DeciTrustNET: A graph based trust and reputation framework for social networks

Raquel Ureña*[a], Francisco Chiclana*[a,b], Enrique Herrera-Viedma*[b,c]

[a] *Institute of Artificial Intelligence (IAI), School of Computer Science and Informatics, De Montfort University, Leicester, UK*

[b] *Andalusian Research Institute on Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, Spain*

[c] *Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation*

Abstract

The world wide success of large scale social information systems with diverse purposes, such as e-commerce platforms, facilities sharing communities and social networks, make them a very promising paradigm for large scale information sharing and management. However the anonymity, distributed and open nature of these frameworks, that, on the one hand, foster the communication capabilities of their users, may contribute, on the other hand, to the propagation of low quality information, attacks and manipulations from users with malicious intentions. All of these risks could end up decreasing users' confidence in these systems and in a reduction of their utilisation. With these issues in mind, the objective of this contribution is to create DeciTrustNET, a trust and reputation based framework for social networks that takes into consideration the users relationships, the historic evolution of their reputations and their profile similarity to develop a tamper resilient network that guarantees trustworthy communications and transactions. An extensive experimental analysis of the developed framework has been carried out confirming that the proposed approach supports robust trust and reputation establishment among the users, even in social network under the presence of malicious users.

Keywords: Trust, Reputation, Influence, Social networks, Decision Making, Opinion dynamics

1. Introduction

Word of mouth (WOM) constitutes one of the oldest ways to disseminate opinions about products or services and, at the same time, the one that has the highest impact in the consumer behaviour as a consequence of the high reliability and credibility transmitted by interactions with known persons[1, 2]. Thanks to the world wide presence of the Internet based technologies, which allow the communication between users from all corners of the planet, a new form of WOM has emerged, the electronic WOM (eWOM). This new source of information is considered as the most influential among the users of information before, during, and after consuming a given product or service [1, 3]. For example, a recent study place eWOM as the most influential pre-purchase source of travel information [4]. The main differences between WOM and e-WOM lie in credibility and speed of propagation.

Nowadays there is a huge offer of on-line communities with different characteristics and purposes ranging from platforms to share pictures and opinions with friends and followers such

*Corresponding author

Email addresses: raquel.urena@dmu.ac.uk (Raquel Ureña*), chiclana@dmu.ac.uk; inv.chiclana@ugr.es (Francisco Chiclana*), viedma@decsai.ugr.es (Enrique Herrera-Viedma*)

as Facebook[1], Instagram[2] or Twitter[3], to e-commerce based frameworks like Amazon or E-bay; Crowdsourcing platforms to share knowledge and expertise, such as Wikipedia[4], Slashdot[5], or Quora[6]; and On-line facilities sharing networks such as UBER[7] and Blablacar[8] for cars or Airbnb[9] for accommodation. In spite of the diverse purposes of these on-line communities, they all have a large number of users that exchange information using a virtual identity. Thus, these systems can be recognised for their open, self-supervised and dynamic nature, characteristics that, on the one hand, foster the information sharing and knowledge discovery but, on the other hand, incentivise malicious users and fraudulent behaviours [5]. In these scenarios, in order to reduce the uncertainty as well as to increase the expectations of positive exchanges or transactions, eWOM, and consequently Reputation and trust play a key role. Indeed, according to the theory of reasoned action [6], the belief influences the attitude and the latter shapes the behavioural intention, thus, when a person acquires a positive attitude toward a given behaviour it is more likely that he/she will engage in such behaviour. In this sense, "reputation can be understood as a predictor of future behaviour based on previous interactions" [7]. For example, an agent will be considered as highly regarded if it has consistently performed satisfactorily in the past and so this agent will be assumed as trustworthy in the future. Thus, reputation and trust may ultimately discourage poor and dishonest agents while motivating reliable and trustworthy ones [8, 9].

In consequence, developing methodologies for assessing reputation and trust of the peers as well as flagging and isolating dishonest behaviour pose a real challenge. With this regard, some commercial examples can be found in the aforementioned e-commerce platforms, Amazon and E-bay, or in facility sharing platforms, Airbnb, Blablacar or UBER, where a public reputation score is calculated based on the feedback from both, hosts and guests. However these examples rely exclusively on the feedback provided by other users, feedback that in many occasions is very hard to get or may be affected by malicious or fraudulent intentions as we will analyse in the next section. Moreover this approach is not operational with new users entering the network due to the well known cold start problem [10].

From the research point of view, we can find several frameworks to propagate trust in Internet based scenarios [8, 11–14] and in group decision making approaches [15–17]. With these regard a comprehensive review of these approaches have been carried out recently by Urena et al. in [18] concluding that there is still a need of tamper-resilient trust and reputation based mechanisms that allow the estimation and propagation of trust not only based on explicit feedback but also on users' relations and behaviours. These systems should be effective in sparse scenarios with new users and malicious intended ones. Sparse scenarios in the context refer to sparse social networks, i.e. networks in which the number of edges are very few in comparison with the total possible number of edges, as opposed to dense or complete networks. The node distribution in sparse connected networks has a scale free, power law distribution [19].

With these challenges in mind, the objective of this contribution is to develop a framework for robust trust establishment in on-line communities even in the presence of malicious and new agents. The proposed system, named DeciTrustNET is graph based reputation-trust aggregation and propagation framework that allows to establish tamper-resilient trust relationships

[1]https://www.facebook.com/
[2]https://www.instagram.com/
[3]https://www.twitter.com/
[4]https://www.wikipedia.com/
[5]https://www.slashdot.com/
[6]https://www.quora.com/
[7]https://www.uber.com/
[8]https://www.blablacar.com/
[9]https://www.airbnb.com/

in on-line communities. The main characteristics and novelties of the proposed approach are enumerated as follows:

- **Double supervised personalised feedback from other members of the network:** DeciTrustNET allows users to rate their interactions, to avoid users' malicious behaviour, will affect the reputation of both rater and rating users.
- **Distinguishing between user's global reputation and pairwise trust:** We propose two measures for each user: (i) the Global reputation, based in all users' interactions, the obtained and provided feedback, their behaviour and the quality of their relationships; (ii) The pairwise trust to measure the level of trust or confidence that could be between two users based on previous interactions [15, 20].
- **Exploiting user relationships and their position in the network to asses users Reputation:** this is based on the premise that individuals with trustworthy friends are more likely to be trustworthy. [13, 21, 22].
- **Tracking user behaviour over time:** DeciTrustNET incorporates the evolution and trajectory of users trust rating in order to motivate the users to engage in long term good behaviour and to penalise those users with sudden change in behaviour.

We believe that DeciTrustNET may be specially useful in three specific scenarios: to carry out trust based negotiations in consensus reaching processes, such as the ones in [23, 24] ; in e-health platforms to provide recommendations on how to keep a healthy lifestyle; and in recommender systems for e-commerce and e-marketing. This contribution is organised as follows: Section 2 surveys the main trust and reputation approaches in the literature as well as the main malicious user behaviours and attacks reported against these type of systems. Section 3 describes the proposed framework, explaining in detail how the trust and reputation are calculated and propagated along the network for each user. Then, in Section 4 the network architecture of the proposed system is explained while the experimental results and the discussion are included in Section 5. Finally we explain the conclusions of our work in Section 6.

2. Trust and reputation in social networks

This section is devoted to present the concepts of trust and reputation, analysing the main existing mechanisms to calculate, propagate and leverage these two measurements in on-line network based scenarios [18]. Moreover, an analysis of the main security threats that these type of systems may present is carried out to ascertain their mean weaknesses in order to propose approaches to tackle them.

2.1. Social Networks

A social network is composed of a set of agents that shares diverse types of information with various purposes such as friendship, e-commerce, information dissemination or business [19]. The Social network theory consist on "the analysis of the different structures in the network to understand which are the underlying pattern that may either facilitate or impede the knowledge creation in this type of interconnected communities" [20]. In comparison with a random graph of nodes, a social network presents some specific characteristics [25] among them we can point out the most remarkable two:

A small-world network: This property implies that even if most of the nodes in a social network are not direct neighbours, they are likely neighbours each other and so every node can be reached from every other node within a few number of hops. The two main structural properties that allow to recognise this type of networks are a high clustering coefficient and

the average path length that scales the logarithm of the number of nodes. A high clustering coefficient is associated with the idea that a friend of one's friend is likely to be one's friend too, while the average path length refers to the average number of nodes in the paths joining two nodes in the network [19].

Scale Free network means that only a few nodes are the ones holding a high number of connections, which are referred to as connections hubs, while the majority of the nodes are connected to very few nodes [19].

2.2. Trust and reputation systems

In spite of being almost omnipresent in our daily routine, modelling formally the concepts of trust and reputation represents a challenge due to ambiguity of these two terms. According to Josang et al. in [7] "Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible." Regarding to reputation, according to the Concise Oxford dictionary, "**reputation** is what is generally said or believed about a person's or thing's character or standing." This definition agrees with the one given by social network researchers that claims that "reputation is a quantity measure derived from the underlying social network which is globally visible to all members of the network" [26].

In the light of both definitions, we will consider trust as the pairwise level of confidence that an entity may have on another one based on previous interactions, while reputation is considered as the "global perception that an agent creates through past actions about its intentions and norms in a global level" [18]. Prior to the analysis of the main existing mechanisms for trust and reputation, we consider necessary to point out the main characteristics that any of these systems should comply with [12], which has been taken into consideration in developing the proposed approach herein.

1. **Self policing**: The system will only use the information, feedback and ratings, provided by other users.
2. **Long lasting entities**: It is assumed that additional interactions will take place in future.
3. **Trust and reputation based on the behaviour over time:** Reputation and trust has to be in terms of long term behaviour, without giving any advantage to newcomers.
4. **Reduced computation cost:** Trust and reputation calculation should not require an effort in terms of computational power.
5. **Robust to malicious users:** Users trying to take advantage or manipulate the system have to be immediately flagged and isolated.

2.2.1. Trust and Reputation calculation

One of the simplest and effective techniques to compute reputation is based on *counting* [27, 28], which is mainly used by eBay and Amazon. In the case of eBay this technique consists of a summation of the positive ratings minus the negative ones whereas in the case of Amazon a weighted average of the ratings is carried out considering other factors such as the rater trustworthiness and the number of provided ratings for example. A statistical approach presented in [7, 29] that estimates the probability that a future transaction would be positive or negative given the historic of the previous transactions. There are also the systems that use fuzzy numbers or linguistic ratings modelled as fuzzy sets in which the membership function describes to what extend an agent can be trustworthy or not. Some examples of these systems are the Regret System presented in[30] and some trust based group decision making methodologies developed in [17, 31, 32].

2.2.2. Propagation approaches

These approaches deal with the situation where there might not be direct trust relationship between all the agents in the network. Therefore their goal is to estimate an unknown trust value between two agents using existent indirect trust paths between them. Flow propagation models are the most frequent used; they assume that an user is likely to trust the statements coming from a trusted user, and make use of a transitive property to estimate the trust score through iterative aggregation along transitive chains until they become stable for all agents [11].

The most representative of these approaches is the one proposed by Guha in [11], which carries out atomic propagation of the trust in four different ways:

1. If agent i trusts agent j ($t_{ij} = 1$), and agent j trusts agent k ($t_{jk} = 1$), then agent i will trust agent k ($t_{ik} = 1$). This is known as **direct propagation** of trust.

2. In agent i_1 trusts agents j_1 and j_2, and agent i_2 trusts agent j_2, the **co-citation** propagation of trust assumes that agent i_2 may trust agent j_1.

3. Given that agent i trusts agent j then the **transpose trust** propagation implies that agent j might present some level of trust towards agent i.

4. Given that agent i trusts agent j then the **trust coupling** propagates to agent k if agent j and agent k trust agents in common.

In the same line, Kamvar et al. propose in [12] a methodology to compute a universal value of trust for each node, in contrast with the pairwise one in [11], with two objectives: (i) To isolate malicious agents from the network by encouraging agents to interact with reputable ones; (ii) To motivate agents to interact by rewarding reputable ones.

2.2.3. Existent Trust and reputation frameworks

In this subsection the characteristics of the main trust and reputation based systems are analysed:

- **RateWeb**[8] is decentralised and unstructured framework applied to web services. This system operates as follows: Each agent stores a personal perception of the services it has interacted with. In order to select a partner, the trusting entity queries the community obtaining a set of eligible services providers including a list of past entities that used the service. The reputation of each service provider is calculated based on the obtained feedback in the following way:

$$Rep_i = \frac{\sum_{j=1}^{L} t_{ij} \lambda_f Cr_j}{\sum_{j=1} Cr_j}$$

  Where L denotes the set of trusting agents which have interacted with the service provider i; t_{ij} represents the pairwise trust value that an agent j has towards agent i; $Cr_j \in [0, 1]$ is the credibility of each agents, as viewed by the inquiring entity, and $\lambda_f \in [0, 1]$ is a trust decay factor over time.

- **R2Trust** has been proposed in [33] as a fully distributed reputation system in which the reputation of an agent is estimated as an aggregation of the obtained feedback weighted by local pairwise trust values. These trust values, calculated using social relationships, consist in probabilistic ratings computed as a function of the past interaction. One of the main advantages of this approach is its capacity of fast reaction in case of an irregular variation on the behaviour of an agent.

- The **GRAft** distributed reputation system [9] is characterised by the use of both explicit reputation information such us feedback, scores and rating, and implicit structural information of the given node in the social network, i.e. the in-degree and out-degree.

- **Random Walk** trust measure is based on the well known Page Rank algorithm [34]. In this system, a random walker surfs the network in a similar way that in the web with the popular Google's algorithm.

- **SocialTrust** [13] is a Random Walk based framework that combines the following factors to model trust between the users: trust group feedback, user's relationship quality and user behaviour over time

- **PCR** was proposed in [14] in the same line than SocialTrust [13]. This system consists in a multigraph based social network where users are characterised and interconnected keeping track of various criteria such as the behavioural activities and social relationship in order to build trust relationships between them even when there is a lack of first hand information. Moreover, in order to discard bad-mouthing and personalised direct distrust propagation, a deception filtering approach has been proposed.

2.3. Malicious User Behaviour

Trust and reputation systems, as other informatics systems, suffer from vulnerability to attacks. In this subsection we present an overview of the most frequent attacks to these systems and some of the approaches to address them. One of the most common malicious behaviours consist in the **unfair rating**. That is, the feedback provided by a given user after a transaction is deliberately false with a malicious intention, for example to manipulate the score towards the benefit of certain entities. Some of the most frequent ways of carrying out this type of attack are [18]:

1. **Self promoting:** This kind of attacks is based on a group of agents that collude in order to highly rate between each others to artificially boost their personal reputations.

2. **Slandering or bad mouthing:** This attack is the opposite to self promoting, that is, it is based in a group of users that agree in giving unfair bad ratings to other users in order to destroy their reputation.

3. **Whitewashing:** This is a short term attack in which the perpetrators intentionally behave in an unfair way in order to obtain a certain benefit even if their reputations gets degraded. Afterwards, they re-enter the system with a new identity.

4. **Orchestrating:** In this case several attackers agree on using one of the aforementioned techniques simultaneously. When there are malicious opinion contributors as well as raters who behave inconsistently when providing their ratings, the dispersion in the feedback provoked by differences in taste can be difficult to discern from that induced by other factors such us unfair ratings. Therefore it is a challenging task to distinguish the legitimate feedback and to filer out the malicious one.

5. **Ballot Box Stuffing:** This consists in obtaining more votes than the expected ones.

In order to recognise and reduce these unfair procedures some approaches have been identified [18]:

1. **Endogenous discounting**: In this case, the statistical properties of the ratings are used to give less importance or even exclude ratings that are suspected to be unfair.

2. **Exogenous discounting:** These mechanisms are based on the idea the raters with low reputation are more likely to provide unfair feedback. Therefore the rater's reputation is taken into consideration to weight the ratings.

3. **Restrict ratings provision:** In order to avoid ballot box stuffing, ratings are only allowed after the transaction has been fully accomplished.

Other vulnerabilities that reputation and trust frameworks may present given their distributed and open nature are:

- **Nearby threats:** In many on-line social networks, even when participants keep strict control on who their friends are, the small world effect means that malicious participants might be at a short distance in the network from a given user [13].

- **Limited network view:** This phenomenon, which is another consequence of the small world effect, implies that agents have no information about the trustworthiness of the majority of the members in the network since they are connected only with a small portion of the network users.

- **Low incentive for providing ratings:** A frequent issue in this type of networks is the absence of mechanisms to encourage users to provide feedback (ratings) following the completion of a transaction.

In summary, the existing systems mostly rely on the users providing an explicit feedback about the interactions with other users [9, 35] to obtain either a global reputation score or a pairwise based trust measure. Nevertheless, as aforementioned, direct rating suffers from various vulnerabilities such as unfair ratings and low user motivation to rate. However, in on-line interconnected systems, other sources can be leveraged to implicitly obtain information about trust and reputation. For example, the quality of the users' relationships and their behaviours and interactions over time. With this in mind, in this contribution we propose a new framework that estimates both pairwise trust and reputation and merge both in an user centric influence measure that weights the degree of impact of each user in the network.

3. The DeciTrustNET framework

The proposed approach is composed of a social network composed of agents who, based on trust and reputation grounded recommendations or behaviours, both influence and influenced by peer agents. In this framework, users can either actively query the agents in their vicinity/neighbourhood or receive information about their trusted neighbours in a passive way. In these cases, although the most straightforward approach would be to take into consideration the most recurring recommendation, agents might feel more confident and comfortable with recommendation coming from trusted agents [2] and not necessarily the most recurring or frequent recommendation. The main building blocks of the proposed DeciTrustNET framework are depicted in Figure 1. One of the main novelties of the proposed system is its dynamic approach to modelling the network of peers, as opposed to the assumed static modelling of implemented in existent systems in the literature [13]. Indeed, in DeciTrustNET both the creation and evolution of the network is carried out dynamically based on the users interactions, the developed trust and their profiles similarities. For every agent in the network the following information is stored:

User Profile: Typically a profile is a user-controlled compilation of information that contains some descriptive data about the person it represents including some characteristics that defines unequivocally a user. The profile features are stored out in a digital way to allow

the estimation of similarities/differences/distances between users. These features will vary depending on the application of the proposed network. For example if we take into consideration a professional network, such as LinkedIn, a profile consists of the different professional features that characterize a member of the network: experience, education, training, spoken languages, etc...

User Interactions: These consist of all the pairwise interactions of a user with others peers (posts, shares, likes, ...). The analysis of these interactions will prove key in the dynamic modelling approach of the directed graph of interactions as detailed later on in this section. To give an example, in the case of Twitter all these interactions can be obtained via its Twitter API [36].

User Received Ratings: These are all the feedback than an agent has received.

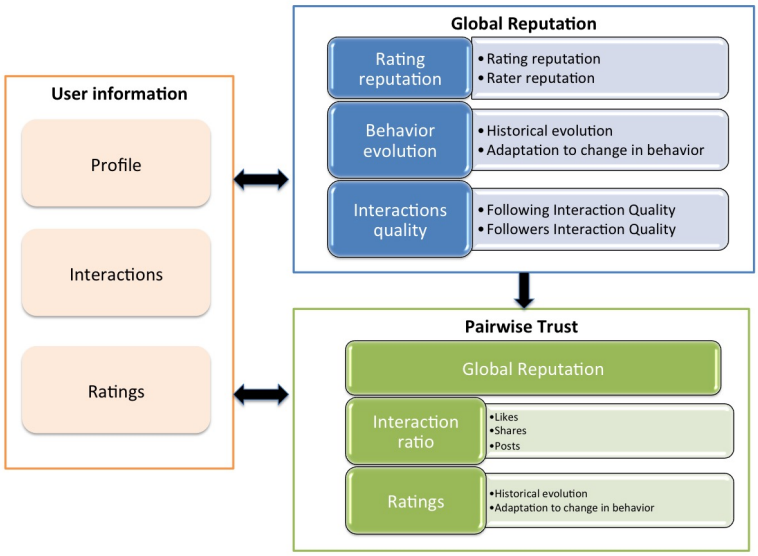

Figure 1: DeciTrustNET main conceptual components

Due to the sparsity of the majority of the social networks and the constant affluence of new users, DeciTrustNET proposes the following measures of trust:

Pairwise Trust. The trust value between two agents in the network depends on their history of interactions, their similarity and their received feedback.

Global Reputation. The trust value of an agent in the network depends on his/her history of interactions with all the users in the network.

To facilitate the understanding of the formalisation of the DeciTrustNET framework, Table 1 presents all the parameters and their symbols used to represent them in the mathematical model developed herein. The variable time, t, is to be considered in the network in order to model, study and analyse the evolution of each agent's profile state after receiving a rating or after an interaction with another agent.

3.1. Pairwise trust assessment

In computing the pairwise trust values between two network agents (users), (i, j) and (j, i), at time t, the following two possible cases are possible: (i) Agents i and j have interacted in the past and there is a stored record of their ratings $[r(i, j, t)]$ and interactions between them $[IR(i, j, t)]$; (ii) There is no previous interaction between the two agents and the only

Table 1: Used parameters

Name	Symbol
Pairwise Trust	PT
Global Reputation	GR
Interaction Ratio	IR
Weighting factor for PT	α, β
Average rating between two agents	$\overline{r}$
Rated reputation	R
Rating reputation	RR
Similarity threshold between RR and r	α_{RR}
Global Interaction quality	GIq
Following Interactions quality	Fwi
Followers Interactions quality	Fwe
Historic behaviour evolution	$Hist$
Adaptation to change in behaviour	AC
Behaviour evolution	BE

information available are their respective global reputation values [$GR(i,t)$ and $GR(j,t)$]. The pairwise trust value between agents in a network can be measured as a normalised weighted average of their ratings, interactions and global reputation values. This is captured in the following definition.

Definition 1 (Pairwise trust). The pairwise trust between (i,j) at time t, $PT(i,j,t)$ is computed as follows:

$$PT(i,j,t) = \alpha \cdot IR(i,j.t) + \beta \cdot \overline{r}(i,j,t) + (1-\alpha-\beta) \cdot GR(j,t) \tag{1}$$

where $0 \leq \alpha + \beta \leq 1; 0 \leq \alpha, \beta \leq 1$

In case (ii) above $\alpha, \beta = 0$, while in case (i) it is expected to have both $\alpha, \beta > 0$. Below, we elaborate on each one of the elements involved in the computation of PT in Definition 1.

Interaction Ratio (IR). This is computed based on the set I of all the exchanges or interactions that user i has carried out with user j. An interaction between two agents i and j happens when one of the following three types of events happen:

- User i gives a like to user j's posting, sharing, etc. [$\#likes(i,j)$].
- User i posts something directly on user j [$\#posts(i,j)$].
- User i shares something with user j [$\#shares(i,j)$].

In social networks such as Facebook, Twitter or Instagram there is another type of interaction between agents denominated "following", which consist in an user voluntarily choosing to receive the updates coming from a given user. In our proposed framework we do not take into consideration this type of interaction in the computation of IR because one of the objectives of DeciTrusNET is to identify agents to follow based on trust and reputation.

Definition 2 (Interaction Ratio). The interaction ratio (IR) of users (i,j) at time t, $IR(i,j,t)$, measures the ratio of interactions between user i with user j with respect to the total number

of interactions carried out with all the users at time t, as per a linear combination of the three interaction events Likes ratio (LR), the Shares ratio (S) and the Posts ratio (PR):

$$IR(i,j,t) = \lambda_1 LR(i,j,t) + \lambda_2 SR(i,j,t) + \lambda_3 PR(i,j,t) \tag{2}$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and

$$LR(i,j,t) = \begin{cases} \dfrac{\#likes(i,j,t)}{\#likes(i,t)} & \text{if } \#likes(i,t) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$SR(i,j,t) = \begin{cases} \dfrac{\#shares(i,j,t)}{\#shares(i,t)} & \text{if } \#shares(i,t) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$PR(i,j,t) = \begin{cases} \dfrac{\#posts(i,j,t)}{\#posts(i,t)} & \text{if } \#posts(i,t) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The above definition of IR would allow impact studies of each of the three individual ratios might have on IR. In addition, the case $IR(i,j,t) = 0$ $(\forall j)$ means that user i has been inactive in the network regarding likes, posting and sharing up to the time t.

Ratings. In contrast with the interactions, direct feedback rating consists in the explicit opinion an agent received on his/her interactions with other agents.

Definition 3 (Average rating). In addition to interacting with agent j, agent i may also desire to feedback about this interaction via an explicit rating with a value $r(i,j) \in [-1,1]$, where -1 means a totally unsatisfactory interaction and 1 represents a fully satisfactory one. The average at time t of all ratings provided by agent i to agent j will be denoted by $\overline{r}(i,j,t)$.

3.2. Global user reputation

In the previous subsection we examined the computation of the trust between two users based on their direct interactions. However, in sparse networks, users often interact with other users without having previous mutual relationship experiences. Thus, it is necessary to devise a computation approach of users' reputation taking into consideration their behaviour in the network, the provided and received ratings and their network interactions. This is the focus of this subsection.

3.2.1. Global reputation assessment

The overall Global Reputation of a user i at time t, $GR(I,t)$, can be calculated as the combination of the following elements:

- Agent's Rating-Reputation $RR(i,t) \in [-1,1]$
- Agent's Rated-Reputation $R(i,t) \in [-1,1]$.
- Agent's Behaviour Evolution $BE(i,t) \in [0,1]$,
- Agent's Interactions Quality $GIq(i,t) \in [0,1]$,

Global reputation can be defined as follows:

Definition 4 (Global reputation for agent i).

$$GR(i,t) = w_{RR}\frac{RR(i,t)+1}{2} + w_R\frac{R(i,t)+1}{2} + w_{Giq}GIq(i,t) + w_{BE}BE(i,t) \tag{6}$$

where $w_{Gig}, w_{BE}, w_{RR}, w_R$, $w_{GIg} + w_{BE} + w_{RR} + w_R = 1$ are the weighting factors to aggregate the different elements. In practice the GR values will be kept between $[0,1]$, truncating to these values if needed. $GR(i,t) = 0$ means a very poor reputation while $GR(i,t) = 1$ means an excellent reputation. Note that when an agent enters the network, there is not any transaction registered and consequetly there is neither information related with the agent's behaviour nor obtained ratings. Consequently the new agents' Global reputation is calculated based on the quality of his Interactions, and asumming the maximal Rating Reputation, that is the agent is assumed to be fair in the provided ratings when entering the network (optimistic approach).

The following subsections are devoted to detail each of the elements in this definition.

3.2.2. Rating based reputation

Direct feedback ratings are key for assessing users' fulfilment. However, in many occasion direct feedback is difficult to get because users do not feel motivated to provide it or, as aforementioned, because it can be related with malicious behaviours. Motivating users to rate others users in a fair way is a challenge, which could be addressed by linking ratings with impact on in both the rated and the rating agents' reputation. In other words, rating agents providing fair ratings would have a positive impact in their associated rating reputation (increase), while the provision of unfair ratings would impact negatively in their rating reputation (decrease). This is the approach implemented in the DeciTrustNET framework as elaborated below.

When agent i rates another agent j, the difference between the given rating value and the agent j's reputation based on previous ratings received could be interpreted as follows: (i) agent i provides a fair rating when it is close to an agent j's reputation (difference is 'small'); (ii) agent i provides a malicious rating when it is outside the above agent j's reputation range. Thus, each agent can be associated two intertwined reputation measures: the Rated-Reputation and the Rating-Reputation. The initial values of an agent Rated-Reputation value $[R(j,0)]$ and Rating-Reputation $[RR(j,0)]$ value will evolved taking into account: (i) the rating values received from other agents and their Rating-Reputation values in the case of the Rated-Reputation update; and (ii) the ratings the agent provides to other agents and their Rated-Reputation values in the case of the Rating-Reputation update. This is reflected in the following definition.

Definition 5 (Rated-Reputation/Rating-Reputation). When an agent j received a new rating value at the instant t from agent i: $r(i,j,t) \in [-1,1]$, then agent j's Rated-Reputation is updated from $R(j,t-1)$ to $R(j,t)$ as follows:

$$R(j,t) = \frac{R(j,t-1)\cdot(\#ratings(j,t)-1) + \left(\frac{RR(i,t)+1}{2}\right)\cdot r(i,j,t)}{\#ratings(j,t)} \tag{7}$$

where $\#ratings(j,t)$ is the total number of ratings received by agent j in the period of time $[0,t]$; $R(j,0) = 0$ is the initial Rated-Reputation of agent j when joining the network. Note that this equation calculates the Rated-Reputation of agent i at moment t as the weighted average of the (previous) Rated-Reputation of agent i at moment $t-1$ based on the ratings received before instant t and the Rating-Reputation $RR(i,t)$ based value in $[0,1]$ of the agent i who is providing the new rating value for agent j at instant t. The Rating-Reputation $RR(i,t)$ of the agent i at instant t is also updated from $RR(i,t-1)$ as follows:

$$RR(i,t) = \begin{cases} RR(i,t-1)\cdot\left(1 - \frac{|R(j,t-1)-r(i,j,t)|}{\#givenRatings(i,t)}\right) & \text{if } |R(j,t-1)-r(i,j,t)| > 2\alpha_{RR} \\ RR(i,t-1) & \text{otherwise} \end{cases} \tag{8}$$

where $\#givenRatings(i,t)$ is the number of ratings provided by agent i in period $[0,t]$; $RR(i,0) = 1$ is the initial Rating-Reputation of the agent i when joining the network; and $\alpha_{RR} \in [0,1]$ is a threshold value to discriminate fair ratings from malicious ratings as described above.

In Definition 5, it is observed that at the initial Rated-Reputation state ($R(j,0) = 0$), any first rating value $r(i,j,1)$ could lead to a substantial decrease of the initial Rating-Reputation of the agent providing such rating. Indeed, an extreme value for such first rating would decrease the initial Rating-Reputation to 0 for any $\alpha_{RR} < 0.5$, and it would remain there forever. It is evident though, that the expressions for updating both the Rated-Reputation and Rating-Reputation values of agents in a network make sense when there is a minimum number of ratings received by agents in the network. Thus, an additional $minReceivedRatings$ parameter is declared as a threshold value to set the minimum number of ratings that users in the network have to receive for the Rated-Reputation and Rating-Reputation values of agents in the network to be updated by the proposed DeciTrustNET framework. This value will depend on the number of users in the network of interest, and the sparsity of the users' relationships. In the simulation presented in the last part of the present paper, this threshold value is set as $minReceivedRatings = 10$.

3.2.3. Behaviour evolution

Many of the approaches mentioned above estimate user's reputation based on a snapshot of the current state of the network, and consequently, they do not motivate users to maintain a long-term good behaviour nor prevent them from repeatedly leaving and re-entering the network to whitewash their trust ratings. Therefore, with the double objective of getting a robust trust measure and ensuring a long lasting good behaviour, it is necessary to asses the historical evolution of users' behaviour during their interactions.

The average of a user's global reputation during a number of past interactions can be considered as a starting point to measure a user's behaviour evolution. The following definition captures the measurement of a use's historical behaviour on a number of past periods of time.

Definition 6 (Historical behaviour evolution). Agents' historical behaviour evolution during the last $M(>1)$ periods of time or updates of their profile is measured as follows:

$$Hist(i,t) = \frac{\sum_{k=t-1-M}^{t-1} GR(i,k)}{M} \tag{9}$$

where $GR(i,k)$ is the Global Reputation of agent i at moment $k \in \{t-1-M, t-M, \ldots, t-1\}$.

Notice that there is nothing to prevent malicious users from suddenly shifting their behaviour to achieve a good Rated-Reputation and start acting improperly again afterwards, i.e. the use of the user's historical behaviour evolution on its own is not sufficient to capture effectively their behaviour evolution. Therefore, following the approach presented in the SocialTrust framework [13], there is also a need in the proposed framework to incorporate a mechanism to detect and penalise sudden changes in a user's behaviour, which is captured in the following definition.

Definition 7 (Adaptation to change in behaviour AC). Agents' adaptation to change in behaviour, $AC(i,t)$, considers the difference between the last rating received by the agent $r(i,t)$, and its historical behaviour evolution, $Hist(i,t)$, i.e.

$$AC(i,t) = -\frac{|r(i,t) - Hist(i,t)|}{2} \tag{10}$$

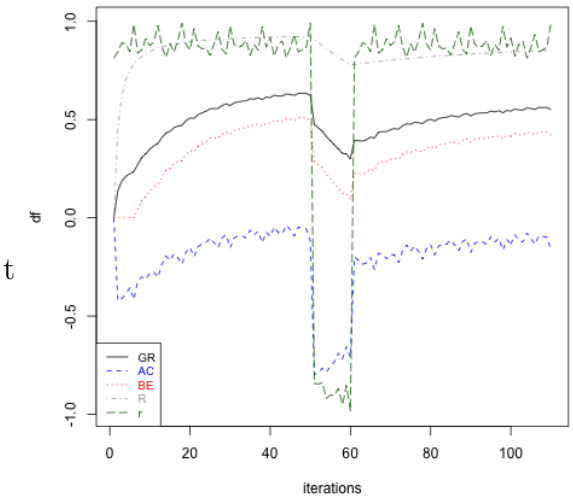

Figure 2: Impact of BE and AC in GR with $w_{bev} = 0.3$

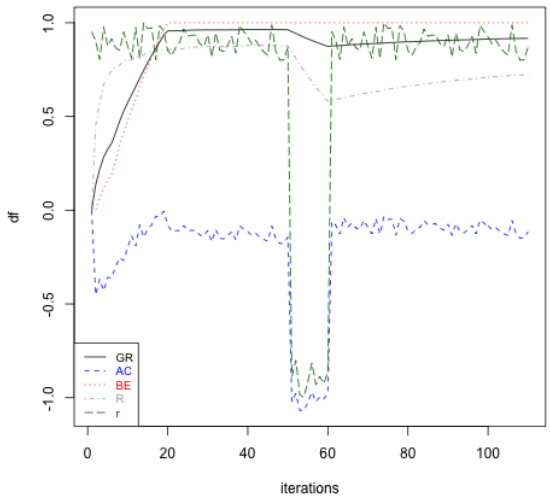

Figure 3: Impact of BE and AC in GR with $w_{bev} = 0$.

Notice that this measure detects and penalizes sudden changes in user's behaviour. Therefore when an agent's behaviour remains relatively stable, the difference between the last obtained rating and his historical evolution will be close to 0 and so his AC value will be close to 0. However if the behaviour of a user changes with the time, either decreases or increases, the absolute value of the difference between the last obtained rating and the historical evolution will increase, and consequently the AC value will decrease. The combination of agents' historical behaviour evolution and adaptation to change in behaviour will be taken into account in measuring their behaviour evolution, which is expressed in the following definition.

Definition 8 (Behaviour Evolution). Agents' behaviour evolution at the instant t is measured as follows:

$$BE(i,t) = w_{bev}AC(i,t) + (1 - w_{bev})Hist(i,t), \tag{11}$$

where $w_{bev} \in [0,1]$ is a trade-off behaviour value between agents' historical behaviour evolution and its change in behaviour.

A high value of w_{bev} reinforces the importance of sudden changes in agents' behaviour, i.e. if a user's Global Reputation experiences a sudden abrupt change, then this will be immediately reflected in such user's behaviour evolution. On the other hand, if w_{bev} is close to 0, then more importance will be given to the users' historical behaviour evolution and sudden changes in the global reputation will be attenuated. Note that given the fact that AC could be negative, BE could be negative as well; however in the simulations its minimum value was set to 0, i.e. the following expression is implemented in the simulations $BE(i,t) = \max\{w_{bev}AC(i,t) + (1 - w_{bev})Hist(i,t)\}$.

In the following, the impact of the AC measure in the global reputation is illustrated. To do so, a simplified scenario where only the new received ratings are considered when computing the Global reputation is used. Assume an scenario where an agent changes his/her behaviour suddenly and shifts from good to bad ratings or from bad to good behaviour as depicted in Figure 2 ($w_{bev} = 0.3$) and Figure 3 ($w_{bev} = 0$) where GR represents the obtained global reputation, AC the adaptation to change, BE the Behaviour evolution, R the rating reputation and r the obtained rating.

In Figure 2, it is observed that AC experiences small changes in the direction of the value 0 while the ratings are constant; however, AC experiences a drastic decrease when the user's

rating behaviour suddenly changes from positive to negative and from negative to positive. In fact, it can be observed that BE and GR variables show a similar relationship with respect to the variable AC, with a sudden decrease of AC reflected in a decrease of GR, and vice versa. Thus, in comparison with only using the Rating reputation, the system reacts faster to changes in behaviour, and so it reacts by immediately decreasing the reputation. In addition, it is noticed that it is not easy to obtain high reputation levels because it requires both the presence of good ratings but also that they maintain stability over time. In Figure 3, and so AC is not considered in BE. In this case, BE and GR remains almost constant and they are not affected at all by sudden changes in the ratings of the user. This is motivated by the fact that the user has a past history of good behaviour and therefore this scenario misses this important change of behaviour. Thus, the presence of the variable AC is necessary to achieve a system sensible to changes even when the agent has a previous good behaviour.

3.2.4. Interactions quality

As aforementioned, trust propagation relies on the idea that users' trustworthiness depends on their trustworthiness connections, i.e. a person can be considered trustworthy if his/her friends are trustworthy. This resembles the idea used by Bonachi et al. in [21, 22] to propose the concept of centrality measure. The interaction quality measure is composed of two main values: following interaction quality and follower interaction quality.

Definition 9 (Following interaction quality). Let $FWI(i)$ be the set of agents who agent i has interacted with in an active way until moment t, i.e. agent i has liked, shared, or posted something coming agents in FWI. Agent i's **following interaction quality** at time t, $Fwi(i,t)$, is computed as follows:

$$Fwi(i,t) = \begin{cases} \dfrac{\sum_{j \in FWI(i)} IR(i,j,t-1)GR(j,t-1)}{\#FWI(i)} & \text{if } \#FWI(i) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\#FWI(i)$ the cardinality of the set $FWI(i)$; $IR(i,j,t-1)$ is the interaction ratio of users (i,j) at time $t-1$; and $GR(j,t-1)$ is agent j's global reputation at time $t-1$.

Definition 10 (Follower interaction quality). Let $FWE(i)$ be the set of agents who who have interacted with agent i in an active way until moment t, i.e. agents in set FWE have liked, shared, or posted something coming from agent i. Agent i's **follower interaction quality** at time t, $Fwe(i,t)$, is computed as follows:

$$Fwe(i,t) = \begin{cases} \dfrac{\sum_{j \in FWE(i)} IR(i,j,t-1)GR(j,t-1)}{\#FWE(i)} & \text{if } \#FWE(i) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where $\#FWE(i)$ the cardinality of the set $FWE(i)$; $IR(i,j,t-1)$ is the interaction ratio of users (i,j) at time $t-1$; and $GR(j,t-1)$ is agent j's global reputation at time $t-1$.

Definition 11 (Global Interaction quality). Agents' Global Interaction quality at time t, $Glq(i,t)$. is measured as a linear combination of $Fwi(i,t)$ and $Fwe(i,t)$:

$$GIq(i,t) = \gamma Fwi(i,t) + (1-\gamma)Fwe(i,t) \tag{14}$$

In comparison with other baseline trust approaches, as is the case of the random walk trust based model in [34], the proposed approach measures on average the quality of the relationships of a user by considering not only the average of quality of the relationships of the agents that engage in a relationship with such user (via $FWI(i)$) but also the own user's average of quality relationship with other users (via $FWE(i)$).

4. DeciTrustNet Architecture and user cases

Based on the defined users' trust and reputation, this section presents a new trust based social network where agents' information propagation is used to provide recommendations to other 'similar' agents with high trust and global reputation levels. Depending on the agents' requirements the proposed framework covers two different **user cases**:

1. **The agent passive mode:** In this case, an agent i only receives information based on its profile similarity and the PairwiseTrust and Global Reputation of the k top users in his/her network of peers. This case could be applicable in a network that aims to increase healthy lifestyles of its members. Let suppose that all users in the network register and share their daily physical activity(ies), let say for example walking during 20 minutes, or doing yoga during 1 hour. Given that DeciTrustNET will identify the most similar agents to agent i, the k agents amongst them with higher reputation will be selected and their main activities will be presented to agent i ranked by order of influence. Agent i will be able to respond to the given recommendation by interacting with the users presented yo him/her by liking, sharing or posting in their publications, or rating them.

2. **The agent active mode:** In this case, an agent i will directly target his/her network of peers to seek for some information. For example, an agent i may ask about what sport is the best to practice every day to keep fit. The user's request may be answered by the top k agents in his/her network and the information will arrive to the user as coming from each peer individually or as a summary of his/her peers' opinions via an intermediate fusion process by considering their reputation degrees, as it is elaborated later.

The DeciTrustNET network architecture is composed of various elements that are shown in Figure 4.

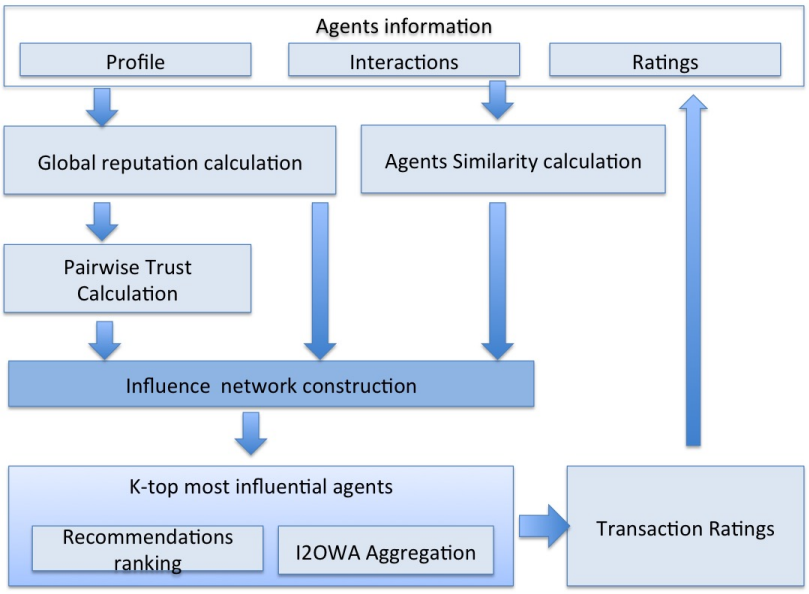

Figure 4: DeciTrustNET network architecture

4.1. Network construction

People tend to interact with others with similar profiles [37–39]. Below, a similarity based influence network in which every agent will be considered as a node of a directed graph is presented [20].

A directed graph is represented by an ordered duplet $G =< H, M >$ where H is a set of nodes connected by a set of directed edges M that interconnect the nodes in pairs with a set of weights attached to it, that is denoted $M = (m_{kl})_{HxH}$ and is referred to as the the adjacency matrix of graph G.

In our context, each node in H corresponds to an agent h , who is unequivocally characterised by its profile $P(h)$ and global reputation $GR(h)$, while $m(k, l)$ will represent the degree of influence that agent k has over agent l, and it will be calculated as a combination of the similarity of the agents' profiles $sim(P^k, P^l)$, and the Pairwise Trust between these agents $PT(k, l, t)$, which considers the Global reputation of the destination agent, $GR(l, t)$. For simplicity, the parameter t, which represents the present time/state of the network, will be dropped from the notation.

When calculating the influence that agent k has over agent l there could be three possible scenarios:

- The two agents already interacted and continue to be connected: their existent PT value can be updated as per expression (1).
- It is the first time agents k and l connect and interact: as discussed previously, $PT(k, l) = GR(l)$.
- Agents' interaction ends up with the disconnection between them, i.e. $m(k, l) = 0$, which can happen if one of the following conditions occurs:
  1. the agents'$PT(k, l)$ value is negative;
  2. the target agent $GR(l)$ value is very low;
  3. the similarity between both agents is very low.

With all these constraints in mind, an influence network is defined as follows:

Definition 12 (Directed influence network). A directed influence network is a directed graph $G_S =< H, M >$ where m_{kl} represents the influence that agent k has over agent l is calculated as follows

$$m_{kl} = \begin{cases} GR(l) \cdot T(sim(kl), PT(k, l)) & \text{if } PT(k, l) > 0) \\ 0 & Otherwhise \end{cases} \tag{15}$$

subject to the constraints $m_{kl} \in [0, 1] \forall k, l, \wedge \sum_k m_{kl} = 1 \ \forall l$; and T is a t-norm operator, i.e. function $T : [0, 1] \times [0, 1] \to [0, 1]$ satisfying the properties:

- Commutativity: $T(a, b) = T(b, a)$
- Monotonicity: $T(a, b) \leq T(c, d)$ if $a \leq c$ and $b \leq d$
- Associativity: $T(a, T(b, c)) = T(T(a, b), c)$
- The number 1 acts as identity element: $T(a, 1) = a$

The Hamacher product [40]

$$T(a, b) = \begin{cases} 0 & \text{if } a = b = 0 \\ \frac{ab}{a+b-ab} & \text{otherwise} \end{cases} \tag{16}$$

outputs low values when one of the inputs is a low value, and therefore it fits the third scenario described above for disconnecting two agents in an influence network.

4.2. Profile similarity assessment

A similarity measure between two objects is usually quantified via the use of a real-valued function measuring the distance between feature vectors characterising the two object [41]. What similarity measure to use as well as which features to take into consideration depends on whether the data type involved in the assessment is binary, qualitative or quantitative [42].

For binary type data, among the suitable similarity measures available, the Jacquard index is widely used [43] . For quantitative type data, an analysis of the impact of different distance measures in the consensus process carried out in [44] showed that the Manhattan and the Euclidean distances facilitated the increase of consensus when the number of agents increased. The Gower's General Similarity Coefficient [45] is one of the most widely used for the three data types mentioned, and an example of it use in assessing the similarity between elderly profiles can be found in [42], which is also used in this contribution to compute the similarity between agents' profiles.

4.3. Agents information fusion

As depicted in Fig. 4, in the DeciTrustNet framework an agent i will receive information coming from the k-top most trustworthy and liked-minded agents in his vicinity. This subsection explains the procedure to fuse an agent's vicinity information.

Dong et al.'s have recently reported in [46] results that demonstrates that Yager's Induced Ordering Weighting Averaging (IOWA) operator [47] is more effective in dealing with malicious behaviour or manipulation in multi person decision making scenarios than the WA operator used in opinion dynamics models [38, 48]. Motivated by this result, in this paper k-top most trustworthy and liked-minded agents's information is summarized via the application of an IOWA operator guided by the agents' influence degrees as described below, and that we call the Influence IOWA (I2OWA) operator.

Definition 13. "An IOWA operator [47] of dimension k is a function $\Phi_W \colon (\mathbb{R} \times \mathbb{R})^k \to \mathbb{R}$, with expression

$$\Phi_W \left(\langle u_1, p_1 \rangle, \ldots, \langle u_k, p_k \rangle \right) = \sum_{i=1}^{k} w_i \cdot p_{\sigma(i)},$$

where $W = (w_1, \ldots, w_k)$ is the vector of weights that represent the agents' influence degrees subject to the constraints $w_i \in [0, 1]$ and $\Sigma_i w_i = 1$, and $\sigma \colon \{1, \ldots, k\} \to \{1, \ldots, k\}$ is the following permutation of the set of inducing values $\{u^1, \ldots, u^k\}$: $u_{\sigma(i)} \geq u_{\sigma(i+1)}$, $\forall i = 1, \ldots, k-1$."

The proposed I2OWA operator is defined as follows:

Definition 14 (I2OWA operator). Let $E_h = \{e_1, \ldots, e_k\}$ be the set of k most trustworthy similar agents connected to a given agent e_h, who provide their opinions, $\{B_1, \ldots, B_k\}$, about a given topic. The Influence IOWA (I2OWA) operator of dimension k, Φ_W^{KD}, is the IOWA operator with the set of influence values, $M = \{m_{1h}, \ldots, m_{kh}\}$, associated with the set agents E_h.

DecitrusNET allocates different importance degrees, $\{w_1, \ldots, w_k\}$, to the different agents based on their influence computed using expression (15), which is done by implementing Yager's approach [47] :

$$w_i = Q\left(\frac{T(i)}{T(k)}\right) - Q\left(\frac{T(i-1)}{T(k)}\right)$$

where Q is the membership function of the linguistic quantifier to be used to implement the (soft) majority concept, $T(i) = \sum_{l=1}^{i} m_{\sigma(l)h}$, and σ is the permutation that orders the induce values from largest to lowest.

5. Analysis of performance

A validation study of DeciTrustNET, as a framework to investigate the influence of the different agents and the evolution of their trust and Global Reputation, has been carried out using computer simulations in R. These simulations aim at: (i) evaluating DeciTrustNET regarding its dealing of strategies that attempt to subvert its effectiveness, including deceptive and malicious users; (ii) comparing DeciTrustNET with alternative trust and reputation models.

5.1. Experimental set up

With the objective of simulating a multiuser graph scenario that supports a cold-start system, we synthetically generate a network of 500 agents who interact for 1000 rounds. The simulation begins with a cold-start by assigning a default global reputation score of zero to each user in the community. To ensure small world properties, an interaction graph representing the social connections of the agents in the network is generated, i.e. the given likes, post and shares are also been generated, following the WattsStrogat model [49] . The agents' profile are randomly set to guarantee a diversity of profiles and a validation of the interaction graph has been performed in order to ensure the link distribution follows power-law and has the adequate clustering coefficient. In the simulations two different types of users have been considered: 1. *Malicious users* who present an intentionally malicious behaviour in the majority of their interactions; 2. *Average users* who sometimes provide an irrelevant response but most of the time their behaviours are fair.

In other to assess the effectiveness of trust ratings the used benchmark measure is the *relative precision @ n*, $Prec_n$, that considers the quality of the top n responses for an user's query q [13]:

$$Prec_n = \frac{\#(R^+ \cap R_n)}{min(\#R_n, n)} \tag{17}$$

R^+ is the set of relevant users for a query q throughout the entire space of users and R_n is the set n top-ranked candidate users (by trust value). Note that the traditional precision and recall measures may not be effective in this context since malicious users may overwhelm a user with several poor quality responses [13].

5.2. Experimental results

Firstly, the quality of the proposed approach is evaluated in various scenarios by increasing the proportion of malicious users from 10% to 80%. When the percentage of malicious users is over 80% the precision drops dramatically, and when the proportion of malicious users is total (100%) trust ratings make no sense and the overall precision is zero. In order to keep the simulation as realistic as possible, malicious users providing irrelevant response is assigned probability value of 1 while the non malicious users provide a corrupted answer with a probability of 0.05. Simulation results of precision value against % of malicious users, depicted in Figure 5, show that the prevision value decreases as the % of malicious user increases; however, it is noticeable that the rate of decrease of precision is relatively low and precision values remains mostly high and stable even when the percentage of malicious users increases over 50%. This stability in the precision can be mainly attributed to the fact that the quality of the relationships and the users feedback are accounted for when calculation the global reputation. Therefore malicious users are immediately detected and set aside.

In the following a comparative analysis between the proposed approach and the most recent trust based frameworks discussed in section 2, SocialTrust [13] and PCR [14], is presented with both the presence and the absence of trust. In this last case a user randomly selects another user among his/her connections. The results of this study, depicted in Figure 6, show that when no trust is considered the precision results decrease at a high rate when the rate of malicious

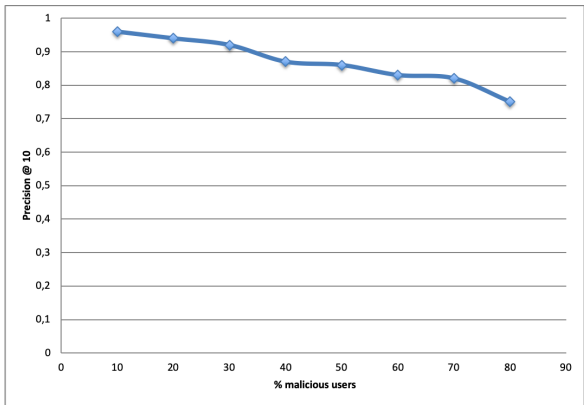

Figure 5: DecitrustNet precision with increasing malicious user

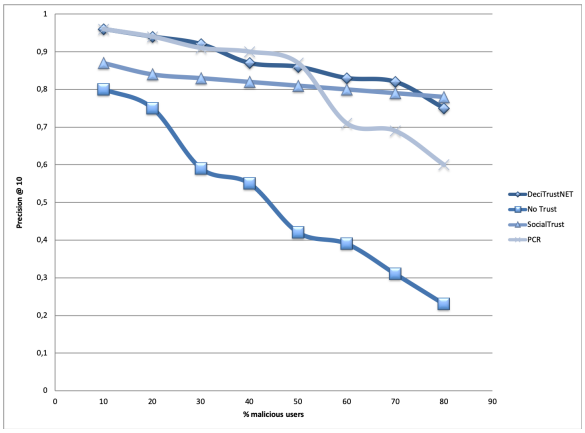

Figure 6: Malicious users scenario with and without trust

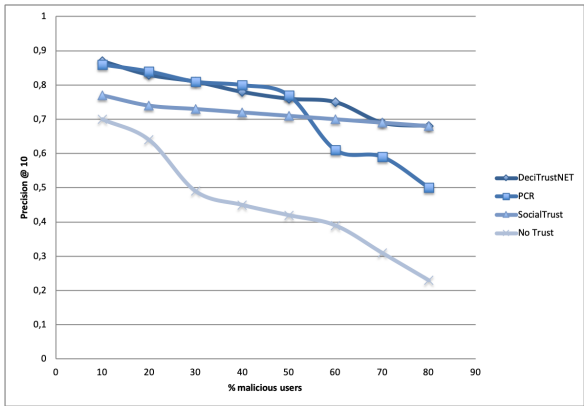

Figure 7: Malicious association of users

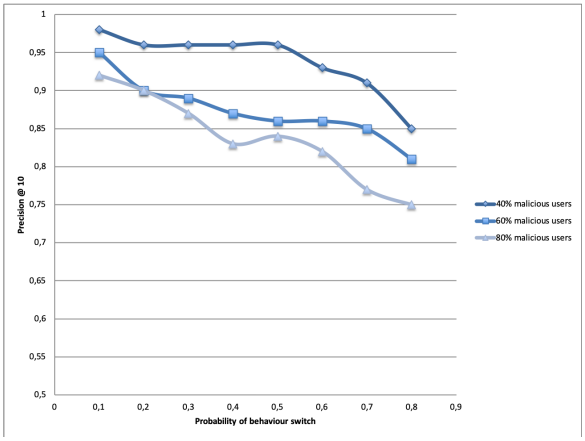

Figure 8: Switch in users behaviour

users is above 30%; on the contrary, when trust is considered, the precision remains high even when more than half of the users are considered malicious users. DeciTrustNET outperforms the state of the art approaches with precision values above 75% even when there is a percentage of malicious users as high as 80%.

Malicious association of users: In the previous experiments malicious users are randomly selected. In the following, a scenario is simulated with various users deliberately associating together in cliques to overpower the proposed trust framework [13]. The objective of this simulation is to test the system even when the users relationships are impacted. To do so, firstly a node is randomly selected as malicious, then up to three more nodes in his network are set as well to be malicious. This process is repeated until the desired percentage of malicious users is obtained. The results of this experiment, presented in Figure 7, show how DeciTrustNET, even in the presence of malicious association of users, is able to keep a high level of precision. This success is mainly due to the consideration of the users' behaviour over time as well as the similarity between the users profiles.

Sudden switch in agent behaviours: In the previous experiments the agents stay either malicious or fair during the whole simulation. In the next experiment, a scenario is simulated in which users may change their behaviour during the simulations with an increasing probability. The results of this experiment, depicted in Figure 8, show how DeciTrustNET precision remains mostly stable until the probability of an agent to change its behaviour reaches 0.6, above which it decreases. This stability can be attributed to the adaptation to change in behaviour evaluation within DeciTrustNET, which add evidence to the usefulness of this measure in allowing a fast detection of variations in users' behaviour.

6. Conclusion

In this contribution we have presented DeciTrustNET, a novel framework that allows robust trust and reputation based communication between agents in a network even in the presence of malicious and new users. The main contributions of the proposed framework with respect to the state of the art are listed bellow:

- **Double supervised personalised feedback from other members of the network:** DeciTrustNET allows users to rate their interactions in such a way that the provided rating will affect both the rating and rated users' reputation, which can be exploited to avoid abuses when rating.
- **Distinguishing between user's global reputation and pairwise trust:** In the proposed framework we proposed two measures for each user. The first one is the Global reputation based in all the user's interactions, feedback given to other users and feedback received from other users, the user's behaviour and the quality of his/her relationships. The second one, the pairwise trust is a measure that indicates the level of trust or confidence between two users based on their previous interactions [15, 20].
- **Exploiting user relationships and their position in the network to asses users' Reputation:** This is based on the premise that an individual with trustworthy friends is more likely to be trustworthy [13, 21, 22].
- **Tracking user behaviour over time:** DeciTrustNET incorporates the evolution and trajectory of user's trust rating in order to motive users' engagement in long term good behaviours and to penalise those whose behaviour experiences sudden changes.

As future research work, we are interested in the context aware customization of the proposed system to different scenarios. More concretely, we believe that DeciTrustNET may be specially useful in three specific scenarios: Consensus achievement in group decision making processes such as the ones carried out in e-democracy; e-health platforms to provide recommendations on how to keep a healthy lifestyle; and in recommender systems for e-commerce and e-marketing. Moreover, in order to improve the feedback provided by the users we will explore the possibility of using multigranular linguistic information. Recent studies in this direction have been presented in [50–52]

7. Acknowledgments

The authors would like to acknowledge the financial support from the EU project H2020-MSCA-IF-2016-DeciTrustNET-746398.

References

[1] T. Daugherty, E. Hoffman, ewom and the importance of capturing consumer attention within social media, Journal of Marketing Communications 20 (2014) 82–102.

[2] N. Huete-Alcocer, A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior, Social Networks 8 (2017).

[3] F. Yang, Effects of restaurant satisfaction and knowledge sharing motivation on ewom intentions: the moderating role of technology acceptance factors., Journal of Hospitality and Tourism Research 41 (2017) 93–127.

[4] M. Sotiriadis, C. Van-Zyl, Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists, Electron. Commer. Res 13 (2013) 103–124.

[5] G. Kambourakis, Anonymity and closely related terms in the cyberspace: An analysis by example, Journal of Information Security and Applications 19 (2014) 2 – 17.

[6] M. Fishbein, I. Ajzen, Beliefs, Attitude, Intention and Behavior: An Introduction to Theory and Research, Addison-Wesley, 1975.

[7] A. Josang, R. Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2007) 618 – 644.

[8] Z. Malik, A. Bouguettaya, Rateweb: Reputation assessment for trust establishment among web services, The VLDB Journal 18 (2009) 885–911.

[9] F. Hendrikx, K. Bubendorfer, R. Chard, Reputation systems: A survey and taxonomy, Journal of Parallel and Distributed Computing 75 (2015) 184 – 197.

[10] J. Bobadilla, F. Ortega, A. Hernando, J. Bernal, A collaborative filtering approach to mitigate the new user cold start problem, Knowledge-Based Systems 26 (2012) 225 – 238.

[11] R. Guha, R. Kumar, P. Raghavan, A. Tomkins, Propagation of trust and distrust, in: Proceedings of the 13th International Conference on World Wide Web, pp. 403–412.

[12] S. D. Kamvar, M. T. Schlosser, H. Garcia-Molina, The eigentrust algorithm for reputation management in p2p networks, in: Proceedings of the 12th International Conference on World Wide Web, WWW '03, ACM, 2003, pp. 640–651.

[13] J. Caverlee, L. Liu, S. Webb, The socialtrust framework for trusted social information management: Architecture and algorithms, Information Sciences 180 (2010) 95 – 112.

[14] S. R. Yan, X. L. Zheng, Y. Wang, W. W. Song, W. Y. Zhang, A graph-based comprehensive reputation model: Exploiting the social context of opinions to enhance trust in social commerce, Information Sciences 318 (2015) 51 – 72.

[15] R. Ureña, F. Chiclana, E. Herrera-Viedma, A new influence based network for opinion propagation in social network based scenarios, Procedia Computer Science 139 (2018) 329 – 337.

[16] J. Wu, F. Chiclana, E. Herrera-Viedma, Trust based consensus model for social network in an incomplete linguistic information context, Applied Soft Computing 35 (2015) 827 – 839.

[17] J. Wu, L. Dai, F. Chiclana, H. Fujita, E. Herrera-Viedma, A minimum adjustment cost feedback mechanism based consensus model for group decision making under social network with distributed linguistic trust, Information Fusion 41 (2018) 232 – 242.

[18] R. Ureña, G. Kou, Y. Dong, F. Chiclana, E. Herrera-Viedma, A review on trust propagation and opinion dynamics in social networks and group decision making frameworks, Information Sciences 478 (2019) 461 – 475.

[19] J. Scott, Social network analysis: a handbook., SAGE Publications, 2000.

[20] R. Ureña, F. Chiclana, G. Melançon, E. Herrera-Viedma, A social network based approach for consensus achievement in multiperson decision making, Information Fusion 47 (2019) 72 – 87.

[21] P. Bonacich., Power and centrality: A family of measures, American Journal of Sociology 92 (1987) 1170–1182.

[22] P. Bonacich, P. Lloyd, Eigenvector-like measures of centrality for asymmetric relations, Social Networks 23 (2001) 191 – 201.

[23] Q. Zha, Y. Dong, H. Zhang, F. Chiclana, E. Herrera-Viedma, A personalized feedback mechanism based on bounded confidence learning to support consensus reaching in group decision making, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2019) 1–11.

[24] H. Zhang, C. Li, Y. Liu, Y. Dong, Modelling personalized individual semantics and consensus in comparative linguistic expression preference relations with self-confidence: An optimization-based approach, IEEE Transactions on Fuzzy Systems (2019) 1–1.

[25] A. Sallaberry, F. Zaidi, G. Melançon, Model for generating artificial social networks having community structures with small-world and scale-free properties, Social Network Analysis and Mining 3 (2013) 597–609.

[26] L. C. Freeman, Centrality in social networks: Conceptual clarification, Social Networks 1 (1979) 215–239.

[27] P. Resnick, R. Zeckhauser, Trust among strangers in internet transactions: empirical analysis of ebays reputation system, The Economics of the Internet and ECommerce 11 (2002).

[28] J. Schneider, G. Kortuem, J. Jager, S. Fickas, Z. Segall, Disseminating trust information in wearable communities, Personal Ubiquitous Comput. 4 (2000) 245–248.

[29] K. Hoffman, D. Zage, C. Nita-Rotaru, A survey of attack and defense techniques for reputation systems, ACM Comput. Surv. 42 (2009) 1:1–1:31.

[30] J. Sabater, C. Sierra, Social regret, a reputation model based on social relations, SIGecom Exch. 3 (2001).

[31] J. Wu, R. Xiong, F. Chiclana, Uninorm trust propagation and aggregation methods for group decision making in social network with four tuple information, Knowledge-Based Systems 96 (2016) 29 – 39.

[32] J. Wu, L. Dai, F. Chiclana, H. Fujita, E. Herrera-Viedma, A new consensus model for social network group decision making based on a minimum adjustment feedback mechanism and distributed linguistic trust, Information Fusion 41 (2018) 232 – 242.

[33] C. Tian, B. Yang, R2trust, a reputation and risk based trust management framework for large-scale, fully decentralized overlay networks, Future Generation Computer Systems 27 (2011) 1135 – 1141.

[34] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web., Technical Report 1999-66, Stanford InfoLab, 1999. Previous number = SIDL-WP-1999-0120.

[35] C. Dellarocas, The digitization of word of mouth: Promise and challenges of online feedback mechanisms, Manage. Sci. 49 (2003) 1407–1424.

[36] Twitter, `https://developer.twitter.com/`, 2020. Online Accessed: 2020-02-04.

[37] Y. Dong, Q. Zha, H. Zhang, G. Kou, H. Fujita, F. Chiclana, E. Herrera-Viedma, Consensus reaching in social network group decision making: Research paradigms and challenges, Knowledge-Based Systems 162 (2018) 3 – 13. Special Issue on intelligent decision-making and consensus under uncertainty in inconsistent and dynamic environments.

[38] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence, models, analysis and simulation, J. Artif. Soc. Social Simul 5 (2002) 1–33.

[39] Y. Dong, Z. Ding, F. Chiclana, E. Herrera-Viedma, Dynamics of public opinions in an online and offline social network, IEEE Transactions on Big Data (2017) 1–11.

[40] E. Klement, R. Mesiar, E. Pap, Triangular Norms, Dordrecht: Kluwer, 2000.

[41] B. Everitt, S. Landau, M. Leesme, I. Stah, Cluster Analysis, Wiley, 2011.

[42] A. Bilbao, A. Almeida, D. L. de Ipiña, Promotion of active ageing combining sensor and social network data, Journal of Biomedical Informatics 64 (2016) 108 – 115.

[43] D. Lewis, V. Janeja, An empirical evaluation of similarity coefficients for binary valued data, Int J Data Wareh Min 7 (2011) 44–66.

[44] F. Chiclana, J. T. Garcia, M. del Moral, E. Herrera-Viedma, A statistical comparative study of different similarity measures of consensus in group decision making, Information Sciences 221 (2013) 110 – 123.

[45] J. C. Gower, A general coefficient of similarity and some of its properties, Biometrics 27 (1971) 857–871.

[46] Y. Dong, Y. Liu, H. Liang, F. Chiclana, E. Herrera-Viedma, Strategic weight manipulation in multiple attribute decision making, Omega 75 (2018) 154 – 164.

[47] R. R. Yager, Induced aggregation operators, Fuzzy Sets and Systems 137 (2003) 59–69.

[48] Y. Dong, Z. Ding, L. Martinez, F. Herrera, Managing consensus based on leadership in opinion dynamics, Information Sciences 397-398 (2017) 187 – 205.

[49] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world'networks, Nature 393 (1998) 440–442.

[50] Z. Zhang, C. Guo, L. Martnez, Managing multigranular linguistic distribution assessments in large-scale multiattribute group decision making, IEEE Transactions on Systems, Man, and Cybernetics: Systems 47 (2017) 3063–3076.

[51] Z. Zhang, W. Yu, L. Martinez, Y. Gao, Managing multigranular unbalanced hesitant fuzzy linguistic information in multiattribute large-scale group decision making: A linguistic distribution-based approach, IEEE Transactions on Fuzzy Systems (2019) 1–1.

[52] W. Yu, Z. Zhang, Q. Zhong, Consensus reaching for magdm with multi-granular hesitant fuzzy linguistic term sets: a minimum adjustment-based approach, Annals of Operations Research (2019).