OpinionWalk: An Efficient Solution to Massive Trust Assessment in Online Social Networks

Abstract—Massive trust assessment (MTA) in an Online Social Network (OSN), i.e., computing the trustworthiness of all users in the network, is crucial in various OSN-related applications. Existing solutions are either too slow or inaccurate in addressing the MTA problem. We propose the OpinionWalk algorithm that accurately and efficiently conducts MTA in an OSN. OpinionWalk models trust by the Dirichlet distribution and uses a matrix to represent the direct trust relations among users. From the perspective of a user, other users’ trustworthiness is stored in a column vector that is iteratively updated when the algorithm “walks” through the network, in a breadth-first search manner. We identify the overlapping subproblems property in MTA and prove OpinionWalk is a more efficient solution. The accuracy and execution time of OpinionWalk are evaluated and compared to benchmark algorithms including EigenTrust, TrustRank, MoleTrust, TidalTrust and AssessTrust, using two real-world datasets (Advogato and Pretty Good Privacy). Experimental results indicate that OpinionWalk is an efficient and accurate solution to MTA, compared to previous algorithms.

Keywords—Massive trust assessments, online social networks, three-valued subjective logic, trust model

I. INTRODUCTION

A. Motivations and Problem Statement

An online social network (OSN) is a type of online service that facilitates the building of social relations/connections among people who have similar interests or backgrounds, and most importantly trust each other in real life. Trust is a hidden fabric in OSNs that enables reciprocal online interactions among users [1]. Besides fostering social interactions, the trust embedded in OSNs can also be used to improve various aspects of our life [2], [3]. For example, the Lending Club¹ company has leveraged the trust relations among users in Facebook.com² to improve its online peer-to-peer lending service. It offers a mechanism to evaluate the trustworthiness of potential borrowers, from the perspective of a lender. A fundamental research problem here is to identify the most trustworthy borrowers based on their social connections to the lender(s). To address this problem, we need a mechanism to conduct massive trust assessment (MTA) in OSNs, i.e., computing the trustworthiness of all users (e.g., potential borrowers), from the perspective of a user (e.g., the lender).

An OSN is usually modeled as a social graph where vertices represent users, and edges denote the trust relations among users. An edge in the graph typically reflects a user’s opinion about the trustworthiness of another user. Therefore, in this paper, the terms opinion and trust relation are used interchangeably to refer to an edge. An opinion can be either a direct or an indirect opinion. While the former is designed for users who have direct interactions between each other, the latter is applied to users who do not know each other in the network. If a user trusts another user, the first user who is regarded as the trustor can rely on the second, known as the trustee, to perform expected actions [4]. Therefore, the MTA problem is to compute the trustworthiness of all trustees, from the point of view of the trustor, based on the trust social graph.

B. Limitations of Prior State of the Art

There are many algorithms applicable to massive trust assessment in OSNs, e.g., EigenTrust (ET) [5], TrustRank (TR) [6]), MoleTrust (MT) [7] and TidalTrust (TT) [8]. While EigenTrust and TrustRank focus on ranking users based on their relative trustworthiness, MoleTrust and TidalTrust are able to compute users’ absolute trustworthiness. By searching a social graph with weighted averages, above-mentioned algorithms compute the trustworthiness of all trustees based on their social connections to the trustor. The time complexity of these algorithms is \(O(n^2)\) where \(n\) is the number of nodes in the graph. One limitation of these algorithms is that they tend to give inaccurate assessments as trusts are modeled by a single real number.

To achieve more accurate trust assessment results, trust is modeled as statistical distributions in subjective logic (SL) [9] and three-valued subjective logic (3VSL) [4]. Based on 3VSL, the AssessTrust (AT) algorithm is proposed to accurately compute the trustworthiness between any two users in [4]. Although AT gives accurate trust assessment results, applying it to MTA yields a time complexity of \(O(n^K)\) where \(K\) is the network’s diameter. In summary, existing algorithms are either inaccurate or very slow in solving the MTA problem.

C. Proposed Approach

Based on 3VSL, we propose the OpinionWalk algorithm to efficiently and accurately address the MTA problem. We first use an opinion matrix to represent the topology of a trust social network. Each entry in the opinion matrix is a direct opinion of two corresponding users, which is modeled as a Dirichlet distribution based on 3VSL. We then design a set of matrix-like operations where traditional multiplication and summation operations are replaced by the discounting and combining operations. The discounting and combining operations are used to model trust propagation and fusion by 3VSL. Based on these newly defined operations, OpinionWalk starts from the trustor, searches the network, and compute users’ trustworthiness in an iterative manner.

We prove that OpinionWalk is an equivalent implementation of AT and offers a better time complexity of \(O(n^2)\) in

¹http://blog.lendingclub.com/facebook-and-lending-club-looks-like-its-working/
²http://www.techinsider.io/facebooks-patent-may-change-your-credit-2015-8
solving the MTA problem. We also compare the execution time and accuracy of OpinionWalk to existing algorithms, including AT, ET, MT, TR and TT, with two real-world datasets. Experimental results validate the accuracy and efficiency of OpinionWalk in solving the MTA problem.

D. Technical Challenges and Solutions

The MTA problem can be accurately solved by applying the AT algorithm \( n \) times, which yields an overall time complexity of \( O(n^K) \). Therefore, the first technical challenge is to keep the accuracy of AT, at the same time, reduce the time complexity. To address this challenge, we design the OpinionWalk algorithm that searches the network in a BFS manner, and compute users’ trustworthiness iteratively. To enable the iterative computations, we use an opinion matrix to represent the network and an individual opinion vector to store the trustworthiness of all users. When OpinionWalk “walks” through the network, the individual opinion vector is updated accordingly. When OpinionWalk finishes, the trustworthiness of all users are obtained. This design could reduce the time complexity from \( O(n^K) \) to \( O(n^2) \).

To enable BFS-based search in OpinionWalk, it is critical to ensure the overlapping subproblems property exists in MTA, if trust is modeled by the 3VSL model. In other words, to compute the trustworthiness of the truster’s \((k+1)\)-hop friends, based on 3VSL, can we reuse the trustworthiness of his \(k\)-hop friends? This is a challenge because the \(k\)- and \((k+1)\)-hop friends could be connected in various ways, resulting in different computations, according to 3VSL. We define the correct rules of applying \(k\)-hop friends’ trustworthiness to compute the trustworthiness of \((k+1)\)-hop friends, based on 3VSL. Therefore, OpinionWalk develops solutions to sub-problems and reuse them to obtain the solution of the original MTA problem. The existence of overlapping subproblems property in MTA supports our adoption of BFS-based search in OpinionWalk.

The third technical challenge is to prove the correctness of OpinionWalk, i.e., it gives the same trust assessment results as AT, no matter what the network topology is. This is a challenge because these two algorithms are designed for different purposes and implemented in completely different ways. To address this challenge, we first prove the operations between opinion matrix and individual opinion vector equivalently implement the discounting and combining operations in simple network topologies. We then extend the proof into an arbitrary topology and recursively show that all sub MTA problems are equivalently solved by OpinionWalk and AT. Finally, we conclude that both OpinionWalk and AT provide the same MTA result.

E. Key Contributions

The key contributions of this paper are concluded as follows. First, an efficient 3VSL-based algorithm, OpinionWalk, is proposed to cope with the MTA problem. Second, the correctness of OpinionWalk is proved and its time complexity is analyzed. Third, trust assessment accuracy and the execution time of OpinionWalk are evaluated in real-world datasets and compared to benchmark algorithms.

The rest of this paper is organized as follows. In section II, we introduce the related work. In section III, we give problem statement and background knowledge of this paper. In section IV we introduce the OpinionWalk algorithm and analyze its time complexity. In section V, we prove the correctness of OpinionWalk. In section VI, we evaluate the runtime and accuracy of OpinionWalk. We conclude our work in section VII.

II. RELATED WORK

Approaches to massive trust assessment in OSNs can be roughly divided into two broad categories, based upon how trust is modeled. Assuming trust is a real number, researchers studied how to compute relative trust [5], [6] and absolute trust [7], [8] in an OSN. On the other hand, trust can be modeled as a statistical distribution [4], [9]–[11], so more accurate trust assessments are realized.

In the first category, relative trust is first studied in peer-to-peer file-sharing networks where peers use trust values to choose trustworthy peers from which they download files [5]. The EigenTrust algorithm proposed in [5] starts from a peer and searches for trustworthy peers based on the following rules. It moves from a peer to another with the probability that is proportional to the other peer’s trust score, i.e., higher the trust score, higher the moving probability. In this way, EigenTrust will more likely reach trustworthy peers than untrustworthy ones. Later on, the relative trust of web pages is investigated in [6] to identify spam pages. The TrustRank algorithm proposed in [6] again employs random walk on the network to rank the trustworthiness of web pages. Both EigenTrust and TrustRank can be viewed as a variant of the PageRank algorithm that is a well-known solution to assigning importance scores to pages on the Internet. These algorithms, however, only generate trust rankings instead of absolute trust values of peers/pages.

Based on the personal opinion of the trustor, for example, MoleTrust [7] proposes a method to compute the trustworthiness of a particular user in a personalized way. MoleTrust starts from the trustor and walks through the trust social network until the trustee is reached. While walking through the network, it only considers incoming edges with trust scores greater than 0.6 and ignores those with lower trust scores. Then, a user’s trust score is computed by averaging all accepted incoming edges weighted by the trust scores of the users from whom the edges originate. Similarly, TidalTrust [8] recursively searches the network with a weighted average approach. The difference between TidalTrust and MoleTrust is that TidalTrust considers only the path(s) with the highest trust score(s) between the truster and trustee, however, MoleTrust uses all paths as long as the trust score of each edge along the paths is greater than 0.6. Recently, the evolution or dynamics of trust in OSNs is studied in FluidRating [12]. FluidRating uses fluid dynamics theory to understand the evolution of trust in OSNs. Specifically, when a new opinion is received, a user refines his/her opinion through fluid exchanges with his/her neighbors.

In the second category, trust is modeled as a statistical distribution, e.g., in subjective logic [9], [10], CertProp [11] and three-valued subjective logic [4]. In this way, trust propagation and fusion are treat as the multiplication and summation
of statistical distributions. Comparing to solutions in the first category, these works achieve a higher accuracy in trust assessments. However, they have difficulty in handling complex networks due to the limitations identified in [4]. To enable trust assessment over large-scale networks, the AssessTrust algorithm is proposed in [4]. AssessTrust is based on the three-valued subjective logic model, and is proven to be able to handle arbitrary networks. A major limitation of AssessTrust is that it is designed to compute the trustworthiness for one user and thus is very slow and inefficient in solving the MTA problem. In summary, existing solutions either have trouble providing accurate trust assessments or are inefficient in solving the MTA problem in a large-scale OSN.

III. PROBLEM STATEMENT AND BACKGROUND

A. Problem Statement

A trust social network is modeled as a directed graph \( G(V, E, \omega) \) where vertex \( i \in V \) represents a user, and edge \( e(i, j) \) denotes user \( i \) trusts user \( j \). The weight on edge \( e(i, j) \) indicates how user \( i \) trusts \( j \). In this way, we can define the MTA problem as follows.

Given a trust social network \( G(V, E, \omega) \), \( \forall i \) and \( j \), s.t. \( i, j \in V \), \( \exists \) at least one path from \( i \) to \( j \), how to compute the trustworthiness of users \( \{ j \in V, j \neq i \} \), from user \( i \)’s perspective.

The weight on an edge can be modeled as a real number or a statistical distribution, resulting in different accuracies in trust assessments. To achieve an accurate MTA, we rely on 3VSL to model the weight of an edge as a Dirichlet distribution.

B. Three-Valued Subjective Logic (3VSL)

Because the shape of a Dirichlet distribution is controlled by a set of parameters, we use these parameters to represent a Dirichlet distribution. Given trustor \( i \) and trustee \( j \), \( i \)’s opinion about the trustworthiness of \( j \) is a Dirichlet distribution denoted by a tuple

\[
(b_{ij}, d_{ij}, n_{ij}, e_{ij})|a_{ij},
\]

where \( b_{ij} + d_{ij} + n_{ij} + e_{ij} = 1 \).

In the definition, \( b_{ij}, d_{ij}, n_{ij} \) represent the probabilities that \( j \) will behave as expected, unexpected, neither expected nor unexpected (i.e., uncertain), respectively. These values are determined by the numbers of evidences supporting the corresponding events. \( e_{ij} \) represents the prior uncertainty about whether \( j \) is trustworthy. Posterior uncertainty \( n_{ij} \) exists due to the fact that evidences are distorted when trust propagates over the network. \( a_{ij} \) is a constant number formed from impressions without solid evidences, e.g. prejudice and preference. As \( a_{ij} \) is not our focus, we omit it and model an opinion as a vector \((b_{ij}, d_{ij}, n_{ij}, e_{ij})\). We use \( \omega_{ij} \) and \( \Omega_{ij} \) to denote \( i \)’s direct and indirect opinions on \( j \), respectively. Note that both direct and indirect opinions are modeled in the same form, a Dirichlet distribution.

Based on 3VSL, we further define a special opinion that will be used to initialize the OpinionWalk algorithm.

\[\frac{\Delta(\omega_{ij}, \omega_{kj})}{\omega_{ij}} \]

Definition 1 (Uncertain Opinion): An uncertain opinion \( \Omega \) is defined as

\[\Omega \triangleq (0, 0, 0, 1),\]

that indicates the trustor is totally uncertain about the trustee’s trustworthiness.

To model trust propagation and fusion in OSNs, 3VSL defines the discounting and combining operations as follows. The discounting operation \( \Delta(\omega_{is}, \omega_{sj}) \) is used to compute \( i \)’s opinion about \( j \)’s trustworthiness based on \( s \)’s opinion on \( j \) where \( s \) is a mutual friend of \( i \) and \( j \). The discounting operation yields a new opinion \( \Omega_{ij} \) where

\[
\begin{align*}
    b_{ij} &= b_{is}b_{sj} \\
    d_{ij} &= d_{is}d_{sj} \\
    n_{ij} &= 1 - b_{ij} - d_{ij} - e_{sj} \\
    e_{ij} &= e
\end{align*}
\]

Here, we have \( e = e_{sj} \) if \( e_{is} \neq 1 \); otherwise, \( e = 1 \). Note the direct opinions \( \omega_{is} \) and \( \omega_{sj} \) can be replaced by indirect opinions in the discounting operation.

The combining operation \( \Theta(\omega_{ij}^l, \omega_{ij}^r) \) is used to combine \( i \)’s two opinions about \( j \)’s trustworthiness. The combining operation will generate a new opinion where

\[
\begin{align*}
    b_{ij} &= \frac{\omega_{ij}^l b_{ij} + \omega_{ij}^r b_{ij}''}{\omega_{ij}^l + \omega_{ij}^r} \\
    d_{ij} &= \frac{\omega_{ij}^l d_{ij} + \omega_{ij}^r d_{ij}''}{\omega_{ij}^l + \omega_{ij}^r} \\
    n_{ij} &= \frac{\omega_{ij}^l n_{ij} + \omega_{ij}^r n_{ij}''}{\omega_{ij}^l + \omega_{ij}^r} \\
    e_{ij} &= \frac{\omega_{ij}^l e_{ij} + \omega_{ij}^r e_{ij}''}{\omega_{ij}^l + \omega_{ij}^r}
\end{align*}
\]

Here, \( \omega_{ij}^l \) or \( \omega_{ij}^r \) could be replaced by \( i \)’s indirect opinions on \( j \) as well. Based on the above definitions, we can easily prove the following corollaries.

**Corollary 1:** Applying the discounting operation on \( \Omega \) and an opinion \( \omega \), we have \( \Delta(\omega, \Omega) = \Omega \) and \( \Delta(\Omega, \omega) = \Omega \).

**Corollary 2:** Applying the combining operation on \( \Omega \) and an opinion \( \omega \), we have \( \Theta(\omega, \Omega) = \Theta(\Omega, \omega) = \omega \).

C. AssessTrust (AT) Algorithm

Fig. 1. An illustration of the AT algorithm based on the bridge topology.

In the following, we use an example shown in Fig. 1 to explain how the AssessTrust (AT) algorithm works. The AT
algorithm starts from the trustee, e.g., D in Fig. 1(a), searches the network backwards and recursively computes the trustworthiness of users one by one. As a result, we get a parsing tree, shown in Fig. 1(b), that describes how discounting and combining operations are applied in computing A’s opinion on D. Traversing the parsing tree in a bottom-up manner, A’s indirect opinion on D ΩAD can be computed as

\[ \Theta \left( \Delta(\omega_{AB}, \omega_{BD}), \Delta(\Theta(\Delta(\omega_{AB}, \omega_{BC}), \omega_{AC}), \omega_{CD}) \right). \]  

(1)

To understand how AT is applied on the graph, we use \(AT^k(i, j)\) to denote that it is the \(k\)-th time that AT is called to compute user i’s opinion on j. At the first time when AT is called, A’s opinion on D is computed from

\[ \Theta \left( \Delta(\Omega_{AB}, \omega_{BD}), \Delta(\Omega_{AC}, \omega_{CD}) \right), \]

where \(\Omega_{AB}\) and \(\Omega_{AC}\) are A’s indirect opinions on B and C, respectively. These two opinions will then be provided by \(AT^2(A, B)\) and \(AT^3(A, C)\), respectively. In \(AT^3(A, C)\), AT computes A’s opinion on C as

\[ \Theta \left( \Delta(\Omega_{AB}, \omega_{BC}), \omega_{AC} \right), \]

where \(\Omega_{AB}\) are computed by \(AT^4(A, B)\). Finally, A’s opinion on D can be computed from Eq. 1. Overall, AT is called four times in this example: \(AT^1(A, D)\), \(AT^2(A, B)\), \(AT^3(A, C)\), \(AT^4(A, B)\). Note that the \(AT^1(A, B)\) is called twice in this example.

IV. DESIGN OF OPINIONWALK

OpinionWalk is essentially a BFS-based algorithm that addresses the MTA problem in a more efficient way. Given a trust social network, OpinionWalk represents it as an opinion matrix

\[ M = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \omega_{n1} & \cdots & \cdots & \omega_{nn} \end{bmatrix}, \]

where element \(\omega_{ij}\) (i, j ≤ n) denotes i’s direct opinion on j. If i has no direction opinion on j, we use ∅ to substitute \(\omega_{ij}\).

From the view of the trustor, e.g., user i, its opinions about the trustworthiness of other users are stored in the individual opinion vector

\[ Y_i^{(k)} = \begin{bmatrix} \Omega_{i1}^{(k)} & \Omega_{i2}^{(k)} & \cdots & \Omega_{in}^{(k)} \end{bmatrix}^T, \]

where \(\Omega_{ij}^{(k)}\) denotes user i’s direct opinion on j after OpinionWalk “walks” \(k\) hops on the network. The initial opinion vector is initialized as

\[ \Omega_{ij}^{(1)} = \begin{cases} \omega_{ij}, & \text{if user } i \text{ directly connects to } j \\ \emptyset, & \text{otherwise} \end{cases} \]

In this way, when OpinionWalk searches the network, the individual opinion vector is updated according to

\[ Y_i^{(k)} = M^T \otimes Y_i^{(k-1)}. \]

A. Operations in OpinionWalk

The operator \(\otimes\) “multiplies” M and \(Y_i^{(k-1)}\) to yield an updated \(Y_i^{(k)}\) as follows.

\[ Y_i^{(k)} = M^T \otimes Y_i^{(k-1)} \]

\[ = \Theta \left( \Delta(\Omega_{i1}^{(k-1)}, \omega_{11}), \cdots, \Delta(\Omega_{in}^{(k-1)}, \omega_{n1}) \right) \]

\[ = \Theta \left( \Delta(\Omega_{i1}^{(k-1)}, \omega_{12}), \cdots, \Delta(\Omega_{in}^{(k-1)}, \omega_{n2}) \right) \]

\[ \cdots \]

\[ = \Theta \left( \Delta(\Omega_{i1}^{(k-1)}, \omega_{1n}), \cdots, \Delta(\Omega_{in}^{(k-1)}, \omega_{nn}) \right) \]

\[ = \begin{bmatrix} \Omega_{11}^{(k)}, \Omega_{12}^{(k)}, \cdots, \Omega_{1n}^{(k)} \end{bmatrix}^T, \]

where \(\Theta\) and \(\Delta\) are the combining and discounting operations, respectively. As shown in Fig. 2, the function of \(\otimes\) is analogous to multiplying a matrix and a vector. The difference lies in the summation and multiplication operations are replaced by the combining and discounting operations, respectively.

Let’s look at element \(\Omega_{ij}^{(k)}\) in \(Y_i^{(k)}\) where \(i \neq j\). It is computed by “multiplying” vectors \([\omega_{1j}, \omega_{2j}, \cdots, \omega_{nj}]\) and \([\Omega_{i1}^{(k-1)}, \Omega_{i2}^{(k-1)}, \cdots, \Omega_{in}^{(k-1)}]^T\).

![Fig. 2. A diagrammatic overview of the \(\otimes\) operation in OpinionWalk.](image)

The principle of operator \(\otimes\) can be better explained by Fig. 3. In the figure, OpinionWalk starts from user i and searches the network level by level. Suppose OpinionWalk is currently searching the \((k-1)\)-th level, and it finds a set of users that are \((k-1)\)-hop away from i. Among these users, we assume \(m\) of them directly connect to user j, i.e., user j is \(k\)-hop away from i. We label these users as \(s_1, s_2, \cdots, s_m\).

When OpinionWalk moves to the next level, i.e., the \(k\)-th level, it updates i’s opinion on j as follows.

\[ \Theta \left( \Delta(\Omega_{is_1}^{(k-1)}, \omega_{s_1j}), \cdots, \Delta(\Omega_{is_m}^{(k-1)}, \omega_{s_mj}) \right) \]

It combines all \(m\) opinions that are computed from discounting \(\omega_{s_1j}\) by \(\Omega_{is_1}\), for all possible \(l = 1, 2, \cdots, m\). Due to Corollaries 1 and 2, this equation\(^3\) can be further generalized to

\[ \Theta \left( \Delta(\Omega_{i1}^{(k-1)}, \omega_{1j}), \cdots, \Delta(\Omega_{in}^{(k-1)}, \omega_{nj}) \right). \]

\(^3\)The equation here is generalized for presentation purposes. When it is implemented, only \(m\) opinions will be considered.
that are currently computed from $\omega_{ij}$ than $\Omega$ indirect opinion

\begin{algorithm}
\SetKwInOut{Input}{Require}
\KwRequire{A directed graph $G$ with a truster $i$ and the maximum searching level $H$}
\KwEnsure{$i$'s opinion $j$ where $j \neq i$.}
\begin{algorithmic}[1]
\State Initialize $M$ and $Y^{(1)}_i$ based on $G$
\State $k \leftarrow 1$
\While{$k < H$}
\ForAll{columns $c_j \in M$ s.t. $j \neq i$}
\ForAll{direct opinions $\omega_{sj} \in c_j$ s.t. $\omega_{sj} \not\in \emptyset$}
\If{$\Omega_{ij}^{(k+1)} \not\in \emptyset$}
\If{$\Omega_{ij}^{(k+1)} = \emptyset$}
\State $\Omega_{ij}^{(k+1)} \leftarrow \Delta(\Omega_{ij}^{(k)}, \omega_{sj})$
\Else
\State $\Omega_{ij}^{(k+1)} \leftarrow \Theta(\Omega_{ij}^{(k+1)}, \Delta(\Omega_{ij}^{(k)}, \omega_{sj}))$
\EndIf
\EndIf
\EndFor
\EndFor
\State $k \leftarrow k + 1$
\EndWhile
\State Return $Y^{(k)}_i$
\end{algorithmic}
\end{algorithm}

The pseudo-code of the OpinionWalk algorithm is shown in Algorithm 1. In the algorithm, line 3 controls how many hops OpinionWalk will search on the network. Lines 4-16 update the indirect opinion $\Omega_{ij}$ iteratively. Line 4 considers all users, other than $i$, as the trustees. Lines 5-14 combine all opinions derived from $\omega_{sj} \not\in \emptyset$. Line 6 obtains $i$'s indirect opinion on one of its $k$-hop friends, e.g., $s$. If this opinion already exists, $i$ discounts $s$'s opinion on $j$ to update $\Omega_{ij}^{(k+1)}$ in line 9. Otherwise, it checks another $k$-hop friend. Line 11 combines all opinions that are currently computed from $\omega_{sj} \not\in \emptyset$. Note that line 11 essentially combines opinions one by one, so $\Omega_{ij}^{(k+1)}$ equals to $\Theta(\Delta(\Omega_{ij}^{(k)}, \omega_{1j}), \cdots, \Theta(\Delta(\Omega_{ij}^{(k)}, \omega_{n-1j}), \Delta(\Omega_{ij}^{(k)}, \omega_{nj})))$.

Because the combining operation is associative [4], the above equation is the same as Eq. 2. After processing all users connecting to $j$, at line 14, the newly computed $\Omega_{ij}$ is used to update the corresponding element in the individual opinion vector. When $i$'s opinions on all possible $j$'s are updated, at line 16, OpinionWalk searches the next level. Finally, the vector $Y^{(k)}_i$ will contain $i$'s opinions about the trustworthiness of all other users.

C. Overlapping Subproblems Property

In the following, we will use the bridge topology in Fig. 1 to show the overlapping subproblems property in MTA when it is solved by OpinionWalk. Using the same example, we also show OpinionWalk is a more efficient algorithm, comparing to AT.

According to the OpinionWalk algorithm, user A's opinions on $B, C$ and $D$ are updated as follows. OpinionWalk starts from $A$ and searches $A$'s 1-hop neighbors. After this search, $A$'s opinions on $B$ and $C$ are updated to

$$\Omega_{AB}^{(1)} = \omega_{AB}$$

and

$$\Omega_{AC}^{(1)} = \omega_{AC}.$$  

When OpinionWalk searches the next level, i.e., $k = 2$, $A$'s opinion on $B$ does not change but its opinion on $C$ and $D$ are changed to

$$\Omega_{AC}^{(2)} = \Theta \left( \Delta(\Omega_{AB}^{(1)}, \omega_{BC}), \omega_{AC} \right)$$

and

$$\Omega_{AD}^{(2)} = \Theta \left( \Delta(\Omega_{AB}^{(1)}, \omega_{BD}), \Delta(\Omega_{AC}^{(1)}, \omega_{CD}) \right).$$

Inserting $\Omega_{AB}^{(1)} = \omega_{AB}$ and $\Omega_{AC}^{(1)} = \omega_{AC}$ into the above equations, we have

$$\Omega_{AC}^{(2)} = \Theta \left( \Delta(\omega_{AB}, \omega_{BC}), \omega_{AC} \right)$$

and

$$\Omega_{AD}^{(2)} = \Theta \left( \Delta(\omega_{AB}, \omega_{BD}), \Delta(\omega_{AC}, \omega_{CD}) \right).$$

When OpinionWalk reaches the third level, i.e., $k = 3$, only $A$'s opinion on $D$ is updated to

$$\Omega_{AD}^{(3)} = \Theta \left( \Delta(\Omega_{AB}^{(2)}, \omega_{BD}), \Delta(\Omega_{AC}^{(2)}, \omega_{CD}) \right).$$

Replacing $\Omega_{AB}^{(2)}$ and $\Omega_{AC}^{(2)}$ with the results obtained when $k = 2$, we have $A$'s opinion on $D$ as

$$\Theta \left( \Delta(\omega_{AB}, \omega_{BD}), \Delta(\omega_{AC}, \omega_{CD}) \right).$$

This equation is exactly the same as Eq. 1 that is obtained by the AT algorithm. Note that the solutions to sub-problems are re-used in OpinionWalk; however, the same sub-problem may be solved over and over again in AT. For example, the trustworthiness between $A$ and $B$ are computed twice in $AT^{(3)}(A, B)$ and $AT^{(4)}(A, B)$. Leveraging the overlapping subproblems property, OpinionWalk yields a better time complexity than AT, which will be formally analyzed in the next subsection.
D. Time Complexity Analysis

In this section, we analyze the time complexities of OpinionWalk and AssessTrust, and find OpinionWalk yields a better time complexity.

AT is a recursive algorithm, and the recurrence equation of its time complexity is

\[ T(n) = (n - 1) \cdot T(n - 1) + C(n - 1), \]

where the first \((n - 1)\) denotes the maximum number of users connecting to the trustee. If there are \(n\) users in the network, there are at most \((n - 1)\) users connecting to the trustee. To find the trustworthiness of a user in the \((n - 1)\) users, it takes \(T(n - 1)\). After that, the opinions derived from the \((n - 1)\) users are combined, which takes \(C(n - 1)\). Expanding the above equation, we obtain AT’s time complexity as

\[ O \left( \sum_{i=1}^{K-1} \frac{(n - 1)!}{(n - 1 - 1)!} \right) = O\left(n^{K-1}\right), \]

where \(K\) is the diameter of the network, measured by number of hops. To solve the MTA problem, AT needs to compute the trustworthiness of every user in the network, so its time complexity goes to

\[ O \left( (n - 1) \cdot n^{K-1} \right) = O\left(n^K\right). \]

In the OpinionWalk algorithm, there are two nested loops: lines 4-16 and lines 5-14. The number of iterations in each loop is \(O(m)\), so the time complexity of the nested loops is \(O(n^2)\). The parameter \(H\) controls the maximum number of levels to be searched by OpinionWalk. As \(H\) is usually a constant number, OpinionWalk’s time complexity becomes \(O(n^2)\). In the extreme case where \(H = K\), OpinionWalk’s time complexity will be \(O(Kn^2)\) that is still much better than AT’s \(O(n^K)\).

In addition, because OpinionWalk uses iterative procedures, it offers a faster execution time compared to the recursive implementation of AT. Because large number of stack operations are involved, AT consumes much more system memory. Faster execution time is extremely useful when the network gets larger and more complex, which are common in current OSNs.

V. CORRECTNESS OF OPINIONWALK

To prove OpinionWalk equivalently implements the AT algorithm, we first show that both AT and OpinionWalk generate the same result if the network topology is either serial or parallel. Then, we show this is true for arbitrary network topologies.

If we zoom into a trust social network, two edges can be connected in series if they are incident to a vertex of degree 2, or in parallel if they join the same pair of distinct vertices. Therefore, two users can be connected in a serial topology shown in Fig. 4(a), or a parallel topology shown in Fig. 4(b). Note that the paths from \(i\) to \(s_1, s_2, \ldots, s_m\) in Fig. 4(b) are disjoint, i.e., no sharing edges along the paths.
B. Parallel Network Topology

Lemma 2: Given two users $i$ and $j$ who are connected by $m$ users in a parallel topology, the opinion $\Omega_{ij}$ computed from AT and OpinionWalk will be the same.

Proof: We use $s_1, s_2, \cdots, s_m$ to denote the $m$ users based on their distances to $i$, i.e., $s_m$ is the farthest away from $i$. As shown in Fig. 4(b), because the paths from $i$ to $s_1, s_2, \cdots, s_m$ are disjointed, the AT algorithm computes $i$’s opinion on $j$ as

$$\Theta(\Delta(\Omega_{is_1}, \omega_{s_1j}), \Delta(\Omega_{is_2}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}, \omega_{s_mj})).$$

(5)

According to the OpinionWalk algorithm, the opinion matrix will be

$$M = \begin{bmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \omega_{s_1j} & \cdots & \cdots & \cdots \\
\cdots & \cdots & \omega_{s_2j} & \cdots & \cdots \\
\cdots & \cdots & \cdots & \omega_{s_mj} & \cdots
\end{bmatrix}.$$

Let’s assume after $k_1, k_2, \cdots, k_m$ searches, the OpinionWalk algorithm reaches users $s_1, s_2, \cdots, s_m$, respectively. After $k_1$ searches, OpinionWalk obtains $i$’s opinion on $s_1$ as $\Omega_{is_1}^{(k_1)}$. Here, $\Omega_{is_1}^{(k_1)}$ is equal to $\Omega_{is_1}$ computed from AT in Eq. 5, due to Lemma 1.

During the $(k_1 + 1)$-th search, OpinionWalk reaches user $j$, and it updates $i$’s opinion on $j$ to

$$\Omega_{ij}^{(k_1+1)} = \Delta(\Omega_{is_1}, \omega_{s_1j}).$$

After $k_2$ searches, OpinionWalk gets $i$’s opinion on $s_2$ as $\Omega_{is_2}$. At the $(k_2 + 1)$-th search, OpinionWalk hits user $j$ and updates $i$’s opinion on $j$ to

$$\Omega_{ij}^{(k_2+1)} = \Theta(\Omega_{ij}^{(k_2)}, \Delta(\Omega_{is_2}, \omega_{s_2j})).$$

From $(k_1 + 1)$-th to $k_2$-th searches, opinion $\Omega_{ij}$ is not updated, so we have $\Omega_{ij}^{(k_2)} = \Omega_{ij}^{(k_1+1)}$. Therefore, we have

$$\Omega_{ij}^{(k_2+1)} = \Theta(\Delta(\Omega_{is_1}, \omega_{s_1j}), \Delta(\Omega_{is_2}, \omega_{s_2j})).$$

Similarly, after $k_m + 1$ searches, $i$’s opinion on $j$ is updated to

$$\Theta(\Delta(\Omega_{is_1}, \omega_{s_1j}), \Delta(\Omega_{is_2}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}, \omega_{s_mj})).$$

For every $l = 1, 2, \cdots, m$, we have $\Omega_{ij}^{(k_l)} = \Omega_{ij}^{(k_m)}$ as it does not change after the $k_l$-th search. Therefore, the above equation becomes

$$\Theta(\Delta(\Omega_{is_1}, \omega_{s_1j}), \Delta(\Omega_{is_2}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}, \omega_{s_mj})).$$

For any $l = 1, 2, \cdots, m$, we know that $\Omega_{ij}^{(k_l)}$ is equal to $\Omega_{ij}$ computed by AT, due to Lemma 1. Therefore, OpinionWalk computes $i$’s opinion on $j$ as

$$\Theta(\Delta(\Omega_{is_1}, \omega_{s_1j}), \Delta(\Omega_{is_2}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}, \omega_{s_mj})).$$

(6)

Because Eqs. 5 and 6 give the same result, we conclude that OpinionWalk equivalently implements AT on a parallel topology.

C. Arbitrary Topology

Given an arbitrary network topology shown in Fig. 5, we assume $m > 1$ nodes $\{s_1, s_2, \cdots, s_m\}$ directly connects to $j$. We use $k_1, k_2, \cdots, k_m$ to denote the last times when users $s_1, s_2, \cdots, s_m$ are visited by the OpinionWalk algorithm, respectively. Let $\{t_1, t_2, \cdots, t_r\}$ denote the users who are connected from $i$ via either a serial or parallel topologies.

Proof: We prove the theorem in a recursive manner, i.e., reducing the original network into sub-network(s) and keep reducing the sub-network(s) until the base case is reached, i.e., two users are connected via either a serial or parallel topology.

Reduction rules

Case 1: There is only one user connecting to $j$, i.e., $m = 1$. In this case, according to the AT algorithm, $\Omega_{ij}$ is computed as

$$\Delta(\Omega_{is_1}, \omega_{s_1j})$$

where $\Omega_{is_1}$ denotes $i$’s opinion on $s_1$, and it can be computed by recursively calling the AT algorithm. For OpinionWalk, at the $(k_1 + 1)$-th search, it reaches $j$ and updates $i$’s opinion on $j$ as

$$\Delta(\Omega_{is_1}^{(k_1)}, \omega_{s_1j}).$$

Case 2: There are more than one user connecting to $j$, i.e., $m > 1$. In this case, AT computes $\Omega_{ij}$ as

$$\Theta(\Delta(\Omega_{is_1}, \omega_{s_1j}), \Delta(\Omega_{is_2}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}, \omega_{s_mj})).$$

After $\hat{k} + 1$ searches where $\hat{k} = \max(k_1, k_2, \cdots, k_m)$, OpinionWalk updates $i$’s opinion on $j$ to

$$\Theta(\Delta(\Omega_{is_1}^{(k_1)}, \omega_{s_1j}), \Delta(\Omega_{is_2}^{(k_2)}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}^{(k_m)}, \omega_{s_mj})).$$

For any user $s_l$, where $l = 1, 2, \cdots, m$, because $k_l$ is the last time that $s_l$ was visited by OpinionWalk, $\Omega_{is_l}$ was not updated after the $k_l$-th search. Therefore, the above equation can be rewritten as

$$\Theta(\Delta(\Omega_{is_1}^{(k_1)}, \omega_{s_1j}), \Delta(\Omega_{is_2}^{(k_2)}, \omega_{s_2j}), \cdots, \Delta(\Omega_{is_m}^{(k_m)}, \omega_{s_mj})).$$

Summarizing the above two cases, we know that if $\Omega_{is_l} = \Omega_{is_l}^{(k_l)}$ for every $l = 1, 2, \cdots, m$, then OpinionWalk and AT yield the same result of $\Omega_{ij}$. For any user $s_l$, $\Omega_{is_l}$ and $\Omega_{is_l}^{(k_l)}$ actually denote $i$’s opinion on $s_l$ computed by AT.
and OpinionWalk, respectively. Both AT and OpinionWalk will work on the same sub-network \( G' = G - e(s_l, j) \) that connects \( i \) to \( s_l \). The sub-network \( G' \) can be further reduced by removing the edge connecting to \( s_l \). Continuing this process, the original network \( G \) will eventually be reduced to the following base case.

**Base Case**

In the base case, user \( i \) connects to users \( t_1, t_2, \cdots, t_r \) via either a serial or parallel topology. Based on Lemmas 1 and 2, AT and OpinionWalk give the same values of \( \Omega_{i, t_1}, \Omega_{i, t_2}, \cdots, \Omega_{i, t_r} \). Overall, we prove that AT and OpinionWalk give the same result.

### VI. Evaluation

We compare the performance of OpinionWalk (OW) to a few benchmark algorithms, including EigenTrust (ET), TrustRank (TR), MoleTrust (MT), TidalTrust (TT) and AssessTrust (AT), in two real-world datasets, Advogato [13] and Pretty Good Privacy (PGP) [11].

Advogato is obtained from an online software development community where an edge from a user to another user represents the user’s opinion about the other’s ability in software development. The trustworthiness between users are categorized into four levels. PGP is collected from a public key certification network where an edge from a user to another user indicates that the user certifies the trustworthiness of the other user. The trustworthiness between users are also classified into four categories. Statistics of these datasets are summarized in TABLE I.

**TABLE I. STATISTICS OF THE ADVOGATO AND PGP DATASETS**

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of Vertices</th>
<th># of Edges</th>
<th>Avg Deg</th>
<th>Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advogato</td>
<td>6,541</td>
<td>51,127</td>
<td>19.2</td>
<td>4.82</td>
</tr>
<tr>
<td>PGP</td>
<td>38,346</td>
<td>31,127</td>
<td>16.3</td>
<td>7.3</td>
</tr>
</tbody>
</table>

**TABLE II. TRANSFORMING ORDINAL VALUES TO REAL NUMBERS**

<table>
<thead>
<tr>
<th>Trust Level</th>
<th>Trust Opinion</th>
<th>Trust Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(0.09, 0.81, 0.11)</td>
<td>0.09</td>
</tr>
<tr>
<td>1</td>
<td>(0.36, 0.54, 0.11)</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>(0.63, 0.24, 0.11)</td>
<td>0.63</td>
</tr>
<tr>
<td>3</td>
<td>(0.8, 0, 0.11)</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Since the edge weights in Advogato and PGP are all ordinal data, we use the linear transformation technique [8], [11] to convert ordinal values into real numbers. As shown in TABLE II, trust data from original datasets are transformed to opinion vectors and real numbers, ranging from 0 to 1. All of these algorithms are implemented by python 2.7. The computer used for evaluations has an Intel CORE i7-3770 3.4 GHz CPU, 16G memory and 256G SSD hard drive.

**A. Accuracy**

The goal of this subsection is to check if OpinionWalk offers more accurate trust assessment results. MoleTrust and TidalTrust output trust values, EigenTrust and TrustRank give trust rankings; therefore, we divide the experiments into two groups. In the first group, we compare OpinionWalk to MoleTrust and TidalTrust by looking at the errors between computed trust values and the ground truth. In the second group, we compare OpinionWalk to EigenTrust and TrustRank by checking the Kendall’s tau ranking correlation coefficient between the rankings generated by the algorithms and the ground truth. The AssessTrust algorithm is not compared here because it provides the same results as OpinionWalk, which is proved previously.

The experiments are carried out as follows. First, we randomly select a pair of users \( u \) and \( v \) where \( u \) connects to \( v \) in the datasets. We treat the direct opinion from \( u \) to \( v \) as the ground truth. Then, we remove the edge \((u, v)\) from the datasets. We run the above-mentioned algorithms to compute \( v \)'s trustworthiness, from \( u \)'s point of view. Finally, we compare the computed results to the ground truth.

To make a fair comparison, we use the expected beliefs of computed opinions [4] to represent the trust values computed by OpinionWalk. In the first group, we use the mean absolute percentage error (MAPE) to measure the errors in computed trust values. In the second group, we rank users based on their trust values obtained from these algorithms. The ranking errors are then measured by the Kendall’s tau ranking correlation coefficients between computed results and the ground truth. We repeat every experiment 100 times to get statistically significant results.

**Fig. 6.** Errors in trust assessments of different algorithms, using the (a) Advogato and (b) PGP datasets.

In Fig. 6, we find OpinionWalk achieves a higher accuracy, compared to MoleTrust and Tidal Trust, in both datasets. In Advogato, the MAPEs of OpinionWalk, MoleTrust and TidalTrust are 0.15, 0.26, 0.23, respectively. In PGP, the MAPEs of OpinionWalk, MoleTrust and TidalTrust are 0.77, 2.89 and 2.76, respectively. OpinionWalk outperforms other algorithms by 30% to 70% in different datasets. This is because OpinionWalk uses 3VSL to model trust, and 3VSL is believed to be the most accurate model for trust assessments in OSNs.

In Fig. 7, OpinionWalk also gives more accurate ranking results, compared to other algorithms. With the Advogato dataset, the Kendall’s tau correlation coefficients between the ranking results generated from OpinionWalk and the ground truth are always greater than 0. Nearly 20% of ranking results are exactly the same as the ground truth, i.e., coefficients equal to 1. Using the PGP dataset, the Kendall’s tau correlation coefficients of OpinionWalk are greater than 0.2. Moreover, about 40% of ranking results are the same as the ground truth. We clearly see that OpinionWalk offers more accurate ranking results than EigenTrust and RankTrust. From Figs. 6 and 7, we conclude that OpinionWalk gives more accurate
to the rapid development of OSN-related applications and services, OpinionWalk will attract more research interests in the future.

**REFERENCES**


