An Evolutionary Model for Constructing Robust Trust Networks

Siwei Jiang Jie Zhang Yew-Soon Ong
School of Computer Engineering
Nanyang Technological University, Singapore, 639798
{sjiang1, zhangj, asysong}@ntu.edu.sg

ABSTRACT

In reputation systems for multiagent-based e-marketplaces, buying agents model the reputation of selling agents based on ratings shared by other buyers (called advisors). With the existence of unfair rating attacks from dishonest advisors, the effectiveness of reputation systems thus heavily relies on whether buyers can accurately determine which advisors to include in trust networks and their trustworthiness. In this paper, we propose a novel multiagent evolutionary trust model (MET) where each buyer evolves its trust network. In each generation, each buyer acquires trust network information from its advisors and generates a candidate trust network using evolutionary operators. Only trust networks providing more accurate seller reputation estimation shall survive to the next generation. Experimental results demonstrate MET is more robust than the state-of-the-art trust models against various unfair rating attacks.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Intelligent agents; Multiagent systems

Keywords

Trust and Reputation Systems; Unfair Rating Attacks; Robust Trust Network; Multiagent Evolutionary Computation

1. INTRODUCTION

Trust plays a vital role in open, large, distributed and dynamic multiagent systems where self-interested agents may be deceptive and strategic. For example, in multiagent-based e-marketplaces, dishonest selling agents may not deliver products with the quality as what they promised. Reputation systems are thus designed for buying agents to model the reputation of sellers based on ratings shared by other buyers (called advisors) and make decisions on which sellers to transact with. However, unfair rating attacks from dishonest advisors, such as the *Sybil*, *Camouflage* and *Whitewashing* attacks, render reputation systems vulnerable to mislead buyers to transact with dishonest sellers [2, 10]. Strategic dishonest advisors may also employ sophisticated attacks, such as a combination of unfair rating attacks.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013),* Ito, Jonker, Gini, and Shehory (eds.), May 6–10, 2013, Saint Paul, Minnesota, USA.

Various trust models [3, 6, 7, 8, 9] have been proposed to cope with unfair ratings. Dishonest advisors' ratings are either filtered out (e.g., BRS [7] and iCLUB [3]) or discounted (e.g., TRAVOS [6], ReferralChain [8] and Personalized [9]) before aggregating advisors' ratings to estimate seller reputation. However, these models are not completely robust against various strategic attacks. In particular, when dishonest advisors occupy a large proportion in e-marketplaces (i.e., Sybil), BRS [7] becomes inefficient and iCLUB [3] is unstable because they both employ the "majority-rule". When dishonest advisors adopt strategic attacks, TRAVOS [6] does not work well because it assumes an advisors' rating behavior is consistent. ReferralChain [8] assigns trust value 1 to every new buyer (advisor) which provides a chance for dishonest advisors to abuse the initial trust (i.e., Whitewashing). Personalized [9] is vulnerable when buyers have insufficient experience with advisors and the majority of advisors are dishonest (i.e., combination of Sybil and Whitewashing). Thus, we need a more robust trust model.

Evolutionary computation, as a search methodology mimicking the natural organism, inherits the characteristic of robustness to enable individuals (agents) to survive in a broad variety of environments [5]. To resist various unfair rating attacks, we thus propose a multiagent evolutionary trust model (MET) for each buyer to evolve its trust network (consisting of information about which advisors to include in the network and their trustworthiness) over time and finally construct accurate and robust trust networks. Specifically, in each generation, each buyer acquires trust network information from its advisors and generates a candidate trust network using evolutionary operators [1] (e.g., crossover and mutation). Based on fitness evaluation, only trust networks providing more accurate seller reputation estimation will be kept for the buyer (i.e., survive to the next generation).

Compared to the state-of-the-art trust models [3, 6, 7, 8, 9], MET has the following unique characteristics: 1) MET allows a buyer to ask its advisors about their trust network information. But, differing from trust propagation in ReferralChain [8], the buyer only uses advisors' information to generate candidate trust networks that will still go through fitness evaluation; 2) evolutionary operators enable buyers to explore diverse forms of trust networks, to alleviate the impact of false information provided by dishonest advisors; 3) different from most of the trust models that aim to accurately model each individual advisor's trustworthiness, MET finds out the optimal trust network that gives the most accurate seller reputation estimation through fitness evaluation, even certain advisor trustworthiness is not very accurate.

We carry out experiments in a simulated multiagent e-marketplace where dishonest sellers try to obtain high transaction volume by colluding with dishonest advisors to launch unfair rating attacks. Results show that MET is more robust than the state-of-the-art trust models and can construct more accurate trust networks to estimate seller reputation.

2. RELATED WORK

In a large e-marketplace, direct experience between a buyer and a seller is often insufficient or even does not exist. In such a case, the buyer has to rely on indirect experience – opinions of other buyers (playing the role of advisors) towards a target seller. However, various forms of cheating behavior (attacks) from advisors have been observed and studied in the trust research community [2, 10]. We provide brief descriptions for the typical unfair rating attacks that will be studied in this paper, including **Constant** where dishonest advisors constantly provide unfairly positive/negative ratings to sellers; **Camouflage** where dishonest advisors camouflage themselves as honest advisors by providing fair ratings to build up their trustworthiness first and then gives unfair ratings; **Whitewashing** where a dishonest advisor is able to *whitewash* its low trustworthiness by starting a new account with the initial trustworthiness value; **Sybil** where a dishonest buyer creates several accounts to constantly provide unfair ratings to sellers [2, 10]; **Sybil Camouflage** (the combination of Sybil and Camouflage) where a number of dishonest advisors perform Camouflage attacks together; and **Sybil Whitewashing** where a number of dishonest advisors perform Whitewashing attacks together.

Various trust models [3, 6, 7, 8, 9] have been proposed to handle unfair ratings from advisors, but they are all vulnerable to certain attacks. Specifically, BRS [7] iteratively filters out unfair ratings based on the "majority-rule" where an advisor is considered as an outlier if its ratings are located outside of the acceptable range of all advisors' accumulated ratings. This rule renders BRS vulnerable to Sybil-based attacks because honest buyers' (the minority) ratings will be incorrectly filtered out. In [3], iCLUB is proposed to handle multi-nominal ratings by applying clustering to divide buyers into different clubs. When having little evidence about sellers, a buyer relies on the club with the maximum number of advisors. In this scenario, Sybil attackers forming a club with many members will mislead the buyer to follow their opinions. Thus, iCLUB is also vulnerable to Sybil-based attacks. TRAVOS [6] discounts advisors' ratings by setting weights of their ratings according to their trustworthiness. In some cases, such weights cannot punish dishonest advisors to a large extent when a buyer has insufficient experience with the sellers which the advisors has encountered in the past. In addition, TRAVOS assumes an advisor's behavior is consistent, making it vulnerable to Camouflage attacks. Yu and Singh propose referral chains to propagate trust through advisors [8]. The initial trustworthiness values of advisors are set to 1 in the range of $[0, 1]$, and these values will be decreased if the advisors' ratings deviate from the buyer's direct experience. This provides an opportunity for Whitewashing attackers to abuse their initial trustworthiness values. In the Personalized approach [9], the trustworthiness of advisors is calculated based on both private and public trust aspects. The private part is vulnerable to Whitewashing attacks because a buyer cannot have many commonly rated sellers with a whitewashing attacker. The public part cannot work well when the majority advisors are dishonest (Sybil attacks). Thus, this approach is vulnerable to the combination of Sybil and Whitewashing attacks.

In contrast, our MET model takes advantage of the robust ability of evolutionary computation in solving dynamic and complex problems [5]. It also has several unique characteristics compared to the state-of-the-art trust models, as summarized in Section 1. All these specific design decisions make MET robust against various unfair rating attacks, which will be demonstrated through experiments in Section 4.

Evolutionary techniques have been widely used to design intelligent agents and multiagent systems. In multi-robot learning [4], each agent is configured with a single-layer neuron network. The evolutionary technique of particle swarm optimization is adopted to reproduce new parameters of the neural network for the agent by allowing it to model neighboring agents. This work assumes that each agent shares its true knowledge. MET addresses false information from agents by constructing accurate and robust trust networks.

3. THE MET MODEL

In e-marketplaces, when a buyer wants to evaluate the reputation of a target seller but does not have sufficient personal experience with the seller, it needs to ask for opinions (ratings) from other buyers (advisors) towards the seller. To cope with possible unfair ratings from dishonest advisors, each buyer in our MET model maintains a trust network consisting of a set of advisors, each of which is assigned with a trust value[1]. The buyer then evolves its trust network over time in order to obtain a high quality trust network.

To measure the quality of trust networks, a specific fitness function is designed by considering simultaneously the following two aspects: 1) suitability of selected advisors; 2) accuracy of trust values assigned to these advisors. In each generation, each buyer asks some randomly selected advisors from its trust network for information about their trust networks and fitness values[2]. By comparing its own trust network with those provided by the advisors, the buyer selects three appropriate trust networks to produce a candidate trust network using evolutionary operators. Finally, between the candidate trust network and the buyer's current trust network, the one with higher fitness value will survive to the next generation. The details of the MET model will be provided in the following sections.

3.1 Fitness Function

Assume that in an e-marketplace, the set of buyers is denoted as $B = \{B_i | i = 1, \cdots, l\}$ and the set of sellers is denoted as $S = \{S_j | j = 1, \cdots, m\}$. We also denote the trustworthiness of an advisor $A_k \in B$ from the view of a buyer B_i as $T_{B_i}(A_k) \in [0, 1]$. In the buyer B_i's trust network TN_{B_i}, the trustworthiness values of advisors connected with B_i is then denoted as $T_{B_i}(A) = \{T_{B_i}(A_k) | A_k \in TN_{B_i}\}$.

A rating provided by buyer B_i to seller S_j is denoted as r_{B_i,S_j}, which can be a binary, multi-nominal or real value. Two types of reputation values of S_j can then be derived by buyer B_i. One type of reputation is derived based on

[1]For a new buyer without any knowledge of the e-marketplace environments, it can randomly select a set of other buyers as its advisors, each of which is assigned with a randomly generated trust value.

[2]How to reach a specific advisor to obtain such information as well as seller ratings is out the scope of this work.

the buyer's personal experience with the seller, denoted as $R_{B_i}(S_j)$. And, another type is calculated based on the experience with the seller (i.e., ratings) shared by the advisors in the buyer's trust network, denoted as $\tilde{R}_{B_i}(S_j)$.

In our MET model, a specific fitness function is designed for buyers to measure the quality of their trust networks by comparing the two types of derived reputation values of sellers. Formally, the fitness value of buyer B_i's trust network $T_{B_i}(A) = \{T_{B_i}(A_k) | A_k \in TN_{B_i}\}$ is calculated as:

$$f(T_{B_i}(A)) = \frac{1}{m'} \sum_{j=1}^{m'} |R_{B_i}(S_j) - \tilde{R}_{B_i}(S_j)| \quad (1)$$

where $R_{B_i}(S_j)$ and $\tilde{R}_{B_i}(S_j)$ are the two types of reputation of a seller S_j respectively. In addition, $m' \leq m$, indicating that sellers with which either buyer B_i or its advisors have no experience will not be considered in fitness evaluation.

Suppose that rating type is real, i.e., $r_{B_i,S_j} \in [0,1]$, and the default value $r_{B_i,S_j} = 0.5$ means that B_i has no experience with S_j. The two types of reputation values of the seller S_j are then calculated respectively as:

$$\begin{cases} R_{B_i}(S_j) & = & \text{mean}(\{r_{B_i,S_j}\}) \\ \tilde{R}_{B_i}(S_j) & = & g(T_{B_i}(A_k) \cdot \text{mean}(\{r_{A_k,S_j}\})) \end{cases} \quad (2)$$

where $R_{B_i}(S_j)$ is the average of B_i's ratings to S_j on different transactions, and $r_{B_i,S_j} \neq 0.5$. The general function $g(\cdot)$ can be a discounting or Dempster-Shafter operator [8]. For simplicity, we use the discounting operator as follows:

$$\tilde{R}_{B_i}(S_j) = [\sum_{k=1}^{n} T_{B_i}(A_k) \times \text{mean}(\{r_{A_k,S_j}\})] / \sum_{k=1}^{n} T_{B_i}(A_k) \quad (3)$$

where $A_k \neq B_i$ (ignore B_i's own opinion), $A_k \in TN_{B_i}$, $n = |TN_{B_i}|$ is the total number of advisors in the trust network ($n \leq l-1$), and $r_{A_k,S_j} \neq 0.5$.

A smaller fitness value indicates that the buyer's trust network is in higher quality, because the combination of advisors' ratings is more similar to the buyer's own opinions regarding the common sellers. In other words, the fitness function measures the suitability of the selected advisors and the accuracy of the trust values assigned to these advisors simultaneously. In addition, it is worth noting that MET does not rely on the trust transitivity or propagation despite the fact that we use the concept of trust network. The accuracy of the trustworthiness values of all advisors in a buyer's trust network is directly measured by the buyer itself, based on the buyer's own experience with sellers.

3.2 Trust Network Comparison

In each generation, each buyer will ask some randomly selected advisors in its trust network to share information about their trust networks and fitness values. However, some dishonest advisors may provide false information. In some cases, honest advisors may unintentionally provide noisy or useless information, because they do not have sufficient experience with sellers in the e-marketplace in order to assign precise trust values to the advisors in its trust network. To alleviate this problem, the buyer will compare trust networks shared by advisors with its own trust network, to choose appropriate trust networks, which will further be used to generate candidate trust networks (see Section 3.3).

More specifically, suppose that buyer B_i has the trust network $T_{B_i}(A) = \{T_{B_i}(A_1), \cdots, T_{B_i}(A_x)\}$ and the fitness value is $f(T_{B_i}(A))$. Buyer B_i randomly chooses an advisor A_r from its trust network to ask for the information about the advisor's trust network and its fitness value. Suppose that advisor A_r provides the trust network information as $T_{A_r}(A) = \{T_{A_r}(A_1), \cdots, T_{A_r}(A_y)\}$ and the fitness value $f(T_{A_r}(A))$. The difference between the trust networks of buyer B_i and advisor A_r is calculated as follows:

$$\text{diff}(T_{B_i}(A), T_{A_r}(A)) = \frac{1}{n'} \sum_{k=1}^{n'} |T_{B_i}(A_k) - T_{A_r}(A_k)| \quad (4)$$

where $n' \geq x, y$ is the number of advisors appearing in either the trust networks of B_i or that of A_r, that is the union of advisors in the two trust networks. In some cases, B_i and A_r may not have common advisors. If an advisor A_k appears in one trust network but not the other, the default trust value of A_k in the second trust network will be assigned as 0.5. The difference between the fitness values of B_i and A_r is:

$$\text{diff}(f(T_{B_i}(A)), f(T_{A_r}(A))) = |f(T_{B_i}(A)) - f(T_{A_r}(A))| \quad (5)$$

Buyer B_i chooses to use the information shared by advisor A_r only when the following condition is satisfied:

$$\begin{aligned} &(\text{diff}(T_{B_i}(A), T_{A_r}(A)) - 0.5) \\ &\quad \times (\text{diff}(f(T_{B_i}(A)), f(T_{A_r}(A))) - 0.5) > 0 \end{aligned} \quad (6)$$

The rationale of Eq. 6 can be explained by analyzing the following three scenarios. In the first scenario where advisor A_r provides both the real trust network and the real fitness value, Eq. 4 and Eq. 5 are smaller than 0.5 because they are similar with B_i's own experience. Then B_i will treat A_r as an honest advisor to use its information. In the second scenario where only one type of A_r's information is false, either Eq. 4 or Eq. 5 is smaller than 0.5. Then the result of Eq. 6 is smaller than 0. Buyer A_r will not be selected as an honest advisor by buyer B_i. In the third scenario where A_r provides both a false trust network and a false fitness value, Eq. 4 and Eq. 5 are both larger than 0.5. Then, A_r's fitness value indicates that A_r's trust network is different from B_i's own opinion, and this information from A_r suggests B_i to avoid such type of trust networks.

3.3 Evolutionary Operators

After choosing three[3] advisors by trust network comparison, the buyer will generate a candidate trust network using evolutionary operators. By comparing the candidate trust network with the buyer's own trust network, the one with higher fitness value measured by Eq. 1 will survive to the next generation. Two widely used evolutionary operators (DE crossover and polynomial mutation [1]) are utilized to produce candidate trust works in MET.

The operator "DE/ran/1/bin" in Differential Evolutionary (DE) [1] is adopted for crossover. Let us denote buyer B_i's trust network in generation $g-1$ as $T_{B_i,g-1}(A)$. At first, the crossover operator generates a new vector $V_{B_i,g}$ as:

$$V_{B_i,g}(A) = T_{A_1,g-1}(A) + F \cdot [T_{A_2,g-1}(A) - T_{A_3,g-1}(A)] \quad (7)$$

where F is the scaling factor which amplifies or shrinks the difference vectors. A_1, A_2, A_3 are the three advisors selected in Section 3.2. $T_{A_1,g-1}(A), T_{A_2,g-1}(A), T_{A_3,g-1}(A)$ are the

[3]If insufficient advisors satisfy Eq. 6, some advisors will be randomly selected from the buyer's trust network.

trust networks shared by them respectively. After that, operator "DE/rand/1/bin" applies the binomial crossover operation to produce the new trust network.

$$T_{B_i,g}(A_k)=\begin{cases} V_{B_i,g}(A_k) & \text{if } rand_k()\leq CR \,||\, k=k_{rand} \\ T_{B_i,g-1}(A_k) & \text{otherwise} \end{cases} \quad (8)$$

where $rand_k() \in [0,1]$ is an uniformly distributed random number and $k_{rand} \in [1,n]$ is a randomly chosen integer. The control parameter CR is the probability for crossover. If $T_{B_i,g}(A_k) < 0$, it is set to 0. If $T_{B_i,g}(A_k) > 1$, it is set to 1.

In general, buyer B_i uses information from the advisors in its own trust network, which is considered as the *local view*. But, such mechanism may lead solutions to get stuck at local optimum. To balance between exploitation and exploration, our MET model designs a probability variable P_{local} (usually close to 1) to control the buyer's view. It means that the buyer has probability $1 - P_{local}$ to obtain information from all advisors in the e-marketplace as the *global view*.

The polynomial mutation is used to add perturbation to B_i's trust network, which is beneficial to generate different solutions and boost the evolutionary process, as follows:

$$T_{B_i,g}(A_k) = T_{B_i,g}(A_k) + \delta \quad \text{if } rand_k() \leq p_m \quad (9)$$

$$\delta = \begin{cases} (2 \cdot rand())^{1/(\eta_m+1)} - 1 & \text{if } rand() < 0.5 \\ 1 - |2(1 - rand())|^{1/(\eta_m+1)} & \text{otherwise} \end{cases} \quad (10)$$

where η_m is used to control the polynominal probability distribution, and p_m is the mutation probability. After evolutionary operations, the number of advisors in $T_{B_i}(A)$ is retained by discarding advisors with smaller trust values.

Input : $T_{B_i,0}(A)$, buyer B_i's current trust network;
G, the maximum number of generations;
P_{local}, probability of local view;
Output: The optimal trust network $T_{B_i}(A)$;

1 Calculate the fitness $f(T_{B_i,0}(A))$ using Eq. 1;
2 **for** $g = 1$ *to* G **do**
3 **if** $rand() < P_{local}$ **then**
4 Randomly select advisors A_{r1}, A_{r2}, A_{r3} that satisfy Eq. 6 from $T_{B_i,g-1}(A)$;
5 **else**
6 Randomly select advisors A_{r1}, A_{r2}, A_{r3} that satisfy Eq. 6 from all possible advisors;
7 Generate $T_{B_i,g}(A)$ by DE with A_{r1}, A_{r2}, A_{r3};
8 Apply the polynomial mutation to $T_{B_i,g}(A)$;
9 Calculate the fitness $f(T_{B_i,g}(A))$;
10 **if** $f(T_{B_i,g}(A)) < f(T_{B_i,g-1}(A))$ **then**
11 Replace $T_{B_i,g-1}(A)$ by $T_{B_i,g}(A)$;
12 Output the optimal trust network $T_{B_i,G}(A)$;

Algorithm 1: Multiagent Evolutionary Trust Model

3.4 Pseudo-Code Summary of MET

The pseudo-code summary of the MET model is given in Algorithm 1. When a buyer has some new experience with certain sellers, the buyer will evolve its trust network to capture advisors' dynamic behavior patterns by Algorithm 1. More specifically, the buyer B_i firstly evaluates its current trust network (in generation 0) based on both the new and old experience with sellers (Line 1). The buyer then acquires trust network information from three randomly selected advisors of its own trust network with the probability P_{local} or from all possible advisors with the probability $1 - P_{local}$. The trust networks shared by all the selected advisors should satisfy Eq. 6 (Lines 3-6), to control the quality of the shared trust networks. With the shared trust networks, the buyer then generates a new trust network using the DE operator and polynomial mutation (Lines 7-8). If the newly generated trust network is better than the buyer's current trust network, the better one will survive to the next generation (Lines 9-11). After generations of evolution, an optimal trust network will be finally obtained for the buyer to accurately model seller reputation.

Table 1: Key Parameters in the Testbed

Key parameters	Values
Number of dishonest duopoly sellers	1
Number of honest duopoly sellers	1
Number of dishonest common sellers	99
Number of honest common sellers	99
Number of dishonest buyers ($\|B^D\|$)	12/28*
Number of honest buyers ($\|B^H\|$)	28/12*
Simulation days ($Days$)	100
Dominance Ratio ($Ratio$)	0.5

* Non-Sybil-based Attack/Sybil-based Attack

4. EXPERIMENTATION

We carry out a rich set of experiments to evaluate MET. In this section, we introduce a multiagent-based e-marketplace testbed firstly, and then evaluate the robustness of MET by simulating different strategies of advisors for sharing their trust networks and examining parameter settings on MET. After that, we compare MET with five existing trust models to show its advantages of being more robust again various unfair rating attacks and more accurately modeling seller reputation and advisor trustworthiness.

4.1 Multiagent-Based E-Marketplace Testbed

As mentioned in [2, 10], the existing testbeds, such as the Agent Reputation and Trust (ART) testbed, are not suitable for carrying out experiments to compare the robustness of trust models under unfair rating attacks. We thus design a multiagent-based e-marketplace testbed to incorporate different trust models and simulate unfair rating attacks from advisors. In the testbed, we simulate a scenario of "Duopoly Market" where two sellers occupy a large portion of the total transaction volume in the market. The *dishonest duopoly seller* tries to compete with the *honest duopoly seller* to gain larger transaction volume by recruiting dishonest buyers to perform unfair rating attacks. The other sellers (*common sellers*) include 99 honest sellers and 99 dishonest sellers, and their reputation are uniformly distributed alone [0, 1]. Typically, trust models are most effective when only 30% of buyers are dishonest [7]. Thus, we add 12 dishonest buyers (attackers) and 28 honest buyers in the market for non-Sybil-based Attacks, and switch their numbers for Sybil-based Attacks. The entire simulation lasts for 100 days. On each day, each buyer chooses to transact with one seller once.

Since most trust models are more effective when every advisor has transaction experiences with many different sellers, we assume that buyers will transact with the duopoly sellers with the probability 0.5 while transacting with each common seller randomly. This implies that duopoly sellers occupy half of transactions in the market, which is called the *dominance ratio*. When deciding on which duopoly seller to transact with, honest buyers use trust models to calculate their reputation and transact with the one with the higher value, while dishonest buyers choose sellers according to

their attacking strategies. After each transaction, honest buyers provide fair ratings, whereas dishonest buyers provide ratings according to their attacking strategies. The key parameters are summarized in Table 1.

To evaluate trust models, we compare the transaction volumes of the duopoly sellers. The robustness of a trust model (defense, Def) against an attack model (Atk) is:

$$\mathcal{R}(Def, Atk) = \frac{|Tran(S^H)| - |Tran(S^D)|}{|B^H| \times Days \times Ratio} \quad (11)$$

where $|Tran(S^D)|$ and $|Tran(S^H)|$ are transaction volumes of the dishonest and honest duopoly sellers, respectively. $\mathcal{R}(Def, Atk) = 1$ or -1 means Def is *complete robust* or *complete vulnerable* to Atk, respectively. The larger value indicates the trust model is more robust against the attack.

The mean absolute error (MAE) of seller reputation is also adopted to measure the accuracy of trust models in modeling seller reputation:

$$\text{MAE}(S_j) = \frac{\sum_t \sum_{B_i} |R^t(S_j) - \tilde{R}^t_{B_i}(S_j)|}{|B^H| \times Days} \quad (12)$$

where $R^t(S_j)$ is the actual reputation of a seller S_j in day t ($t \in [0, Days]$), and $\tilde{R}^t_{B_i}(S_j)$ is the estimated reputation of S_j by a trust model based on experience of a honest buyer B_i's advisors ($B_i \in B^H$). A smaller MAE indicates that the trust model predicts seller reputation more accurately.

In the testbed, each trust model is tested against every attack over 50 independent runs. The experimental results show the mean and standard deviation ($mean \pm std$), and the best results are in bold font. The ratings to sellers are set as the real type. For parameters of trust models, we use the values suggested by their authors. The parameter settings of our MET are outlined as following. The number of advisors in a buyer's trust network is $n = 25$. The maximum generation is $G = 10$. The probability of local view is $P_{local} = 0.8$. The DE operator has $CR = 0.6, F = 0.3$, and the polynomial mutation has $\eta_m = 20$ and $p_m = 0.05$.

4.2 The Influence of Trust Networks

In MET, buyers exchange information about seller ratings and trust networks. Honest buyers always provide truthful information to others. Besides unfair ratings generated by their various attacking strategies, dishonest buyers may also provide three types of trust networks to other buyers:

- Truthful trust network: a dishonest buyer provides truthful information about the trustworthiness of advisors in its trust network. Such information will help other buyers to obtain optimal trust networks quickly.
- Noisy trust network: a dishonest buyer shares a trust network with randomly generated trust values for advisors, implying that some buyers do not share their evolved trust networks but only the initial networks.
- Collusive trust network: a dishonest buyer provides false information about the trustworthiness of some advisors because they are in the same colluding group. Such collusive information is more difficult to detect.

In this section, our model is tested with the three types of trust networks. A rating to a seller from a buyer is a real value. Since BRS [7], TRAVOS [6] and Personalized [9] are designed to deal with binary ratings, the rating is converted to [*negative*, *positive*] when it is in the range of $[0, 0.5)$ and $(0.5, 1.0]$, respectively. The iCLUB approach [3] is proposed for multi-nominal ratings, so the rating is converted to [*worst*, *bad*, *neutral*, *good*, *best*] for the ranges of $[0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), [0.8, 1.0]$, respectively.

Table 2: Robustness of MET with the Existence of Truthful, Noisy and Collusive Trust Networks

	Constant	Camouflage	Whitewashing
MET– truthful	**0.99±0.02**	**0.99±0.03**	**0.99±0.02**
MET– noise	0.98±0.02	**0.99±0.03**	**0.99±0.03**
MET– collusive	0.98±0.02	**0.99±0.02**	0.98±0.04
	Sybil	Sybil Cam*	Sybil WW*
MET– truthful	**0.96±0.07**	**0.99±0.07**	**0.98±0.08**
MET– noise	0.91±0.08	0.96±0.08	0.94±0.08
MET– collusive	0.87±0.15	0.94±0.06	0.82±0.11

*Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

In Table 2, we can see that MET is highly robust against the Constant, Camouflage and Whitewashing attacks no matter whether dishonest buyers (advisors) provide truthful, noisy or collusive trust networks. Under the other three Sybil-based attacks, although the robustness values of MET decreases when dishonest buyers provide noisy and collusive trust networks, it still exhibits reasonably high robustness values ($\mathcal{R}(\text{MET}, Atk) \geq 0.82$). Hereafter, we test MET under the most challenging case (collusive trust networks) unless explicitly indicated. This will also affect the ReferralChain model [8] but not other trust models.

4.3 Impact of Parameter Settings on MET

In this section, we investigate the impact of the parameter setting (i.e., the number of advisors in a buyer's trust network n) on the performance of MET. When a buyer acquires information about sellers from its trust network, the smaller/larger value of n means the buyer can consult less/more advisors, respectively. If the buyer has fewer advisors, the evaluation of seller reputation is difficult because it receives less information on target sellers. On the other hand, when the buyer has too many advisors, MET may require more generations to evolve the buyer's trust network.

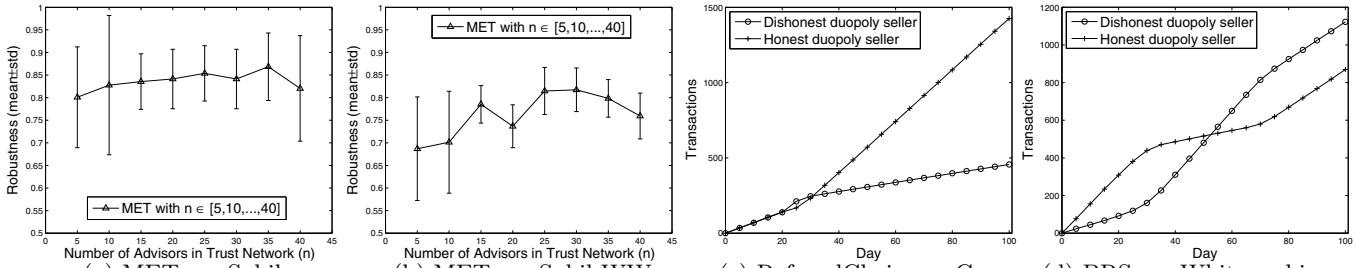
Fig. 1(a-b) show the average robustness of MET against Sybil and Sybil Whitewashing, respectively. MET is tested with $n \in [5, 10, \cdots, 40]$ advisors. It demonstrates that the parameter settings has certain impact on MET and $n = 25$ is a good choice in our experiments. As shown in Fig. 1(b), it is also worth to notice that MET with different parameters exhibits reasonably high average robustness even under the strongest attack (i.e., $\mathcal{R}(\text{MET}, \text{Sybil WW}) > 0.68$).

4.4 Comparison of Robustness

We also carry out a set of experiments to compare the robustness of MET with that of other trust models. The results are presented in Table 3 and Fig. 1(c-d) and Fig. 2. Next, we will describe the results under each type of attacks.

From Table 3, we can see that all the trust models are robust against the **Constant** attack. Consistent with the authors' own experimental results [7], the table also shows that BRS is not completely robust against the Constant attack ($\mathcal{R}(\text{BRS}, \text{Constant}) = 0.87$).

The **Camouflage** attackers provide fair ratings to common sellers to establish their trustworthiness before day 20, and then give unfair ratings to all sellers. In Table 3, ReferralChain's robustness is low with respect to Camouflage. Before day 20, ReferralChain is unable to decrease the trustworthiness of dishonest advisors as they provide fair ratings.



(a) MET vs. Sybil (b) MET vs. Sybil WW (c) ReferralChain vs. Cam (d) BRS vs. Whitewashing

Figure 1: (a-b): MET with Different Parameter Settings; (c-d): Transactions along Days

Table 3: Robustness of Trust Models vs. Attacks

	Constant	Camouflage	Whitewashing
BRS	0.87±0.03	0.89±0.02	-0.18±0.07
iCLUB	0.98±0.02	**0.99±0.02**	0.77±0.13
TRAVOS	0.97±0.02	0.82±0.03	0.87±0.03
ReferralChain	0.89±0.04	0.69±0.04	-0.95±0.08
Personalized	**0.99±0.03**	**0.99±0.03**	**0.98±0.03**
MET	0.98±0.02	**0.99±0.02**	**0.98±0.04**
	Sybil	Sybil Cam*	Sybil WW*
BRS	-0.99±0.08	-0.47±0.07	-0.30±0.07
iCLUB	0.23±0.35	0.90±0.09	0.20±0.29
TRAVOS	0.16±0.09	-0.57±0.07	-0.98±0.07
ReferralChain	0.82±0.06	0.63±0.08	-0.98±0.07
Personalized	0.74±0.45	**0.94±0.08**	-1.00±0.08
MET	**0.87±0.15**	**0.94±0.06**	**0.82±0.11**

*Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

Buyers receive seller information from their advisors with equal chance. Thus, the two duopoly sellers obtain similar transaction volume before day 20 (see Fig. 1(c)). After that, ReferralChain gradually decreases the trustworthiness of dishonest advisors when they give unfair ratings. The transaction volume of the honest duopoly seller becomes larger than its competitor after day 30.

Each **Whitewashing** attacker provides one unfair rating on each day and starts with a new buyer account on the next day. In Table 3, the value $\mathcal{R}(\text{BRS}, \text{Whitewashing}) = -0.48$ shows that BRS is vulnerable to this attack. According to Fig. 1(d), the honest duopoly seller has more transactions than the dishonest one in the beginning. However, after some time (around 52 days), the dishonest duopoly seller's transaction volume exceeds its competitor. To explain, after day 52, the accumulated reputation of a seller will more easily fall in the rejection area of the beta distribution of an honest buyer rather than a Whitewashing attacker. This means that honest buyers will be incorrectly filtered out by BRS. Listening advice from Whitewashing attackers misleads buyers to transact with the dishonest duopoly seller. ReferralChain is vulnerable to Whitewashing because the initial trustworthiness of advisors (buyers) is set to 1 in that model [8], and it is difficult for buyers to select reliable advisors between honest buyers and Whitewashing attackers.

MET allows buyers to exchange their advisor information to generate candidate trust networks. Through fitness comparison using Eq. 1, trust networks with the most suitable advisors will be kept for buyers. It is also difficult for Whitewashing attackers to get into buyers' trust networks. Thus, MET is able to obtain the high robustness of 0.98.

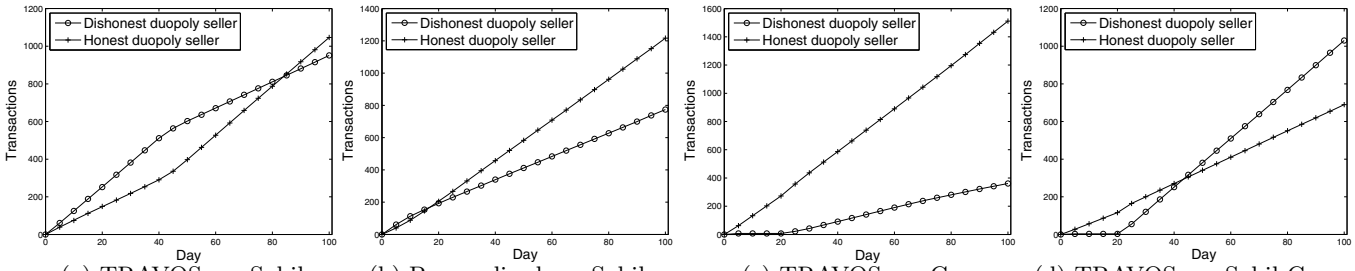
BRS is completely vulnerable to the **Sybil** attack due to its employed "majority-rule". The robustness of iCLUB is not stable with the standard deviation of ($std = 0.35$). To explain, in Sybil, the majority of buyers are dishonest. When a buyer only relies on its own experience to model the trustworthiness of advisors, it still has the accurate modeling. If it also relies on opinions of majority advisors (which are dishonest), it will have incorrect modeling of advisors. In consequence, the modeling of seller reputation will be inaccurate. Personalized also has large standard deviation ($std = 0.45$) because it has the similar design as iCLUB.

TRAVOS is not completely robust against Sybil attacks. In the early period, TRAVOS cannot find enough reference sellers so the discounting of advisors' ratings is not effective (referred to as *soft punishment*). For instance, suppose that the trustworthiness of dishonest/honest advisors is 0.4/0.6, and each advisor (12 honest ones and 28 dishonest ones) gives one rating to a seller. An honest seller's reputation is $0.40 \approx (0.6 \times 12 + 1)/(0.4 \times 28 + 0.6 \times 12 + 2))$ and that of the dishonest seller is $0.60 \approx (0.4 \times 28 + 1)/(0.4 \times 28 + 0.6 \times 12 + 2)$. However, if a trust model is able to set the dishonest/honest advisors' trustworthiness as 0.1/0.9, the evaluation of seller reputation will become more accurate. The Personalized approach, in the beginning, also suffers from the *soft punishment* when the buyer relies on public trust to evaluate advisors' trustworthiness. In Fig. 2(a-b), as buyers have more experience, TRAVOS and Personalized become more effective after day 80 and day 15, respectively.

MET generates diverse trust networks with several candidate trust values of advisors using evolutionary operators, and keeps the best trust network with the most accurate trust assigned to advisors. This increases the chance for buyers to punish Sybil attackers to a large extent. Thus, MET obtains the high robustness as $\mathcal{R}(\text{MET}, \text{Sybil}) = 0.87$.

Unlike the Sybil attack, **Sybil Camouflage** is unable to render BRS completely vulnerable. This is because in the beginning attackers camouflage themselves as honest ones by providing fair ratings where BRS is always effective. After attackers stop camouflaging, the dishonest duopoly seller's transaction volume will soon exceed its competitor. Under Camouflage and Sybil Camouflage, the robustness of ReferralChain is similar because the buyer aggregates only its local advisors' ratings to predict sellers' reputation.

Comparing with Camouflage, TRAVOS becomes vulnerable to Sybil Camouflage. Although TRAVOS will inaccurately promote the trustworthiness of a Camouflage attacker (most are slightly larger than 0.5), when majority of buyers are honest, the aggregated ratings from attackers are still not able to overweigh honest buyers' opinions. However, under Sybil Camouflage, when majority are dishonest buyers, these attackers' aggregated ratings will easily overweigh honest buyers' opinions and render TRAVOS vulnerable. Fig. 2(c-d) clearly show the difference in the robustness of TRAVOS against Camouflage and Sybil Camouflage.



(a) TRAVOS vs. Sybil (b) Personalized vs. Sybil (c) TRAVOS vs. Cam (d) TRAVOS vs. Sybil Cam

Figure 2: Transaction Volume along Days for Dishonest and Honest Duopoly Sellers

Sybil Whitewashing is the strongest attack among the six investigated attacks. It can defeat BRS, TRAVOS, ReferralChain and Personalized as shown in Table 3. Similar to Sybil, the robustness of iCLUB against Sybil Whitewashing is still unstable. Comparing with Whitewashing, BRS is still vulnerable to Sybil Whitewashing, while TRAVOS and Personalized change dramatically from being robust to completely vulnerable. For TRAVOS, since each whitewashing attacker provides only one unfair rating, buyers cannot find reference sellers to discount attackers' trustworthiness to a large extent. When majority are softly punished attackers, TRAVOS will always suggest honest buyers to transact with the dishonest duopoly seller. For Personalized, the buyer cannot find enough commonly rated sellers and will heavily rely on public trust to evaluate an advisor's trustworthiness, which is inaccurate when majority of buyers are dishonest. Thus, similar to TRAVOS, whitewashing attackers' trustworthiness cannot be discounted to a large extent and *soft punishment* renders Personalized completely vulnerable.

ReferralChain is completely vulnerable to Sybil Whitewashing, whereas MET is sufficient robust to this attack ($\mathcal{R}(\text{MET}, \text{Sybil WW}) = 0.82$ in Table 3). Although both ReferralChain and MET allow buyers to ask their advisors about other advisors, MET applies evolutionary operators (e.g., crossover and mutation) to generate new trust networks. The candidate trust networks will go through the fitness evaluation. Only when the selected advisors and trustworthiness values for those advisors show better accuracy, they will be kept by buyers. Besides, unlike ReferralChain, MET does not assign high initial trust values to advisors.

In summary, experimental results show that iCLUB and the Personalized approach have large perturbation under Sybil attacks. BRS, TRAVOS and ReferralChain are vulnerable to Sybil, Camouflage and Whitewashing, respectively. It demonstrates that our MET model is more robust than these other trust models against typical attacks.

4.5 MAE of Modeling Seller Reputation

In this section, we carry out further experiments to compare trust models in term of mean absolute error (MAE) of duopoly sellers' reputation. The smaller MAE indicates the trust model is more accurate in modeling seller reputation.

In Tables 4-5, under Constant, Camouflage and Whitewashing, our MET is able to obtain the best results for both duopoly sellers' reputation. Under other Sybil, Sybil Camouflage and Sybil Whitewashing, MET and iCLUB provide the best MAE values. In most cases, other trust models (except MET) obtain smaller MAE for the honest duopoly seller reputation while larger MAE for the dishonest duopoly seller reputation, implying that it is more difficult to obtain accurate dishonest sellers' reputation because they recruit attacker to perform strategic attacks.

Table 4: Mean Absolute Error (MAE) of Reputation Estimation for Dishonest Duopoly Sellers

	Constant	Camouflage	Whitewashing
BRS	0.52±0.05	0.50±0.04	0.74±0.02
iCLUB	0.83±0.21	0.73±0.14	0.80±0.09
TRAVOS	0.42±0.03	0.56±0.01	0.57±0.02
ReferralChain	0.05±0.01	0.15±0.01	0.59±0.03
Personalized	0.45±0.08	0.46±0.07	0.83±0.03
MET	**0.02±0.01**	**0.02±0.01**	**0.03±0.01**
	Sybil	Sybil Cam*	Sybil WW*
BRS	0.73±0.03	0.67±0.03	0.61±0.01
iCLUB	**0.06±0.01**	0.70±0.10	**0.06±0.01**
TRAVOS	0.29±0.01	0.54±0.02	0.56±0.02
ReferralChain	0.08±0.02	0.19±0.02	0.68±0.04
Personalized	0.24±0.07	0.59±0.08	0.24±0.02
MET	0.07±0.04	**0.11±0.02**	0.20±0.06

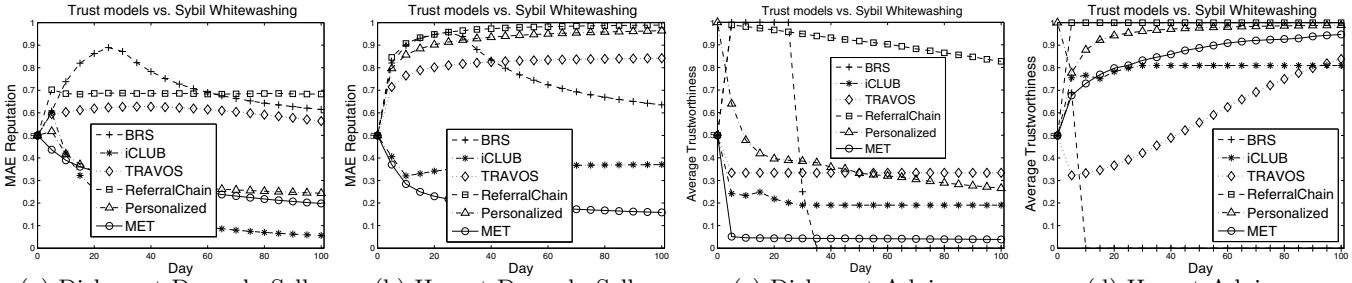
*Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

Table 5: Mean Absolute Error (MAE) of Reputation Estimation for Honest Duopoly Sellers

	Constant	Camouflage	Whitewashing
BRS	0.19±0.06	0.11±0.04	0.58±0.02
iCLUB	**0.01±0.00**	**0.01±0.00**	0.11±0.07
TRAVOS	0.17±0.01	0.25±0.01	0.28±0.01
ReferralChain	0.06±0.02	0.16±0.01	0.97±0.03
Personalized	0.02±0.00	**0.01±0.00**	0.06±0.01
MET	**0.01±0.00**	**0.01±0.00**	**0.05±0.03**
	Sybil	Sybil Cam*	Sybil WW*
BRS	0.99±0.00	0.73±0.01	0.64±0.01
iCLUB	0.35±0.17	**0.01±0.00**	0.37±0.14
TRAVOS	0.44±0.02	0.57±0.01	0.84±0.01
ReferralChain	0.10±0.02	0.19±0.02	0.99±0.01
Personalized	0.20±0.21	0.05±0.00	0.96±0.02
MET	**0.09±0.06**	0.08±0.02	**0.16±0.11**

*Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

For Sybil and Sybil Whitewashing, iCLUB gets the best results on the dishonest duopoly seller ($\text{MAE}(S^D) = 0.06$), whereas it is unable to accurately estimate the honest duopoly seller's reputation ($\text{MAE}(S^H) = 0.37$). It is consistent with the results of robustness comparison in the previous section. To explain, when a buyer conducts enough transactions with the dishonest duopoly seller, iCLUB adopts the buyer's local knowledge to calculate the dishonest seller's reputation. However, in some cases, the buyer has little evidence about the honest duopoly seller, and iCLUB has to rely on global knowledge to calculate the honest seller's reputation. When majority of advisors are dishonest, iCLUB suggests the buyer to transact with the dishonest duopoly seller rather the honest one, and then a rating will give to the dishonest duopoly seller. The consequence is that the buyer will still not have sufficient experience with the honest duopoly seller to accurate model this seller's reputation.



(a) Dishonest Duopoly Seller (b) Honest Duopoly Seller (c) Dishonest Advisors (d) Honest Advisors

Figure 3: (a-b): MAE of Duopoly Sellers' Reputation; (c-d): Average Trustworthiness of Advisors

Fig. 3(a-b) show the MAE of duopoly sellers' reputation along days by trust models against Sybil Whitewashing. MET and iCLUB can get small MAE values for modeling both duopoly sellers' reputation. Thus, iCLUB and our MET model are more effective than other four trust models in predicting sellers' reputation.

4.6 Trustworthiness of Advisors

We also show how the different trust models can accurately model the trustworthiness of both honest and dishonest advisors, under the Sybil Whitewashing attack[4]. Fig. 3(c-d) show the average trustworthiness of dishonest and honest advisors modeled by honest buyers, respectively. BRS assigns both dishonest and honest advisors' trustworthiness as zero after day 35 because it filters out all the advisors, which is similar to the reason why BRS is vulnerable to Whitewashing in Section 4.4. Although ReferralChian assigns the trustworthiness value 1 to the honest advisors, it maintains the average trustworthiness of dishonest advisors at a high level (i.e., larger than 0.8 in Fig. 3(c)). As shown in Fig. 3(c-d), iCLUB, TRAVOS and Personalized are able to reduce or increase the average trustworthiness of dishonest or honest advisors, respectively. However, these three trust models cannot enforce the difference in dishonest and honest advisors' average trustworthiness to a large extent. In contrast, MET is more effective than those trust models for modeling the trustworthiness of advisors.

5. CONCLUSION AND FUTURE WORK

In this paper, a novel multiagent evolutionary trust (MET) model is proposed for constructing robust and accurate trust networks for buyers to accurate model seller reputation in the presence of unfair rating attacks. Each buyer in our model evolves its own trust network by asking its advisors to provide their trust network information. By doing so, the buyer is able to select suitable advisors into its trust network, and assign accurate trustworthiness to these advisors simultaneously. Experimental studies confirm that MET is more robust and effective than the state-of-the-art trust models against various unfair rating attacks.

For future work, we will examine how the combination of sellers' cheating behaviors and advisors' unfair ratings will impact trust models. We also plan to build a comprehensive testbed to evaluate the robustness of trust models by incorporating more intelligent attacks.

[4]We choose the Sybil Whitewashing attack because it is the strongest attack among the six typical attacks.

6. ACKNOWLEDGEMENT

This work is supported by the Ministry of Education Academic Research Fund Tier 1 Grant Singapore (M4010265 RG15/10) awarded to Dr. Jie Zhang.

7. REFERENCES

[1] S. Jiang, J. Zhang, and Y. Ong. A multiagent evolutionary framework based on trust for multiobjective optimization. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 299–306, 2012.

[2] A. Jøsang. Robustness of trust and reputation systems: Does it matter? In *Proceedings of the 6th IFIP International Conference on Trust Management (IFIPTM)*, pages 253–262, 2012.

[3] S. Liu, J. Zhang, C. Miao, Y. Theng, and A. Kot. iCLUB: an integrated clustering-based approach to improve the robustness of reputation systems. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011.

[4] J. Pugh and A. Martinoli. Multi-robot learning with particle swarm optimization. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2006.

[5] R. A. Sarker and T. Ray. *Agent-Based Evolutionary Search*. Springer, 2010.

[6] W. Teacy, J. Patel, N. Jennings, and M. Luck. TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

[7] A. Whitby, A. Jøsang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of the 3rd International Joint Conference on Autonomous Agenst Systems Workshop on Trust in Agent Societies (AAMAS)*, 2004.

[8] B. Yu and M. Singh. Detecting deception in reputation management. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2003.

[9] J. Zhang. *Promoting Honesty in E-Marketplaces: Combining Trust Modeling and Incentive Mechanism Design*. PhD thesis, University of Waterloo, 2009.

[10] L. Zhang, S. Jiang, J. Zhang, and W. Ng. Robustness of trust models and combinations for handling unfair ratings. In *Proceedings of the 6th IFIP International Conference on Trust Management (IFIPTM)*, volume 374, pages 36–51, 2012.