Indirect Trust is Simple to Establish

Elham Parhizkar[*] , **Mohammad Hossein Nikravan** and **Sandra Zilles**
Department of Computer Science, University of Regina, Canada
e.parhizkarabyaneh@gmail.com, nikravam@uregina.ca, zilles@cs.uregina.ca

Abstract

In systems with multiple potentially deceptive agents, any single agent may have to assess the trustworthiness of other agents in order to decide with which agents to interact. In this context, indirect trust refers to trust established through third-party advice. Since the advisers themselves may be deceptive or unreliable, agents need a mechanism to assess and properly incorporate advice. We evaluate existing state-of-the-art methods for computing indirect trust in numerous simulations, demonstrating that the best ones tend to be of prohibitively large complexity. We propose a new and easy to implement method for computing indirect trust, based on a simple *prediction with expert advice* strategy as is often used in online learning. This method either competes with or outperforms all tested systems in the vast majority of the settings we simulated, while scaling substantially better. Our results demonstrate that existing systems for computing indirect trust are overly complex; the problem can be solved much more efficiently than the literature suggests.

1 Introduction

Effective collaboration among agents in multi-agent systems (MAS) often requires the agents to assess one another's trustworthiness. Typically, one represents an agent's trustworthiness as a numeric value, e.g., denoting the probability that an interaction with the agent is successful. A truster is an agent that tries to estimate the trustworthiness of another agent, called the trustee, and typically derives its estimate partly from the following types of information [Jøsang *et al.*, 2007]: *direct trust* information, which is based on the past experience from the truster's direct interactions with the trustee; *indirect* trust information, which is based on recommendations on the trustee, provided by other agents, called advisers.

Direct trust information may be unavailable or unreliable in the following situations: (i) in MAS with a large number of trustees, (ii) when a truster or a trustee joins the system as a new user, (iii) when the behavior of trustees changes dynamically. In such cases, the trustworthiness of a trustee may be estimated based on recommendations. This poses the problem of detecting and handling unreliable recommendations made by advisers that are subjective, deceptive, or negligent, as these can have an adverse effect when estimating the trustworthiness of trustees [Jøsang *et al.*, 2007].

In this paper, we therefore focus exclusively on the problem of computing indirect trust, cast as follows. A truster tries to estimate the trustworthiness of trustees based on indirect trust when it has *no history of interaction* in the MAS and the only information shared by advisers are trustworthiness estimates of individual trustees. In each step, the truster receives recommendations about trustees, picks one trustee to interact with, observes the outcome, and updates its indirect trust estimate for that trustee. The goal is (a) to estimate the trustworthiness of trustees, where individual advisers may vary in terms of their reliability; and (b) to minimize the number of negative outcomes from interactions with trustees.

To study the effectiveness of computing *indirect trust*, we deliberately ignore additional information, such as stereotypes [Burnett *et al.*, 2010] or organizational information [Kollingbaum and Norman, 2002].

Various indirect trust models were proposed to cope with unreliable advisers [Teacy *et al.*, 2006; Regan *et al.*, 2006; Teacy *et al.*, 2012; Jiang *et al.*, 2013; Yu *et al.*, 2014; Whitby *et al.*, 2004]. We simulate one new and four existing systems over a large variety of scenarios. Our results show that *on average* most systems behave equally well. In particular, this raises doubts in published simulations. For example, the system MET [Jiang *et al.*, 2013] claims to beat TRAVOS [Teacy *et al.*, 2006], based on simulations with a fixed distribution of trustee behavior, while our simulations, averaged over 100 random trustee behavior distributions, show that both methods fare equally well on average, for each of a large variety of chosen attack scenarios. For each pair of systems, one can find specific distributions in which one outperforms the other and vice versa; our first contribution hence is to show that the simulations presented in published studies are not comprehensive enough to assess how well a system can handle a specific kind of attack.

While, in our simulations, MET and TRAVOS *on average* clearly outperform the other two existing methods we tested, neither MET nor TRAVOS would be practical in large-scale

[*] Contact Author

MAS: MET's runtime is prohibitively large, while TRAVOS' need to store all past recommendations by all advisers makes it infeasible. This raises the question whether indirect trust can be computed effectively by a method that scales well both in terms of runtime and in terms of memory.

The answer is yes. Our second contribution is to demonstrate that existing systems solving the indirect trust problem are overly complex, by providing a very simple and efficient, yet highly effective indirect trust method. Our proposed method ITEA (Indirect Trust with Expert Advice) establishes indirect trust through an online learning algorithm. More specifically, ITEA is inspired by the *predicting from expert advice* (PEA) model [Littlestone and Warmuth, 1994], in which the learner aggregates predictions made by a group of experts (advisers) in a weighted average; the weights are updated based on the most recent predictions. PEA is conceptually very simple, so that ITEA is much simpler than all existing methods for deriving indirect trust, easier to implement, and more efficient in terms of both runtime and memory.

In our simulations, ITEA on average competes with MET and TRAVOS, suggesting that, against the trend in the literature, it is possible to solve the indirect trust problem with a method that is practical for large-scale MAS.

2 Related Work

Various indirect trust mechanisms have been proposed [Teacy *et al.*, 2006; Regan *et al.*, 2006; Jiang *et al.*, 2013; Yu *et al.*, 2014; Teacy *et al.*, 2012; Yu and Singh, 2003; Irissappane and Zhang, 2017; Liu *et al.*, 2017; Liu *et al.*, 2011; Cohen *et al.*, 2018; Weng *et al.*, 2010; Jøsang and Ismail, 2002]. Some assume that reliable advisers are in the majority, and thus consider all advisers whose recommendations display different statistical properties than the majority as unreliable [Jøsang and Ismail, 2002]. Others identify unreliable recommendations by comparing them with a truster's direct experience [Teacy *et al.*, 2006; Yu and Singh, 2003]. Unfair recommendations are then filtered out [Jøsang and Ismail, 2002; Jiang *et al.*, 2013; Yu *et al.*, 2014], discounted [Teacy *et al.*, 2006], or re-interpreted by learning correlations between a truster's direct experience and an adviser's ratings [Regan *et al.*, 2006; Teacy *et al.*, 2012]. In the following, we give a brief overview of some of the state-of-the-art techniques.

TRAVOS [Teacy *et al.*, 2006] models trustworthiness probabilistically, through a Beta distribution computed from the outcomes of all the interactions a truster has observed. It discounts unfair recommendations by assigning weights according to the probability of their accuracy, which is estimated based on the similarity between current and past ratings.

The Bayesian method BLADE [Regan *et al.*, 2006] models correlations between the direct experience of a trustee and recommendations of an adviser. HABIT [Teacy *et al.*, 2012] extends BLADE by analyzing correlations of the behavior within groups of trustees. It was claimed to outperform BLADE, even without correlation between trustees' behavior.

By contrast, MET [Jiang *et al.*, 2013] uses an evolutionary model to assess advisers. In MET, every truster has a network consisting of a set of advisers and their trustworthiness; an evolutionary algorithm develops these networks over time.

Our newly designed system, ITEA, was most inspired by ACT [Yu *et al.*, 2014], which infers a trust model by reinforcement learning. We conjectured that the reinforcement learning approach deployed by ACT is more complex than needed and that similar tasks as ACT's can be solved more easily with a method that uses a very simple online learning approach. Consequently, we chose ACT as one of the methods to which to compare ITEA.

Some of the principles behind reinforcement learning are akin to those used in evolutionary algorithms, and hybrid methods were proposed for various applications, cf. [Drugan, 2019]. Therefore, we selected MET's evolutionary approach as a competing model as well. MET was claimed to outperform iCLUB [Liu *et al.*, 2011], ReferralChain [Yu and Singh, 2003], and Zhang's Personalized model [Zhang, 2009].

The online learning paradigm on which ITEA is built makes no assumptions about stochasticity of the information sources, and thus stands in sharp contrast to probabilistic modelling. We therefore decided to further compare ITEA to a state-of-the-art probabilistic approach. To the best of our knowledge, HABIT is the best-performing such method to date. While TRAVOS was claimed to be inferior to MET [Jiang *et al.*, 2013], it is an influential and highly cited system, so that we include it in our evaluation (it turns out that on average, it actually is on par with MET). With HABIT, MET, and ACT, our empirical tests include three systems that have, to the best of our knowledge, so far not been defeated in an experimental study.

Some of the most highly cited trust models, such as EigenTrust [Kamvar *et al.*, 2003], PeerTrust [Xiong and Liu, 2004], and PowerTrust [Zhou and Hwang, 2007], cannot be included in our study since they are too restrictive to handle the settings in our simulations. For example, Eigentrust relies on a set of so-called "pre-trusted peers", i.e., trustees known to be trustworthy. Moreover, it assumes that the probability with which a specific adviser makes a reliable recommendation on a trustee is independent of the trustee. Neither of these two restrictive assumptions applies to our simulation testbed.

3 Indirect Trust With Expert Advice

We will compare existing indirect trust systems to a new and very simple one proposed in this section. Our method is called *Indirect Trust with Expert Advice (ITEA)*, as it is based on the following online learning paradigm, dubbed *prediction from expert advice (PEA)* [Littlestone and Warmuth, 1994].

Suppose a learner aims to predict an unknown sequence $o_1, \ldots, o_T$ of outcomes from an outcome space $\mathcal{O}$. At time t, the learner has access to predictions $f_{1,t}, \ldots, f_{K,t}$ made by a set of K experts, where $f_{k,t}$ refers to the prediction made by the kth expert at time t. The learner then makes a prediction $\hat{p}_t$. All predictions belong to the same decision space $\mathcal{D}$. The true outcome o_t is revealed after the learner's prediction.

The learner's and the experts' predictions are then evaluated using a loss function $\ell : \mathcal{D} \times \mathcal{O} \to \mathbb{R}_{\geq 0}$, so that the total loss accumulated over the first t steps is calculated as follows:

$$L_t = \sum\nolimits_{r=1}^{t} \ell(\hat{p}_r, o_r), \quad L_{k,t} = \sum\nolimits_{r=1}^{t} \ell(f_{k,r}, o_r) \,,$$

where L_t and $L_{k,t}$ are the cumulative loss of the learner and the kth expert, resp. The learner's goal is to minimize the *regret* $R_T = L_T - \min_{1\leq k\leq K} L_{k,T}$ over T time steps, i.e., the difference between its own cumulative loss and that of the best expert in hindsight. Sub-linear regret, $R_T = o(T)$, would let R_T become insignificant relative to T, for large T.

Littlestone and Warmuth (1994) proposed a simple weighted average method for solving this problem. We use a popular special case of this method, namely the *Exponential-Weighted Average* strategy, in which, at time t, the learner predicts

$$\hat{p}_t = \frac{\sum_{k=1}^{K} w_{k,t-1} f_{k,t}}{\sum_{k=1}^{K} w_{k,t-1}} ,$$

where $w_{k,t} = w_{k,t-1} \cdot \exp^{-\eta\ell(f_{k,t},y_t)}$, $\eta > 0$, is the weight assigned to the expert k at time t. Note that larger weights are assigned to experts with lower regrets. When ℓ is convex in its first argument and has values in $[0,1]$, by choosing $\eta = \sqrt{8 \cdot \ln(K)/T}$, the sublinear regret bound $R_T \leq \sqrt{T \ln(K)/2}$ is achieved [Cesa-Bianchi and Lugosi, 2006].

In our MAS framework, we denote the set of trusters (e.g., service consumers) by $C = \{c_i \mid i = 1, \ldots, N\}$, the set of trustees (e.g., service providers) by $S = \{s_j \mid j = 1, \ldots, M\}$, and the set of advisers by $A = \{a_k \mid k = 1, \ldots, K\}$. A truster c_i in ITEA corresponds to a learner in the PEA model, while an adviser a_k is modeled as an expert in PEA. An adviser $a_k \in A$, at time t, may provide a recommendation $f_{k,t}(j)$ which could be its own direct trust value for s_j (in case of a reliable adviser) or a distorted version thereof (in case of an unreliable adviser). The outcome of an interaction between c_i and s_j at time t, denoted as $o_t(i,j)$, is either successful ($o_t(i,j) = 1$) or unsuccessful ($o_t(i,j) = 0$).

ITEA's pseudo-code is given in Algorithm 1. In each of T rounds, the following is executed. For each trustee s_j, the truster c_i receives a recommendation about s_j from each of the K advisers (line 3) and computes an indirect trust value of s_j using a weighted average (line 4), where the initial weights are all $\frac{1}{K}$. It attempts to minimize its loss by picking the trustee s_{j^*} with the highest indirect trust estimate (line 6). Note that the trust estimate $\hat{p}_t(i,j^*)$ is c_i's estimate of the probability that an interaction with s_{j^*} will be successful, based on indirect trust. The true outcome observed (successful or unsuccessful interaction with s_{j^*}, cf. line 7) is either 1 or 0, and the loss $\ell(f_{k,t}(j^*), o_t(i,j^*))$ of adviser a_k's prediction is observed, for each $k \in \{1, \ldots, K\}$ (line 8). The weights of the advisers are then adjusted depending on their losses (line 9). Note that truster c_i simultaneously solves M PEA problems, one for each trustee.

Interesting properties of this algorithm include:

- ITEA does not need to memorize individual losses incurred by advisers in previous rounds; the weights reflect the past performance of advisers in a cumulative way. This makes ITEA very simple to implement, as well as efficient in terms of memory and runtime.
- The weights assigned to advisers depend on the trustee, so that the system can handle advisers that are more reliable for some trustees than for others. This is important in scenarios where advisers want to, e.g., bad-mouth some trustees while being truthful about some others.

Algorithm 1 The ITEA Algorithm

```
1:  Initialization: $w_{k,0} = \frac{1}{K}, 1 \leqslant k \leqslant K$
2:  for $t = 1$ to $T$ do
3:      for $j = 1$ to $M$ do
4:          $c_i$ receives advisers' recommendations $f_{1,t}(j), f_{2,t}(j), \ldots, f_{K,t}(j) \in [0,1]$.
5:          $c_i$ makes its own prediction $\hat{p}_t(i,j) \in [0,1]$:
                $\hat{p}_t(i,j) = \frac{\sum_{k=1}^{K} w_{k,t-1}(j)\ f_{k,t}(j)}{\sum_{k=1}^{K} w_{k,t-1}(j)}$.
6:      end for
7:      $c_i$ picks $s_j$ with highest $\hat{p}_t(i,j)$ denoted by $s_{j^*}$ for interaction.
8:      $c_i$ observes outcome $o_t(i,j^*) \in \{0,1\}$ of interaction.
9:      $c_i$ suffers loss $\ell(\hat{p}_t(i,j^*), o_t(i,j^*))$;
        adviser $a_k$, $1 \leq k \leq K$, suffers loss $\ell(f_{k,t}(j^*), o_t(i,j^*))$.
10:     $c_i$ updates weights of advisers:
            $w_{k,t}(j^*) = w_{k,t-1}(j^*) \cdot \exp^{-\eta\ell(f_{k,t}(j^*),O_t(i,j^*))}$,
        where $\eta > 0$ is the learning rate, fixed in advance.
11: end for
```

- ITEA does not assume any stochasticity in the behavior of the advisers; it is therefore very flexible in handling even dynamically changing adviser behaviors.

4 Simulation Setup

We compared ITEA to the methods TRAVOS [Teacy *et al.*, 2006], MET [Jiang *et al.*, 2013], HABIT [Teacy *et al.*, 2012], and ACT [Yu *et al.*, 2014] in terms of two measures.

Mean Absolute Error. Typically, the trustworthiness of a trustee is represented by a number in $[0,1]$, where the value 1 (0, resp.) indicates perfect honesty (complete dishonesty, resp.) Trust systems compute estimates of trustworthiness values, and MAE refers to the mean absolute difference between the actual and the estimated values, where the mean is taken over all pairs of truster and trustee. We will report MAE always after a fixed number of interactions.

Relative Frequency of Unsuccessful Interactions. MAE ignores utility aspects such as cost. If a negative (unsuccessful) interaction is more costly than a positive (successful) interaction, MAE alone is not sufficient for assessing a system. Hence, we will evaluate each system based on how many negative interactions it makes in order to complete a fixed target number of positive interactions. Specifically, we report the fraction of the number of negative interactions over the total number of interactions, which we call Relative Frequency of Unsuccessful Interactions (RFU). Yu *et al.* (2014) reported a measure called NAUL that is in essence equivalent to RFU.

4.1 Trustees and Trustworthiness

In our simulations, interactions with a trustee are either positive or negative, i.e., the outcomes are binary. All interactions are random events that are independent of one another.

Each trustee has a constant trustworthiness value in $[0, 1]$ corresponding to the probability of a positive outcome when interacting with that trustee.

Our simulations use one truster, 10 trustees, and 100 advisers. The trustworthiness value of each individual trustee is sampled uniformly at random from the values $0.1, 0.2, \ldots, 0.9$; each reported result is the average over 100 such sampled trustee combinations. For preprocessing, we let all advisers (but not the truster) interact with the trustees so that they can establish direct trust information about the trustees. We want that information to be near-accurate, to be able to simulate reliable advisers alongside unreliable ones. Hence, in preprocessing, we execute 300,000 interactions each of which involves an adviser randomly chosen from the pool of 100 and a trustee randomly chosen from the pool of 10. Each adviser records, for each trustee s_j, the number of positive and the number of negative interactions it has had with s_j.

Advisers then compute their (direct) trust in a trustee using the Beta Reputation System (BRS) [Jøsang and Ismail, 2002], which models the trust of an agent in the trustee s_j as

$$\mathrm{brs}(p_j, n_j) = \frac{p_j + 1}{p_j + n_j + 2} \ (\in (0, 1)) ,$$

where p_j (n_j, resp.) is the number of positive (negative, resp.) interactions between the agent and s_j from the preprocessing phase. An honest adviser asked for a recommendation on trustee s_j will simply report the pair (p_j, n_j).[1]

4.2 Adviser Settings

Our evaluation comprises various settings of how advisers can distort the actual values of p and n.

Setting 1: Partly Random Advisers. A partly random adviser first picks trustees for which it will provide randomly distorted recommendations; each trustee has a 50% chance of being picked. About all other trustees, that adviser will always be honest. For each trustee s_j that was picked, the adviser randomly selects a number $z \in (0, 1)$, computes any pair (p'_j, n'_j) of non-negative integers such that $\mathrm{brs}(p'_j, n'_j) = z$, and will subsequently always report (p'_j, n'_j) about s_j, irrespective of s_j's actual trustworthiness.

Setting 2: Badmouthing (BM)/Ballot-Stuffing (BS) Advisers. A BM/BS adviser first picks trustees for which it will always provide distorted recommendations; each trustee has a 50% chance of being picked. About all other trustees, that adviser will always be honest. The distorted recommendation (p, n) is computed as follows: the adviser compares the pairs (p_j, n_j) recorded in the preprocessing phase for each trustee s_j and returns the pair with the lowest (in case of BM) or the highest (in case of BS) value of $\mathrm{brs}(p_j, n_j)$.

Setting 3: Additive BM/BS Advisers. This scenario is adapted from [Yu *et al.*, 2014]. An additive BM adviser does the following independently for each trustee s_j: it first samples a random number $z \in [0.8, 1]$ and subtracts z from its own direct trust in s_j, i.e., it computes $z^* = \frac{p_j+1}{p_j+n_j+2} - z$. If $z^* > 0$, a pair (p, n) with $\mathrm{brs}(p, n) = z^*$ is returned, else $(0, p_j + n_j)$ is returned, in which case the adviser says all its interactions with s_j were negative. An additive BS adviser uses $z^* = \frac{p_j+1}{p_j+n_j+2} + z$. If $z^* < 1$, a pair (p, n) with $\mathrm{brs}(p, n) = z^*$ is returned, else $(p_j + n_j, 0)$ is returned, as if all interactions of the adviser with s_j had been positive.

Setting 4: All-Negative/All-Positive Advisers. An all-negative (all-positive, resp.) adviser reports $p = 0$ and $n = 1,000,000$ ($p = 1,000,000$ and $n = 0$, resp.) for each trustee, irrespective of the adviser's interaction history.

Setting 5: Fully Random Advisers. These advisers act like partly random advisers (see Setting 1), except that they randomly distort the recommendations for *all* trustees.

Setting 6: Selective BM/BS Advisers. A selective BM adviser is honest (i.e., reports (p_j, n_j)) for each trustee s_j with $\mathrm{brs}(p_j, n_j) < 0.5$. For the remaining s_j, it reports $(0, p_j + n_j)$, as if each interaction with s_j had been negative. In the BS case, trustees s_j with $\mathrm{brs}(p_j, n_j) > 0.5$ are reported honestly; all others are reviewed with $(p_j + n_j, 0)$ as if each interaction with them had been positive.

Whitewashing and Camouflage Attacks. The above settings were tested separately with the whitewashing and camouflage attacks as detailed in [Jiang *et al.*, 2013]. Whitewashing requires the system to handle new advisers at any point in time; in ITEA, the weight assigned to a new adviser is set to the mean of the weights of all advisers after the previous step, weights are then normalized as usual.

4.3 Further Details

In our experiments, the truster computes direct trust (from its actual interactions with trustees) using BRS. It starts off with an empty history, i.e., its direct trust information for each trustee s_j is $(p^*_j = 0, n^*_j = 0)$. The experiment then proceeds in rounds, each round consisting of the following steps:

1. The truster simulates a system (e.g., MET) to interrogate advisers and to compute indirect trust for each trustee.
2. The trustee s_j for which the computed indirect trust is largest is chosen for interaction (in the case of more than one candidate, one will be chosen at random).
3. The outcome of the interaction with s_j is used to update the direct trust information (p^*_j, n^*_j); if the interaction was positive, p^*_j is incremented, else n^*_j is incremented.

We force the truster to prioritize trustees for interaction only by indirect trust. Direct trust is used only to initialize advisers and to detect unreliable advisers. Our reason for ignoring direct trust at the selection step is that we want to evaluate the effect of *indirect trust* as computed by the systems.

We did not tune any parameters for any of the systems. For the state-of-the-art approaches, we used the same parameter settings as reported in the experiments in the respective publications, with one small exception: For MET [Jiang *et al.*, 2013], the authors report using $n = 25$ advisers for the size of what they call a trust network. This was out of a total of $K = 40$ advisers. Since our simulations used $K = 100$

[1] In Algorithm 1, an adviser reports a single number in the range $[0, 1]$, not a pair of integers. In this case, and whenever we talk about indirect trust estimates as numbers in the range $[0, 1]$, we implicitly map the pair $(p, n) \in \mathbb{N} \times \mathbb{N}$ to the number $\mathrm{brs}(p, n) \in (0, 1)$.

advisers, we set $n = 55$ for MET to make sure a trust network always contains more than half of the available advisers. HABIT can be set up in various ways; we chose the DP-Dirichlet model [Teacy *et al.*, 2012].

ITEA requires a loss function and a learning rate. As loss function, we chose squared-error loss. As mentioned before, the learning rate $\eta = \sqrt{8\,ln(K)/T}$ yields good guarantees, where K is the number of advisers and T the number of rounds of interaction with the trustee whose trustworthiness is being estimated. In our simulations, $K = 100$ and T is not known in advance, so that we replaced T by the total number of interactions (when measuring MAE) or the target number of positive interactions (when measuring RFU).

5 Results

For each setting, we report simulations with three different percentages of unreliable advisers (40%, 70%, 90%), which were chosen at random from the set of all advisers.

5.1 RFU

To calculate RFU, we ran each system with a target number S of successful interactions, i.e., after S successful interactions, the ratio of the number U of unsuccessful interactions over $U + S$ was recorded. The average such ratio over 100 runs is reported in Table 1 for $S = 50$. We also tried $S = 100$ in most settings, but the results were almost identical. Every table entry is a pair *a*/*b*, where a refers to the chosen setting without whitewashing, and b with whitewashing.

Due to space constraints, results for HABIT are not listed; it significantly lost to all other methods in all settings, except Setting 6-BM, where it beat the four other methods by far. We conjecture that HABIT's strength lies mainly in its handling of direct trust and less in its indirect trust calculation.

We ran paired t-tests for ITEA in comparison to each of the other methods. If ITEA is never significantly beaten, its entry is shown in bold. An entry for ACT, TRAVOS, or MET is in bold if it is the single significant winner or, if no single winner exists and it is not significantly worse than ITEA.

All methods handled the camouflage attack very well. In all settings, the RFU values with and without camouflage attack were almost identical, so that we do not list the camouflage results. The only exception was that in Settings 5 and 6, TRAVOS behaved *better* with camouflage than without.

None of the methods showed strong negative reactions to whitewashing attacks; sometimes RFU worsened slightly for ACT, for MET in Setting 4, and for ITEA in Setting 6-BS. Occasionally, whitewashing affects RFU positively, e.g., for MET in Setting 6-BS. We did not study individual trustee distributions in which the RFU for whitewashing was much worse than that without. Given that on average whitewashing did not cause any serious trouble for any of the tested methods, we leave a more detailed analysis for future work.

Settings 1 and 2 were easy to handle for all methods (except HABIT), both with and without whitewashing attack; only ACT had slight difficulties with Setting 2-BM. These difficulties become more prominent in Setting 3. In Setting 4, MET significantly beats all other methods, while ACT is the clear loser again; TRAVOS and ITEA are on par with each other. Setting 5 is won by ITEA and TRAVOS, which substantially beat MET and ACT for large pools of deceptive advisers. Finally, Setting 6 witnesses that all of the methods that we tried can be fooled greatly. ITEA and TRAVOS though handle Setting 6-BS better than ACT and MET.

We do not report standard deviation in Table 1, as it is not meaningful given the variety of trustworthiness distributions. RFU Scatterplots for MET and ITEA (omitted due to space constraints) show that, except for Setting 5, these methods mostly agree on which trustee distributions are difficult to handle, but typically with a few outliers in each direction.

We declare ITEA, TRAVOS, and MET the winners in terms of RFU, among the methods tested here. We proceed with a study of MAE only for these three methods.

5.2 MAE

Table 2 shows MAE results after a total of 50 interactions (for 90% deceptive advisers only), which seem more erratic. Without whitewashing, ITEA wins in Settings 1, 2-BM, and 5, while TRAVOS wins in Settings 3, 4, and 6; MET doesn't compete. With whitewashing, MET wins for smaller percentages of deceptive advisers (not displayed), as well as in Settings 1, 2-BS, and 5, whereas TRAVOS wins for 90% whitewashing advisers in Settings 3, 4, and 6. Whitewashing negatively affects ITEA's MAE. Overall, there is no single winner.

A comparison of Tables 1 and 2 shows that good MAE is not necessary for good RFU. Intuitively, a system may perform many successful interactions if it can identify even a single highly trustworthy trustee. To do so, it is not required to estimate the trustworthiness of *all* trustees well. In many applications (e.g., e-market places) it is of crucial importance to perform mostly successful interactions. In such applications, one should put more emphasis on RFU than on MAE.

5.3 Scalability

Asymptotically, ITEA's runtime scales optimally with the number M of trustees and the number K of advisers, under the intuitive assumption that a system should consider, for each trustee, a number of recommendations that grows linearly with the number of advisers. ITEA's runtime is in $\Theta(MK)$, as is easily verified from Algorithm 1. The same appears to be true for TRAVOS and ACT. By comparison, MET requires $\Omega(MK^2)$ runtime in order to compare and select advisers for a trust network, which requires evaluating a fitness function. For HABIT, which takes the group behavior of trustees into account, the runtime is in $\Omega(M^2K)$.

We measured CPU time on a MacBook Pro/2.3 GHz Intel Core i7 for various settings. As the trends overall were identical, Table 3 reports the results averaged over 100 runs for Setting 2. As expected, ITEA substantially beats our implementations of the other methods in terms of runtime. ACT is also very fast, but could not compete with ITEA in terms of RFU. The RFU competitors TRAVOS and MET are 50 times and 238 times slower than ITEA, resp.

While TRAVOS is slower than ITEA, asymptotically both systems scale well in terms of runtime. However, TRAVOS stores all recommendations by all advisers over time, which makes it infeasible in terms of memory consumption. ITEA's memory consumption, by comparison, is negligible.

S1	90%	70%	40%	S4-BM	90%	70%	40%
ITEA	**0.150 / 0.144**	**0.139 / 0.136**	0.142 / **0.141**	ITEA	0.153 / **0.150**	0.151 / **0.147**	0.145 / **0.142**
ACT	**0.165 / 0.150**	0.153 / **0.139**	0.151 / **0.141**	ACT	0.203 / 0.182	0.191 / 0.176	0.179 / **0.146**
TRAV	0.160 / **0.144**	0.160/ **0.140**	0.149 / **0.141**	TRAV	0.153 / 0.161	0.153 / 0.159	0.153 / 0.159
MET	**0.149 / 0.142**	**0.140 / 0.143**	**0.127 / 0.130**	MET	**0.127 / 0.146**	**0.131 / 0.141**	**0.130 / 0.142**
S2-BM	90%	70%	40%	S4-BS	90%	70%	40%
ITEA	**0.138 / 0.133**	**0.151 / 0.147**	**0.135 / 0.133**	ITEA	0.159 / **0.142**	0.149 / **0.140**	0.142 / **0.140**
ACT	0.179 / 0.152	0.171 / **0.145**	0.166 / **0.140**	ACT	0.433 / 0.520	0.267 / 0.331	0.159 / 0.169
TRAV	**0.144 / 0.133**	0.168 / 0.157	**0.146** / 0.143	TRAV	0.145 / **0.145**	0.144 / **0.144**	0.143 / **0.143**
MET	0.155 / 0.149	**0.140 / 0.137**	**0.136 / 0.136**	MET	**0.130 / 0.143**	**0.131 / 0.142**	**0.131 / 0.141**
S2-BS	90%	70%	40%	S5	90%	70%	40%
ITEA	**0.129 / 0.129**	0.145 / 0.145	**0.133 / 0.133**	ITEA	**0.168 / 0.197**	**0.163** / 0.155	**0.136 / 0.133**
ACT	**0.139** / 0.141	**0.140 / 0.143**	**0.138 / 0.136**	ACT	0.254 / **0.196**	0.206 / 0.165	0.174 / **0.139**
TRAV	0.133 / **0.129**	0.150 / 0.145	**0.135 / 0.133**	TRAV	**0.169 / 0.192**	**0.172** / 0.171	0.142 / 0.141
MET	**0.136 / 0.138**	**0.131 / 0.133**	**0.134 / 0.135**	MET	0.294 / **0.213**	**0.169 / 0.140**	**0.138 / 0.135**
S3-BM	90%	70%	40%	S6-BM	90%	70%	40%
ITEA	**0.145 / 0.143**	0.151 / 0.147	**0.149 / 0.141**	ITEA	0.616 / 0.618	0.593 / 0.579	**0.161 / 0.148**
ACT	0.204 / 0.163	0.184 / 0.175	0.168 / **0.152**	ACT	**0.531 / 0.535**	**0.246** / 0.208	**0.184 / 0.145**
TRAV	**0.152** / 0.164	0.155 / 0.166	**0.154** / 0.159	TRAV	0.589 / 0.585	0.522 / 0.551	**0.159** / 0.163
MET	**0.139 / 0.145**	**0.136 / 0.132**	**0.138 / 0.140**	MET	0.602 / 0.593	0.561 / **0.183**	**0.146 / 0.141**
S3-BS	90%	70%	40%	S6-BS	90%	70%	40%
ITEA	**0.154 / 0.141**	0.149 / **0.141**	**0.142 / 0.140**	ITEA	0.233 / 0.553	**0.150** / 0.202	**0.141 / 0.145**
ACT	0.313 / 0.476	0.260 / 0.313	0.171 / 0.159	ACT	0.597 / 0.620	0.421 / 0.360	0.164 / 0.163
TRAV	**0.164** / 0.145	0.158 / **0.143**	**0.146 / 0.140**	TRAV	**0.208 / 0.275**	**0.152 / 0.152**	**0.143 / 0.141**
MET	**0.141 / 0.142**	**0.136 / 0.132**	**0.137 / 0.140**	MET	0.594 / 0.350	0.201 / **0.152**	**0.138 / 0.142**

Table 1: RFU without/with whitewashing; Sx refers to Setting x

	ITEA	TRAVOS	MET
S1	**.090** / .101	.148 / .147	.102 / **.093**
S2-BM	**.137** / .162	.147 / **.150**	.165 / **.152**
S2-BS	**.162** / .162	**.157** / .159	**.162 / .148**
S3-BM	.375 / .430	**.218 / .230**	.447 / .406
S3-BS	.434 / .440	**.235 / .249**	.438 / .398
S4-BM	.379 / .433	**.219 / .232**	.453 / .412
S4-BS	.439 / .445	**.113 / .114**	.444 / .404
S5	**.178** / .199	.188 / .193	.202 / **.185**
S6-BM	.329 / .330	**.266/ .268**	.329 / .300
S6-BS	.267 / .305	**.180 / .238**	.320 / .292

Table 2: MAE without/with whitewashing; ‘Sx’ refers to Setting x

ITEA	ACT	TRAVOS	MET	HABIT
0.228	0.795	11.36	54.27	27.40

Table 3: CPU Time (sec.) for 50 interactions, S2-BM, 90% (values for BS and other percentages were nearly identical)

6 Conclusions

We revealed two shortcomings in the *indirect trust* literature.

First, we showed that empirical studies employing fixed distributions of trustee behavior (as, e.g., the study claiming that MET beats TRAVOS [Jiang *et al.*, 2013]) are misleading. We found distributions of trustees' trustworthiness values (e.g., $[.9, .6, .4, .4, .4, .4, .2, .2]$ for the ten trustees in our simulations) for which TRAVOS was substantially outperformed by MET, ITEA, and HABIT, while *on average*, TRAVOS, MET and ITEA are on par, and outperform HABIT. This suggests that individual trustworthiness distributions are not indicative of a system's performance in general, and many published claims are based on too narrow a set of simulations.

Second, we showed that the best-performing methods from the literature (MET and TRAVOS, based on our simulations) are overly complex, as seen either in runtime or in memory consumption. The very simple method ITEA that we proposed shows that such complexity is unnecessary for solving the indirect trust problem. It is obvious from the design of ITEA that it scales very well; arguably, its asymptotic complexity is optimal for the task for which it is designed.

Our evaluation has some limitations: (i) We did not simulate dynamic changes in the trustworthiness of trustees or in the reliability of advisers. However, since ITEA makes no stochasticity assumptions, we expect it to handle such cases well. (ii) We did not analyze individual distributions of trustees' trustworthiness values that are difficult to handle for one method but easy for another; our tables above present only averages. Examining such distributions will shed more light on the strengths and weaknesses of individual methods.

Note that, for simplicity, we focused on binary interaction outcomes. However, ITEA can be adapted in a straightforward way to non-binary (even continuous) outcomes. More generally, with appropriate modifications, the approach behind ITEA may be of value in various applications, including e-marketplaces, P2P networks, and car-to-car networks, where cars exchange information on, e.g., road trafficability.

Acknowledgements

The authors thank M. Bowling, R. Sutton, and R. Holte for fruitful discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[Burnett *et al.*, 2010] Chris Burnett, Timothy J. Norman, and Katia Sycara. Bootstrapping trust evaluations through stereotypes. In *Proceedings of the 9th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 241–248, 2010.

[Cesa-Bianchi and Lugosi, 2006] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[Cohen *et al.*, 2018] Robin Cohen, Peng F. Wang, and Zehong Hu. Revisiting public reputation calculation in a personalized trust model. In *Proceedings of the 20th International Trust Workshop*, pages 13–24, 2018.

[Drugan, 2019] Madalina M. Drugan. Reinforcement learning versus evolutionary computation: A survey on hybrid algorithms. *Swarm and Evolutionary Computation*, 44:228–246, 2019.

[Irissappane and Zhang, 2017] Athirai A. Irissappane and Jie Zhang. Filtering unfair ratings from dishonest advisors in multi-criteria e-markets: a biclustering-based approach. *Autonomous Agents and Multi-Agent Systems*, 31:36–65, 2017.

[Jiang *et al.*, 2013] Siwei Jiang, Jie Zhang, and Yew-Soon Ong. An evolutionary model for constructing robust trust networks. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 813–820, 2013.

[Jøsang and Ismail, 2002] Audun Jøsang and Roslan Ismail. The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*, pages 2502–2511, 2002.

[Jøsang *et al.*, 2007] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43:618–644, 2007.

[Kamvar *et al.*, 2003] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International Conference on World Wide Web*, pages 640–651, 2003.

[Kollingbaum and Norman, 2002] Martin J. Kollingbaum and Timothy J. Norman. Supervised interaction: creating a web of trust for contracting agents in electronic environments. In *Proceedings of the 1st International Conference on Autonomous Agents and Mmulti-Agent Systems*, pages 272–279, 2002.

[Littlestone and Warmuth, 1994] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[Liu *et al.*, 2011] Siyuan Liu, Jie Zhang, Chunyan Miao, Yin-Leng Theng, and Alex C. Kot. iCLUB: an integrated clustering-based approach to improve the robustness of reputation systems. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1151–1152, 2011.

[Liu *et al.*, 2017] Yuan Liu, Jie Zhang, Quanyan Zhu, and Xingwei Wang. CONGRESS: A hybrid reputation system for coping with rating subjectivity. *IEEE Transactions on Computational Social Systems*, pages 163–178, 2017.

[Regan *et al.*, 2006] Kevin Regan, Pascal Poupart, and Robin Cohen. Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 1206–1212, 2006.

[Teacy *et al.*, 2006] W. T. Luke Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12:183–198, 2006.

[Teacy *et al.*, 2012] W. T. Luke Teacy, Michael Luck, Alex Rogers, and Nicholas R. Jennings. An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling. *Artificial Intelligence*, 193:149–185, 2012.

[Weng *et al.*, 2010] Jianshu Weng, Zhiqi Shen, Chunyan Miao, Angela Goh, and Cyril Leung. Credibility: How agents can handle unfair third-party testimonies in computational trust models. *IEEE Transactions on Knowledge and Data Engineering*, 22:1286–1298, 2010.

[Whitby *et al.*, 2004] Andrew Whitby, Audun Jøsang, and Jadwiga Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of the 7th International Workshop on Trust in Agent Societies*, pages 106–117, 2004.

[Xiong and Liu, 2004] Li Xiong and Ling Liu. Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16:843–857, 2004.

[Yu and Singh, 2003] Bin Yu and Munindar P. Singh. Detecting deception in reputation management. In *Proceedings of the 2nd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 73–80, 2003.

[Yu *et al.*, 2014] Han Yu, Zhiqi Shen, Chunyan Miao, Bo An, and Cyril Leung. Filtering trust opinions through reinforcement learning. *Decision Support Systems*, 66:102–113, 2014.

[Zhang, 2009] Jie Zhang. *Promoting honesty in electronic marketplaces: Combining trust modeling and incentive mechanism design*. PhD thesis, University of Waterloo, 2009.

[Zhou and Hwang, 2007] Runfang Zhou and Kai Hwang. Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing. *IEEE Transactions on Parallel and Distributed Systems*, 18:460–473, 2007.