Dinh Que Tran
Phuong Thanh Pham

TREEXTRUST: TOPIC-AWARE COMPUTATIONAL TRUST BASED ON INTERACTION EXPERIENCE, REPUTATION OF USERS WITH SIMILARITY, AND PATH ALGEBRA OF GRAPH IN SOCIAL NETWORKS

Abstract *The trust measure is the confidence or reliability among users or peers, which has been studied widely in online social networks. Most trust models are currently based on the concepts of interaction trust and reputation trust; however, various forms of interactions and analyses of the interaction contexts have not been considered fully for trust estimation. Moreover, the mechanism for computing reputation trust based on propagation lacks a clear foundation and is expensive in computation. The purpose of this paper is to present a family of computational trust models (called TreeXTrust) to estimate the trust degree of a user* truster *on another user* trustee. *Our model is a mathematical formulation that is based on an aggregation of topic-aware experience trust with various forms of interactions and topic-aware reputation trust with users' similarity and operators on path algebra in a graph. We conducted experiments to evaluate the impacts of interaction forms and users' interests on experience trust and the correlation of experience trust and reputation trust on overall trust estimation. Our experimental results demonstrated the following: (i) interest degrees influenced experience trust more than interaction ones did; (ii) a community's evaluation of some trustee affected an overall trust estimation more than a truster's individual evaluation did. Our family of models outperformed the state-of-the-art methods that have been presented in the literature and is a framework for selecting and implementing a suitable model of computational trust for our problem at hand.*

Keywords computational trust, social network, interaction, interests, reputation

Citation Computer Science 26(2) 2025: 5–31

1. Introduction

In today's technological landscape, distributed computational systems have evolved from closed structures into dynamic frameworks with various autonomous components. These systems include e-commerce platforms, recommender systems, multi-agent systems [7, 13, 30, 34], and online social networks [4, 6, 8, 10, 15, 16, 24, 26, 29, 31, 33, 37, 38, 41, 42]. The common thread is that individual components called "peers" make decisions based on self-interest while collaborating through interactions. This cooperation hinges on trust – a cornerstone for building relationships, sharing information, and fostering connections. Often explored in psychology, social sciences, and computer science, trustworthiness among peers plays a crucial role for such activities. Marsh's work [28] is considered to be an initiated computational trust model in computer science and has been further developed in various research areas like multi-agent systems and complex online networks. In online networks, trust is often seen as the confidence that a "truster" has in a "trustee" during their interactions. Modern models draw from individual experiences and community evaluations to determine trustworthiness, accounting for dynamics, context, and communal factors; this leads to changing levels of trustee trustworthiness over time. Methods for measuring trust encompass statistical approaches, machine learning, heuristics, and behavioral analysis. These techniques focus on network structures, transmitted message content, and past interactions (like feedback). The purpose of this article is to introduce the TreeXTrust framework – a set of trust models that merge topic-aware experience trust and topic-aware reputation trust based on user interactions and community dynamics. Topic-aware experience trust estimation considers two main aspects: interaction trust with familiarity, dispatching, and response degrees, and user interests from message exchanges ("entries"). Topic-aware reputation trust uses community evaluations based on similarity peers or propagation computation via path algebra. We also acknowledge Hamdi's work [14] on trust quantification in network structures and the concept of "Path Strength." In addition, we incorporate the principle of "homophily" that can be observed in online social networks to measure similarities within communities and exploit path-algebra operators to gauge trustworthiness from a community.

2. Related work

The first attempt to propose a computational trust model in computer science was made by Marsh [28]. Trust computational techniques [37] can be classified into statistical, machine learning, heuristic-based, and behavior-based techniques. Nearly all techniques focus their efforts on constructing models by means of network structures, the contents of messages that are transmitted among users, and the interaction types that have occurred between peers in the past (such as sending, feedbacking, and forwarding).

Along with these studies, this paper proposes a family of models of computational trust called TreeXTrust to estimate the trust degree of a user *truster* on another user *trustee*. Our model is a mathematical formalization for estimating *topic-aware trust* degrees among peers based on the aggregation function of *topic-aware experience trust* with users' topic interests by means of past interactions and *topic-aware reputation trust* being evaluated from some community or group of users. In order to estimate topic-aware experience trust, we defined a function with two parameters: (i) degrees of experience trust (which is defined by means of forms of interactions such as familiarity degrees, degrees of responses, and dispatching among users); and (ii) the degrees of users' interests (where interest degrees are determined from entries that have been dispatched among the users). We shared the computing of the interaction score by relying on the assumption of the frequency of the interactions of neighbors with the other works such as TidalTrust [11], TrustWalker [17], SWTrust [18], and LoTrust [19]. However, our work goes further by investigating various types of interactions and analyzing the contents of messages to determine users' interests. Then, topic-aware experience trust was estimated by means of an aggregation function of the interaction scores of a truster with a trustee and the trustee's interest degrees. In contrast with other studies such as [19] in building user's interests (which made use of the SPARQL query language), we utilized the approach of the semantic extension of words by means of Wikipedia that was proposed by Gabrilovich et al. [9] and Kang et al. [20]. We analyzed entries into words by the tf-idf technique [27] to compute the weights of the words in a document for representing vectors of entries and topics and then defined the interest degrees in the topics. The estimation of trustworthiness was considered to be a refinement of the experience trust computation, which we proposed in our previous work [32, 41].

Reputation trust [13,30,34,37] is defined as the reliability of a peer on another one that is inferred from some community or group of peers. Some works have made use of the propagation of trust via the graph structure of the network to construct reputation trust; e.g., TidalTrust [11], TrustWalker [17], and SWTrust [18]. Their approach was to select some paths for computations in order to avoid computational complexity; for example, selecting the shortest path that connects a truster and a trustee. However, the problem with this approach is that it lacks the basics for such a computation. In this paper, we provide techniques for estimating trustworthiness from a community by using similarity measures or operators in path algebra.

For similarity in a community, we accept the characteristic of *homophily* in online social networks [10, 21, 22, 26, 37]. This concept means that peers tend to associate and interact with similar ones; it has been widely studied in the literature; e.g., [2,3,22]. Khanam et al. [22] stated that the study of homophily can provide important insights into the flow of information and the behaviors of users and is extremely useful in analyzing the formations of online communities. However, their research lacked formularization and, thus, resulted in difficulties in implementation.

According to Golbeck [11], multiple routes of trust interconnect two users from a source to a destination within trust networks. In order to quantify the trustworthiness that is inherent in such relationships, Hamdi [14] introduced the concept of "path strength" in his thesis and established the notion of a "perfect" path; by the means of this, one can achieve a maximum path strength value.

In this paper, we define a similarity measure to be an aggregation of two factors: a profile similarity, and an interest similarity. While an interest similarity results from the similarity of users' entries with topics, a profile similarity is based on the similarity of entries among the users. We also provide the parameters to measure the similarity between users for the proposed trust formula; then, we obtain alternative evaluation models from communities depending on the various similarity degrees (w.r.t. truster or trustee). The details of this are presented in Section 5. In addition, we develop a technique for computing the trustworthiness from a community based on the path algebra operators of a graph in order to estimate the trust propagation [15, 25, 35, 40, 43].

3. Major contribution and organization

First, we constructed a hierarchical structure of neighbors (w.r.t. some peer truster u_i). Layer L_i^1 consisted of neighbors that were connected directly to u_i, and L_i^2 consisted of users that were directly connected to users in L_i^1 but not to those in L_i^1, and so on. The recursive definition is formulated in Section 2. Then, estimating the trustworthiness of u_i on trustees $u_j \in L_i^k$ $k \geq 1$ was computed by means of a concatenation operator along a path as well as an aggregation of one that combined various paths. The details are given in Section 6. In the scope of this paper, we merely focused on considering trustees who had some direct interaction with the truster; then, the community was composed of those peers in L_i^1 who interacted directly with both the u_i and u_j trusters. A full paper of integrating similarity and path algebra for estimating the trustworthiness of a truster on those trustees that belong to layers L_i^k where $k \geq 2$ will be presented in another work of ours.

We conducted experiments on two distinct data sets in order to evaluate our TreeXTrust model. The purpose was to validate how trustworthiness was affected by factors of interaction experience, user's interests, and evaluations from the community. This paper is an extension, upgrade, and update of our previous work [32,39,41]. Our contributions are summarized as follows:

- Proposing a computational model of topic-aware interaction experience trust with various types of interactions and user's interest degrees. The model is an aggregation function of the trustworthiness of a truster on trustees, with the interaction forms being familiarity, response and dispatching, and the degrees of the trustee's interests on the topics.
- Proposing similarity measures that are based on the similarity of the users' profiles and interest measures. In turn, the similarity in interest according to the topics and in the user profiles are constructed from natural language-processing techniques for representing the vectors of the topics and the vectors of the entries. Similarity measures contribute to the formulation of computing topic-aware reputation trust.

- Proposing techniques for estimating reputation trust from communities, including *repmaX, repaP, repeS, repeeS.* While repeS results from the similarity neighbors of the truster, repeeS is based on the similarity with the trustee. The formulation with repmaX and repaP are constructed from path algebra operators that connected the truster and trustee.
- Proposing the overall trust-estimation formula by weightedly aggregating topic-aware experience trust and reputation trust. This formula defines a family of models for computing topic-aware trust when combining various methods of estimating experience trust with interest measures and reputation estimation according to different techniques.
- Performing experiments on two distinct data sets (one in Vietnamese, and one in English) to demonstrate the following:
  - how parameters of users' interests and interaction degrees affect experience trust computation in our proposed model;
  - how experience and reputation trust contributes to overall trust in evaluating partners;
  - which forms of reputation have stronger impact while evaluating trustworthiness from community.

The remainder of this paper is structured as follows. Section 4 describes a model of a social network and its hierarchy structure. Section 5 presents an analysis method of the entry data of users' interests in topics and similarity measures. Section 6 presents path algebra. Section 7 is first devoted to representing types of user interactions with familiarity, response, and dispatching. Then, we present a formula for topic-aware experience trust computation based on the types of the proposed interaction and the user's interest degrees. Section 8 presents a reputation-trust computation from communities that is inferred from similarity measures and path algebra. Then, it presents the overall topic-aware trust, which aggregates experience trust and reputation trust. Section 9 is devoted to describing the experimental evaluation. Section 10 offers our conclusions.

4. Social network and hierarchical structure

4.1. Model of social network

A social network is defined as a directed graph $\mathcal{S} = (\mathcal{U}, \mathcal{I}, \mathcal{E}, \mathcal{T})$, in which:

- $\mathcal{U} = \{u_1, \ldots, u_n\}$ is a set of users whose elements are autonomous entities that are called *peers*. In this paper, the terms "peer" and "user" are used interchangeably.
- $\mathcal{I}$ is a set of all interactions or connections u_{ij} from u_i to u_j; $|I_{ij}|$ is denoted as the number of such interactions. Each interaction between users u_i and u_j is a transaction at an instant time, which occurs when u_i sends messages such as posts, comments, likes, opinions, etc. to u_j via some "wall."

- $\mathcal{E} = \{E_1, \ldots, E_n\}$ is a set of entries that are dispatched by users u_i, where $E_i = \{e_{i1}, \ldots, e_{in_i}\}$. Each *entry* is a brief piece of text that is given by some user u_i to make a description or to post information/comments/opinions about an item (such as a paper, a book, a film, a video, etc.).
- $\mathcal{T} = \{t_1, \ldots, t_p\}$ is a collection of topics in which each topic is defined as a set of terms or words.

4.2. Neighbor-based hierarchy structure

This section presents an update of the hierarchy structure of users based on layers of users, which was proposed by us in our previous work [39]. The structure is intuitive for describing a community that is useful in reputation-trust computation with a similarity measure and path algebra in the next section. In the following, we will formalize the concept of levels based on the neighbors of peers. If u_i has some direct interaction with u_j, then u_j is called a neighbor of Layer 1 (or *1-neighbor* of u_i). We make a convention that the 0-neighbor of u_i is u_i. The concept of the *k-neighbor* of u_i is defined recursively as follows.

Definition 1. *A user u_j is a neighbor of a user u_i if there is an interaction from u_i to u_j. Let $I_{i\rightarrow}$ be a set of all u_j such that there is an interaction from u_i i to u_j.*

Definition 2. *Given a peer u_i, a peer u_j is a k-neighbor of u_i ($k \geq 2$) if the two following conditions are satisfied:*

(i) u_j has no direct interaction from any peer of l-neighbor of u_i for all $l \leq k-2$;

(ii) there is at least a peer of (k-1)-neighbor of u_i, which has some direct connection with u_j.

Denote L_i^k for all $k \geq 1$ to be a set of k-neighbors of u_i. It is easy to prove the following proposition:

Proposition 1. *Given a source peer u_i, there exists a number n_i such that $L_i^1 \ldots, L_i^{n_i}$ are k-neighbors of u_i and satisfy the following conditions:*

(i) for each $v \in L_i^k$ ($k = 2, \ldots, n_i$), v not being interacted directly with any one in $\cup_{l=0}^{k-2} L_i^l$;

(ii) $L_i^k \cap (\cup_{l=0}^{k-1} L_i^l) = \emptyset$, for all $k \geq 1$.

Thus, we have a taxonomy of neighbors of u_i and $L_i^1 \ldots, L_i^{n_i}$ is then called a *taxonomy* or a *hierarchy* of the neighbors of u_i. Estimating the trust value of a source peer u_i on a sink peer u_j depends on whether the sink one belongs to which layer of taxonomy (w.r.t the source). When a sink peer belongs to the hierarchy, the trust estimation is based on the interaction experience, similarity, and path algebra. A sink peer that is not of the hierarchy of a source peer is called its ∞-neighbor. We have the following definition:

Definition 3. *A peer u_j is called a p-neighbor (w.r.t. a taxonomy of a source peer u_i) if $u_j \in L_i^p$ for all $p = 1, \ldots, n_i$.*

Definition 4. *A peer u_j is called a ∞-neighbor (w.r.t. a taxonomy of a source peer u_i) if $u_j \notin L_i^k$ for all $k = 1, \ldots, n_i$.*

Definition 5. *Given two peers u_i and u_j, path $p(i,j)$ connects two peers if there exists a sequence of peers u_k ($k = 0,\ldots,q$) that has a connection that is coupled with each other: $u_i = u_0$ connects with u_1, u_1 connects with $u_2,\ldots,u_{q-1}$ connects with $u_j = u_q$.*

We have the following proposition:

Proposition 2. *Given a source peer u_i, suppose u_j is a p-neighbor of u_i where $1 \le p \le n_i$. There always exists a path $p(i,j)$ that connects u_i and u_j. Denote $\Phi(i,j)$ to be a set of all paths $p(i,j)$ that connect u_i and u_j.*

Our problem is how to estimate the topic-trust values of truster u_i on trustees u_j. There are three cases: (i) there is a direct interaction between u_i and u_j; (ii) there is not any direct interaction between truster u_i and trustee u_j, but there exists a path $p(i,j)$ that connects u_i and u_j; (iii) there is no path that connects u_i and u_j, which means that u_j is a ∞-neighbor (w.r.t u_i). The scope of this paper focuses on investigating the first case in which there is a direct interaction between source peer u_i and sink peer u_j. The detail of the topic-trust model and the techniques of computation will be presented in the next sections. In the two remaining cases, we make use of the integration of path algebra and similarity measures for reputation estimation; the research results are presented in another of our works.

5. User interests and similarity measures

5.1. Vectorial representation of texts

The vectorial model for representing texts by means of tf-idf has been widely used in various fields of computer science, such as information retrieval and text mining [27]. Along with the works that are related to extending semantics ([9] and [20]), we applied the approach for extracting entries into words and enriching these bags of words into semantics words based on Wikipedia (https://vi.wikipedia.org/wiki/). Then, we vectorized the entries and topics with word weights in the texts.

From then on, a document was always considered to be a set of terms. We made use of the $tf - idf(d, D_i) = tf(d, D_i) \times idf(d, \mathcal{D})$ technique for the vectorial representation of such entries and topics, where $tf(d, D_i)$ was the frequency that term d appeared in D_i, and $idf(d, \mathcal{D}) = log(\frac{|\mathcal{D}|}{1+\{|D_i||d \in D_i\}})$. The vector representation in its general form is described as follows.

Given a collection of documents $D = \{D_1,\ldots,D_p\}$, each is represented as set of terms or words $D_i = \{d_{i1},\ldots,d_{i_{pi}}\}$. Let $V = \{v_1,\ldots,v_q\}$ be a set of all of the distinct terms in the whole collection. The weight of term $d \in V$ (w.r.t. D_i) is defined by $w_d = tf(d, D_i) \times idf(d, D)$. And then, each D_i is represented as a q-dimension vector $\mathbf{D_i} = (w_1,\ldots,w_q)$, where $w_k = tf(v_k, D_i) \times idf(v_k, D)$, $k = 1,\ldots,q$. We utilized the technique to represent the entries and topics in the vectors; these are described in the rest of this subsection.

5.1.1. Representing vectors of entries

Suppose $V_E = \{e_1, \dots, e_r\}$ is a set of r distinct terms in all entries $e_{ij} \in E_i$ in $\mathcal{E}$. An entry vector $\mathbf{e_{ij}}$ (w.r.t. the entry $e_{ij} \in E_i$) is defined as follows:

$$\mathbf{e_{ij}} = (e_{ij}^1, \dots, e_{ij}^{|V_E|}),\ i = 1, \dots, n,\ j = 1, \dots, n_i, \tag{1}$$

where $e_{ij}^k = tf(e_l, e_{ij}) \times idf(e_l, E_i)$, where $e_l \in V_E,\ l = 1, \dots, r$ and $k = 1, \dots, |V_E|$.

5.1.2. Representing vectors of topics

Suppose that $V_T = \{v_1, \dots, v_q\}$ is a set of q distinct terms of all $t_i \in \mathcal{T}$. A topic vector (w.r.t. each topic t_i) is a weighted one, which is defined as follows:

$$\mathbf{t_i} = (w_{i1}, \dots, w_{iq}), \tag{2}$$

where $w_{ik} = tf(v_k, T_i) \times idf(v_k, \mathcal{T})$, $v_k \in V_T$.

Given an entry e_{il} being dispatched by u_i, an entry vector (of w.r.t. topics $\mathcal{T}$, or briefly – a topic entry vector) is a weighted one, which is defined as follows:

$$\mathbf{e_{il}^t} = (e_{il}^1, \dots, e_{il}^p), \tag{3}$$

where $e_{il}^k = tf(v_k, e_{il}) \times idf(v_k, E_i)$, $v_k \in V_T$.

Thus, we can define a sequence of topic entry vectors $\mathbf{e_{il}^{t_1}}, \dots, \mathbf{e_{il}^{t_p}}$ (w.r.t. each entry) and a sequence of entry vectors $\mathbf{e_{i1}}, \dots, \mathbf{e_{in_i}}$ (w.r.t. entries $e_{ij} \in E_j$). These vectors will be utilized for constructing the measures of the user's similarity and interests; these are presented in the next subsection.

5.2. User interest degrees

The traditional Pearson correlation degrees $\mathrm{cor}(\mathbf{e_{ij}^t}, \mathbf{t_k})$ among entries e_{ij} given by u_i (w.r.t. topics t_k) are as follows:

$$\mathrm{cor}(\mathbf{u}, \mathbf{v}) = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2} \times \sqrt{\sum_i (v_i - \bar{v})^2}}, \tag{4}$$

where $\bar{u} = \frac{1}{n}(\sum_{i=1}^n u_i)$, and $\bar{v} = \frac{1}{n}(\sum_{i=1}^n v_i)$. Since the values of function $\mathrm{cor}(x, y)$ belong to $[-1, 1]$, we may make use of function $f(x) = \frac{(x+1)}{2}$ to bound the values of function $\mathrm{cor}(x, y)$ into unit interval $[0, 1]$. This means that, instead of Formula (4), the following formula (5) will be applied in the paper:

$$\mathrm{cor}(\mathbf{u}, \mathbf{v}) = \frac{\frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2} \times \sqrt{\sum_i (v_i - \bar{v})^2}} + 1}{2}. \tag{5}$$

Definition 6. *Let $\mathcal{P}(E_i)$ be a set of all subsets of entries E_i given by $u_i \in \mathcal{U}$ and $\mathcal{P}(\mathcal{E}) = \bigcup_{u_i \in \mathcal{U}} \mathcal{P}(E_i)$. A function $f : \mathcal{U} \times \mathcal{P}(\mathcal{E}) \times \mathcal{T} \to [0, 1]$ is called an interest measure if it satisfies condition $f(u_i, Y_1, t) \le f(u_i, Y_2, t)$ for all $Y_1, Y_2 \in \mathcal{P}(E_i)$ such that $Y_1 \subseteq Y_2$.*

It is easy to prove the following proposition:

Proposition 3. *A function* $f_{\text{interest}} : \mathcal{U} \times \mathcal{P}(\mathcal{E}) \times \mathcal{T} \rightarrow [0,1]$ *is an interest measure if and only if it satisfies the following conditions:*

(i) If $\text{cor}(\mathbf{e_{i,k}}, \mathbf{t_j}) \geq \text{cor}(\mathbf{e_{i,k}}, \mathbf{t_l})$*, then* $f_{\text{interest}}(u_i, e_i, t_j) \geq f_{\text{interest}}(u_i, e_i, t_l)$,

(ii) If $\text{cor}(\mathbf{e_{i,k}}, \mathbf{t_h}) \geq \text{cor}(\mathbf{e_{j,l}}, \mathbf{t_h})$*, then* $f_{\text{interest}}(u_i, e_i, t_h) \geq f_{\text{interest}}(u_j, e_i, t_h)$.

Definition 7. *An entry* e_{ij} *is called* θ-entry *(w.r.t. topic* t_k*) if and only if* $\text{cor}(\mathbf{e^t_{ij}}, \mathbf{t_k}) \geq \theta$*, where* $0 < \theta \leq 1$ *is a given threshold.*

We have the following proposition:

Proposition 4. *Suppose* $|E_i|$ *is the number of elements in* E_i *and* n_i^t *is the number of* θ*-entries that are concerned with the topic* t *given by* u_i*. The following are the interest measures:*

$$\text{intMax}(u_i, t) = \max_j(\text{cor}(\mathbf{e^t_{ij}}, \mathbf{t})), \tag{6}$$

$$\text{intCor}(u_i, t) = \frac{\sum_j \text{cor}(\mathbf{e^t_{ij}}, \mathbf{t})}{|E_i|}, \tag{7}$$

$$\text{intSum}(u_i, t) = \frac{1}{2}\left(\frac{n_i^t}{\sum_{l \in \mathcal{T}} n_i^l} + \frac{n_i^t}{\sum_{u_k \in \mathcal{U}, l \in \mathcal{T}} n_k^l}\right). \tag{8}$$

For easy presentation, we denote $\text{intX}(u_i, t)$ to be one of the above formulas in which X may be $\text{Sum}, \text{Cor}, \textit{or}\ \text{Max}$. The interest vector of the users in various topics is defined by the following formula:

$$\mathbf{u_i^t} = (u_i^1, \ldots, u_i^p), \tag{9}$$

in which $u_i^k = \text{intX}(u_i, t_k)$ is the interest degree of user u_i in topics $t_k \in \mathcal{T}$ $(k = 1, \ldots, p)$, X may be $\text{Sum}, \text{Cor}, \text{Max}$ as defined in Proposition 4. The definition of the vectors with interest degrees is utilized for constructing the similarity of the users; this is presented in the next subsection.

5.3. Similarity measure

The similarity measure has been widely used to construct the recommendations of items and the services in the recommender system [11] as well as in social networks [6,12]. Golbeck [12] showed that there was a strong and significant correlation between trust and user similarity: the more similar two people were, the greater the trust between them. In contrast to the similarity that was inferred from the ratings of films, however, the similarity that we constructed in this paper resulted from the profiles of user comments that were dispatched on social networks. Since there is no clear definition of similarity in the literature, we formalized the definition of similarity based on the usual metric measure as follows.

Definition 8. *Given a vector space V, a function $sim : V \times V \to [0,1]$ is a similarity measure if it satisfies the following conditions:*

(i) $sim(u,u) = 1$, for all $u \in V$;
(ii) $sim(u,v) = sim(v,u)$ for all $u, v \in V$;
(iii) $sim(u,w) + sim(w,v) - sim(u,v) \leq 1$ for all $u, v, w \in V$.

5.3.1. Similarity of users' interest

The interest similarity of two peers u_i and u_j in topic t is defined to be a similarity of two vectors $\mathbf{u_i^t}$ and $\mathbf{u_j^t}$ as follows:

$$sim_t^X(i,j) = \frac{< \mathbf{u_i^t}, \mathbf{u_j^t} >}{\|\mathbf{u_i^t}\| \times \|\mathbf{u_j^t}\|}, \tag{10}$$

in which $< u, v >$ is the scalar product, $\times$ is the usual multiple operations, and $\|.\|$ is the Euclidean length of a vector; X is Max, Cor or Sum up on the selection of interest degree as defined in Proposition 4.

5.3.2. Profile similarity

Given two peers u_i and u_j, the profile similarity of two peers u_i and u_j is defined as an overall similar function of all vectors $\mathbf{e_{ik}}$ and $\mathbf{e_{jl}}$ where $k = 1, \ldots, n_i$, $l = 1, \ldots, n_j$. This is defined by the following formula:

$$sim_{prof}(i,j) = F\left(\frac{< \mathbf{e_{ik}}, \mathbf{e_{jl}} >}{\|\mathbf{e_{ik}}\| \times \|\mathbf{e_{jl}}\|}\right), \tag{11}$$

in which F may be a usual *min* or *average*.

5.3.3. User similarity

Based on the definition of the similarity of interest and profile, we have the definition of the similarity of the users as follows.

Definition 9. *The similarity between two users u_i and u_j is defined by the weighted composition of their partial similarities and given by the following formula:*

$$sim^X(i,j) = \alpha \times sim_{prof}(i,j) + \beta \times sim_t^X(i,j), \tag{12}$$

where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$.

Thus, there are three similarity measures among two users: $sim^{max}(i,j)$, $sim^{cor}(i,j)$, and $sim^{sum}(i,j)$ (w.r.t. X is Max, Cor, or Sum). We have the following proposition:

Proposition 5. *$sim_t^X(i,j)$, $sim_{prof}(i,j)$ and $sim^X(i,j)$ are similarity measures.*

6. Path algebra-based trust computation

This section presents an application of path algebra [15, 25, 35, 40, 43] for the computation of the reputation trust from a community. We reformalized the necessary formulas from the work that was given by Wang et al. [43] for the purposes of our paper. We proceeded to construct a class of functions for estimating topic trust by means of community as follows.

Definition 10. *Given a set of natural numbers* $\mathbb{N}$*, an operator* $f_{path}^{trust} : \cup_{n\in\mathbb{N}}[0,1]^n \to [0,1]$ *is called an aggregation operator if it fulfills the following conditions:*

(i) $f_{path}^{trust}(0,\ldots,0) = 0$ *and* $f_{path}^{trust}(1,\ldots,1) = 1$*;*

(ii) For all k*,* $x_1 \le y_1 \ldots x_n \le y_n \Rightarrow f_{path}^{trust}(x_1,\ldots,x_n) \le f_{path}^{trust}(y_1,\ldots y_n)$*.*

It is easy to prove the following proposition.

Proposition 6. *Mappings* $f : [0,1]^n \to [0,1]$*, which are defined by the following formulas, are aggregation operators:*

(i) $f(x_1,\ldots,x_n) = \max(x_1,\ldots,x_n)$*;*

(ii) $f(x_1,\ldots,x_n) = \min(x_1,\ldots,x_n)$

(iii) $f(x_1,\ldots,x_n) = \Pi_{i=1}^{n} x_i$*;*

(iv) $f(x_1,\ldots,x_n) = \frac{\sum_{k=1}^{n} x_k}{n}$*.*

Based on the paths that connect u_i and u_j, we were able to compute the topic-trust values for this couple by means of the aggregation operators. We applied the aggregation operators to various paths and along a path. The overall trust value is then called the path-based topic-aware reputation trust. We formulate these statements as follows.

Definition 11. *Suppose that* $p(i,j)$ *is a path with a length of* m *connecting* u_i *and* u_j*. The topic trust of* u_i *on* u_j *along the path is defined by the following formula:*

$$trust_{topic}^{p(i,j)}(u_i,u_j) = f_{path}^{trust}(u_{i1},\ldots,u_{mj}), \tag{13}$$

where u_{kl} *are the topic-trust values* u_k *that are assigned to* u_l*, and* $f_{path}^{trust}(p)$ *is an aggregation operator.*

Definition 12. *Suppose that* $\Phi(i,j)$ *to be the set of paths* $p(i,j)$ *from* u_i *to* u_j*. Then, the path-based topic-aware reputation trust of* u_i *on* u_j *of* t *is defined by the following formula:*

$$trust_{topic}^{path}(i,j,t) = f_{p(i,j)\in\Phi(i,j)}(trust_{topic}^{p(i,j)}(i,j,t)), \tag{14}$$

in which $trust_{topic}^{p(i,j)}(i,j,t) = f_{path}^{trust}(u_{i1},\ldots,u_{mj})$ *is the topic trust of* i *on* j *along path* $p(i,j)$*.*

For convenience in presentation, we utilized two operators (*concatenation* $\otimes$ and *aggregation* $\oplus$), which are operations along a path and combining various paths, respectively, that were proposed by Hang et al. [15]. We reformulated the trust-computation formula as follows.

Definition 13. *Suppose that $\Phi(i,j)$ is the set of paths $p(i,j)$ that connect u_i and u_j. The path-based topic-aware reputation trust of u_i on u_j of t is defined by the following formula:*

$$trust_{topic}^{path}(i,j,t) = \oplus_{p(i,j)\in\Phi(i,j)}(\otimes_{k,l} trust_{topic}^{exp}(k,l,t)), \tag{15}$$

where $\otimes$ and $\oplus$ are the concatenation and aggregation operators, respectively.

In this paper, we made use of the operators in Proposition 6 for our implementation; these are presented in Section 8.

7. Topic-aware experience trust

This section is devoted to presenting the experience-based trust computation with the different types of interaction: familiarity, dispatch, and responses. Compared to the solely dispatch interaction type that was previously proposed by us [41], the refinement model emerged from the following observations:

- *Familiarity:* if a user u_i shares with u_j a set of common users more than with u_k, then u_i is more familiar with u_j than with u_k;
- *Response:* a response that is given by a target to a resource is feedback (comment, like, etc.) or sharing with others when the target receives a message from the resource;
- *Dispatching:* when a user u_i sends messages to u_j more than to u_k, then u_j is more reliable than u_k for user u_i.

These concepts are formalized in the next subsection.

7.1. Familiarity, response, and dispatching

Definition 14. *The degree of the familiarity of two peers u_i and u_j is defined as follows:*

$$famil(i,j) = \frac{|I_{i\rightarrow} \cap I_{j\rightarrow}|}{|I_{i\rightarrow} \cup I_{j\rightarrow}|}, \tag{16}$$

where $I_{i\rightarrow}$ is the set of u_k to which u_i sends messages.

Definition 15. *Suppose that I_{ij} is a set of all interactions from i to j, and $|I_{ij}|$ is its number of elements. The dispatching degree from a user u_i to user u_j (denoted as $disp(i,j)$) is defined by the following formula:*

$$disp(i,j) = \frac{|I_{ij}|}{\sum_{k=1}^{n}|I_{ik}|}, \tag{17}$$

where $|I_{ik}|$ is the number of interactions of u_i with each $u_k \in \mathcal{U}$.

Definition 16. *Given that $I_{i\leftarrow j}^{resp}$ is a set of all responses from u_j to u_i, where a response is feedback from u_j when receiving a message that is dispatched from u_i. The degree of the response of u_j to u_i is defined as follows:*

$$resp(i,j) = \frac{|I_{i\leftarrow j}^{resp}|}{|\bigcup_k I_{k\leftarrow j}^{resp}|}. \tag{18}$$

Definition 17. *The interaction experience trust of user u_i on user u_j (denoted as $trust^{exp}(i,j)$) is defined by the following formula:*

$$trust^{exp}(i,j) = w_1 \times famil(i,j) + w_2 \times disp(i,j) + w_3 \times resp(i,j) \tag{19}$$

where $w_1, w_2, w_3 \geq 0,\ w_1 + w_2 + w_3 = 1$.

7.2. Refined topic-aware experience trust

This subsection presents a class of functions of two parameters (including the interaction types and interest degrees) in the topics.

Definition 18. *A function $trust_{topic} : \mathcal{U} \times \mathcal{U} \times \mathcal{T} \to [0,1]$ is called a topic-trust measure in which $[0,1]$ is an unit interval of the real numbers if it fulfills the following conditions:*

(i) $(\ intX(u_j,t_1) \geq intX(u_j,t_2)) \Rightarrow (trust^{exp}_{topic}(u_i,u_j,t_1) \geq trust^{exp}_{topic}(u_i,u_j,t_2))$;

(ii) $((trust^{exp}(i,j) \geq trust^{exp}(i,k))\&(\ intX(u_j,t) \geq intX(u_k,t))) \Rightarrow$
$(trust_{topic}(u_i,u_j,t) \geq trust_{topic}(u_i,u_k,t))$.

Definition 19. *Suppose that $trust^{exp}(i,j)$ is the interaction experience trust of u_i on u_j and $intX(j,t)$ is the interest degree of u_j on topic t. Then, the topic-aware experience trust of u_i on u_j of topic t is defined by the following formula:*

$$trust^{exp}_{topic}(i,j,t) = \lambda \times trust^{exp}(i,j) + \mu \times intX(j,t), \tag{20}$$

where $\lambda, \mu \geq 0,\ \lambda + \mu = 1$.

Then, given a source peer u_i, a sink peer u_j, and a topic t, the value $trust_{topic}(i,j,t) = u^t_{ij}$ means that *truster* u_i trusts *trustee* u_j of topic t (w.r.t. the degree u^t_{ij}). We have the following proposition:

Proposition 7. *the topic-aware experience trust $trust^{exp}_{topic}(i,j,t)$ of u_i on u_j of topic t, which is defined by Formula (20), is a topic-trust measure.*

Proof. Let the parameters λ and μ in Formula (20) be unchangeable. It is easy to confirm that the formula satisfies Conditions (i) and (ii) in Definition 18. □

8. Topic-aware reputation and overall topic-aware trust

When a truster lacks information or is uncertain of his/her own evaluation of some trustee, he/she will utilize opinions or evaluations from some community. Reputation trust is considered to be the trustworthiness degree that some community assigns to a trustee. Overall, topic-aware trust is a composition of topic-aware-experience and topic-aware-reputation trust degrees. We consider four types of computation from a community (w.r.t. selected operators and similarity degrees to a truster or trustee):

- *remaX*: based on community with neighbor peers who evaluate trustees with paths according to usual max and $\times$ operators, respective to aggregation and concatenation;
- *repaP*: based on community with neighbor peers who evaluate trustees with paths according to trust-average and usual $\times$ operators, respective to aggregatiion and concatenation;

- *repeS*: based on community with neighbor peers who are similar with truster;
- *repeeS*: based on community with neighbor peers who are similar with trustee.

We formalize these types in the definitions that are presented in the next subsection.

8.1. Community-based reputation-trust estimation

Definition 20. *Given a source peer u_i and L^1_{ij} is the 1-neighbors of both u_i and u_j. The topic-aware reputation trust of u_i on u_j with repmaX is defined by the following formula:*

$$trust^{repmaX}_{topic}(i,j,t) = \max_{v\in L^1_{ij}}(trust^{exp}_{topic}(i,v,t) \times trust^{exp}_{topic}(v,j,t)) \tag{21}$$

in which $trust^{exp}_{topic}()$ is the topic-aware experience trust that is given in Formula (20).

Definition 21. *Given a source peer u_i and L^1_{ij} is the 1-neighbors of both u_i and u_j, the topic-aware reputation trust of u_i on u_j with repaP is defined by the following formula:*

$$trust^{repaP}_{topic}(i,j,t) = \frac{\sum_{v\in L^1_{ij}}(trust^{exp}_{topic}(i,v,t) \times trust^{exp}_{topic}(v,j,t))}{\sum_{v\in L^1_{ij}} trust^{exp}_{topic}(v,j,t)} \tag{22}$$

in which $trust^{exp}_{topic}()$ is the topic-aware experience trust that is given in Formula (20).

Definition 22. *Given a source peer u_i and L^1_{ij} is the 1-neighbors of u_i and u_j, the topic-aware reputation trust of u_i on u_j with trustee similarity (repeeS) is defined by the following formulas:*

$$trust^{repeeS}_{topic}(i,j,t) = \frac{\sum_{v\in L^1_{ij}} trust^{exp}_{topic}(i,v,t) \times sim(v,j)}{\sum_{v\in L^1_{ij}} sim(v,j)} \tag{23}$$

in which $sim(v,j)$ is the similarity measure of v on u_j being defined by Formula (12).

Definition 23. *Given a source peer u_i and L^1_{ij} is the $1-$level neighbors of u_i and u_j. The topic-aware reputation trust of u_i on u_j with truster similarity (repeS) is defined by the following formulas:*

$$trust^{repeS}_{topic}(i,j,t) = \frac{\sum_{v\in L^1_{ij}} trust^{exp}_{topic}(v,j,t) \times sim(i,v)}{\sum_{v\in L^1_{ij}} sim(i,v)} \tag{24}$$

in which $sim(i,v)$ is the similarity measure of v on u_i that is defined in Formula (12).

Proposition 8. *Functions $trust^{repY}_{topic}()$ given by formulas (21), (22), (23), and (24), where $repY$ is repmaX, repaP, repeS, or repeeS are topic-trust measures.*

Proof. We need to prove that Formulas (21), (22), (23), and (24) satisfy Conditions (i) and (ii) in Definition 18. It is clear that, since $trust^{exp}_{topic}(i,v,t)$ and $trust^{exp}_{topic}(v,j,t))$ satisfy Conditions (i) and (ii), $trust^{exp}_{topic}(i,v,t) \times trust^{exp}_{topic}(v,j,t))$ the max of them are also satisfied. Thus, $trust^{repmaX}_{topic}(i,j,t)$ is a topic-trust measure. Similarly, when we take the sum with all $v \in L^1_{ij}$ of $trust^{exp}_{topic}(i,v,t) \times trust^{exp}_{topic}(v,j,t))$ and then normalize it in order to result to the value in interval $[0,1]$ by dividing $\sum_{v \in L^1_{ij}} trust^{exp}_{topic}(v,j,t)$, Formula (22) also satisfies the conditions of a topic-trust measure. Thus, $trust^{repaP}_{topic}(i,j,t)$ is a topic-trust measure.

In Formula (23), since $trust^{exp}_{topic}(i,v,t)$ is a topic-trust measure and $sim(v,j) \in [0,1]$, $trust^{exp}_{topic}(i,v,t) \times sim(v,j)$ is a topic-trust measure. And then, normalizing the sum on all $v \in L^1_{ij}$ of $trust^{exp}_{topic}(i,v,t) \times sim(i,v)$ by dividing $\sum_{v \in L^1_{ij}} sim(v,j)$, $trust^{repeeS}_{topic}(i,j,t)$ is a topic-trust measure. Similarly, it is easy to prove that, in Formula (24), $trust^{repeeS}_{topic}(i,j,t)$ is also a topic-trust measure. Thus, the proposition is proven. □

8.2. Overall topic-aware trust

The composition of experience and reputation trust is defined as follows.

Definition 24. *Suppose that $trust^{exp}_{topic}(i,j,t)$ and $trust^{rep}_{topic}(i,j,t)$ are the experience and reputation-trust degrees of u_i on u_j, respectively. Then, the overall topic-aware trust of u_i on u_j of topic t is defined by the following formula:*

$$trust_{topic}(i,j,t) = \gamma \times trust^{exp}_{topic}(i,j,t) + \delta \times trust^{repY}_{topic}(i,j,t) \tag{25}$$

where $repY$ may be $repmaX$, $repaP$, $repeS$, or $repeeS$ and $\gamma, \delta \geq 0$, $\gamma + \delta = 1$.

9. Experimental evaluation

This section is devoted to stating the problem statements, assessment methods, and experimental results for validating our proposed model. We concentrate on investigating the influences of the degrees of interaction, user's interests, and reputation on the trustworthiness estimation of a truster on trustees.

9.1. Problem statement

Problem 1: Evaluating the impacts of interaction measures and user's interest degrees on experience trust as were presented in Section 7. The following issues will be addressed:

- elucidating influence of interactive factors on experience trust; how difference between utilization of single interaction and combinations of interactions forms on trustworthiness estimation;
- delving into trust-computation integration between two users based on their interaction experiences and interests; our objective was to ascertain which factor, experience, or user's interests in particular topic had greater influence on trust.

Problem 2: Comparing similarity-based and path-algebra-based models. To assess and compare models with the different techniques, we focus on two main aspects:

- evaluating user's compatibility metrics by performing comparison repeeS, repeS, and combined (w.r.t. repmaX and repaP);
- comparing our trust models with Hamdi's models [14], which determine reliability using path algebra and similarity.

Problem 3: Assessing the impact of the parameters on the overall trust computation, including the interaction types ($disp, resp, famil$), users' interests ($intX$, where X is Max, Sum, Cor), and methods of determining reputation trust ($repmaX, repaP, repeeS, repS$). By means of our experimental results, we were able to establish which trust model in the family of models was suitable for our application at hand.

9.2. Evaluation methods

We developed a comprehensive test scenario to address the research questions that were mentioned earlier. In the context of large social networks where posts are constantly being published, keeping track of all of the posts (especially those of interest) becomes quite a challenge. As a result, our main focus was on exploring whether we could suggest posts to a specific member (referred to as "x") in the group that matched their preferences and interests. To achieve this goal, we looked into the following aspects:

- Understanding user interests and interactions: By analyzing the data that is available from a group, we investigated what kinds of content a user preferred and how they had interacted with other members in the past.
- Importance of post content in predicting user's interest: We examined whether the content of an post (particularly, its topics) played a significant role in predicting whether a member will be interested in that post.
- Evaluating credibility of post authors: Since posts also include information about their authors, we explored how historical-interaction data could help us determine how much trust a user placed in a specific author. This trust factor then affected how motivated the user was to consider information from this post.

When thinking of a group member as a user on an e-commerce site and a post as a product, we encountered substantial challenges that were analogous to those that had been encountered by the recommendation systems that were employed by major platforms (like Netflix and Amazon) [1] [5] [23] [36]. These predicaments inherently entailed inherent complexities, thus rendering pinpoint predictions an elusive feat. For assessing the efficacy of a system, we utilized conventional metrics such as accuracy, recall, precision, and F1-score. Additionally, the inclusion of a Recall@k metric was utilized to further enhance the comprehensive evaluation of a system's performance. To address these aspects, we proposed methods for calculating the level of a user's interest in a topic ($intX$), finding out how similar two users were in their interests regarding a topic ($sim(i, j, t)$), and establishing the level of trust between two users

($repmaX, repaP, repeeS$, and $repeS$). These methods contributed to input factors that shaped the outcomes of the scenario that we investigated. Precision and recall are given in the following formulas:

$$Precision = \frac{TruePositives}{TruePositives + TrueNegatives} = \frac{TP}{TP + TN} \tag{26}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} = \frac{TP}{TP + FN} \tag{27}$$

The F_1-score was determined by means of precision and recall as follows:

$$F_1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{28}$$

We used the concept "relevant users" to define those who actively engaged by commenting on a post. The notation "**@k**" denotes a precise reference to a specific number of users (being indicated by the integer value "k") that are being considered.

$$Recall@k = \frac{\text{Number of users that are relevant @k}}{\text{Number of all relevant users}} \tag{29}$$

9.3. Experimental data

We utilized two distinct data sets for testing and evaluation in this paper. The first data set, named "DAR – DONG ANH RUNNERS" (https://m.facebook.com/groups/370942430322164/), or DAR[1] for short, consisted of data from a community of running enthusiasts in Vietnam. Vietnamese discussions within this data set revolved around topics that were related to running, such as running fashion, the appropriate diets for running (health in running), running competitions, and various running genres (such as road running and trail running), along with technical aspects of running. According to the statistics as of April 30, 2021, the "DAR – DONG ANH RUNNERS" running group consisted of 497 members; 89 members actively participated in posting a total of 442 posts from 2018 through April 2021. The interactions within the group (including likes and comments) involved 218 members, with nearly 10,000 comments. The second data set that was obtained for this study was sourced from Kaggle and pertained to a Facebook group named "Cheltenham's Facebook Groups" (https://www.kaggle.com/datasets/mchirico/cheltenham-s-facebook-group), or CG for short. The discussions within this group were in English and revolved around everyday issues that were faced by residents of Cheltenham, Pennsylvania, USA, such as traffic problems, sewer issues, and pet-related concerns (dogs, cats) as well as significant matters (like Bill Cosby's lawsuit). The relationships among the users in the two data sets are illustrated in Figure 1.

[1] https://github.com/ThanhPhamPhuong/DARDataset

a) b)

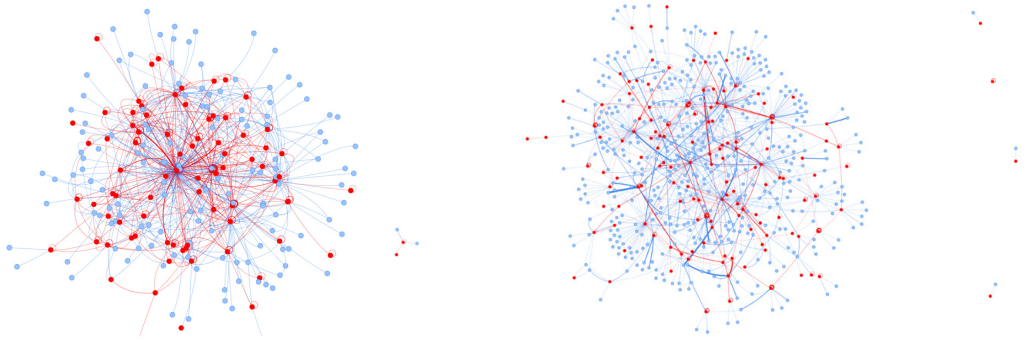

Figure 1. User-relationship-distribution comparison: a) user relationships on DAR data set; b) user relationships on CG data set

When observing the two data sets, we recognized that there were corresponding comments for each post from other users. Each comment could be viewed as a response that reflected the commenter's familiarity with the user being commented on, and the content of the post also indicated the level of the relevance to a particular topic. The interactions that formed $disp(), resp(), famil()$, interest degrees $intX$, and computation methods $repmaX, repaP, repeeS, repeS$ were the parameters that we needed to evaluate their influences on trustworthiness. We made use of a training and testing method that used cross-validation with K-fold, where the main input parameters were the aforementioned parameters. The test data (with $K = 7$) is provided in Table 1. The experimental results of the models will be presented in the next subsection.

Table 1

Parameters of two data sets (DAR and CG)

Data sets	Total posts	Total members	Active members	Train data	Test data	Observation	Labeled data
DAR	441	497	224	378	63	441 × 224 = 98,784	3957
CG	4049	2000	1035	3470	579	4049 × 1035 = 4,190,715	16,601

9.4. Experimental results

To understand how the different types of interactions influenced the outcomes, we examined a scenario in which one type of interaction was used and one where all three were combined. The results are presented in Tables 2 and 3 (w.r.t. the two data sets).

Upon a closer examination, it was clear that the first three models (Models 1, 2, and 3) produced positive outcomes by combining the three different types of interactions when assessing the trust degrees from an experiential perspective.

Table 2
Influence of interaction types on trust CG data set

Model	Resp	Disp	Famil	intMax	intSum	intCor	Recall	Precision	F1	Recall@1	Recall@5	Recall@10	Recall@20
1	x	x	x	x	–	–	0.297	0.291	0.294	0.29	0.61	0.70	0.77
2	x	x	x	–	x	–	0.289	0.296	0.292	0.30	0.62	0.73	0.81
3	x	x	x	–	–	x	0.292	0.291	0.291	0.30	0.61	0.72	0.79
4	–	x	–	x	–	–	0.295	0.264	0.279	0.28	0.61	0.75	0.86
5	–	x	–	–	x	–	0.296	0.256	0.275	0.28	0.54	0.62	0.68
6	–	x	–	–	–	x	0.296	0.259	0.276	0.27	0.53	0.62	0.67

Table 3
Influence of interaction types on trust DAR data set

Model	Resp	Disp	Famil	intMax	intSum	intCor	Recall	Precision	F1	Recall@1	Recall@5	Recall@10	Recall@20
1	x	x	x	x	–	–	0.167	0.634	0.263	0.07	0.35	0.58	0.77
2	x	x	x	–	x	–	0.185	0.602	0.283	0.08	0.33	0.56	0.79
3	x	x	x	–	–	x	0.185	0.61	0.283	0.07	0.33	0.56	0.79
4	–	x	–	x	–	–	0.115	0.523	0.188	0.09	0.31	0.50	0.72
5	–	x	–	–	x	–	0.146	0.456	0.215	0.08	0.24	0.39	0.60
6	–	x	–	–	–	x	0.136	0.536	0.215	0.10	0.29	0.47	0.75

This was different from the results of considering only one type of interaction; this is shown in the next three models (Models 4, 5, and 6). Specifically, these later models performed less favorably when focusing on the dispatch interaction. As previously mentioned, the task of recommending the posts of potential interest to users posed a considerable challenge in terms of achieving precise determinations. To address this challenge, we incorporated a measure of Recall@k where parameter k was evaluated at various values: 1, 5, 10, and 20. The ensuing results are meticulously presented in Tables 2 and 3 when corresponded to the CG and DAR data sets. Upon a closer examination of the findings, it became evident that achieving pinpoint accuracy in the predictions of posts that captured user interest ($k = 1$) yielded an accuracy rate of merely 10% for both the CG and DAR data sets. However, the accuracy rate remains at the same level when confronted with the task of recommending ten posts that were aligned with user preferences (around 10%). It was noteworthy, however, that the accuracy notably surged to 75% and attained an impressive 86% when tasked with suggesting 20 posts that aligned with potential user interest. Remarkably, this heightened accuracy was achieved within the context of a substantial data set that was comprised of more than 4000 posts for the CG data set.

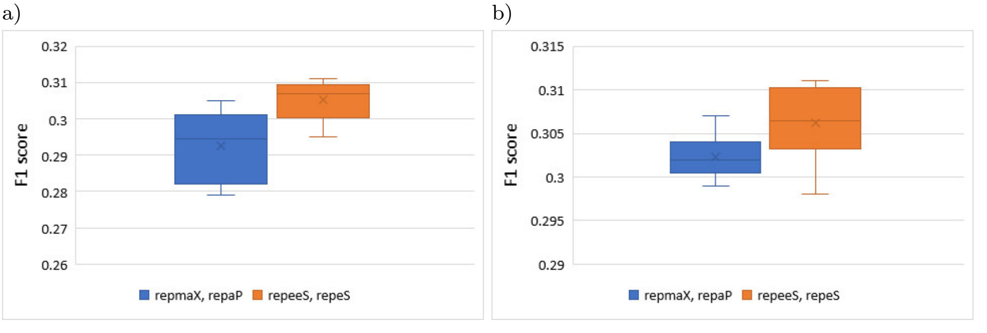

Figure 2. Comparison: path algebra-based trust (repmaX, repaP) and similarity-based trust (repeeS, repeS): a) DAR data set; b) CG data set

Another aspect we needed to explore was the impact of the similarity measures on the reliability between two users. We used the K-fold model to evaluate this, using inputs "repeeS" and "repeS" for similarity-based trust and "repmaX" and "repaP" for path-algebra-based trust. The results are summarized in Tables 4 and 5 and Figures 2a and 2b. It was evident that the inclusion of the similarity measure within the trust model (represented by the orange block) yielded higher F1-score results when compared to calculating the reliability without considering similarity (indicated by the blue block). Consequently, the presence of the similarity factor exerted a substantial influence on the formulation of dependable metrics. This observation held true for both the DAR and CG data sets, thus confirming the significant impact of the similarity factor on the reliability assessment.

Table 4

F1-score measure values in two cases: trust based on path algebra and trust based on similarity CG data set

Model	Resp	Disp	Famil	intMax	intSum	intCor	repmaX	repaP	repeeS	repeS	Recall	Precision	F1
1	x	x	x	x	–	–	x	–	–	–	0.313	0.295	0.303
2	x	x	x	–	x	–	x	–	–	–	0.303	0.296	0.299
3	x	x	x	–	–	x	x	–	–	–	0.305	0.298	0.301
4	x	x	x	x	–	–	–	x	–	–	0.316	0.298	0.307
5	x	x	x	–	x	–	–	x	–	–	0.305	0.298	0.301
6	x	x	x	–	–	x	–	x	–	–	0.306	0.3	0.303
7	x	x	x	x	–	–	–	–	x	–	0.323	0.299	0.31
8	x	x	x	–	x	–	–	–	x	–	0.312	0.298	0.305
9	x	x	x	–	–	x	–	–	x	–	0.32	0.303	0.311
10	x	x	x	x	–	–	–	–	–	x	0.309	0.302	0.305
11	x	x	x	–	x	–	–	–	–	x	0.32	0.298	0.308
12	x	x	x	–	–	x	–	–	–	x	0.305	0.291	0.298

Table 5

F1-score measure values in two cases: trust based on path algebra and trust based on similarity DAR data set

Model	Resp	Disp	Famil	intMax	intSum	intCor	repmaX	repaP	repeeS	repeS	Recall	Precision	F1
1	x	x	x	x	–	–	x	–	–	–	0.184	0.578	0.279
2	x	x	x	–	x	–	x	–	–	–	0.202	0.563	0.296
3	x	x	x	–	–	x	x	–	–	–	0.199	0.558	0.293
4	x	x	x	x	–	–	–	x	–	–	0.187	0.595	0.283
5	x	x	x	–	x	–	–	x	–	–	0.204	0.569	0.3
6	x	x	x	–	–	x	–	x	–	–	0.207	0.583	0.305
7	x	x	x	x	–	–	–	–	x	–	0.204	0.579	0.302
8	x	x	x	–	x	–	–	–	x	–	0.213	0.562	0.309
9	x	x	x	–	–	x	–	–	x	–	0.21	0.563	0.306
10	x	x	x	x	–	–	–	–	–	x	0.199	0.576	0.295
11	x	x	x	–	x	–	–	–	–	x	0.215	0.566	0.311
12	x	x	x	–	–	x	–	–	–	x	0.213	0.564	0.308

Moreover, we assessed the performance of two similarity-based trust models: "repeeS," and "repeS." We applied Hamdi's similarity-based recommendation model to the two DAR data sets (CG) to compare the results. Notably, our proposed model displayed improved performance when compared to Hamdi's model in calculating the indirect trust between two users based on their similarity. This is shown in Figures 3a and 3b. Additionally, we compared our novel path-algebra-based trust model ("repmaX" and "repaP") with Hamdi's model by using the "most trustable path" algorithm. When conducted on the DAR and CG data sets, this comparison consistently favored our model's F1-score performance – even in scenarios that were akin to Hamdi's "perfect" path. In Figures 4a and 4b, it is evident that the outcomes from both sets of favorable data demonstrated that our trust model that was based on path algebra exhibited superior performance as compared to Hamdi's proposed model. This even held true in those scenarios where Hamdi anticipated achieving the "perfect" trust path.

a) b)

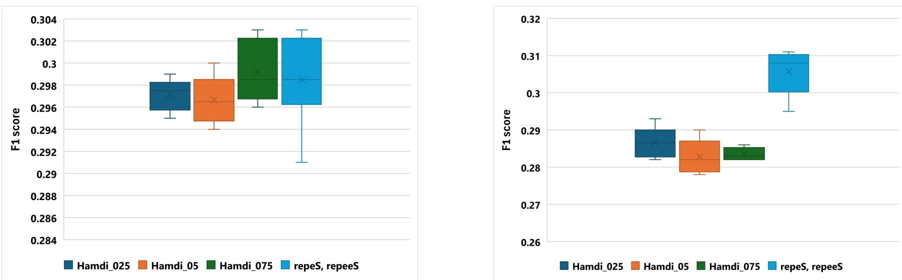

Figure 3. Comparison: proposed model and Hamdi's model in three thresholds: a) CG data set; b) DAR data set

a) b)

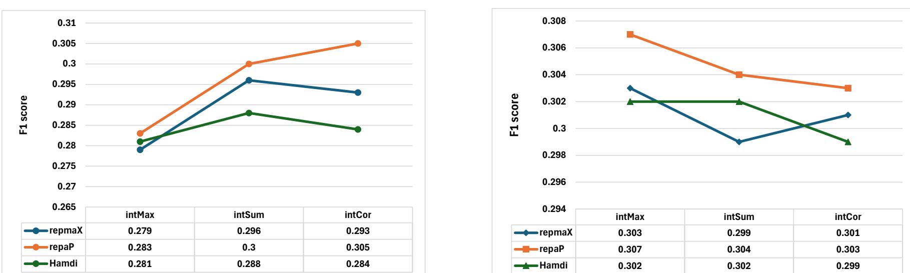

Figure 4. F1-score comparison: repmaX, repaP, and Hamdi's "perfect" path: a) DAR data set; b) CG data set

10. Conclusions

This paper has introduced a family of computational trust models which aggregates trustworthiness degrees of topic-aware experience trust and topic-aware reputation trust from community. The former trust computation is constructed from users' interest degrees and three forms of interaction including dispatching, familiarity and responses. The latter trust estimation is inferred from communities. The various techniques of inference trust from the community have been considered, including the similarity to truster or trustee and operators of the path algebra. And then the overall trust is resulted from the aggregation of experience trust and reputation trust. Our conducted experimental results on the family of proposed models with two distinct data sets indicated that the topic-aware experience trust with the refinement of interaction is better than to use one form of interaction. It is demonstrated that the trustworthiness inferred from community affects the overall trust computation more than the topic-aware experience trust. In addition, the integration of experience and reputation trust is better for estimating the trustworthiness of a partner than using merely one form. However, our work has some limitations. We merely consider the case of computing the trustworthiness of a truster on trustees which have direct interaction. The proposed model may be extended to estimate trustworthiness of a peer truster to any peer trustee in any level of the hierarchy structure or to a peer without any interaction with each other by combining techniques of similarity measure and path algebra. These issues need to be investigated furthermore and these research results will be presented in our future work.

References

[1] Albora G., Rossi-Mori L., Zaccaria A.: Sapling Similarity: a performing and interpretable memory-based tool for recommendation, *Knowledge-Based Systems*, vol. 275, 110659, 2023. doi: 10.1016/j.knosys.2023.110659.

[2] Anderson A., Huttenlocher D., Kleinberg J., Leskovec J.: Effects of user similarity in social media. In: *WSDM '12: Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 703–712, 2012. doi: 10.1145/2124295.2124378.

[3] Bhuiyan T.: A survey on the relationship between trust and interest similarity in online social networks, *Journal of Emerging Technologies in Web Intelligence*, vol. 2(4), pp. 291–299, 2010. doi: 10.4304/jetwi.2.4.291-299.

[4] Bingöl K., Eravcı B., Etemoğlu Ç.Ö., Ferhatosmanoğlu H., Gedik B.: Topic-based influence computation in social networks under resource constraints, *IEEE Transactions on Services Computing*, vol. 12(6), pp. 970–986, 2016. doi: 10.1109/tsc.2016.2619688.

[5] Choi J., Hong S., Park N., Cho S.B.: Blurring-Sharpening Process Models for Collaborative Filtering. In: *SIGIR'23: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1096–1106, Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3539618.3591645.

[6] Crandall D., Cosley D., Huttenlocher D., Kleinberg J., Suri S.: Feedback effects between similarity and social influence in online communities. In: *KDD'08: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 160–168, 2008. doi: 10.1145/1401890.1401914.

[7] De Siqueira Braga D., Niemann M., Hellingrath B., Buarque De Lima Neto F.: Survey on computational trust and reputation models, *ACM Computing Surveys*, vol. 51(5), 101, 2018. doi: 10.1145/3236008.

[8] Fan J., Qiu J., Li Y., Meng Q., Zhang D., Li G., Tan K.L., Du X.: Octopus: An online topic-aware influence analysis system for social networks. In: *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 1569–1572, IEEE, 2018. doi: 10.1109/icde.2018.00178.

[9] Gabrilovich E., Markovitch S.: Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In: *IJCAI'07: Proceedings of the 20th International Joint Conference on Artifical Intelligence*, vol. 7, pp. 1606–1611, 2007.

[10] Ghafari S.M., Beheshti A., Joshi A., Paris C., Mahmood A., Yakhchi S., Orgun M.A.: A survey on trust prediction in online social networks, *IEEE Access*, vol. 8, pp. 144292–144309, 2020. doi: 10.1109/access.2020.3009445.

[11] Golbeck J.: Trust on the world wide web: a survey, *Foundations and Trends® in Web Science*, vol. 1(2), pp. 131–197, 2008. doi: 10.1561/1800000006.

[12] Golbeck J.: Trust and nuanced profile similarity in online social networks, *ACM Transactions on the Web (TWEB)*, vol. 3(4), 12, 2009. doi: 10.1145/1594173.1594174.

[13] Gómez Mármol F., Martínez Pérez G.: Trust and reputation models comparison, *Internet Research*, vol. 21(2), pp. 138–153, 2011. doi: 10.1108/10662241111123739.

[14] Hamdi S.: *Computational models of trust and reputation in online social networks*, Ph.D. thesis, Université Paris-Saclay (ComUE), 2016.

[15] Hang C.W., Wang Y., Singh M.P.: Operators for propagating trust and their evaluation in social networks, Technical report, North Carolina State University, Department of Computer Science, 2008.

[16] Huynh T.D., Jennings N.R., Shadbolt N.R.: An integrated trust and reputation model for open multi-agent systems, *Autonomous Agents and Multi-Agent Systems*, vol. 13, pp. 119–154, 2006. doi: 10.1007/s10458-005-6825-4.

[17] Jamali M., Ester M.: TrustWalker: a random walk model for combining trust-based and item-based recommendation. In: *KDD'09; Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 397–406, 2009. doi: 10.1145/1557019.1557067.

[18] Jiang W., Wang G.: SWTrust: Generating trusted graph for trust evaluation in online social networks. In: *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 320–327, IEEE, 2011. doi: 10.1109/trustcom.2011.251.

[19] Kalaï A., Wafa A., Zayani C.A., Amous I.: LoTrust: A social Trust Level model based on time-aware social interactions and interests similarity. In: *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, pp. 428–436, IEEE, 2016. doi: 10.1109/pst.2016.7906967.

[20] Kang J., Lee H.: Modeling user interest in social media using news media and wikipedia, *Information Systems*, vol. 65, pp. 52–64, 2017. doi: 10.1016/j.is.2016.11.003.

[21] Khan J., Lee S.: Online social networks (OSN) evolution model based on homophily and preferential attachment, *Symmetry*, vol. 10(11), 654, 2018. doi: 10.3390/sym10110654.

[22] Khanam K.Z., Srivastava G., Mago V.: The homophily principle in social network analysis, *arXiv preprint arXiv:200810383*, 2020.

[23] Kim D., Suh B.: Enhancing VAEs for collaborative filtering. In: *RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems*, ACM, 2019. doi: 10.1145/3298689.3347015.

[24] Li K., Zhang L., Huang H.: Social influence analysis: models, methods, and evaluation, *Engineering*, vol. 4(1), pp. 40–46, 2018. doi: 10.1016/j.eng.2018.02.004.

[25] Liu G., Wang Y., Orgun M.A., Lim E.P.: Finding the optimal social trust path for the selection of trustworthy service providers in complex social networks, *IEEE Transactions on Services Computing*, vol. 6(2), pp. 152–167, 2013. doi: 10.1109/tsc.2011.58.

[26] Manchala D.W.: E-commerce trust metrics and models, *IEEE internet computing*, vol. 4(2), pp. 36–44, 2000. doi: 10.1109/4236.832944.

[27] Manning C.D., Raghavan P., Schütze H.: *Introduction to information retrieval*, Cambridge University Press, 2008. doi: 10.1017/cbo9780511809071.

[28] Marsh S.P.: Formalizing trust as a computational concept, 1994. University of Stirling, Ph.D. thesis. https://www.cs.stir.ac.uk/~kjt/techreps/pdf/TR133.pdf.

[29] Nefti S., Meziane F., Kasiran K.: A fuzzy trust model for e-commerce. In: *Seventh IEEE International Conference on E-Commerce Technology (CEC'05)*, pp. 401–404, IEEE, 2005. doi: 10.1109/ICECT.2005.4.

[30] Nguyen M.H., Tran D.Q.: A combination trust model for multi-agent systems, *International Journal of Innovative Computing, Information and Control*, vol. 9(6), pp. 2405–2420, 2013.

[31] Patel J., Teacy W.T.L., Jennings N.R., Luck M.: A probabilistic trust model for handling inaccurate reputation sources. In: *Trust Management: Third International Conference, iTrust 2005, Paris, France, May 23–26, 2005. Proceedings*, pp. 193–209, Springer, 2005. doi: 10.1007/11429760_14.

[32] Pham P.T., Nguyen M.H., Tran D.Q.: Incorporation of Experience and Reference-Based Topic Trust with Interests in Social Network. In: *Advances in Information and Communication Technology: Proceedings of the International Conference, ICTA 2016*, pp. 286–293, Springer, 2017.

[33] Podobnik V., Striga D., Jandras A., Lovrek I.: How to calculate trust between social network users? In: *SoftCOM 2012, 20th International Conference on Software, Telecommunications and Computer Networks*, IEEE, 2012.

[34] Ramchurn S.D., Sierra C., Jennings N.R., Godo L.: A computational trust model for multi-agent interactions based on confidence and reputation. In: *6th International Workshop of Deception, Fraud and Trust in Agent Societies*, pp. 69–75, Melbourne, Australia, 2003.

[35] Richardson M., Agrawal R., Domingos P.: Trust management for the semantic web. In: *The Semantic Web – ISWC 2003: Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20–23, 2003. Proceedings*, pp. 351–368, Springer, 2003. doi: 10.1007/978-3-540-39718-2_23.

[36] Sachdeva N., Dhaliwal M.P., Wu C.J., McAuley J.: Infinite Recommendation Networks: A Data-Centric Approach, 2022. doi: 10.48550/arXiv.2206.02626.

[37] Sherchan W., Nepal S., Paris C.: A survey of trust in social networks, *ACM Computing Surveys (CSUR)*, vol. 45(4), 47, 2013. doi: 10.1145/2501654.2501661.

[38] Tang J., Sun J., Wang C., Yang Z.: Social influence analysis in large-scale networks. In: *KDD'09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 807–816, 2009. doi: 10.1145/1557019.1557108.

[39] Tran D.Q.: Computational topic trust with user's interests based on propagation and similarity measure in social networks, *Southeast Asian Journal of Sciences*, vol. 7(1), pp. 18–27, 2019.

[40] Tran D.Q., Pham P.T.: Path Algebra for Topic Trust Computation based on References of Users in Social Network, *Southeast Asian Journal of Sciences*, vol. 5(1), 2017.

[41] Tran D.Q., Pham P.T.: Modeling computational Trust based on Interaction Experience and Reputation with user interests in Social Network, *Journal of Computer Science and Cybernetics*, vol. 38(2), pp. 147–163, 2022. doi: 10.15625/1813-9663/38/2/16749.

[42] Uddin M.G., Zulkernine M., Ahamed S.I.: CAT: a context-aware trust model for open and dynamic systems. In: *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pp. 2024–2029, 2008. doi: 10.1145/1363686.1364176.

[43] Wang S.: The generalized path algebras over standardly stratified algebras, *Algebra and Discrete Mathematics*, vol. 5(3), pp. 119–126, 2006.

Affiliations

Dinh Que Tran
Department of Information Technology, Posts, and Telecommunications Institute of Technology, Ha Noi, Viet Nam, tdque@yahoo.com

Phuong Thanh Pham
Department of Informatics, Dai Nam University, Ha Noi, Viet Nam, thanhpp@dainam.edu.vn

Received: 4.09.2023
Revised: 16.06.2024
Accepted: 12.09.2024