Saeidi *Comput Soc Netw* (2020) 7:3
https://doi.org/10.1186/s40649-020-00077-6

Computational Social Networks

RESEARCH Open Access

A new model for calculating the maximum trust in Online Social Networks and solving by Artificial Bee Colony algorithm

Shahram Saeidi*

*Correspondence:
sh_saeidi@iaut.ac.ir
Department of Industrial Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Abstract

The social networks are widely used by millions of people worldwide. The trust concept is one of the most important issues in Social Network Analysis (SNA) which highly affects the quantity and quality of the inter-connections, decisions, and interactions among the users in e-commerce or recommendation systems. Many normative algorithms are developed to calculate the trust which most of them are complicated, depend on the network structure, and need lots of critical information that makes them hard to use. The aim of this paper is proposing a descriptive, simple and effective method for calculating the maximal trust and the trust route between any two users of an Online Social Network (OSN). For this purpose, four new models for estimating the trust mechanism of the users are proposed and analyzed using Kolmogorov–Smirnov and Anderson–Darling statistical hypothesis tests to identify and validate the best-fitted model based on 20,613 empirical results gathered from 4552 social network volunteers. Due to the time–complexity of the problem, a meta-heuristic algorithm based on the Artificial Bee Colony (ABC) optimization method is also developed for solving the best-fitted model. The proposed algorithm is simulated in Matlab® over six larger test cases adopted from the Facebook dataset. In order to evaluate the performance of the developed algorithm, the Ant Colony Optimization (ACO) and Genetic Algorithm (GA) based meta-heuristics are also simulated on the same test cases. The comparison of the computational results shows that the ABC approach performs better than the ACO and GA as the size of the network increases.

Keywords: Social Network Analysis, Maximum trust, Trust route, Artificial Bee Colony Optimization, Genetic Algorithm, Ant Colony Optimization

Introduction

Online Social Networks are generally web-based services on a platform in which people can share their ideas, favorites, photos, information and events with each other [1]. Virtual SNs can be divided into public and private groups. In the former type, every kind of users may exist with different aims and motivations and follow the networks using related websites or applications. The number of users in such a group often reaches to a hundred million people. For example, Facebook, Orkut, and Myspace are some of the most popular public social networks. The latter group is formed around a specific

subject and consists of fewer users, the Last.fm on music; GoodReads on book study and Flicker on photography are some famous examples.

Social Networks (SN) are growing rapidly and some important aspects like ideas, behaviors, etc., are diffusing [25]. In these conditions, making the correct decision under a dynamic situation plays an important role for SN users. The Online Social Networks Analysis (OSNA) consists of many different important issues like trust, security, between-ness, determining the leaders, centrality, prestige, finding maximum cliques, determining the malignant nodes and so on. Nowadays, SN users are very interested in sharing data and information. By developing the mobile platforms and their applications, the social networks are also growing rapidly and most parts of the interactions are performed by anonymous users. So, the trust concept plays an important key role in the construction of the relations among the users [26]. Besides, Deng et al. [7] claim that trust is the most important factor in the decision-making process. Hence, calculating the trust value has raised the interest of many researchers and due to the time complexity and NP-Hard nature of the problem, plenty of different heuristic and meta-heuristic approaches are developed.

The users of the SNs have clearly fewer imitations and constraints in choice ability and making decisions since they are not affected by some factors like location, time, culture, government, etc., which do exist in the real world and the users can choose in a worldwide scale. The OSNs provide a platform for users to make better decisions without getting affected by these factors. Therefore, it is necessary to study the variety and capabilities of each type of SNs as social media. Trust means accepting the risk of being misused and transferring part of assets or privacy with the goal of cooperating with another actor. In each trust-based relationship, there are at least two components, the truster, and the trustee. It is assumed that both components are targeted in action and seek to satisfy their needs. The trusting party must decide whether or not to engage with the other (i.e., accept the risk), and the trusted party should also choose between maintaining trust or breaking it; therefore, a trust-based relationship is a bilateral act which is based on the principle of maximizing the benefits under hazardous condition. Under such conditions, the SN users influence each other and change their behavior. Kumar et al. [17] evaluated social influence metrics and calculated the probability of an individual becoming influenced.

In social networks, trust represents the level of confidence about the reliability and correctness of the entity's behaviors [34]. The trust issue is a central concept, people do this for social action to meet their needs through social transfers, and these exchanges have a key role in building social action [4]. In OSNs the users can perform many activities while trust is one of the most important factors needed for making the decisions. The trust is a mechanism for promoting and propagation of the collaboration among the users and also plays a role as a security operator which has been widely implemented on computer networks [7], because keeping or protecting the privacy of the users highly depends on the amount of the trust they evaluate. A small mistake in such an evaluation, i.e., trusting to a hacker or giving access permissions to a spyware program, may lead to big failure in the collaboration process. The main challenge in calculating the trust between two users that do not know each other is how and to what measure the trust value transfers along a social route. This issue requires the evaluation of the trust

between two users along the social trust route based on the transitivity property of the trust, for example, if A trusts B, and B trusts C, then A can trust C [5, 20].

The transitivity property is an important assumption defined or used in many types of research and some proofs are provided for this issue. Christianson and Harbison [3] published the paper "Why isn't trust transitive?" which the title may confuse the readers so that the trust is not really transitive. A complete definition and discussion on transitivity property is provided by Liu et al. [20, 21], which classify the transitivity property and prove that trust transitivity exists among the SN users in different ways. The current research is performed under the transitivity assumption and the empirical results informally confirm this property.

The current study starts at applied orientation and inductive approach. It means that the researchers of this study reach the overall results from minor cases. The research method is descriptive including surveys, questionnaires, and content-analysis. The environment of research includes both library and field study. The strategies used contain surveys, psychology, sociology, and combined concepts. The rest of the paper is organized as follows: the related work is stated in "Literature review" section. The proposed trust calculation models and statistical evaluation for finding the best model are discussed in "The proposed approach" section. The proposed ABC algorithm for solving the best model is explained in "The proposed ABC algorithm" section, and the conclusions are considered in "Conclusions" section.

Literature review

In this section, a comprehensive review of the literature is studied and discussed. Next, the research aim and question are stated based on the common weakness of the previous researches. The research methodology is explained in the last subsection.

Related work

Several studies on SNs are performed by researchers and some methods are proposed for calculating the trust in the literature. Based on the application type and available information, these methods can be divided into three groups: the graph-based, mutual trust, and hybrid models. The more recent researches are studied and reported in this section.

Guha et al. [13] incorporated the distrust concept in calculating the trust propagation and showed that a small number of expressed trust/distrust by the network users increases the accuracy of calculating the trust between any two users. Dwyer et al. [9] compared Facebook and Myspace social networks considering attitude and behavior point of view. Based on the research results, they concluded that Facebook users have more trust in this site and its members. Yet, Myspace users are more active in developing relations and making new friends. Taherian et al. [29] proposed a new trust inference algorithm named RN-Trust using a resistive network concept. They evaluated and analyzed this algorithm and reported that it calculates the trust more accurately than previous approaches. Evans and Wensley [10] focused on the viewpoint that the trust is a necessity and essential pre-condition for sharing knowledge. The main aim of their study was discussing the stochastic relationship between social network principles, network structure, and trust. They analyzed the trust concept and

investigated the power of fiducial relations in societies for making the trust operative. Zhan et al. [35] proposed a trust maximization algorithm based on task-oriented social networks and evaluated the performance of the algorithm by some extensive experiments.

Trust is an essential condition for collaboration in peer-to-peer (p2p) systems. Liu et al. [18] proposed a model for promoting trust in social networks for p2p systems. They first proposed a three-layer trust propagation framework consisting of the promotion layer, awareness layer, and calculation layer. Next, they developed an algorithm for implementing and forming trust networks. Their simulation results show that the trust promotion model can effectively increase the safety and stability of p2p systems and improve resource availability. Podobnik et al. [23] proposed a model to convert a user's individual social graph structure into a more general weighted graph. They verified their model using a Facebook application named "Closest Friends" and evaluated the proposed approach with 150 Facebook users. Fong et al. [11] used data mining techniques to determine the relative importance factors affecting trust in SNs. They used Feature Selection algorithms prior to constructing a decision tree to classify the predicted class. Dehghan et al. [6] investigated the presented methods and algorithms for inferring trust in OSNs and reported their advantages and disadvantages. They concluded that due to the fact that web-based SNs indeed perform on the trust concept, more accurate and fast algorithms should be developed to help users gain more valuable information. Daneshmand and Daneshmand [5] presented a definition of trust based on sociological grounds. They also described several aspects of trust like transitivity and composability.

A comprehensive review of trust is done by Sherchan et al. [26] in SNs. They studied the trust from the social and computer science point of view and defined the concept of social trust in the context of social networks for the first time. Three different aspects of the trust including trust information collection, trust evaluation, and trust dissemination are covered in this survey. Liu et al. [18] also proposed a new network structure including trust, social relationships, and recommendation roles and introduced a new concept Quality of Trust (QoT). Next, they modeled the optimal social trust path selection problem with multiple end-to-end Quality of Service (QoS) constraints as a multi-constrained optimal path selection problem, which is shown to be an NP-Complete problem. The preprocessing of a SN using a trustable familiarity chain detection based on user domain using online SNs' microworld network properties and benefits of weak connections is discussed by Jiang et al. [16]. The authors proposed a method of generating trustable graphs and inspecting real information adopted from OSNs in order to select strong neighbors using a breadth-first search algorithm and concluded the effectiveness of this approach. Situm [28] on her master's thesis focused on trust among the users on p2p social networks. She proposed some algorithms to calculate the trust on Facebook and evaluated the presented algorithms.

Nuñez-Gonzalez et al. [22] used a machine learning method for predicting trust. Their proposed method used training techniques to gain the trust value based on reputation features obtained from volunteer users called witness trustors. The machine learning method for predicting the trust works in a fixed size space, so the variable

size of information should be reduced to a constant size volume. Sanadhya and Singh [24] used the Ant Colony Optimization method for calculating the trust based on the structural and behavioral properties of the OSNs. They used the Facebook dataset for simulating their proposed algorithm.

Wang et al. [34] developed a new trust calculation method based on game theory concepts. They divided the nodes of the network into four categories as service nodes, feedback nodes, recommendation nodes, and managed nodes to describe the trust degree more accurately. The authors used service reliability (the trustworthiness of service that service nodes provide), feedback effectiveness (the trustworthiness of feedback that feedback nodes return), and recommendation credibility (the trustworthiness of recommendation that recommendation nodes give) concepts to estimate the trust quantity and showed the effectiveness of their proposed method. Frikha et al. [12] considered the time factor for estimating trust between the social network users for recommendation systems. They developed an application for Facebook users to demonstrate the importance of time affecting the users' interaction for determining social friends. Hamdi [14] on her Ph.D. thesis proposed a trust management model named IRIS considering social activities of users including their social relationships, preferences, and interactions. Singh and Chin [27] proposed a conceptual framework for calculating the trust among OSN users and claimed it may be not possible to use mathematical models in offline networks where users do not have the previous relationship. Takalkar and Mahalle [30] in their research, reviewed different metrics and methods for calculating the trust value on SNs and proposed a trust-based approach for discussing the confidentiality in OSNs.

Wang et al. [33] proposed a new trust evaluation scheme based on evidence theory. The authors considered the risk of privacy leakage by information flow prediction to make the trust evaluation more comprehensive and compared their method with some previous algorithms by considering accuracy, mean error, and *F*-score and concluded the superiority of the proposed method.

The brief review of related work indicates that all previously developed techniques for calculating the trust in OSNs are mostly theoretical, impractical, and actually impossible or hard to implement. For example, Takalkar and Mahalle [30] define a dynamicity index as the ratio of the number of times the user has logged-in, to the amount of his/her activities. Obviously, a specific user has no idea about the number of times that other users have logged-in the network, hence the dynamicity index cannot be calculated.

The research question

The users of the SNs in the real world, construct their interactions based on a mental trust amount to other users which can be expressed as a real value in [0…1]. As discussed in "Related work" section, all the previous researches can be considered as normative approaches that try to construct methods or formulas to indicate how this trust should be calculated, rather than explaining how people trust each other in fact. In other words, the social media users, their opinions, their mental and experimental methods for evaluating the trust are ignored. The main weakness of the previous researches, is trying to answer the question "How SN users should trust each other?" This research aims to answer the question "How the SN users trust each other in the real world?" or "What is the trust mechanism (model) of the real SN users?"

In this paper, based on the information gathered by questionnaires from the OSN users, the more important metrics and the trust mechanism used by individuals are inspired and extracted and the best-fitted model is selected among the four proposed models. The statistical analysis shows that the first proposed model better fits the behavior of the SN users of the statistical society. Therefore, this model is announced as an answer to the research question.

The proposed approach

In this section, first, the structure of the social network, the questionnaire structure, the proposed trust calculation models, and the statistical analysis are defined and discussed.

The trust calculation models

The social network considered in this paper is similar to Facebook consisting of millions of connected people. This structure can be defined as graph $G=(V, A)$, where the V is the set of vertices (users) where $\|G\| = N$ and the A is the set of arcs or the inter-connections among the users. This graph is logically undirected because the friendship concept in such a network is not directional. But the trust concept and its value between any two users are directional and the corresponding structure should be defined as a directed graph.

As mentioned in "Literature review" section, the trust concept has a transitivity property among the users of the network [7]. It means that if Alice trusts Bob, and Bob trusts Charlie, so Alice can indirectly trust Charlie via Bob. It is obvious that trust value will decrease as many as the number of intermediate nodes increases. In this paper, the trust value of the source user v_s to the destination user v_d is calculated by multiplying the trust values along the route using the iterative multiplication strategy. If there are multiple feasible routes between these nodes, the route with maximal trust value is desired. This concept is shown in Fig. 1.

In this paper, the trust values between any two friends are treated as probability values. Considering Fig. 1, the trust of the user s to the user m is given as 0.8 which somehow means that the user s is happy of his/her trust to user m, in 80% of the cases, or the user m disappoints the user s in 20% of the interactions. Equation (1) represents this issue:

$$Trust\ (s \rightarrow m)\ =\ \text{Prob.}\ (m\ \text{satisfies}\ s)\ = 1 -\ \text{Prob.}\ (m\ \text{disappoints}\ s). \tag{1}$$

In other words, in case of making a decision about a new interaction between these two users, the user s imagines that the results will be satisfying with a probability of 80%.

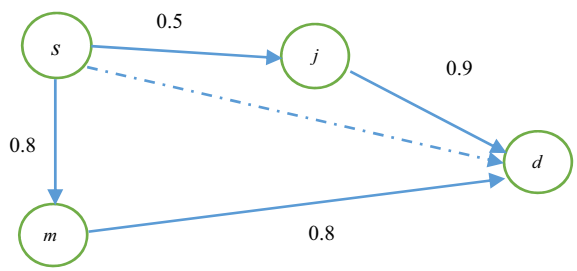

Fig. 1 Obtaining the maximal trust

The same is true from user m to user d. So, the indirect trust from the user s to the user d, is equal to the satisfaction probability of two independent events which is calculated as the multiplication of the consequent probability values.

In Fig. 1, the aim is to calculate the trust value from user s to user d (dashed arrow). These users are not directly connected, but have two friends in common: m and j. So, there are two possible trust routes. Using the first route ($s \rightarrow j \rightarrow d$) the trust is calculated as $0.5 \times 0.9 = 0.45$ and on the second route ($s \rightarrow m \rightarrow d$) is $0.8 \times 0.8 = 0.64$. So the second route is reported as the trust route with $Trust(s \rightarrow d) = 0.64$. The trust between two arbitrary nodes via a path is the product of the trusts along the edges, and trust between any two parties is the maximal value of trust along any path. So, the general form of this issue can be written as Eq. (2). The transitivity property of the trust helps us to aggregate the trust value along with a path from source to destination nodes. Hence trust is discounted with the increase of transitivity hops [3]. This strategy has been widely used in the literature as a feasible trust aggregation method [19, 32].

$$Trust(s \rightarrow d) = \text{Max} \prod_{i=s}^{j=d} LinkValue(v_i, v_j) \quad \forall (v_i, v_j) \in A \tag{2}$$

where $LinkValue(V_i,V_j)$ is a positive real value in [0...1] demonstrating the direct trust of node V_i to node V_j. This value is known for the users of the SNs, that is, every user believes in a specific amount of trust to each of his/her connected friends, although he/she does not know how this value is formed. As mentioned in "The research question" section, the main objective of this paper is providing a model to represent the evaluation mechanism of this value. The parameters affecting the *LinkValue* are not stable and most of them change in the time, so it can be concluded that the trust of V_i to V_j will change (increase or decrease) depending on the positive or negative feedback.

Figure 2 demonstrates a small sample graph with 10 nodes and the trust value of user v_6 to user v_2 is requested. The route highlighted in red shows the best-obtained path with maximal trust value. It should be noticed that the reverse trust route from user v_2 to user v_6 may be different because of the asymmetric property of the trust.

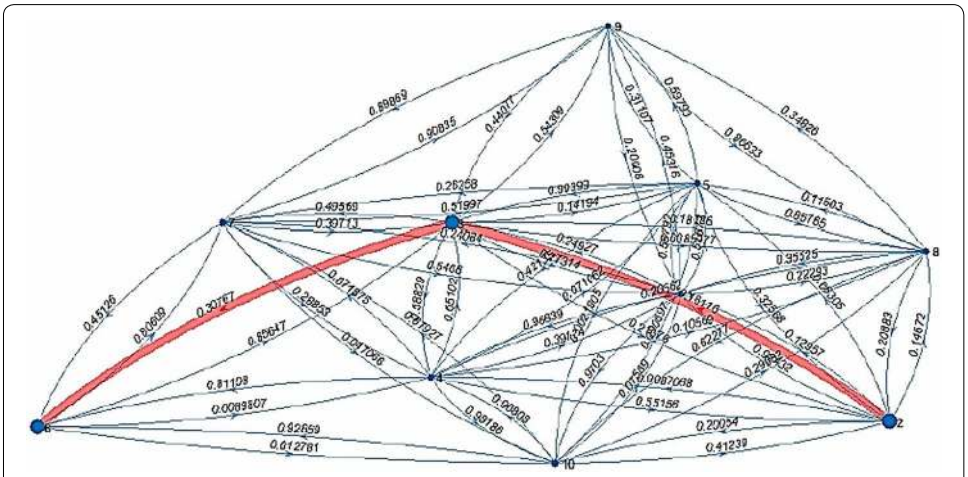

Fig. 2 A sample trust route on a small graph with 10 nodes

It is obvious that the problem will be complicated in large networks as there are numerous nodes, arcs, and feasible routes; and the problem changes to a difficult combinatorial optimization problem which needs a long computational time to be solved [12]. So, meta-heuristic approaches are developed by the researchers to tackle this obstacle.

The main objective of this paper is providing a model for calculating the maximal trust value and the trust route (containing some intermediate users) from user v_i to user v_j (v_i, $v_j \in V$) with a maximal calculated trust value. In other words, the user v_i needs (or wants) to know what is the maximal trust he/she can rely on user v_j and via which users this value can be obtained. This is also known as the maximal trust problem which is shown to be an NP-Hard problem [18].

The trust calculation model is constructed in two stages. At the first stage, a dynamic individual node value for every user on the network is calculated based on the user's characteristics and personal information and at the second stage, the inert-personal trust value is estimated considering the node values obtained at the first stage using the proposed model. This information makes up the statistical society of the research and is used to estimate the trust mechanism of the OSN users and compare the proposed models with empirical data to figure out the best fitting model.

In order to gather needed information from the statistical society, a two-part questionnaire is designed using Google Docs and the URL of the questionnaire was distributed online by WhatsApp, Instagram, and Telegram social media users. The users were requested to participate in the research and resend the URL address to their own friends. In the first part of the questionnaire, as shown in Table 1, the volunteers were requested to score ten listed F_i factors (F_1, F_2, …, F_{10}) based on a 5-point Likert scale to indicate their opinion about the most important factors affecting the individual's trustiness.

In the second part of the questionnaire, the responders were asked to state at most five numbers of their social friends, give values to related F_i factors, and declare how much they trust those friends. Table 2 shows a sample response from a volunteer scoring her

Table 1 The part #1 questionnaire structure with a sample response

Part 1	Rate the factors by choosing one of the items					Your own status
Score	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	
Factor						
F_1: age				√		27
F_2: education					√	B.Sc.
F_3: gender		√				F
F_4: job title			√			N/A
F_5: number of received likes					√	1772
F_6: amount of activity (posts)				√		128
F_7: number of friends			√			304
F_8: marital status	√					Married
F_9: received bad reports				√		3
F_{10}: profile having own photo					√	Yes

Table 2 The part #2 questionnaire structure with a sample response

Part 2	**Please fill the following table for a maximum of 5 different friends of yours which you have enough information about. In the last column, identify the amount of your mental trust to him/her with a number in [0…1] range**										
Factor	F_1	F_2[a]	F_3	F_4[b]	F_5	F_6	F_7	F_8[c]	F_9	F_{10}	**Your trust to him/her**
Nickname											
Nilay	45	1	F	1	6420	410	1277	1	0	Yes	0.85
Mary	22	4	F	4	211	96	137	2	3	Yes	0.25
John	29	2	M	2	5547	702	963	1	0	Yes	0.70
Michael	31	4	M	5	3512	588	540	2	1	No	0.45
Dominique	37	3	F	2	4593	432	725	1	0	Yes	0.60

[a] 1: Ph.D. or equivalent; 2: M.Sc.; 3: B.Sc.; 4: diploma; 5: others

[b] See the attached SOC classification

[c] 1: married; 2: single; 3: separated; 4: divorced

five friends. After the assessment and refinement of the responses, $N=4552$ completely answered sheets remained out of 6167 total responded questionnaires. The incomplete or inconsistent sheets were ignored. The consistency of the responses was checked by analyzing the relationships among the answers, for example, a responder of age 16, cannot hold a Ph.D. degree, or get married. Out of 4552 response sheets, 2797 volunteers addressed five friends, 1369 volunteers addressed four friends, and the rest addressed only three friends. Thus, a total number of $N'=2794\times5+1369\times4+\times389\times3=20$, 613 users, their related F_i factor values, and trustiness value were gathered. Considering the Cochran's formula, a sample of size 385 would be enough for a society of size 1000,000,000 or more ($\alpha=0.05$ error), so the statistical society is large enough to rely on the results.

By assessing the first part of the questionnaire, it is concluded that the responders imply that the personal node value (impact factor) of a network user is highly affected by the user's activity, the number of his/her friends, job title (occupation position), level of education and the reports (negative comments, or dislikes) against the user. So, the *node-value* calculation formula can be written as Eq. (3):

$$\begin{aligned} NodeValue(v_i) = {} & w_1 Acitivity(v_i) + w_2 Friends(v_i) + w_3 Likes(v_i) \\ & + w_4 JobTitle(v_i) + w_5 Education(v_i) - w_6 Reports(v_i) \quad \forall v_i \in V, \end{aligned} \tag{3}$$

where w_i coefficients are related importance weights of the parameters.

This equation implies that the users with more activities (new posts or sharing others'), a higher number of friends and likes, better job positions, higher education, and less obtained negative reports will have higher prestige, hence they are potentially more trustable users. The job title is a term that returns some information about the position and responsibilities of the people. In this paper, these titles are adopted from the United States' Bureau of Labor Statistics (http://www.bls.gov/soc) and are scored based on questionnaire filled by volunteers. In order to evaluate the weight coefficients, the Part 1 section of the questionnaire is analyzed. This section can be considered as a $[F_{ij}]10\times5$

matrix, where F_{ij} is the sum of tick marks of factor i on scale j voted by a total of 455 responders. The corresponding numeric value of the scale attributes Sj are assumed as $\{-2,-1, 0, 1, 2\}$ for {Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree}. Next, the weighted sum of each row of the matrix is calculated as:

$$\text{RowSum}(i) = \sum_{j=1}^{5} F_{ij} S_j \quad \forall i = 1, 2, \ldots 10. \tag{4}$$

The factors having small RowSum values are ignored and the remaining factors (F_2, F_4, F_5, F_6, F_7, and F_9) are normalized and evaluated as $w_1 = 0.098$, $w_2 = 0.101$, $w_3 = 0.197$, $w_4 = 0.204$, $w_5 = 0.298$, and $w_6 = 0.102$, respectively.

After calculating the personal node values of the users, these values are normalized. Next, the direct trust value between two connected nodes (friends) like v_i and v_j should be calculated. For this purpose, four calculation models (estimators) are proposed and defined as Eqs. (5), (6), (7) and (8). The main idea of proposing these models is based on three simple concepts: (1) the amount I trust you, completely depends on who you are (your *NodeValue*), and who I am (my *NodeValue*). (2) The trust between the users is asymmetric, that is, I don't have to trust you as much as you trust me. (3) The trust of a person with higher *NodeValue* (social prestige) to a person with lower *NodeValue*, would be less than the opposite direction. Obviously many mathematical equations can be proposed to preserve these properties, some simple models are considered in this paper as follows:

Proposed Model #1

$$Trust\left(v_i \rightarrow v_j\right) = \frac{NodeValue(v_j)}{NodeValue(v_i) + NodeValue(v_j)} \quad \forall v_i, v_j \in V. \tag{5}$$

Proposed Model #2

$$Trust\left(v_i \rightarrow v_j\right) = 1 - \frac{NodeValue(v_i) \times NodeValue(v_j)}{NodeValue(v_i) + NodeValue(v_j)} \quad \forall v_i, v_j \in V. \tag{6}$$

Proposed Model #3

$$Trust\left(v_i \rightarrow v_j\right) = 1 - \sqrt{NodeValue(v_i) \times NodeValue(v_j)} \quad \forall v_i, v_j \in V. \tag{7}$$

Proposed Model #4

$$Trust\left(v_i \rightarrow v_j\right) = 1 - NodeValue(v_i) \times NodeValue(v_j) \quad \forall v_i, v_j \in V. \tag{8}$$

Evaluation and analysis

In order to evaluate the goodness-of-fit of the proposed models, the second section of the questionnaire is used which the volunteers were asked to consider some real social friends and express how much they trust to each of them. The obtained empirical results for $N' = 20613$ samples of the questionnaire along with the calculated trust values using Eqs. (5), (6), (7) and (8) are depicted in Fig. 3. The details behind the calculation method of these values are quite simple. For example, the user v_j (the volunteer filling

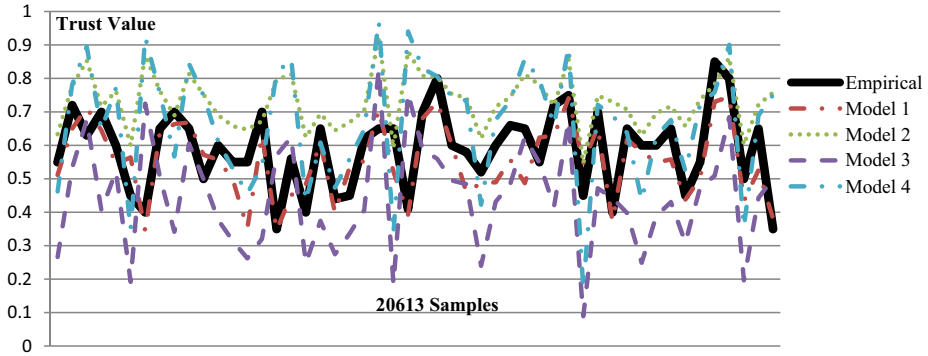

Fig. 3 The empirical results in comparison with the proposed trust calculation models

the questionnaire) has also mentioned his/her trust to all the v_j friends on the last column of Table 2, e.g., trust of user v_i to the listed friends *Nilay*, *Mary*, *John*, *Michael*, *and Dominque*, are declared as 0.85, 0.25, 0.7, 0.045, and 0.60 consequently. These are the $Trust(v_i \rightarrow v_j)$ values which are referred to as *Empirical* trust values in Fig. 3.

Next, considering the F_1 through F_{10} values given in Tables 1 and 2, the *NodeValue* parameter is calculated by Eq. (1) for v_i using data in Table 1; and also for his/her five connected v_j friends using data in Table 2. Finally, the $Trust(v_i \rightarrow v_j)$ estimations are performed using the proposed models which are depicted along with the empirical trust values in Fig. 3.

As Fig. 3 shows, the calculated (estimated) trust values using Model #1 have a better fitness (similarity) to the empirical trust values gathered from the questionnaire respondents. For making a precise decision, some important statistics are calculated which are given in Table 3.

$$\mathrm{EE} = EmpiricalTrustvalue - EstimatedTrustValue \tag{9}$$

$$\mathrm{MAD} = \frac{\sum |\mathrm{EE}|}{N'} \tag{10}$$

$$\mathrm{MSE} = \frac{\sum (\mathrm{EE})^2}{N'} \tag{11}$$

Table 3 The calculated statistics for proposed model

Statistics	Model #1	Model #2	Model #3	Model #4
$\overline{\mathrm{EE}}$	0.000878624	− 0.00100121	0.000971717	− 0.001480456
MAD	0.742702413	0.347216907	0.498407248	0.351791115
MSE	0.005860388	0.035265108	0.051411847	0.040121825
TS	2.401509025	− 5.853566014	3.957777025	− 8.542924408
UCL	2.792561073	1.305535571	1.874011254	1.322734594
LCL	− 2.792561073	− 1.305535571	− 1.874011254	− 1.322734594

EE, MAD, MSE and TS represent *Estimation Error*, *Mean Absolute Deviation*, *Mean Square Error* and *Tracking Signal*, respectively, and are calculated using Eqs. (9) through (12)

$$\mathrm{TS} = \frac{\sum \mathrm{EE}}{\mathrm{MAD}} \tag{12}$$

The TS statistics is an indicator for monitoring the forecast validity. It is most often used when the validity of the forecasting model might be in doubt. As long as the tracking signal is within the limits, the estimation process is in control. Limits are usually between 2 to 5 standard deviations. Because 1 standard deviation is approximately equivalent to 1.25 MAD, a common boundary of 3 standard deviations (or ± 3.75 MAD) is used for Upper Control Limit (UCL) and Lower Control Limit (LCL) [31].

Considering the calculated statistics reported in Table 1, Model #1 consists of the least estimation deviation in comparison with the other three models. The value of *TS* statistics obtained for the proposed Model #1 is between the control limits of ± 3.75 MAD ($-2.7925 < 2.4015 < 2.7925$), whereas the Models #2 and #4, both have a $\mathrm{TS} < -3.75 * \mathrm{MAD}$ meaning a persistent under-forecasting and the Model #3 with $\mathrm{TS} > 3.75 * \mathrm{MAD}$ suffers over-forecasting. The trust values gathered by the empirical method along with the values obtained by the proposed models are divided into 10 intervals of length 0.1 which are given in Table 4. The related cumulative probability functions are depicted in Fig. 4.

Table 4 The cumulative probability function of the empirical and proposed models

Interval	Model				
	Empirical	Model #1	Model #2	Model #3	Model #4
0–0.1	0	0	0	0	0
0.1–0.2	0	0.03	0	0.02	0
0.2–0.3	0	0.05	0.08	0.04	0
0.3–0.4	0.07	0.09	0.14	0.08	0.07
0.4–0.5	0.21	0.16	0.25	0.24	0.21
0.5–0.6	0.39	0.33	0.5	0.37	0.39
0.6–0.7	0.79	0.61	0.87	0.52	0.79
0.7–0.8	0.91	0.87	0.96	0.81	0.91
0.8–0.9	1	0.96	1	0.95	1
0.9–10	1	1	1	1	1

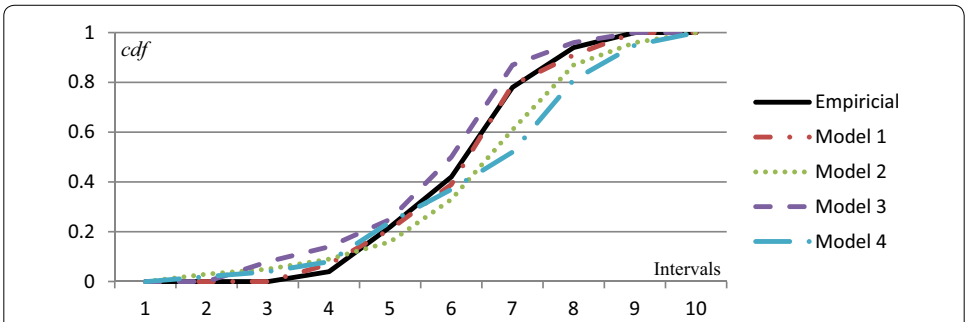

Fig. 4 The cumulative probability functions

Table 5 The Kolmogorov–Smirnov test results

Test	Model #1	Model #2	Model #3	Model #4
KS statistic	0.03	0.17	0.13	0.24
Test result	Cannot be rejected	Rejected	Rejected	Rejected

Figure 4 and the statistical calculations confirm the Model #1 is a good candidate for estimation the trust in real SNs. In order to validate and justify this hypothesis, at the next step, the "goodness of fit" test is performed over the proposed Model #i ($i=1$, 2, 3, 4) considering the following hypothesis using both Kolmogorov–Smirnov (KS) and Anderson–Darling methods.

$\mathbf{H_0}$ The obtained trust values by the proposed Model #i follow the empirical distribution.

$\mathbf{H_1}$ The obtained trust values by the proposed Model #i do not follow the empirical distribution.

Using SPSS software and assuming the significance level $\alpha=0.05$, the calculated p-value using the Anderson–Darling test is calculated as 0.32 for Model #1. So the hypothesis H_0 cannot be rejected because p-value $> \alpha$. The calculated KS statistics for the proposed models and the hypothesis test results are given in Table 5. At the significance level $\alpha=0.05$, the acceptance critical value is 0.044 which means that the null hypothesis is rejected for Models #2, #3 and #4, but cannot be rejected for Model #1.

Hence, the validity of the proposed Model #1 is clearly concluded and this model can be used for calculating the trust between any two connected users. For example, considering Alice and Bob as two friends with normalized node values equal to 0.135 and 0.284, respectively, the trust between these users would be calculated as:

$$Trust(\text{Alice} \rightarrow \text{Bob}) = \frac{0.284}{0.135 + 0.284} = 0.678, \tag{13}$$

$$Trust(\text{Bob} \rightarrow \text{Alice}) = \frac{0.135}{0.284 + 0.135} = 0.322. \tag{14}$$

Considering Eq. (4) and according to calculated trust values in the above example, it can be concluded that the trust concept between two friends is not bidirectional and it completely depends on the node value of both users. So, by identifying the best-fitted trust model, the proposed maximal trust problem can be formulated as the following mathematical programming model:

$$\text{Max} \prod_{i=s}^{j=d} \frac{NodeValue(v_j)}{NodeValue(v_i) + NodeValue(v_j)} \tag{15}$$

s.t.

$$NodeValue(v_i) = w_1 Acitivity(v_i) + w_2 Friends(v_i) + w_3 Likes(v_i) + w_4 JobTitle(v_i) + w_5 Education(v_i) - w_6 Reports(v_i) \quad \forall v_i \in V, \tag{16}$$

$$\sum_{i=1}^{6} W_i = 1, \tag{17}$$

$$W_i > 0. \tag{18}$$

Considering the objective function, the model is known as an unconstrained single-ration hyperbolic programming problem [15] which is shown to be NP-Hard [2]. Hence, a meta-heuristic approach based on ABC algorithm is proposed for solving the model.

The proposed ABC algorithm

The Artificial Bee Colony algorithm which was first introduced in 2005 is a population-based intelligent search method adopted from real honey bee colonies searching for food. The colonies spread around the hives and try to collect nectar. The initial search is performed by employed bees randomly in order to explore the larger neighbor source. Next, they return back to the home and inform the onlooker bees by dancing over the hive. The onlooker bees watch the dance of the employed bees and choose the food source. The employed bee whose food source has been abandoned becomes a scout and starts to search for finding a new food source. The algorithm continues until stopping conditions are met.

A social network graph plays the role of the flower garden (food source). Each flower (SN user) contains a specific amount of nectar (the user's social prestige) which is extracted by honey bees. The nectar extraction means the calculation of the user's *NodeValue* using Eq. (2). The bees begin their tour from the nest (source node) to the last flower field (destination node) in such a way to maximize total collected nectar, or in other words, to maximize the trust value over a trust route on the SN. Hence, each of the bees travels (generates) a route from the source (s) to the destination (d). The honey bees inform the others about the amount of food they have discovered on their path by dancing around the nest, which corresponds to calculating the $Trust(s \rightarrow d)$ using Eq. (15) and selecting the best route among all the traveled routes so far. The summary of adaptation of the maximal trust problem and the ABC algorithm concepts is given in Table 6.

The proposed ABC algorithm is programmed in Matlab® using a personal computer with a 4.2 GHz processor and 2 GB of RAM. The specifications of the simulated test

Table 6 ABC and related trust concepts

ABC concept	Adapted maximal trust problem term
Food source	Social Network
Flower	Node (user of the SN)
Nectar extraction	NodeValue calculation
Nectar quality	NodeValue
Honey bee	The traveled route from source to destination
Dancing	Choosing the route with a higher trust value

Table 7 Specification of simulated test cases

Case #	Graph size		Source and destination		Simulation parameters	
	N	*V*	*s*	*d*	# of iterations	# of bees
1	100	262	1	100	20	10
2	250	774	1	250	20	15
3	500	5640	1	500	50	40
4	750	13711	1	750	70	50
5	1000	34548	1	1000	80	60
6	2000	69436	1	2000	120	80

Artificial Bee Colony

Input: Network Graph G(V, A), source node (s), destination node (d).

Output: Maximal trust from s to d.

- *Begin*
- *Calculate the NodeValues of the graph using equation 3.*
- *Generate initial random population.*
- *Generate initial random routes.*
- *Evaluate the trust value of the routes using equations 15.*
- *Repeat (Main loop)*
  - *a) Repeat for every employed bee*
    - *Randomly select another node on the current route and make a new route.*
    - *Calculate the fitness of new route.*
    - *If the fitness of the new route increased, accept and keep it.*
  - *b) Repeat for every onlooker bee*
    - *Choose the better food source using roulette wheel selection method.*
    - *Make a new route by random changes on the current route and calculate the fitness.*
    - *If the new fitness if better than the global best fitness, replace it as the global best, increase bound parameter.*
  - *c) Repeat for every scout bee*
    - *Inspect the bound parameter of the routes.*
    - *For routes with improvements less than the bound parameter*
      - *Generate new random routes.*
      - *Evaluate the fitness of new route using equation 15.*
      - *Replace the new route if the fitness is better.*
- *Until no improvement in the last ten iterations.*
- *Report the route with maximal trust value.*
- *End.*

Fig. 5 The pseudo-code of the proposed ABC algorithm

cases adopted from the Facebook sample dataset are given in Table 7. The pseudo-code of the proposed ABC algorithm is given in Fig. 5.

The complexity of the proposed ABC algorithm is $O(n^2)$, because the main loop of the algorithm repeats (l times) until no more improvement happens in objective function value during the last 10 iterations. Inside the main loop, three smaller loops are executed sequentially (a, b, and c sections of the pseudo-code) each of which n_1, n_2, and n_3 times consequently. So, the maximum iterations of the algorithm will be $l \times r$ times where $r = \max\{n_1, n_2, n_3\}$. The value of the l and r increases for the larger network sizes, but the complexity remains $O(n^2)$.

Table 8 The simulation parameters of GA and ACO

Case #	GA				ACO		
	Population size	Crossover rate (%)	Mutation rate (%)	No. of iterations	No. of ants	Pheromone evaporation rate (%)	No. of iterations
1	15	80	15	20	10	1	10
2	18	80	15	25	15	1	15
3	40	80	15	50	40	1	40
4	50	80	15	75	50	1	50
5	70	85	18	80	60	2	70
6	80	87	20	150	80	2	140

Table 9 The simulation results

Case #	Compared meta-heuristic approaches					
	ABC		GA		ACO	
	Trust	CT[a]	Trust	CT[a]	Trust	CT[a]
1	0.308	0.293	0.306	0.251	0.306	0.225
2	0.245	1.034	0.243	1.497	0.243	1.176
3	0.274	2.417	0.214	5.607	0.274	6.054
4	0.132	3.95	0.132	8.218	0.132	9.37
5	0.106	19.166	0.099	45.014	0.081	83.51
6	0.232	35.38	0.217	79.293	0.202	142.9

[a] Computation time (s)

The idea of the Genetic Algorithm (GA) was first introduced by John Holland in 1960 and next extended by his student David Goldberg in 1989. The GA method is a population-based approach and is made up of some chromosomes each of which represents a solution to the problem being solved. The Ant Colony Optimization first developed by Dorigo et al. [8] is inspired by real ants and their behavior. Real ants which live in colonies, leave the nest to find food and come back again at every time. Based on observations, these ants always choose the shortest path to reach the food.

The simulation parameters of the GA and ACO methods are given in Table 8. These parameters are tuned-up using the trial-error method and the best values are obtained. Each of the sample test cases is simulated for 10 times and the best-obtained solution and average computation time (s) in comparison with GA and ACO results are reported in Table 9. The simulation results show that the computation time of the proposed ABC is considerably less than that of GA and ACO algorithms, and as the size of the graph increases, this difference highly increases. The obtained trust values are better or equal to the results calculated by GA and ACO approaches.

The comparison of the computation time of simulated algorithms for test cases 1 through 4, and test cases 5 and 6 are depicted in Figs. 6 and 7, respectively.

As Figs. 6 and 7 show, the computation time for solving the problem increases exponentially for GA and ACO approaches when the size of the problem increases, but the increasing slope of the proposed ABC remains almost linear. The performance of the ACO is the worst among the others. This is caused by the behavior of artificial

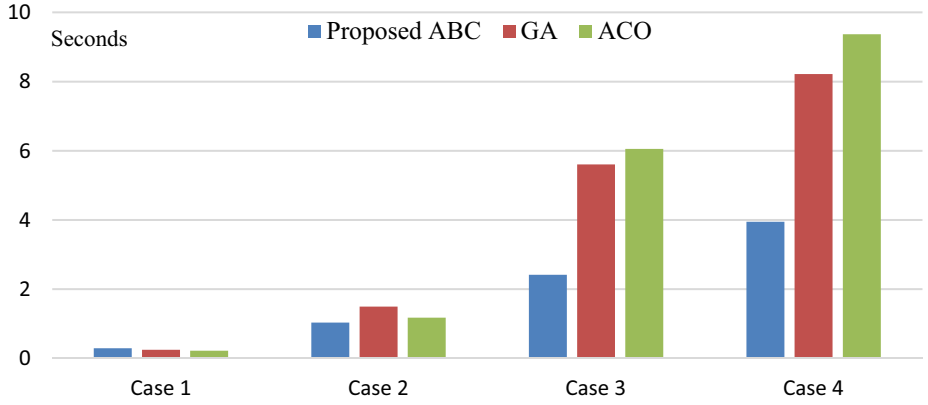

Fig. 6 The comparison of the computation time of the small test cases

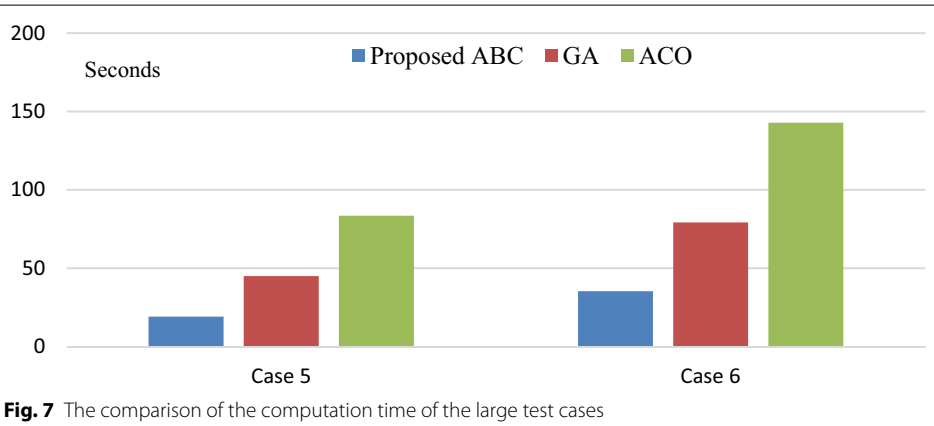

Fig. 7 The comparison of the computation time of the large test cases

ants. Every artificial ant during its local search for finding the better route, just looks at the amount of the pheromone on outgoing arcs of the current node and the path (next arc) selection is made based on local information and the search agents (ants) have no further (global) information about the whole network which makes it difficult for them to find a better solution. In Genetic Algorithm, every chromosome demonstrates a route from the source node (s) to the destination node (d), so the length of the chromosome may reach to total number of the nodes of the network (N) which has a negative impact on the performance of the mutation and crossover operations and affects the computation time. The convergence diagrams of the proposed ABC algorithm along with GA and ACO approaches for test case 5 are depicted in Fig. 8. In this figure, the vertical axis shows the value of the calculated trust from the source to destination nodes and the horizontal axis shows the number of the iterations.

The meta-heuristic algorithms have random nature and the response may change at every execution. In order to evaluate the stability of the ACO, GA, and the proposed ABC algorithms, each of the test cases 5 and 6 is executed 50 times and the standard deviation of the obtained results is calculated and reported in Table 10. The small values for the standard deviations show the high stability of the algorithms.

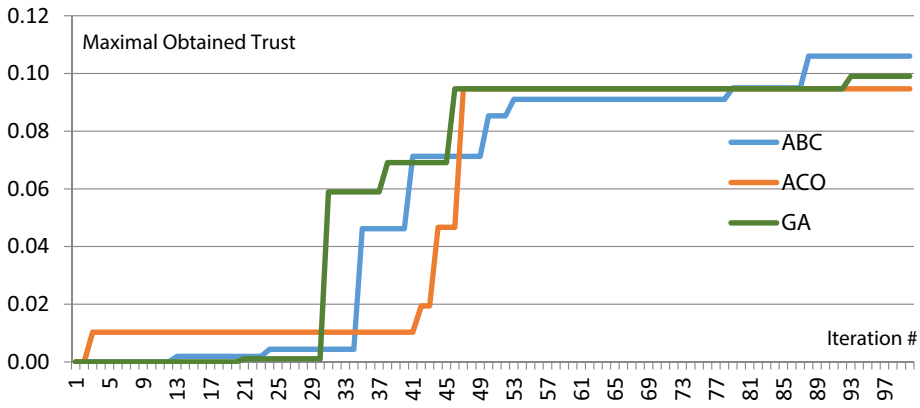

Fig. 8 The convergence diagram of test case 5

Table 10 The standard deviation of ABC, GA, and ACO algorithms

Case #	Standard deviation		
	Proposed ABC	GA	ACO
Case #5	1.24E−02	4.16E−03	2.19E−03
Case #6	6.47E−03	1.06E−02	7.41E−03

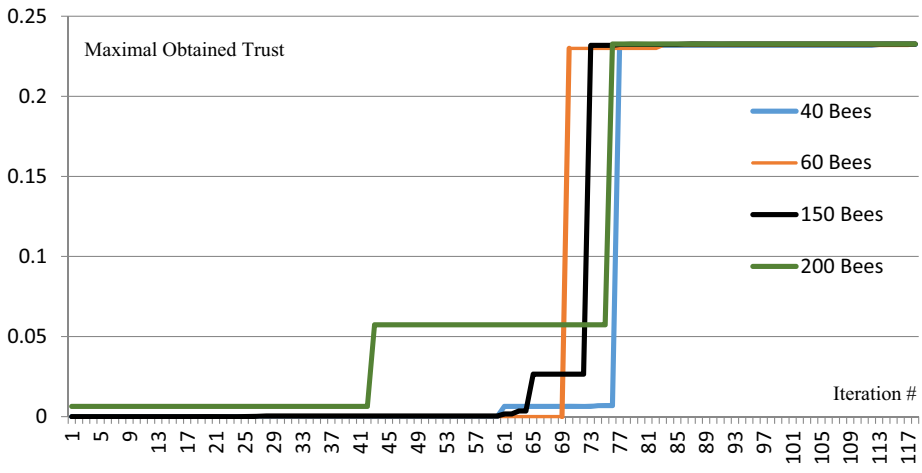

Fig. 9 The effect of increasing the number of artificial bees in test case 6

The convergence speed of the proposed ABC algorithm is affected by the number of initially employed bees. For example, the test case 6 is executed using 40, 60, 150 and 200 number of bees and the related convergence diagrams are given in Fig. 9. As the figure shows, the algorithm converges to the final value in less number of iterations as the number of bees increases.

Conclusions

Social networks and their applications are the necessities of today's life where millions of users are involved. Almost all of the interactions and transactions are performed based on the trust which directly depends on one's personality and the history of his/her previous activities. The study of the previous work shows that almost all of the proposed algorithms are complicated to understand have not considered the way the people trust each other in fact. In other words, these researches aim to show how people should trust each other, rather than how they really do it.

This research in this paper is organized in two phases: the first phase deals with distinguishing the trust mechanism between any two users of a social network, and the second phase proposes a meta-heuristic for obtaining a trust route based on the result of the first phase. The trust calculation concept is considered from another point of view for modeling the trust mechanism of people in the real world. For this reason, using questionnaires the social media users were asked to score the most important personal parameters affecting the trust, and also declare how much they trust their friends. This research proposed four new models for calculating the trust between any two friends of a social network and the best model is chosen using a statistical hypothesis test based on the information gathered by questionnaires. The statistical analysis revealed the proposed Model #1 better fits the empirical trust values and describes the behavior of social users more accurately than the other models. The trust concept provided by this model is easy to understand and its simplicity is a great benefit. However, the calculation of the maximal trust and the trust route on a real network needs lots of computational time and due to the NP-Hard complexity of the problem, a meta-heuristic algorithm based on Artificial Bee Colony approach is also developed. The proposed algorithm is programmed in Matlab® and is simulated using a personal computer running Microsoft Windows 10 with 4.2 GHz processor and 2 GB of RAM.

The proposed algorithm is executed for sample test cases adopted from the Facebook dataset and its efficiency was compared with the Genetic Algorithm and Ant Colony Optimization algorithms. The computational results show that the maximal trust values obtained by the proposed algorithm are better or equal to the values obtained by GA and ACO approaches. Besides, by increasing the size of the problems, the slope of the computation time of the proposed algorithm is clearly less than that of GA and ACO methods.

The national region and the small number of volunteers contributing to this research is the main limitation of this research. As a future work perspective, more volunteers would be involved in different regions of the world. Furthermore, other meta-heuristics can be developed and compared.

Abbreviations

SNA: Social Network Analysis; OSN: Online Social Networks; ABC: Artificial Bee Colony; ACO: Ant Colony Optimization; SN: Social Network; OSNA: Online Social Network Analysis; QoT: Quality of Trust; QoS: Quality of Service; EE: Estimation Error; MAD: Mean Absolute Deviation; MSE: Mean Square Error; UCL: Upper Control Limit; LCL: Lower Control Limit; KS: Kolmogorov–Smirnov; GA: Genetic Algorithm.

Acknowledgements

Not applicable.

Authors' contributions

The author read and approved the final manuscript.

Funding
The author has not received any funding for performing this research.

Availability of data and materials
All data used in the first phase of this research are gathered online from SN volunteers using questionnaires. The dataset used in the second phase of the research is adopted from http://snap.stanford.edu/data.

Competing interests
The author declares no competing interests.

Received: 10 November 2019 Accepted: 30 January 2020
Published online: 13 February 2020

References

1. Backstrom L, Huttenlocher D, Kleinberg J, Lan X. Group formation in large social networks: membership, growth, and evolution. In: Paper presented at the proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; 2006.
2. Boros E, Hammer PL. Pseudo-boolean optimization. Discrete Appl Math. 2002;123(1–3):155–225.
3. Christianson B, Harbison WS. Why isn't trust transitive? Paper presented at the International workshop on security protocols; 1996.
4. Coleman JS, Coleman JS. Foundations of social theory. Cambridge: Harvard University Press; 1994.
5. Daneshmand F, Daneshmand A. Computational algorithms in social trust. Glob J Sci Eng Technol. 2012;2:21–6.
6. Dehghan Z, AlMurtadha Y, Kuen LN, Salam ZA. Current trust inference mechanisms in web based social networks. J Comput Sci. 2012;8(9):1496.
7. Deng S, Huang L, Xu G. Social network-based service recommendation with trust enhancement. Expert Syst Appl. 2014;41(18):8075–84.
8. Dorigo M, Birattari M, Stützle T. Ant Colony Optimization. Comput Intell Mag IEEE. 2006;1:28–39. https://doi.org/10.1109/MCI.2006.329691.
9. Dwyer C, Hiltz S, Passerini K. Trust and privacy concern within social networking sites: a comparison of Facebook and MySpace. In: AMCIS 2007 proceedings; 2007. p. 339.
10. Evans MM, Wensley AK. Predicting the influence of network structure on trust in knowledge communities: Addressing the interconnectedness of four network principles and trust. Electron J Knowl Manag. 2009;7(1):41–54.
11. Fong S, Zhuang Y, Yu M, Ma I. Quantitative analysis of trust factors on social network using data mining approach. In: Paper presented at the first international conference on future generation communication technologies; 2012.
12. Frikha M, Mhiri M, Zarai M, Gargouri F. Time-sensitive trust calculation between social network friends for personalized recommendation. In: Paper presented at the proceedings of the 18th annual international conference on electronic commerce: e-commerce in smart connected world; 2016.
13. Guha R, Kumar R, Raghavan P, Tomkins A. Propagation of trust and distrust. In: Paper presented at the proceedings of the 13th international conference on world wide web; 2004.
14. Hamdi S. Computational models of trust and reputation in online social networks. Saint-Aubin: Université Paris-Saclay; 2016.
15. Hammer PL, Rudeanu S. Boolean methods in operations research and related areas, vol. 7. Berlin: Springer Science & Business Media; 2012.
16. Jiang W, Wang G, Wu J. Generating trusted graphs for trust evaluation in online social networks. Future Gener Comput Syst. 2014;31:48–58.
17. Kumar N, Guo R, Aleali A, Shakarian P. An empirical evaluation of social influence metrics. In: Paper presented at the 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM); 2016.
18. Liu F, Li X, Ding Y, Zhao H, Liu X, Ma Y, Tang B. A social network-based trust-aware propagation model for P2P systems. Knowl Based Syst. 2013;41:8–15.
19. Liu G, Wang Y, Orgun MA. Optimal social trust path selection in complex social networks. In: Paper presented at the twenty-fourth AAAI conference on artificial intelligence; 2010.
20. Liu G, Wang Y, Orgun MA. Trust transitivity in complex social networks. In: Paper presented at the twenty-fifth AAAI conference on artificial intelligence; 2011.
21. Liu G, Wang Y, Orgun MA, Lim E-P. Finding the optimal social trust path for the selection of trustworthy service providers in complex social networks. IEEE Trans Serv Comput. 2011;6(2):152–67.
22. Nuñez-Gonzalez JD, Graña M, Apolloni B. Reputation features for trust prediction in social networks. Neurocomputing. 2015;166:1–7.
23. Podobnik V, Striga D, Jandras A, Lovrek I. How to calculate trust between social network users? In: Paper presented at the SoftCOM 2012, 20th international conference on software, telecommunications and computer networks; 2012.
24. Sanadhya S, Singh S. Trust calculation with ant colony optimization in online social networks. Procedia Comput Sci. 2015;54:186–95.
25. Shakarian P, Bhatnagar A, Aleali A, Shaabani E, Guo R. Diffusion in social networks. Berlin: Springer; 2015.
26. Sherchan W, Nepal S, Paris C. A survey of trust in social networks. ACM Comput Surv. 2013;45(4):1–33. https://doi.org/10.1145/2501654.2501661.
27. Singh MM, Chin TY. Hybrid multi-faceted computational trust model for online social network (OSN). Int J Adv Comput Sci Appl IJACSA. 2016;7(6):11.
28. Situm M. Analysis of algorithms for determining trust among friends on social networks. (M.Sc.). Zagreb, Vienna; 2014.

29. Taherian M, Amini M, Jalili R. Trust inference in web-based social networks using resistive networks; 2008. p. 233–8. https://doi.org/10.1109/iciw.2008.41.
30. Takalkar VK, Mahalle PN. Confidentiality in online social networks; ATRust-based approach. J Cyber Secur Mob. 2016;4(3):125–44.
31. Trigg DW. Monitoring a forecasting system. OR. 1964;15(3):271–4. https://doi.org/10.2307/3007215.
32. Walter F, Battiston S, Schweitzer F. A model of a trust-based recommendation system on a social network. Auton Agent Multi Agent Syst. 2008;16:57–74. https://doi.org/10.1007/s10458-007-9021-x.
33. Wang J, Qiao K, Zhang Z. Trust evaluation based on evidence theory in online social networks. Int J Distrib Sens Netw. 2018;14:155014771879462. https://doi.org/10.1177/1550147718794629.
34. Wang Y, Cai Z, Yin G, Gao Y, Tong X, Han Q. A game theory-based trust measurement model for social networks. Comput Soc Netw. 2016;3(1):2.
35. Zhan J, Fang X, Killion P. Trust optimization in task-oriented social networks; 2011. p. 137–43. https://doi.org/10.1109/cicybs.2011.5949408.

Publisher's Note