mainly comprise the dataset.

Table 5. Datasets.

Notation	Description	# of Fraudsters	# of Non-Fraudsters	Baseline Accuracy (%)
D_{100}	All collected data, described in Section 4.	1212	3327	73.2981
D_0	Subset of D_{100} with $R_a = 0$.	705	1966	73.6054
D_{0+}	Subset of D_{100} with $R_a > 0$.	507	1361	72.8587
D_{15}	The top 15% of D_{100} with the highest R_a.	385	296	56.5345

In this study, we divided the experiment into four tests, and each test used a dataset from Table 5. In each test, we used various combinations of attributes to evaluate whether adding the two proposed privacy-related attributes improves the prediction accuracy. Two classification algorithms from Weka [30], the J48 decision tree and the artificial neural network (ANN), were used in this study to conduct 10-fold cross validation. The experiment adopted the default parameter settings of both algorithms in Weka.

5.2. Experiment Results

5.2.1. Results from Dataset D_{100}

Dataset D_{100} is the dataset collected following the steps described in Section 4. It includes all of the accounts collected, regardless of their anonymous ratio. As indicated in Table 5, D_{100} contains 1212 fraudster accounts and 3327 non-fraudster accounts, yielding a baseline accuracy of 73.2981%.

Tables 6 and 7, respectively, show the performance results of J48 and ANN with D_{100}. When using the nine attributes in S_9, the addition of the two privacy-related attributes, R_a and n_{ab}, improved the prediction accuracy from 75.8537% to 82.5072% for J48 and from 75.0606% to 79.4228% for ANN.

Recall and precision were also significantly improved. Similar results were observed using *k*-core, *k*-core and CW, or ND_r. When using NDA_{mean}, the addition of R_a and n_{ab} improved the prediction accuracy and precision but slightly reduced recall. Notably, among the 4539 accounts in dataset D_{100}, 1868 accounts (approximately 41%) had at least one anonymous rating (i.e., $R_a > 0$). By adding R_a and n_{ab}, the accounts with $R_a > 0$ could be more effectively predicted, which therefore, results in an improved prediction accuracy for both J48 and ANN. However, 59% of the accounts in D_{100} still had $R_a = 0$. Thus, using only R_a and n_{ab} resulted in poor recall, as shown in the last rows of Tables 6 and 7.

For both J48 and ANN, the addition of R_a and n_{ab} reduced the number of false negatives, except when using the attribute NDA_{mean}. However, J48 and ANN produced slightly different results for false positives. For J48, the addition of R_a and n_{ab} always reduced the number of false positives; for ANN, the addition of R_a and n_{ab} occasionally increased the number of false positives.

Table 6. J48 performance with dataset D_{100}.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	75.8537	0.2525	0.6169	190	906
S_9 & (R_a & n_{ab})	82.5072	0.4586	0.8029	137	657
k-core	77.5281	0.3193	0.6649	195	825
k-core & (R_a & n_{ab})	82.2868	0.4381	0.8119	123	681
k-core & CW	86.9354	0.7120	0.7796	244	349
k-core & CW & (R_a & n_{ab})	89.2928	0.7360	0.8431	166	320
ND_r	84.6883	0.8226	0.6750	480	215
ND_r & (R_a & n_{ab})	86.7812	0.8234	0.7211	386	214
NDA_{mean}	88.6980	0.8581	0.7531	341	172
NDA_{mean} & (R_a & n_{ab})	90.3503	0.8267	0.8146	228	210
R_a & n_{ab}	80.2159	0.2855	0.9153	32	866

Table 7. Artificial Neural Network (ANN) performance with dataset D_{100}.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	75.0606	0.1988	0.5995	161	971
S_9 & (R_a & n_{ab})	79.4228	0.3977	0.7026	204	730
k-core	74.4437	0.2335	0.5506	231	929
k-core & (R_a & n_{ab})	78.6737	0.3738	0.6843	209	759
k-core & CW	85.6356	0.5066	0.9192	54	598
k-core & CW & (R_a & n_{ab})	87.0456	0.5611	0.9239	56	532
ND_r	83.8070	0.8069	0.6613	501	234
ND_r & (R_a & n_{ab})	84.2476	0.8119	0.6689	487	228
NDA_{mean}	84.4679	0.7995	0.6771	462	243
NDA_{mean} & (R_a & n_{ab})	84.6001	0.7904	0.6828	445	254
R_a & n_{ab}	78.5636	0.2294	0.8770	39	934

To investigate whether fraudsters use anonymous transactions more often than non-fraudsters do, we compared the R_a distribution of the 1212 fraudster accounts with that of the 3327 non-fraudster accounts in dataset D_{100}. Figure 1 shows the proportion of fraudster (or non-fraudster) accounts with a R_a of more than or equal to a certain threshold among all 1212 fraudster (or 3327 non-fraudster) accounts. Among the 1212 fraudster accounts, 507 (approximately 41.83%) accounts satisfied $R_a > 0$; 429 (approximately 35.4%) satisfied $R_a \geq 0.1$; and 121 (approximately 9.98%) satisfied $R_a = 1$. In each R_a range in Figure 1, we observed a significant proportion of fraudster accounts. By contrast, among the 3327 non-fraudster accounts, 1361 (approximately 40.91%) accounts satisfied $R_a > 0$; 555 (approximately 16.68%) satisfied $R_a \geq 0.1$; and none of the non-fraudster accounts satisfied $R_a = 1$. As the threshold of R_a increased, the proportion of non-fraudster accounts decreased more quickly than the proportion of fraudster accounts did. Overall, for the same threshold of R_a in Figure 1, the proportion of fraudster accounts was always larger than the proportion of non-fraudster accounts. The result indicated that fraudsters use anonymous transactions more often than non-fraudsters do.

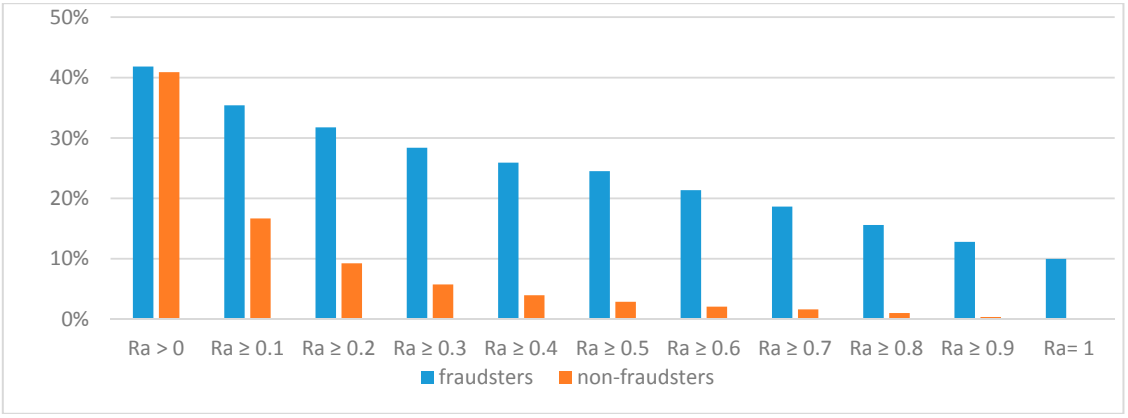

Figure 1. Proportions of fraudster and non-fraudster accounts w.r.t. R_a in dataset D_{100}.

5.2.2. Results from Dataset D_0

Dataset D_0 contained all accounts in D_{100} where $R_a = 0$. This dataset comprised 705 fraudster accounts and 1966 non-fraudster accounts, yielding a baseline accuracy of 73.6054% (Table 5). Note that if $R_a = 0$, then $n_{ab} = 0$. Because R_a and n_{ab} are 0 for all accounts in D_0, adding R_a and n_{ab} to the classification algorithms did not improve performance. With or without R_a and n_{ab}, the results were the same for J48 (Table 8). The decision tree algorithm selects the most discriminating attribute to split the tree at each step. However, because R_a and n_{ab} are 0 throughout the dataset, they are the least discriminating attributes. For ANN, the results were similar with or without R_a and n_{ab} (Table 9). Using only R_a and n_{ab} for dataset D_0 predicted that all accounts were non-fraudster accounts (i.e., the majority class), resulting in a baseline accuracy of 73.6054% and 0 recall, as shown in the last rows of Tables 8 and 9.

Table 8. J48 performance with dataset D_0.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	76.8252	0.2411	0.6693	84	535
S_9 & (R_a & n_{ab})	76.8252	0.2411	0.6693	84	535
k-core	76.3010	0.2199	0.6513	83	550
k-core & (R_a & n_{ab})	76.3010	0.2199	0.6513	83	550
k-core & CW	85.6608	0.6553	0.7674	140	243
k-core & CW & (R_a & n_{ab})	85.6608	0.6553	0.7674	140	243
ND_r	82.2538	0.8511	0.6192	369	105
ND_r & (R_a & n_{ab})	82.2538	0.8511	0.6192	369	105
NDA_{mean}	88.5062	0.8652	0.7421	212	95
NDA_{mean} & (R_a & n_{ab})	88.5062	0.8652	0.7421	212	95
R_a & n_{ab}	73.6054	0.0000	N/A	0	705

Table 9. ANN performance with dataset D_0.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	76.5631	0.2284	0.6626	82	544
S_9 & (R_a & n_{ab})	76.5631	0.2284	0.6626	82	544
k-core	75.2153	0.2667	0.5646	145	517
k-core & (R_a & n_{ab})	75.2153	0.2667	0.5646	145	517
k-core & CW	83.3396	0.4312	0.8736	44	401
k-core & CW & (R_a & n_{ab})	83.4519	0.4567	0.8451	59	383
ND_r	81.1307	0.7887	0.6103	355	149
ND_r & (R_a & n_{ab})	81.5799	0.8000	0.6164	351	141
NDA_{mean}	85.0618	0.8922	0.6607	323	76
NDA_{mean} & (R_a & n_{ab})	84.6499	0.8539	0.6623	307	103
R_a & n_{ab}	73.6054	0.0000	N/A	0	705

5.2.3. Results from Dataset D_{0+}

Dataset D_{0+} contained all accounts where $R_a > 0$ in D_{100}. This dataset comprised 507 fraudster accounts and 1361 non-fraudster accounts, yielding a baseline accuracy of 72.8587% (Table 5). As shown in Tables 10 and 11, with the addition of R_a and n_{ab}, the prediction accuracy improved, and in most cases, precision and recall also improved. Because $R_a > 0$ for all accounts in D_{0+}, the addition of R_a and n_{ab} improved performance. As shown in the last rows of Tables 10 and 11, using only R_a and n_{ab} on D_{0+} yielded accuracies of 89.6146% for J48 and 86.0278% for ANN, which is more than a 7.1% improvement over the corresponding baseline accuracy of 72.8587%. By contrast, using only R_a and n_{ab} on D_{100} yielded an accuracies of 80.2159% for J48 and 78.5636% for ANN, which is a less than 5.3% improvement over the corresponding baseline accuracy of 73.2981% (see Tables 5–7). Because the proportion of anonymous transactions was larger in D_{0+} than in D_{100}, the impact of adding R_a and n_{ab} to the classification accuracy was also larger in D_{0+} than in D_{100}.

Table 10. J48 performance with dataset D_{0+}.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	74.3041	0.1499	0.6080	49	431
S_9 & (R_a & n_{ab})	91.1670	0.7416	0.9171	34	131
k-core	81.0493	0.4813	0.7284	91	263
k-core & (R_a & n_{ab})	91.0600	0.7456	0.9087	38	129
k-core & CW	89.4004	0.6371	0.9585	14	184
k-core & CW & (R_a & n_{ab})	94.7537	0.8501	0.9514	22	76
ND_r	87.6338	0.7929	0.7614	126	105
ND_r & (R_a & n_{ab})	93.4154	0.7949	0.9550	19	104
NDA_{mean}	90.4176	0.8343	0.8166	95	84
NDA_{mean} & (R_a & n_{ab})	93.6831	0.8067	0.9534	20	98
R_a & n_{ab}	89.6146	0.6627	0.9359	23	171

Table 11. ANN performance with dataset D_{0+}.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	73.6081	0.1598	0.5473	67	426
S_9 & (R_a & n_{ab})	86.3490	0.6746	0.7917	90	165
k-core	73.2334	0.1578	0.5229	73	427
k-core & (R_a & n_{ab})	87.2591	0.6844	0.8165	78	160
k-core & CW	88.2762	0.6055	0.9417	19	200
k-core & CW & (R_a & n_{ab})	92.4518	0.8185	0.8944	49	92
ND_r	87.4732	0.8698	0.7241	168	66
ND_r & (R_a & n_{ab})	88.8116	0.8619	0.7587	139	70
NDA_{mean}	89.2398	0.9073	0.7492	154	47
NDA_{mean} & (R_a & n_{ab})	89.9893	0.8462	0.7974	109	78
R_a & n_{ab}	86.0278	0.5740	0.8661	45	216

5.2.4. Results from Dataset D_{15}

Dataset D_{15} contained the top 15% accounts in D_{100} based on R_a, representing a dataset with a large proportion of anonymous transactions. The smallest anonymous ratio of the accounts in D_{15} was 0.2. As indicated in Table 5, dataset D_{15} contained 385 fraudster accounts and 296 non-fraudster accounts, yielding a baseline accuracy of 56.5354%. As shown in Tables 12 and 13, the addition of R_a and n_{ab} improved both the prediction accuracy and precision, but in some cases, recall was decreased.

Table 12. J48 performance with dataset D_{15}.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	56.2408	0.9816	0.5617	291	7
S_9 & (R_a & n_{ab})	88.8399	0.8579	0.9449	20	56
k-core	63.8767	0.4500	0.8221	37	209
k-core & (R_a & n_{ab})	89.1336	0.8597	0.9430	20	54
k-core & CW	83.2599	0.7868	0.9006	33	81
k-core & CW & (R_a & n_{ab})	92.3642	0.8935	0.9690	11	41
ND_r	76.9457	0.8597	0.7627	103	54
ND_r & (R_a & n_{ab})	90.8957	0.8571	0.9792	7	55
NDA_{mean}	84.2878	0.9403	0.8117	84	23
NDA_{mean} & (R_a & n_{ab})	88.8399	0.8545	0.9427	20	56
R_a and n_{ab}	89.7210	0.8649	0.9487	18	52

Table 13. ANN performance with dataset D_{15}.

Attributes	Accuracy (%)	Recall	Precision	False Positives	False Negatives
S_9	53.7445	0.7842	0.5612	233	82
S_9 & (R_a & n_{ab})	79.1483	0.8026	0.8240	66	76
k-core	61.5272	0.4711	0.7458	61	201
k-core & (R_a & n_{ab})	79.5888	0.8442	0.8045	79	60
k-core & CW	74.7430	0.6605	0.8537	43	129
k-core & CW & (R_a & n_{ab})	87.8120	0.8727	0.9081	34	49
ND_r	77.6799	0.9169	0.7463	120	32
ND_r & (R_a & n_{ab})	78.7078	0.8156	0.8093	74	71
NDA_{mean}	82.5257	0.9714	0.7759	108	11
NDA_{mean} & (R_a & n_{ab})	86.9310	0.9117	0.8645	55	34
R_a and n_{ab}	77.9736	0.8052	0.8052	75	75

6. Discussion

The results in Tables 5–13 showed that the addition of R_a and n_{ab} improved the prediction accuracy. In most cases, the addition of R_a and n_{ab} reduced either the number of false positives, the number of false negatives, or both. Except in the experiment with dataset D_0, the addition of R_a and n_{ab} to the attribute NDA_{mean} always reduced false positives but increased false negatives. However, the number of reduced false positives was greater than the number of increased false negatives. Therefore, the prediction accuracy was improved.

To evaluate the performance improvement of adding R_a and n_{ab}, we calculated the difference in the prediction accuracy of datasets evaluated with and without adding R_a and n_{ab} (Table 14). For all attributes in Table 14, the ordering of accuracy improvement was $D_{15} > D_{0+} > D_{100}$. That is, the addition of R_a and n_{ab} had a stronger positive impact on accuracy for datasets with higher percentages of anonymous transactions. Therefore, as using anonymous transactions to hide fraudulent activities becomes more prevalent, the importance of using the privacy-related attributes to detect fraudsters also increases.

Table 14. Percentage of accuracy improvement with the addition of R_a & n_{ab}.

Attributes	J48			ANN		
	D_{100}	D_{0+}	D_{15}	D_{100}	D_{0+}	D_{15}
S_9	6.6535	16.8629	32.5991	4.3622	12.7409	25.4038
k-core	4.7587	10.0107	25.2569	4.23	14.0257	18.0616
k-core & CW	2.3574	5.3533	9.1043	1.41	4.1756	13.069
ND_r	2.0929	5.7816	13.95	0.4406	1.3384	1.0279
NDA_{mean}	1.6523	3.2655	4.5521	0.1322	0.7495	4.4053

In Table 15, the baseline accuracy and the accuracy of using only R_a and n_{ab} were copied from the last column of Table 5 and the last rows of Tables 6–13, respectively. The improvement column was calculated as the accuracy of using only R_a and n_{ab} subtracted from the corresponding baseline accuracy. Notably, ordering the datasets by the accuracy improvement over the baseline accuracy was the same

as ordering them by their proportions of anonymous transactions: $D_{15} > D_{0+} > D_{100} > D_0$. Thus, the importance of R_a and n_{ab} increased with the proportion of anonymous transactions in the dataset.

Table 15. Prediction accuracy (%) of baseline, using only R_a & n_{ab}, and improvement.

Datasets	Baseline Accuracy	J48		ANN	
		Accuracy	Improvement	Accuracy	Improvement
D_{15}	56.5345	89.7210	33.1865	77.9736	21.4391
D_{0+}	72.8587	89.6146	16.7559	86.0278	13.1691
D_{100}	73.2981	80.2159	6.9178	78.5636	5.2655
D_0	73.6054	73.6054	0	73.6054	0

Because dataset D_{0+} contained all accounts where $R_a > 0$ in D_{100}, we chose dataset D_{0+} to evaluate how R_a and n_{ab} are distributed among fraudster and non-fraudster accounts (Table 16). Although the mean value of R_a was smaller for non-fraudster accounts than for fraudster accounts, the reverse was true for the standard deviation of R_a. Similar results were also found for n_{ab}.

Table 16. Mean and standard deviation of R_a and n_{ab} in dataset D_{0+}.

-	R_a		n_{ab}	
	Mean	Stdev	Mean	Stdev
Fraudsters	0.567901819	3.069033531	0.361071348	13.81824708
Non-fraudsters	0.147264717	18.50183688	0.190070867	88.81797154

To indicate how R_a affects fraudster distribution, we calculated the proportions of fraudster and non-fraudster accounts for several subsets of the dataset D_{100}, where each subset only contained the accounts where R_a was more than or equal to a certain threshold (Table 17). The proportions of fraudster accounts in the datasets where $R_a \geq 0$ (i.e., D_{100}) and $R_a > 0$ (i.e., D_{0+}) were 26.7% and 27.14%, respectively; the difference was only 0.44%. However, when the threshold of R_a was increased to $\geq$0.1, the proportion of fraudster accounts in the dataset became 43.6%, a 16.46% increment over the dataset with $R_a > 0$. The proportion of fraudster accounts in the resulting dataset increased with the threshold. Finally, when R_a reached its maximal value of 1, the resulting dataset contained only fraudster accounts. Thus, the fraudster distribution reflected that an account with a higher R_a was more likely to be a fraudster account.

Table 17. Fraudster distribution in datasets with R_a greater than or equal to a certain threshold.

Subset of D_{100}	Fraudsters		Non-Fraudsters	
	Count	Percentage	Count	Percentage
$R_a \geq 0$	1212	26.70%	3327	73.30%
$R_a > 0$	507	27.14%	1361	72.86%
$R_a \geq 0.1$	429	43.60%	555	56.40%
$R_a \geq 0.2$	385	55.56%	308	44.44%
$R_a \geq 0.3$	344	64.30%	191	35.70%
$R_a \geq 0.4$	314	70.40%	132	29.60%
$R_a \geq 0.5$	297	75.57%	96	24.43%
$R_a \geq 0.6$	259	78.96%	69	21.04%
$R_a \geq 0.7$	226	80.71%	54	19.29%
$R_a \geq 0.8$	189	84.75%	34	15.25%
$R_a \geq 0.9$	155	92.81%	12	7.19%
$R_a \geq 1.0$	121	100.00%	0	0.00%

7. Conclusions

A privacy-aware reputation system in online auctions offers the same service to everyone and does not discriminate between honest and fraudulent users. Although it protects the privacy of each

user, it can also be misused to cover criminal activities by enabling a fraudster to hide the fact that all or most of his/her positive ratings are given by accomplices. Without considering this fact, the scores provided by the reputation system can be misleading.

In this paper, we proposed two privacy-related attributes to quantify the proportion of anonymous ratings that a user received. We showed that both attributes improved the performance of the fraudster detection method. Future work should address how to calculate the reputation score to avoid an inflated reputation. The reputation system can employ a more sophisticated method to calculate the reputation score, for example, by assigning lower and higher weights to anonymous and non-anonymous ratings, respectively. Because the reputation score is available to all users to evaluate the trustworthiness of a buyer in real time, its impact can be quite substantial.

On some auction websites (e.g., eBay), anonymity is allowed, not only for giving ratings, but also for placing bids. This anonymous bidding function can also be abused by fraudsters to protect shill bidders, who bid on items with the intent to artificially raise their prices. Previous work on shill bidding detection includes deriving features from the bidding history to calculate the likelihood of a user participating in shill bidding [31], introducing a formal model checking approach to detect shill bidding [32], investigating the relationship between final auction prices and shill activities [33], and so on. Applying privacy-related features similar to the anonymous ratio to detect shill bidding is a potential area for further study.

Supplementary Materials: The following are available online at www.mdpi.com/1099-4300/19/7/338/s1.

Acknowledgments: This research is supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant 105-2632-H-155-022.

Author Contributions: Both authors contributed to the conception and design of the study, the collection and analysis of the data and the discussion of the results. Jun-Lin Lin wrote the manuscript. The contributions of both authors are 80% (Jun-Lin Lin) and 20% (Laksamee Khomnotai).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bin, Z.; Yi, Z.; Faloutsos, C. Toward a comprehensive model in internet auction fraud detection. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 7–10 January 2008; p. 79.
2. Wang, J.C.; Chiu, C.Q. Detecting online auction inflated-reputation behaviors using social network analysis. In Proceedings of the Annual Conference of the North American Association for Computational Social and Organizational Science, Notre Dame, IN, USA, 26–28 June 2005.
3. Morzy, M. New algorithms for mining the reputation of participants of online auctions. *Algorithmica* **2008**, *52*, 95–112. [CrossRef]
4. Wang, J.-C.; Chiu, C.-C. Recommending trusted online auction sellers using social network analysis. *Expert Syst. Appl.* **2008**, *34*, 1666–1679. [CrossRef]
5. Chau, D.; Pandit, S.; Faloutsos, C. Detecting fraudulent personalities in networks of online auctioneers. In *Knowledge Discovery in Databases: PKDD 2006*; Fürnkranz, J., Scheffer, T., Spiliopoulou, M., Eds.; Springer: Berlin, Germany, 2006; pp. 103–114.
6. Pandit, S.; Chau, D.H.; Wang, S.; Faloutsos, C. Netprobe: A fast and scalable system for fraud detection in online auction networks. In Proceedings of the 16th International Conference on World Wide Web, Banff, AL, Canada, 8–12 May 2007; pp. 201–210.
7. Morzy, M. Cluster-based analysis and recommendation of sellers in online auctions. *Comput. Syst. Sci. Eng.* **2007**, *22*, 279–287.
8. Lin, S.J.; Jheng, Y.Y.; Yu, C.H. Combining ranking concept and social network analysis to detect collusive groups in online auctions. *Expert Syst. Appl.* **2012**, *39*, 9079–9086. [CrossRef]
9. Yu, C.H.; Lin, S.J. Web crawling and filtering for on-line auctions from a social network perspective. *Inf. Syst. E Bus. Manag.* **2012**, *10*, 201–218. [CrossRef]

10. Yu, C.H.; Lin, S.J. Fuzzy rule optimization for online auction frauds detection based on genetic algorithm. *Electron. Commer. Res.* **2013**, *13*, 169–182. [CrossRef]
11. Chiu, C.C.; Ku, Y.C.; Lie, T.; Chen, Y.C. Internet auction fraud detection using social network analysis and classification tree approaches. *Int. J. Electron. Commer.* **2011**, *15*, 123–147. [CrossRef]
12. Lin, J.-L.; Khomnotai, L. Improving fraudster detection in online auctions by using neighbor-driven attributes. *Entropy* **2016**, *18*, 11. [CrossRef]
13. Lin, J.-L.; Khomnotai, L. Using neighbor diversity to detect fraudsters in online auctions. *Entropy* **2014**, *16*, 2629–2641. [CrossRef]
14. Tadelis, S. Reputation and feedback systems in online platform markets. *Annu. Rev. Econ.* **2016**, *8*, 321–340. [CrossRef]
15. Gefen, D.; Karahanna, E.; Straub, D.W. Trust and TAM in online shopping: An integrated model. *Manag. Inf. Syst. Q.* **2003**, *27*, 51–90.
16. Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Manag. Sci.* **2003**, *49*, 1407–1424. [CrossRef]
17. Ba, S.; Pavlou, P.A. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *Manag. Inf. Syst. Q.* **2002**, *26*, 243–268. [CrossRef]
18. Melnik, M.I.; Alm, J. Does a seller's ecommerce reputation matter? Evidence from eBay auctions. *J. Ind. Econ.* **2002**, *50*, 337–349. [CrossRef]
19. Jolivet, G.; Jullien, B.; Postel-Vinay, F. Reputation and prices on the e-market: Evidence from a major french platform. *Int. J. Ind. Org.* **2016**, *45*, 59–75. [CrossRef]
20. Laitinen, E.K.; Laitinen, T.; Saukkonen, O. Impact of reputation and promotion on internet auction outcomes: Finnish evidence. *J. Internet Commer.* **2016**, *15*, 163–188. [CrossRef]
21. Rabby, F.; Shahriar, Q. Non-neutral and asymmetric effects of neutral ratings: Evidence from eBay. *Manag. Decis. Econ.* **2016**, *37*, 95–105. [CrossRef]
22. Utz, S.; Matzat, U.; Snijders, C. On-line reputation systems: The effects of feedback comments and reactions on building and rebuilding trust in on-line auctions. *Int. J. Electron. Commer.* **2009**, *13*, 95–118. [CrossRef]
23. Carter, M.; Tams, S.; Grover, V. When do I profit? Uncovering boundary conditions on reputation effects in online auctions. *Inf. Manag.* **2017**, *54*, 256–267. [CrossRef]
24. Chau, D.H.; Faloutsos, C. Fraud Detection in Electronic Auction. Available online: http://www.cs.cmu.edu/~dchau/papers/chau_fraud_detection.pdf (accessed on 2 July 2017).
25. Gregg, D.G.; Scott, J.E. The role of reputation systems in reducing on-line auction fraud. *Int. J. Electron. Commer.* **2006**, *10*, 95–120. [CrossRef]
26. Noufidali, V.; Thomas, J.S.; Jose, F.A. E-Auction Frauds—A Survey. *Int. J. Comput. Appl.* **2013**, *61*, 41–45.
27. Lee, C.-L. Customer Behavior of Using Privacy Protection Mechanism in Online Auctions. Available online: http://etd.lib.nctu.edu.tw/cgi-bin/gs32/ncugsweb.cgi?o=dncucdr&s=id=%22NCU984203031%22.&searchmode=basic (accessed on 2 July 2017).
28. Ruten. Available online: http://www.ruten.com.tw/ (accessed on 2 July 2017).
29. You, W.; Liu, L.; Xia, M.; Lv, C. Reputation inflation detection in a Chinese C2C market. *Electron. Commer. Res. Appl.* **2011**, *10*, 510–519. [CrossRef]
30. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann Publishers: Burlington, MA, USA, 2011; p. 664.
31. Trevathan, J.; Read, W. Detecting shill bidding in online English auctions. In *Handbook of Research on Social and Organizational Liabilities in Information Security*; Information Science Publishing: Hershey, PA, USA, 2005; pp. 446–470.
32. Xu, H.; Cheng, Y.-T. Model checking bidding behaviors in internet concurrent auctions. *Int. J. Comput. Syst. Sci. Eng.* **2007**, *4*, 179–191.
33. Dong, F.; Shatz, S.M.; Xu, H.; Majumdar, D. Price comparison: A reliable approach to identifying shill bidding in online auctions? *Electron. Commer. Res. Appl.* **2012**, *11*, 171–179. [CrossRef]