Social Network Analysis and Mining (2024) 14:75
https://doi.org/10.1007/s13278-024-01235-8

ORIGINAL ARTICLE

Agent-based simulation of fake news dissemination: the role of trust assessment and big five personality traits on news spreading

Radifan Fitrach Muhammad[1] · Shoji Kasahara[1]

Received: 10 October 2023 / Accepted: 27 February 2024

Abstract

The recent development of Social Networking Services (SNS) has changed the way of information sharing and interchanging. With countless amounts of information broadcast daily, many people use it every time, even during a crisis and disaster situation. However, the trustworthiness of information has become one of the issues on SNS. In this paper, we propose a trust model consisting of identity-based, behavior-based, relation-based, feedback factors, and information-based trust, in which the Big-Five personality traits are also considered. We conducted an agent-based modeling simulation for the proposed trust model, investigating users' behavior according to the Big-Five personality traits and several users' aspects: knowledge level and psychopathy. The experiment is based on online surveys and related works representing social network users' behavior. We compare the overall trust and trustworthiness in the numerical results, validating our proposed trust model. Moreover, we systematically compare the occurrence of fake news under conditions where the initial news is either a truthful or fabricated fake news. Numerical results also show that overall trust is sensitive to information-based trust, whereas it is not affected by behavior-based trust. Furthermore, openness, conscientiousness, and extroversion were correlated with overall trust, while the effect of agreeableness and neuroticism on overall trust were insignificant.

Keywords Agent-based modeling · Big-five personality traits · Disaster-related news · Fake news · Social networking service · Trust

1 Introduction

Social networking services (SNS) are platforms for people to share and exchange information related to trends and what current events worldwide. Among its ease of use, the information's credibility supported by the emerging fake news, has become a significant problem. Knowing this concern, people may trust unreliable news, particularly during crises. Therefore, we attempt to understand how the SNS users accept the trustworthiness of this type of information by considering trust, personality factors, and agent-based modeling.

According to Cigi-Ipsos: (2019), 86% of SNS users believed fake news at least once in 2019. In Watson (2022), critical type information such as COVID-19-related content becomes the number one topic, which is often containing misleading and false information. It is also reported in Neubaum et al. (2014) that most users prefer to refrain from contributing to the information and are likely to seek the currently disseminated news during disaster events. These circumstances can be a serious problem for rescue teams if the information distributed immediately after the occurrence of a disaster is fake news. The existence of fake news may prevent the true news from being distributed and, confusing anyone concerned about the disaster. In order to prevent this chaotic situation and identify the disaster situation, trustworthiness calculation have been widely used in several decision-making tools (Gao et al. 2019). The main idea is to quantify the trustworthiness of the news by considering the complexity and vagueness of social networks.

As in many disaster and social network cases, numerous research fields, including social computing, cognitive sciences, and data science, have been applied to elaborate

✉ Radifan Fitrach Muhammad
muhammad.radifan_fitrach.mo1@is.naist.jp

Shoji Kasahara
kasahara@is.naist.jp

[1] Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, Takayama-cho 8916-5, Ikoma-shi, Nara 630-0192, Japan

Published online: 30 March 2024

on fake news and its dissemination on social networks. Burbach et al. (2019) aimed to implement personality traits gathered from an online questionnaire of individuals to understand how each behaves and is connected in social networks. They also gathered information on the user behavior of the network using an online questionnaire. These questionnaire responses determined how individuals' personality traits created different responses while interacting with information, such as, likes, shares, and comments. These personality traits will be the key to identifying who is responsible for fake news creation and dissemination through social networking services. Furthermore, these personality traits were analyzed using agent-based modeling to show how messages are exchanged among users.

On social networks, if widely spread news is fake, most users who receive the news will be deceived. However, investigating this incident in the real world is time-consuming and not practical. This problem can be manifested by modeling how users trust and process news. In this paper, we considered how SNS users trust news from other users. Based on the trust model in Gao et al. (2019), we consider a new trust evaluation, information-based trust. The information-based trust comprises of semantic and surface features. The former represents the content accuracy, while the latter describes the post's exterior point of view, such as photo inclusion, logical degree, and popularity of the post Lucassen and Schraagen (2011). Here, we define information-based trust by three characteristics of the content, the included photo, logic, and its popularity. From this extension, the post also affects the user's judgment of the truth or falsehood regarding distributed information.

We also considered an agent-based simulation model that integrates personality traits and trust in information dissemination in disaster situations. Information dissemination is affected not only by network diffusion and personality traits as introduced in Burbach et al. (2019), but also by trust features, including information-based trust. We introduce an agent-based simulation to model a state called the trusting process, which is performed before the dissemination process (Burbach et al. 2018). This model enabled us to explore how information dissemination works based on individuals' trust and personalities. We developed a trust evaluation system by expanding the existing model with information-based trust by applying an agent-based model using NetLogo simulator. The dissemination process and user behavior were investigated using Big-Five personality traits.

The paper structure is organized as follows. We briefly present the related work in Sect. 2. The explanation of SNS is presented in Sect. 3.The big-five personality traits in SNS usage is presented in Sect. 4, and the proposed model is presented in Sect. 5. The experiments and questionnaire are shown in Sect. 6. While the numerical examples and discussion are discussed in Sect. 7, followed by the conclusion of the paper in Sect. 8.

2 Related work

In this section, we show related work on three different aspects, fake news on SNS, trust evaluation, and the big five personality traits. We then discuss a practical way to measure user trustworthiness and to incorporate both cognitive and social features into the system model.

Fake News is intentionally or unintentionally misleading information presented by one party to the other (Zhou and Zafarani 2020; Allcott and Brennan 2017). Users can share fake information in two ways; through verbal and indirect communication. In verbal communication, identifying a person whose lying is a skill can be learned through experience (Jackson et al. 2006; Ryu et al. 2018). However, this mechanism may not apply in the case of social networking services. In SNS, trust is a primary issue and one of the main criteria for users to accept or discard information.

Trust is one of the factors affected by human cognition, which can be achieved by either short or long-term interaction within a community (Cheng et al. 2017; Rotenberg 2018). In a network, interaction builds cooperation, and each user benefit from the group. This interaction will build perceptions, and then form a trusting decision.

In social networking services, trust is gained from perceived information quality, perceived system reliability, and perceived trust (Mohd Suki 2014; Delone and McLean 1992). This implies that an SNS user will trust a rumor according to how they think about the information quality, system reliability, and rumors' trustworthiness. Since relationships in SNS are regarded as social capital established among society, forming groups follows multidimensional factors consisting of many aspects such as relations, family, and friends (Rotenberg 2018; Erickson 2011). This means that when users gain trustworthiness, they also form a multidimensional way of thinking towards other users. However, due to the complexity of the problems, such as knowledge, point of interest, and personality differences, it might be difficult for users to examine the reliability of online news, especially in the middle of a crisis.

During disaster phases, people are likely to seek the latest news rather than contributing to the information. Only a few groups of people will share the information they have received (Neubaum et al. 2014). The ability to widely share critical information for coordination is the main advantage of using social networks during disasters (Takahashi et al. 2015). However, most information shared by the social network user is regarded as a rumor, a piece of unproven information that can later be corroborated as trusted or fake news (Liu et al. 2014). This unproven information quickly spreads because users will likely fail to collect the desired important facts (Comfort et al. 2004; Liu et al. 2014). This situation splits users into two groups, the trusting group and the distrusting group, resulting from their personal information processing. This polarization among users may lead malicious users to spread rumors, which lead to the existence of fake news and misinformation. Fake news is false fabricated information or a statement in a report on social media (Liu et al. 2014; Lazer et al. 2018). This news issued by an unreliable person is called a rumor, which is classified as a fact or a false rumor (Liu et al. 2014). This false information aims to deceive people, raising trust issues concerning the accuracy and credibility of information. The significance of trust in information dissemination is discussed in Cheng et al. (2017). In Cheng et al. (2017), the authors proposed trust classification, which includes institution-based, knowledge-based, calculative-based, personality-based, and cognition-based in mass communication schemes. The authors pointed out that the research focused on social networking sites, specifically in mass, group, and interpersonal communication. Burbach et al. (2019) pointed out that the personality of an SNS user affects the information dissemination.

Gao et al. (2019) proposed Info-Trust, a trustworthiness-evaluation scheme with multi-criteria trust factors. Trust is classified into four types: identity-based trust, behavior-based trust, relation-based trust, and feedback-based trust. Info-Trust provides an assessment of the trustworthiness of the information sources. However, the model does not consider the importance of the information for trust features, as pointed out in Lucassen and Schraagen (2011). In this paper, we extend the model of Gao et al. (2019) to a model in which users' perception of the information source is considered.

Users' communication styles also depend on their perceptions, preferences, and behaviors (Hawkins et al. 1980). This means that personalities also influence how users behave in social networking services. Burbach et al. (2019) proposed the Big-Five personality traits, dark triad, and regulatory self-efficacy to explain how users react differently towards fake news. The Big-Five personality traits model describes an individual with five characteristics. The model was first developed (Costa and McCrae 1999), consists of five personality traits: openness, conscientiousness, agreeableness, and neuroticism. Each personality trait is a spectrum that differentiates human behavior (McCrae and Costa 1997; Costa and McCrae 1999).

Numerous researchers approach this problem by presenting new fact-checking algorithms, which are based on social properties data such as popularity and link structure, and context property such as the content of the tweet and meta-information (Gao et al. 2019; Esteves et al. 2018). However, there is also a need to consider the personality and users' mental condition during unpredictable disaster crises, which can be very suitable by applying an agent-based modeling method (Burbach et al. 2019; Preston et al. 2021; Rand et al. 2015).

Our contribution in this paper is to add information-based trust into the trust model and the agent-based modeling implementation that includes both trust and personality traits of fake news emerging. Adding information-based trust into the trust model allowed us to examine the credibility of the news. Considering the news features and their types, we can evaluate the change in trend from regular news to fake news which is originally introduced in this paper. This news-changing behavior was originally introduced in this paper using the agent-based modeling method.

3 Social networking service

In this paper, we focus on Twitter as a microblogging platform provided by Twitter, Inc., which is a social networking service that allows users to interact with each member by sending and receiving short posts called tweets inside their system (Murthy 2018). On Twitter, registered users can create and interact with tweets through several actions, such as retweets and likes, which help disseminate news. The action of tweets creation, retweets, and likes have to be modeled in this research to produce accurate system behavioral actions and simulation results.

Tweeting is the action of creating a post and sending it through the global Twitter system Murthy (2018). Users linked to the tweet starter receive the tweet and react according to their personality factors. The reaction would be retweet, like, or comment at the tweet. While retweeting is spreading the received tweets to other linked users, likes and comments support the tweet starter to increase the validity of the information. However, comments provides additional information to agree or disagree with by providing some information, deduction, and evidence.

We assume that a tweet is characterized by five attributes: the number of pictures in the tweet, clarification, comments, shares, and likes. In terms of the number of pictures in a tweet, the user can attach a maximum of four pictures to a single tweet. This number is filled in by the attributes of a tweet. Comments describe how many comments are attached by other users in the tweet, while shares and likes explain how many shares and likes the tweet has received.

4 Big-five personality traits

The information dissemination behavior of users depends on the users themselves. In Burbach et al. (2019), the Big-Five personality traits were used to explain how personalities affect users' behavior when using SNS. This finding also implicitly shows that personality and trust models should

Table 1 Personality Notations

Notation	Description
U	Set of users
N	Number of users
O_i	Openness of user i
E_i	Extroversion of user i
CS_i	Conscientiousness of user i
A_i	Agreeableness of user i
NR_i	Neuroticism of user i
Kn_i	Knowledgeability of user i
Ps_i	Psychopathy of user i

be considered when explaining information dissemination. Moreover, recent trends show many applications of user classifications and clustering based on interests, including Twitter (Twitter: 2023). Applying personality traits to a trust evaluation system enables information propagation effectiveness and combating fake news. In this paper, we assume that each user has his/her own personality that differentiates their ways of receiving information from and sending it to others.

In terms of the characterization of the personality of a user, the Big-Five personality traits were developed by Allport and Odbert (1936), which identify individual differences in choosing the right words in Webster's Unabridged Dictionary. The Big-five personality traits comprise a taxonomy of psychology consisting of openness, extroversion, conscientiousness, agreeableness, and neuroticism.

In this paper, we assume that user action depends on their personality, characterized by the Big-Five personality traits[1] We define $U(= \{1, 2, \dots, N\})$ as the set of users joining an SNS system. For user i ($\in U$), let O_i, E_i, CS_i, A_i, and NR_i denote user i's openness, extroversion, conscientiousness, agreeableness, and neuroticism, respectively. These variables were within the interval of [0,1]. For each variable, the higher the value, the greater is the corresponding characteristic. The parameters are listed in Table 1. In the following subsections, we describe the details of personality traits.

4.1 Openness

Openness personality represents how a user opens towards any new source of information (Costa and McCrae 1999). In this research, the openness users have the characteristics of being creative and tend to find a piece of new information (DeYoung 2015). This tendency gives the openness users a faster reaction time for finding and receiving news than those with low openness traits.

4.2 Extroversion

Extroversion is considered an essential factor that represent the extent nof information spreads (Seidman 2020). Extroversion users are willing to socialize, gather new information, and share positive emotions towards other users (DeYoung 2015). Previous researchers found that users with high extroversion are likely to support any information sent by their related persons (Seidman 2020; Correa et al. 2010). Extroverted users are also likely to have more SNS friends (Wang et al. 2012; Ross et al. 2009). In the SNS environment, if users followed each other, there is a high possibility that the information will be shared and provide high relation-based trust, as described in the following Sect. 5.3.

4.3 Conscientiousness

While receiving new information, users can derive information either rationally or emotionally. They form a rational decision by paying attention to the details of the information and, being cautious not to get trapped by fake information that may spread widely in the social networks. This cautious act is a typical way of information processing of users with conscientiousness personality traits. Conscientiousness personality traits explain the user's tendency to follow rules and refrain from spreading rumors until they have the credibility to be trusted (DeYoung 2015).

4.4 Agreeableness

During a disaster event, the manifestation of empathy and sympathy towards victims influences our decision to trust or distrust the disaster information. This emphatic response represents cooperation and altruism among users who have agreeableness. Agreeableness is typically related to cooperative and altruistic behavior (DeYoung 2015). These two aspects affect dissemination behavior. Agreeableness users tend to spread the news if they believe it and try to share the news with a user with many followers in the network (Buchanan 2021; Buchanan and Kempley 2021).

4.5 Neuroticism

During a disaster event, being negative and anxious about the upcoming event might become a problem. Along with anger, feelings of frustration, and depression, this may lead to the creation and dissemination of fake news over

[1] In this research, we are not specifically focused on how each personality affects each other. However, according to Klimstra et al., the results found in studies by Allemand et al., 2007, Allemand et al., 2008, and Soto and John, 2012 have reported inconsistent results, since the concept itself consists of five different independent traits. This leads us to assume that personality is an independent factor that further affects users' interactions on social networking services. Klimstra et al. (2013)

the network. This mental condition is characterized by the neuroticism trait. People with this personality trait are likely to be closed off, fearful, moody, and jealous of other users (DeYoung 2015). In SNS networks, neuroticism is related to sharing more information, presenting falsehoods, and pursuing personal objectives (Seidman 2020; Twomey and O' Reilly 2017). The desire to be at the center of attention leads this type of user to create fake news for others.

4.6 Additional personality characteristics: knowledgeability and psychopathy

In addition to the Big-Five personality traits, we further consider two personality features: knowledgeability and psychopathy.

In general, knowledgeable users carefully consider the trustworthiness of disseminated news (Landrum et al. 2015). Prior studies show that the lack of careful thinking and decision-making is associated with insufficient and/or inaccurate prior knowledge (Pennycook and Rand 2021). We define $Kn_i \in [0, 1]$ as the knowledgeability of user i. A large Kn_i implies user i has high knowledgeability.

A situation similar to the creation of fake news also occurs if users have a high amount of extroversion, which makes them communicate well with other users. In this paper, we refer to these personalities as psychopathic. Psychopathic users share fake news by distorting the contents and/or adding new information from previous news that might be important for the disaster victims. This type of user shares fake news knowing that the material is incorrect, instead of sending false information accidentally (Buchanan and Kempley 2021). We define $Ps_i \in [0, 1]$ as the psychopathy of user i. A high Ps_i implies that users i has high psychopathy.

5 Trust model

Trust is a fundamental cognitive factor in believing and having faith in an object. This belief involves one party's interest with reliance on the other (Rousseau et al. 1998; Özer and Zheng 2017). Trust can develop over time as its value changes based on interactions between two or more people over time. Trustworthiness is considered a display of behavior by a party that acts in a trustworthy manner (Özer and Zheng 2017). In this paper, trust is considered as a property of users indicating how trustworthy they are, while trustworthiness represents the degree to which users accept news.

In the case of SNS, trust can be formed through community interactions (Cigi-Ipsos: 2019). By engaging in communication that shapes perceptions, the decision to trust or distrust an object emerges within the human brain. Trust is established by at least two users: an information writer and a reader, the trustee and trustor. Because building trust quickly is difficult, we propose a trust evaluation system based on Gao et al. (2019), which considers four trust factors: identity-based trust, behavior-based trust, relation-based trust, and feedback-based trust. We extend the model proposed in Gao et al. (2019) by adding information-based trust to illustrate the significance of information factors that may mislead users.

Table 2 Notations for Trust Model

Notation	Description
$tw_k^{(i)}$	The tweet issued by user i at time k
$D(tw_k^{(i)})$	The dissemination ratio of $tw_k^{(i)}$
$Ntw(n)$	Number of tweet that has been generated at time n
$T_i(n)$	Overall trust of user i at time n
$\overline{T}(n)$	Average trust at time n
IT_i	Identity-based trust of user i
BT_i	Behavior-based trust of user i
RT_i	Relation-based trust of user i
FF_i	Feedback factor of user i
IFT_i	Information-based trust of user i
$C(tw_k^{(i)}, n)$	Accuracy of tweet $tw_k^{(i)}$
$IP(tw_k^{(i)})$	Ratio of number of included pictures in $tw_k^{(i)}$
$PP(tw_k^{(i)}, n)$	The popularity of tweet $tw_k^{(i)}$
Kn_i	Knowledgeability of user i
Ps_i	Psychopathy of user i
$NC_i(n)$	Number of negative comments of tweets created by user i
$PC_i(n)$	Number of positive comments of tweets created by user i
$FP(n)$	Number of users who become the followers of users
$FN(n)$	Number of users who quit the followers of users

In the following, we consider a discrete-time SNS system where time is divided into slots. We assume that at time slot n $(= 0, 1, 2, \ldots)$, user i $(\in U)$ has the overall trust $T_i(n)$, and that user i takes an action such as creating a news, distributing or discarding a forwarded news, according to the value of $T_i(n)$. Table 2 summarizes the notations used for the five trust factors.

We introduce the social popularity function of the variable associated with user i, Var_i, which was originally proposed in Gao et al. (2019)

$$f(Var_i) = \frac{\log(Var_i + 1)}{\log(\max_{j \in U}(Var_j + 1))}. \tag{1}$$

5.1 Identity-based tust

Identity-based trust is the identity profile of the SNS users (Gao et al. 2019). The identity-based trust is formed with the social popularity denoted by the number of followers, the authority factor, and the age factor. Let IT_i and AF_i denote the identity-based trust and the age factor of SNS user i ($\in U$), respectively. The age factor of user i, denoted as AF_i, corresponds to the account's age displayed on the social network profile. For instance, if a user registered on the network in 2019, the age value would be 4 in 2023. The formulation for the age factor of user i, AF_i, is defined as

$$AF_i = \frac{Ag_i}{\overline{Ag} + Ag_i}, \tag{2}$$

where Ag_i is the age value of user i and $\overline{Ag}$ is the average age of all the users.

At each time slot, each user decides to/not to be the follower of other users. We assume that for $i, j \in U$, user j follows user i if his/her overall trust $T_j(n)$ is greater than threshold θ_{trust}. We introduce the following two subsets of U:

$$\begin{aligned} FP(n) &= \{j \in U : T_j(n-1) \le \theta_{trust}, T_j(n) > \theta_{trust}\}, \\ FN(n) &= \{j \in U : T_j(n-1) > \theta_{trust}, T_j(n) \le \theta_{trust}\}. \end{aligned}$$

$FP(n)$ is the set of users that are new followers of any other users at time n, while $FP(n)$ is the set of users that stop follow other users at n.

We also define $followers_i(n)$ as the number of followers of user i at time n. The value of $followers_i(n)$ is updated with $FP(n)$ and $FN(n)$ according to the following equation

$$followers_i(n) = followers_i(n-1) + |FP(n)| - |FN(n)|, \tag{3}$$

where $|\cdot|$ denotes the cardinality of a set.

Let $P_i(n)$ denote the popularity factor of user i at time slot n, defined by

$$P_i(n) = f(followers_i(n)), \tag{4}$$

where $f(\cdot)$ is the popularity function defined in (1).

Then, the identity-based trust, IT_i, is defined as

$$IT_i(n) = w_p \cdot P_i(n) + w_{au} \cdot Au_i + w_{af} \cdot AF_i, \tag{5}$$

where w_p, w_{au}, and w_{af} are weight parameters, and Au_i is the authority score of user i defined by

$$Au_i = \begin{cases} 1, & \text{if the verified badge exists in user } i'saccount, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Here, the verified badge is a verification mark attached to the profiles of some legitimate users. $Au_i = 1$ implies that user i has verified badge on his/her account page.

5.2 Behavior-based trust

Behavior-based trust reflects the cognitive processes that dictate how information spreaders behave on social networking services. This is related to the number of fake news items that the information spreaders follow, implying how controversial they behave in the social network. There are two behavioral responses on a social network that users can draw from a tweet: comments and mentions. Comments are responding to someone's tweet by placing a response tweet into the tweet comment section. On the contrary, mentions are actions of calling someone into another tweet.

In this model, we suppose that comments are more informative to examining trustworthiness than mentions, which was also pointed out in Gao et al. (2019). The main reason is that comments are placed within tweets, whereas mention is an action of sharing with another node that may not have any linked connection to a specific user. In behavior-based trust calculation, the primary focus is only on the factors that belong to the information spreaders, such as comments, shares, and likes, where mentions only show the actions made by the receivers. We also assume that the behavior of information spreaders is more important than the source of the post.

Let $BT_i(n)$ ($i \in U$) denote the quantity of behavior-based trust of user i at time slot n. The evaluation is performed in two different point of view; the tweet creator perspective and the tweet receiver perspective. In terms of the tweet creator perspective, $BT_i(n)$ is calculated based on three factors; senders' likes, shares, and comments. We define $tw_k^{(i)}$ ($i \in U, k = 1, 2, \dots, n$) as the tweet generated by user i at time k.

Now we define the following variables associated with the tweet $tw_k^{(i)}$.

$lk(tw_k^{(i)}, n)$: The number of likes that $tw_k^{(i)}$ receives by time slot n.

$sh(tw_k^{(i)}, n)$: The number of shares that $tw_k^{(i)}$ receives by time slot n.

$co(tw_k^{(i)}, n)$: The number of comments that $tw_k^{(i)}$ receives by time slot n.

Let $Inf(tw_k^{(i)}, n)$ denote the influence value of a single tweet $tw_k^{(i)}$ at time slot n, which is defined by

$$Inf(tw_k^{(i)}, n) = \frac{f(lk(tw_k^{(i)}, n)) + f(sh(tw_k^{(i)}, n)) + f(co(tw_k^{(i)}, n))}{3}. \tag{7}$$

Then, Influence creator factor $IC_i(n)$ is defined as

$$IC_i(n) = \frac{1}{\#_i^{tweet}(n)} \sum_{k=1}^{\#_i^{tweet}(n)} Inf(tw_k^{(i)}, n), \tag{8}$$

where $\#_i^{tweet}(n)$ is the number of tweets created by user i until time slot n.

From the viewpoint of the users who received the tweet from user i, they want to understand how often user i interacted with fake news and true news. This can be evaluated by measuring the number of fake tweets liked, shared, and discarded by user i compared to true news.

Let $Tlk_i(n)$ (resp. $Flk_i(n)$) denote the number of likes on true (resp. fake) news by user i at time slot n. We define $LK_i(n)$ as the degree of likes by user i at time slot n, which is given by

$$LK_i(n) = \frac{f(Tlk_i(n))}{f(Tlk_i(n)) + f(Flk_i(n))}. \tag{9}$$

In order to characterize the share behavior of users, we introduce the degree of share for user i, Sh_i. Let $Tsh_i(n)$ (resp. $Fsh_i(n)$) denote the number of shares on true (resp. fake) news by user i at time slot n. We also define $Tds_i(n)$ (resp. $Fds_i(n)$) as the number of discards on true (resp. fake) news by user i at time slot n. Then Sh_i is defined as

$$Sh_i(n) = \frac{f(Tsh_i(n)) + f(Fds_i(n))}{f(Tsh_i(n)) + f(Fsh_i(n)) + f(Tds_i(n)) + f(Fds_i(n))}. \tag{10}$$

Let $IR_i(n)$ denote the influence receiver factor of user i in time slot n, defined by

$$IR_i(n) = \frac{LK_i(n) + Sh_i(n)}{2}. \tag{11}$$

With the influence creator factor $IC_i(n)$ and influence receiver factor $IR_i(n)$, the behavior based trust of user i in time slot n, $BT_i(n)$, is defined as

$$BT_i(n) = \frac{IC_i(n) + IR_i(n)}{2}. \tag{12}$$

5.3 Relation-based trust

In general, users are ikely to trust the information provided by family members and friends. The relation-based trust is a measure of how closely a user is related to his/her colleagues who provide information. We adopt the relation-based trust of Gao et al. (2019) in which the degree of closeness is characterized by the betweenness centrality and number of shortest paths between users. The major difference between this model and that in Gao et al. (2019) is that users with high extroversion personality trait tend to have high interaction with others, causing high relation-based trust value Correa et al. (2010). The relation-based trust of user i, RT_i, consists of two factors: the local clustering coefficient of user i, LC_i, and the betweenness centrality of user i, $\sigma(i)$. The local clustering coefficient quantifies of how close a node, in this case, a user, is to its neighbors. This was proposed by Watts and Strogatz (1998) with the main goal of determining the proportion of the current number of links divided by the maximum possibility of links that could exist between the nodes. Following (Gao et al. 2019), the local clustering coefficient of user i, LC_i is calculated according to the following equation:

$$LC_i = \frac{2Ln_i}{No_i(No_i - 1)}, \tag{13}$$

where Ln_i is the number of links between user i's neighboring users, and No_i the number of user i's neighboring users.

In contrast, the betweenness centrality addresses the centrality of a graph by measuring the shortest path of every vertex. The betweenness centrality of user i, $\sigma(i)$, is defined as follows:

$$\sigma(i) = \sum_{s \neq i \neq t} \frac{\varsigma_{st}(i)}{\varsigma_{st}}, \tag{14}$$

where $\varsigma_{st}(i)$ $(s, t, i \in U)$ is the number of the shortest connection paths between users s and t via user i, and ς_{st} is the number of the shortest connection paths between users s and t.

With LC_i and $\sigma(i)$, RT_i is defined as

$$RT_i = \frac{LC_i + (1 - \sigma(i))}{2}. \tag{15}$$

5.4 Feedback factor

The feedback factor accounts for the trust given by the receiver of a tweet by providing reviews or rating the information source (Gao et al. 2019). The more positive comments posted in the comment section of the tweet, the more trustworthy the information becomes. The feedback factor was evaluated from the viewpoints of tweet creators and tweet receivers.

From the tweet creator's viewpoint, the feedback factor is measured as the number of positive comments that user i created for true tweets over fake tweets. We define $FC_i(n)$ as the feedback creator factor of user i at time slot n, given by

$$FC_i(n) = \frac{TPC_i(n) + FNC_i(n) + 2}{(TPC_i(n) + 1) + (FPC_i(n) + 1) + (TNC_i(n) + 1) + (FNC_i(n) + 1)}, \tag{16}$$

where $TPC_i(n)$ (resp. $TNC_i(n)$) is the cumulative number of positive (resp. negative) comments on the true tweets created by user i, counted at time slot n, and $FPC_i(n)$ (resp. $FNC_i(n)$) is the cumulative number of positive (resp. negative) comments on the fake tweets created by user i, counted at time slot n.

On the other hand from the tweet receiver viewpoint, it is important to measure the number of positive comments received by user i received over the negative comments. Let

$FR_i(n)$ denote the feedback receiver factor of user i at time slot n, which is defined as

$$FR_i(n) = \frac{PC_i(n) + 1}{(PC_i(n) + 1) + (NC_i(n) + 1)}, \tag{17}$$

where $PC_i(n)$ (resp. $NC_i(n)$) is the cumulative number of positive (resp. negative) comments on the tweets created by user i, counted at time slot n.

With $FC_i(n)$ and $FR_i(n)$, the feedback factor of user i $FF_i(n)$ is formulated as

$$FF_i(n) = \frac{FC_i(n) + FR_i(n)}{2}. \tag{18}$$

5.5 Information-based trust

In social networking service, the information feature plays an important role in information trust during disaster. For instance, the presence of photos significantly improves trustworthiness (Riegelsberger et al. 2003; Lucassen and Schraagen 2011). Based on the 3 S-model of information trust (Lucassen and Schraagen 2011), the information features considered here are semantic features, which represent content accuracy and surface features that include photos, logic, and post popularity.

We focus on the dissemination of tweets, and we need to characterize the trustworthiness of each tweet created by a user. Let $C(tw_k^{(i)}, n)$ denote the content accuracy of tweet $tw_k^{(i)}$ at time slot n, which is defined by

$$C(tw_k^{(i)}, n) = \frac{PF(tw_k^{(i)}, n) + 1}{PF(tw_k^{(i)}, n) + NF(tw_k^{(i)}, n) + 2}, \tag{19}$$

where $PF(tw_k^{(i)}, n)$ (resp. $NF(tw_k^{(i)}, n)$) is the cumulative number of users who give positive (resp. negative) feedback to tweet $tw_k^{(i)}$ until time slot n. We define the accuracy for the contents generated by user i, $C_i(n)$, as the average of $C(tw_k^{(i)}, n)$ taken by all the tweets issued by user i

$$C_i(n) = \frac{1}{\#_i^{tweet}(n)} \sum_{k=1}^{\#_i^{tweet}(n)} C(tw_k^{(i)}, n). \tag{20}$$

Similarly, we define $IP(tw_k^{(i)})$ as the ratio of the number of pictures included in tweet $tw_k^{(i)}$ to the maximum number of pictures in a tweet, which is given by

$$IP(tw_k^{(i)}) = \frac{Pictures(tw_k^{(i)})}{4}, \tag{21}$$

where $Pictures(tw_k^{(i)})$ is the number of pictures in $tw_k^{(i)}$. Note that on Twitter, the maximum number of pictures in a tweet is four. Let $IP_i(n)$ denote the average number of pictures included in a tweet by user i. Here, the average is taken by user i's tweets generated until time slot n. $IP_i(n)$ is given by

$$IP_i(n) = \frac{1}{\#_i^{tweet}(n)} \sum_{k=1}^{\#_i^{tweet}(n)} IP(tw_k^{(i)}). \tag{22}$$

In terms of the popularity of a tweet generated by a user, let $Likes(tw_k^{(i)}, n)$ denote the number of users who gave their likes to tweet $tw_k^{(i)}$ until time slot n. We define the popularity of tweet $tw_k^{(i)}$, $PP(tw_k^{(i)}, n)$, as

$$PP(tw_k^{(i)}, n) = \begin{cases} 1, \text{ if } Likes(tw_k^{(i)}, n) \geq N/2, \\ 0, \text{ otherwise.} \end{cases} \tag{23}$$

Then, the popularity value of $user_i$, $PP_i(n)$, is defined as

$$PP_i(n) = \frac{1}{\#_i^{tweet}(n)} \sum_{k=1}^{\#_i^{tweet}(n)} PP(tw_k^{(i)}, n). \tag{24}$$

With these features, we define the information-based trust for tweets by an SNS user i at time slot n, $IFT_i(n)$, as the following equation

$$IFT_i(n) = w_c \cdot C_i(n) + w_{ip} \cdot IP_i(n) + w_l \cdot LT_i + w_p \cdot PP_i(n), \tag{25}$$

where w_η's ($\eta \in \{c, ip, l, p\}$) are weighting factors of variables.

5.6 Overall trust

Finally, we define the overall trust of user i at time slot n, $T_i(n)$, as the following equation

$$\begin{aligned} T_i(n) = &w_{it} \cdot IT_i(n) + w_b \cdot BT_i(n) + w_r \cdot RT_i \\ &+ w_f \cdot FF_i(n) + w_{if} \cdot IFT_i(n), \end{aligned} \tag{26}$$

where w_ξ's ($\xi \in \{it, b, r, f, if\}$) are weighting factors of component trust values.

To achieve a balance between all trust models, we determined the weights according to the questionnaire results shown in the following section of Simulation and Questionnaire.

6 Simulation and questionnaire

In this section, we present the agent-based simulation experiment for our proposed trust model. First, we illustrate the procedure of the agent-based simulation and explain how the overall trust for each user is calculated. Then, we present the

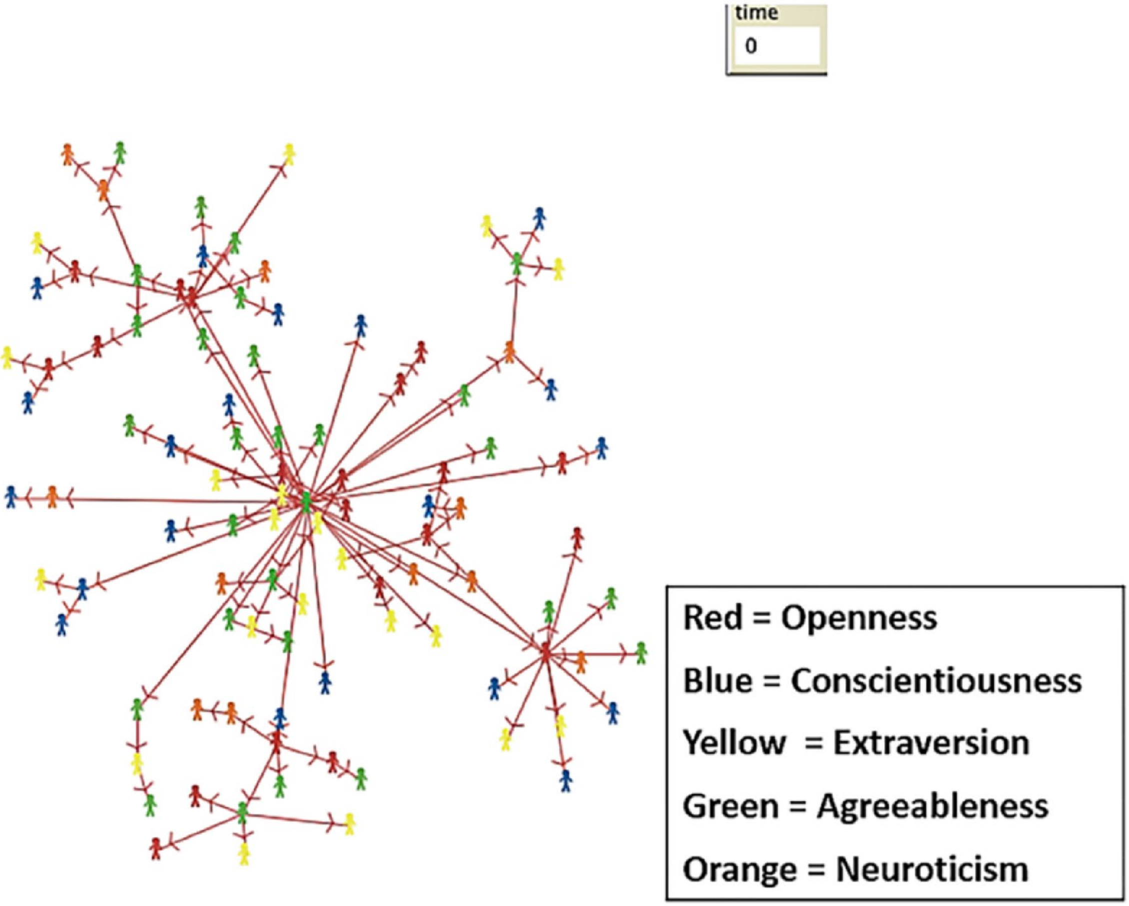


Fig. 1 *Netlogo* Interface: Users are generated and assigned with personalities

questionnaire conducted to set the parameters of our simulation model.

6.1 Agent-based simulation

We developed an agent-based modeling environment for the SNS on NetLogo 6.0.4. In the simulation, SNS users were modeled as agents, and tweets were represented as the delivered materials. Trust changes were observed through the agents' overall trust values.

Our agent-based simulation consisted of four phases. The first phase is user generation, where users are generated and linked according to the Barabási Albert model of scale-free network (Barabási 2002). Then, each user is characterized with personality traits explained in Sect. 4.

Figure 1 shows a sample of a user relation network generated by the Barabási Albert model. In this figure, each user is characterized by the maximum personality attribute in a specific color. For example, the user with the highest openness value is colored red, while the user with the highest conscientiousness value is colored blue. The network relation illustrates how information is disseminated among users. In social networks, these relations are described as social links. For instance, if user 1 follows user 2, the information will propagate from user 2 to user 1.

The second phase involved tweet creation. Figure 2 illustrates this phase, where several tweets are created in the network. In the figure, the blue edges represent the direction of tweet information, while the social links between users are depicted in red. As shown, if user *i* has no followers, the tweet will not be propagated to other users. During this phase, a limited number of users receive the news, and each of them examines the trustworthiness of the news according to the trust model. Subsequently, the news will be disseminated further through the network.

The third phase is the fake-news creation. When the news becomes popular, a fake news appears, as shown in Fig. 3.

The fourth phase is the dissemination process of fake news. The users will receive the information and decide to share according to the CS_i value. When the knowledgeable users received the fake news, he/she will establish clarification based on the news, telling that the news is not correct. This creates social rejection towards all news related to the

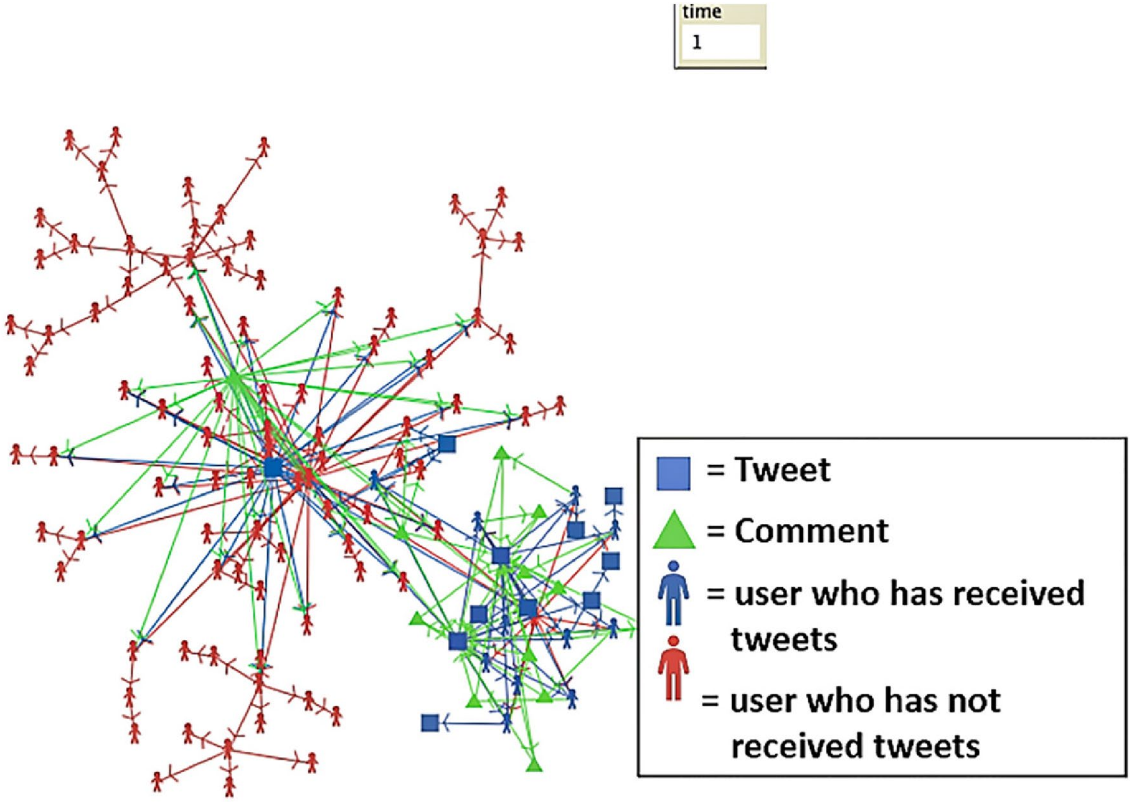

time
1
■ = Tweet
▲ = Comment
= user who has received tweets
= user who has not received tweets

Fig. 2 *Netlogo* Interface: Tweets are generate

same topic. Then, the fake news will be disappeared. See Fig. 4.

Our simulation is a discrete-time simulation with time slot $n \in \{0, 1, 2, \ldots, 50\}$. In one simulation experiment, the initialization process is executed at $n = 0$, then the above four steps are performed and terminated at $n = 50$. This experiment is repeatedly executed 100 times with different seeds, and the average values of performance measures are calculated.

6.2 Trust calculation procedure

In this subsection, we describe the procedure of trust calculation in our agent-based simulation.

First, a randomly selected user with openness personality creates a news and spreads it to the linked users. The main reason for this selection is that a high level of openness, coupled with extroversion, indicates increased usage of social networking sites (SNS) (Correa et al. 2010; Burbach et al. 2019). Furthermore, it has been observed that individuals who are open to experiences generally participate in more groups, resulting in a higher number of followers (Burbach et al. 2019). Therefore, selecting users with high openness is anticipated to generate greater information propagation, which will show the performance of our proposed model. Then, the users who have received the news will examine the trustworthiness of the news according to the trust model. The trustworthiness of the news is examined according to the trust model with five factors of trust: identity-based, relation-based, behavior-based, feedback factor, and information-based.

Assume that user i receives the news at time slot $t = n$. If the conscientiousness of user i, CS_i, is greater than or equal to 0.8, user i rejects the news. The action taken to disseminate the news depends on user's other personality traits. Users with openness, agreeableness, and extroversion are likely to share the received information. If CS_i is smaller than 0.5, user i accepts the news and shares it with his/her followers. If the value of CS_i is equal to or greater than 0.5 but less than 0.8, users will receive the information without sharing it with others.

In terms of the creation of fake news, we assume that a user with psychopathy creates a fake news from a receiving information. Let $D(tw_k^{(i)})$ denote the ratio of the number of users who received tweet $tw_k^{(i)}$ to the number of users N. We call $D(tw_k^{(i)})$ the dissemination ratio of the kth tweet issued by user i.

Consider the case where $D(tw_k^{(i)})$ reaches the value greater than 0.7 and user j ($\neq i$) receives the tweet $tw_k^{(i)}$.

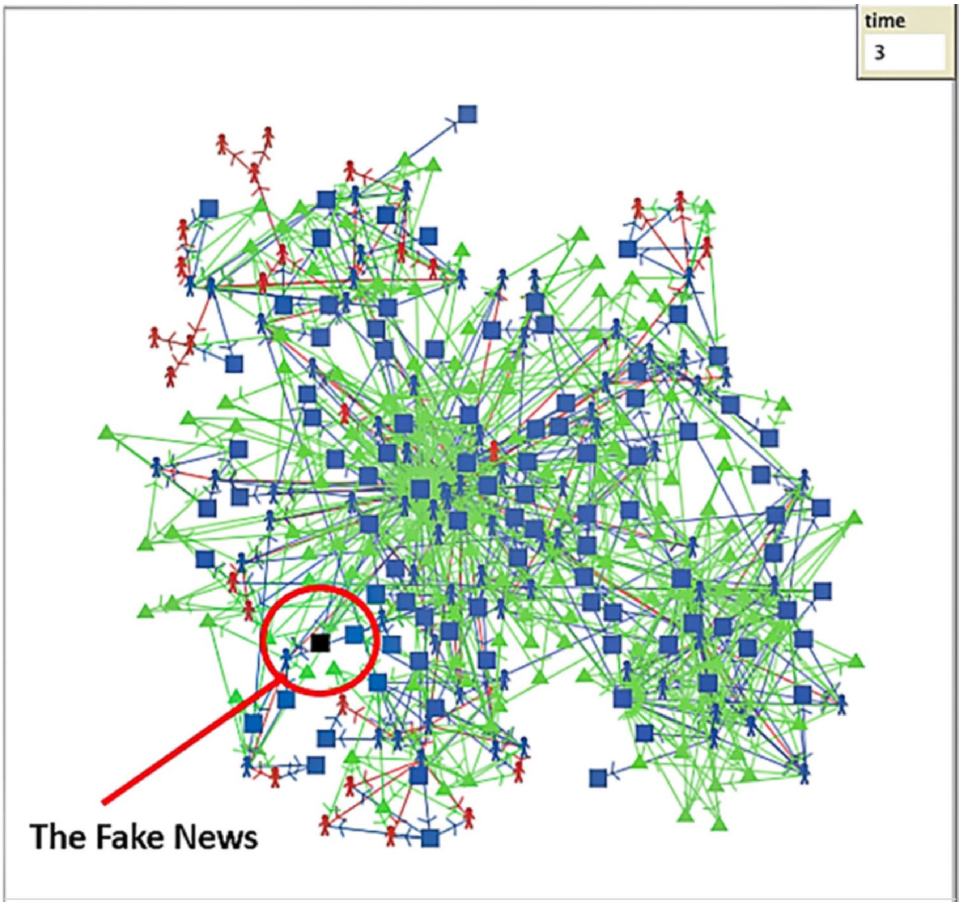


Fig. 3 *Netlogo* Interface: Fake news are generated

If user j is psychopathy ($Ps_j = 1$) and his/her overall trust $T_j(n)$ is greater than or equal to 0.5, user j creates a fake news and disseminates it, independently of $tw_k^{(i)}$. The fake news will be removed if the number of negative comments added to the news is greater than $N/2$. The increase in negative comments is the result of clarification actions by knowledgeable users. The more negative comments are added to the fake news, the more users will counter the fake news.

In order to evaluate the information dissemination over the SNS, we consider the average of the overall trust values of all users, $\overline{T}(n)$.

6.3 Questionnaire

To set the simulation parameters of the trust model, we performed a questionnaire survey of 150 university students in Indonesia. We gathered our respondents' opinions about fake news. We achieve this by presenting questions in two parts, the users‘ opinion on how fake information spread and which factors affect their trusting behavior in social networking services. Examples of the questionnaire are presented below:

- According to this piece of information, what is your opinion?
- According to your experience which factors below, affect your decision on trusting or neglecting the news?

The questionnaire results will be used to determine the weight factor of the trust model evaluation. This online survey was taken from April to May 2020 with the target respondents being mainly Twitter users, who use Twitter frequently daily. We believe that determining this weight factor, in the beginning, will help us evaluate the trustworthiness of users according to what the group of people thinks about how to evaluate a piece of information.

In this survey, we successfully gathered 150 responses, taken from Google Forms, a questionnaire-taking platform.

The results of the questionnaire are shown in Fig. 5. In this figure, 62% of the users believe that information quality is the key factor towards information acceptance. Note that information quality is the main feature off the proposed information-based trust. Then, the other resulting ratio of Identity, behavior, relation, and feedback are 18%, 8%, 8%, and 4%, respectively. These values are used for the weight parameters of the overall trust (26). “Generally, the weight factor can be changed dynamically by

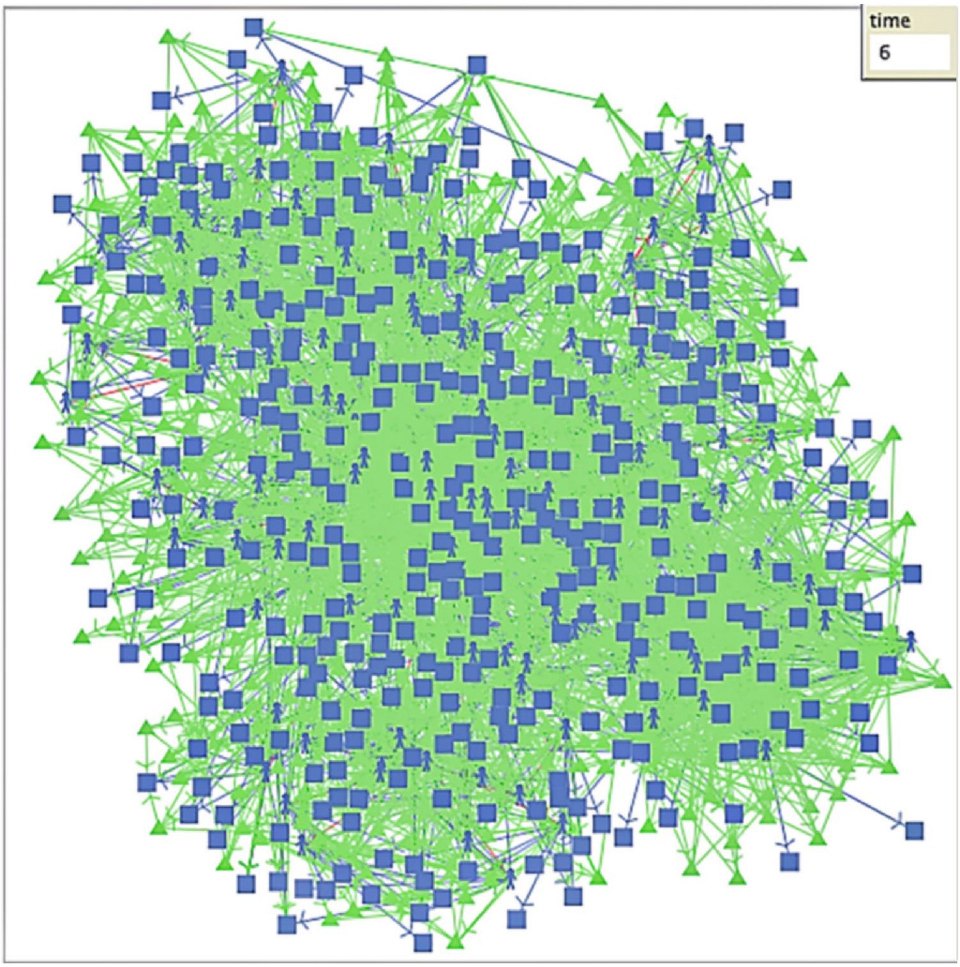


Fig. 4 *Netlogo* Interface: Fake news disappearance

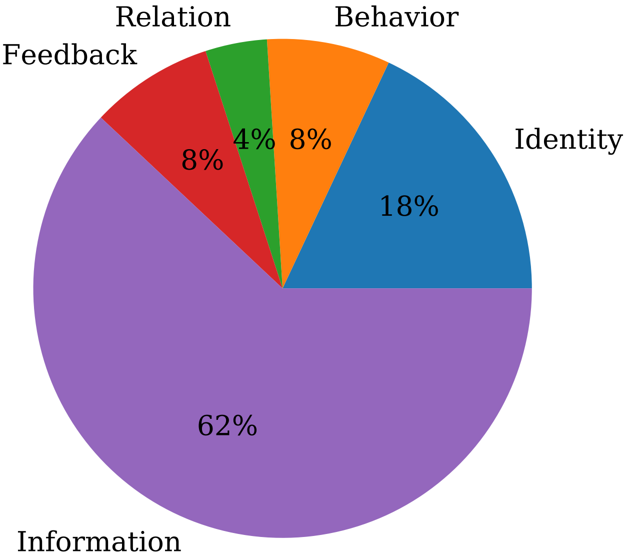


Fig. 5 Questionnaire results for trust model

the SNS manager; in the case of Twitter, the administrator may adjust it. However, with the current progress of Twitter as a community-based service, providing a simple public questionnaire in the community is essential.

6.3.1 Parameter settings

The weights of the overall trust in (26) were determined according to the questionnaire result shown in Fig. 5. We call the weight-value set the baseline value. In order to investigate the impact of each trust factor on the overall trust, we formed five different scenarios to eliminate one trust factor. These scenarios aim to understand how the trust model works on SNS networks. We considered scenarios in which one of the trust factors was set to zero, while the remaining weights were normalized proportionally to the questionnaire result. For example, when the weight of the identity-based trust, w_{it}, is set to 0, w_b is given by

$$w_b = \frac{0.08}{0.08 + 0.04 + 0.08 + 0.62} \approx 0.097.$$

Table 3 Trust Simulation Scenarios

Weight	Baseline	wo IT_i	wo BT_i	wo RT_i	wo FF_i	wo IFT_i
w_{it}	0.18	0	0.195	0.187	0.195	0.473
w_b	0.08	0.097	0	0.083	0.0869	0.210
w_r	0.04	0.048	0.043	0	0.043	0.105
w_f	0.08	0.097	0.086	0.0833	0	0.210
w_{if}	0.62	0.756	0.673	0.645	0.673	0

Table 4 Big-Five Personality Simulation Scenarios

Argument of *UP*	Baseline	Openness	Extroversion	Conscientiousness	Agreeableness	Neuroticism
openness	0.2	1	0	0	0	0
Extroversion	0.2	0	1	0	0	0
conscientiousness	0.2	0	0	1	0	0
agreeableness	0.2	0	0	0	1	0
neuroticism	0.2	0	0	0	0	1

Table 5 Parameter Settings

Description	Value	Source
Age: Ag_i	Sampled from Poisson distribution with mean 3.5	Assumption
Number of users with $Au_i = 1$	Sampled uniformly from [0, 100]	Assumption
Number of users with $Kn_i = 1$	Sampled from Normal distribution $N(49.82, 8.85)$	(Nielsen et al. 2020)
Probability of psychopathy users sending fake news: P	0.5	Assumption
Number of users: N	100	Assumption
Ratio of psychopathy users	20%	Assumption
Openness: O_i	Sampled from $N(48.01, 08.95)$	(Schmitt et al. 2007)
Extroversion: E_i	Sampled from $N(51.25, 8.81)$	(Schmitt et al. 2007)
Conscientiousness: CS_i	Sampled from $N(47.19, 11.24)$	(Schmitt et al. 2007)
Agreeableness: A_i	Sampled from $N(46.38, 9.02)$	(Schmitt et al. 2007)
Neuroticism: NR_i	Sampled from $N(49.73, 9.66)$	(Schmitt et al. 2007)
Threshold: θ_{trust}	0.5	Assumption

Similarly, w_r, w_f, and w_{if} are set to 0.048, 0.097, and 0.756, respectively. Table 3 shows the six scenarios: 1) baseline, 2) case without identity-based trust (wo IT_i), 3) case without behavior-based trust (wo BT_i), 4) case without relation-based trust (wo RT_i), 5) case without feedback factor (wo FF_i), and 6) case without information-based trust (wo IFT_i).

In terms of the Big-Five personality traits, we also consider the six scenarios. In the baseline scenario, all weights of the five traits were the same and equal to 0.2. For the remaining scenarios, the weight of one personality was set to one, and those of the other four personality traits were set to zero. (See Table 4.)

The number of users was set to $N = 100$. Personality traits were assigned according to a previous study (Schmitt et al. 2007). Following (Schmitt et al. 2007), the personality traits of a user were determined according to normal distributions. (See Table 5.) For each personality trait, we generated N positive samples from the corresponding normal distributions. Then, the samples are normalized with the maximum value of the N samples so that the resulting samples are in [0,1]. We conducted 100 simulation experiments with different seeds, taking the averages of the performance measures over 100 samples.

In the simulation, we introduce the probability of a malicious user to disseminate fake news, P. The value of P is taken from 0 to 1. If P is set to 0, then the malicious users will not disseminate the fake news. If the value is set to 1, then the malicious users will disseminate the fake

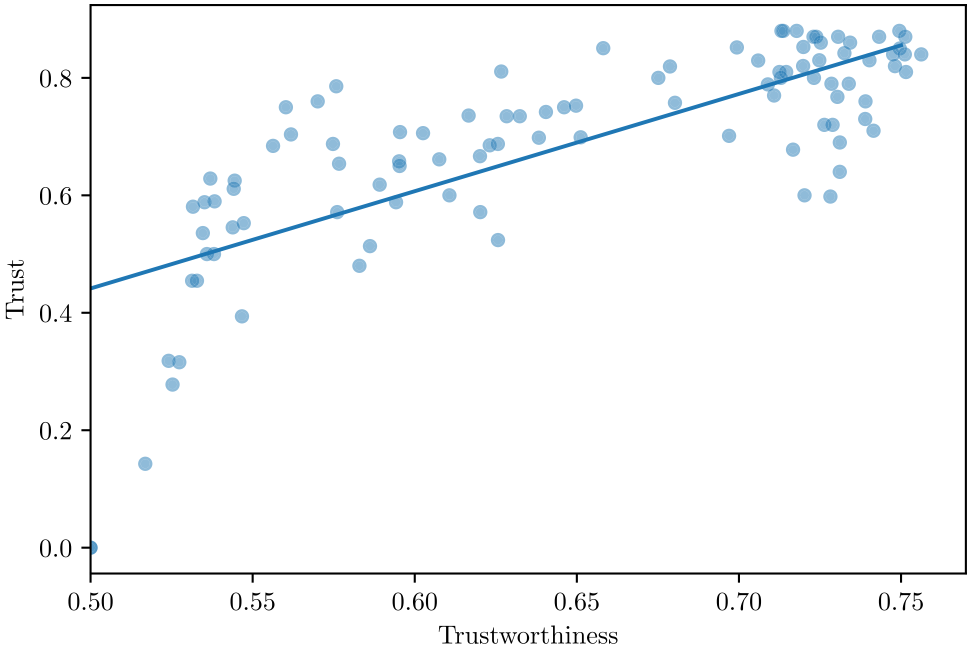


Fig. 6 The overall trust T_i and trustworthiness comparison

news. In the simulation, we set $P = 0.5$. This means the malicious users will not always share fake news, but also keep the true news disseminated.

Table 5 summarizes the parameter setting of the simulation experiments.

7 Results and discussion

In this section, we present our simulation results. We first show our trust model validation and sensitivity, then discussing the effect of Big-Five personality traits on trust values.

7.1 Trust model validation

To validate the model, we introduce the trustworthiness of a user defined by

$$Trustworthiness(i) = \frac{N^{AN}(i)}{N^{AN}(i) + N^{DN}(i)}, \quad i \in U, \tag{27}$$

where $N^{AN}(i)$ (resp. $N^{DN}(i)$) is the number of news items accepted (resp. discarded) by user i. We consider the baseline scenario whose parameter setting is shown in Tables 3 and 4. We also set $P = 0$, following (Antoci et al. 2019).

Figure 6 shows the relation between the trustworthiness and overall trust of 100 users at time slot $n = 10$ for one simulation experiment. In this figure, each point represents $(Trustworthiness(i), T_i(20))$ for user i $(i \in \{1, \dots, 100\})$, while the line is the result of the linear regression analysis for those points. We observe from this figure that the overall trust grows with increase in the trustworthiness value. This tendency is supported by Antoci et al. (2019), validating our trust and system models.

7.2 Component sensitivity of trust model

Figure 7 illustrates the evolution of the mean overall trust for the six scenarios in Table 3. It is observed in this figure that the overall trust values for all cases slightly increase. The mean overall trust in the case without behavior-based trust BT_i achieves the highest, while that in the case without information-based trust IFT_i is the lowest. Note that the case without information-based trust is equivalent to the system of Info-Trust (Gao et al. 2019).

In the parameter settings, we set the weights of the overall trust according to the questionnaire results, which gives the information-based trust the highest weight value. In terms of behavior-based trust, the weight value is similar to feedback factor, but has a lower trust value compared to the feedback factor evaluation.

Figure 8 shows the average along with its standard deviation of the overall trust against P, the probability of a malicious user sending fake news. We conducted simulation

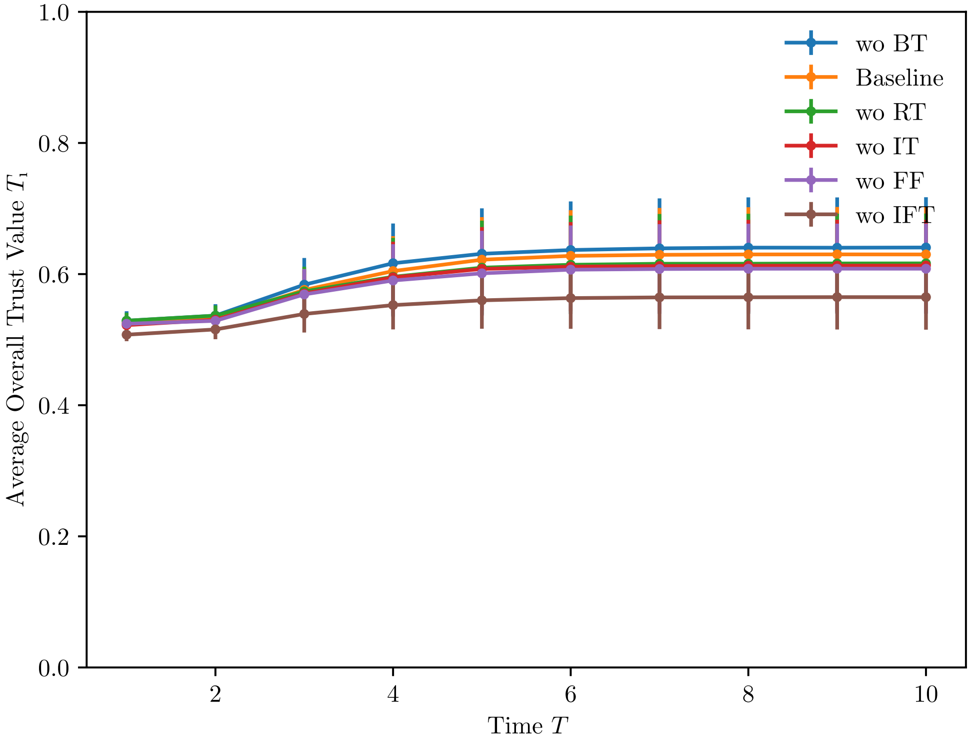


Fig. 7 The overall trust T_i over the time of the trust scenarios

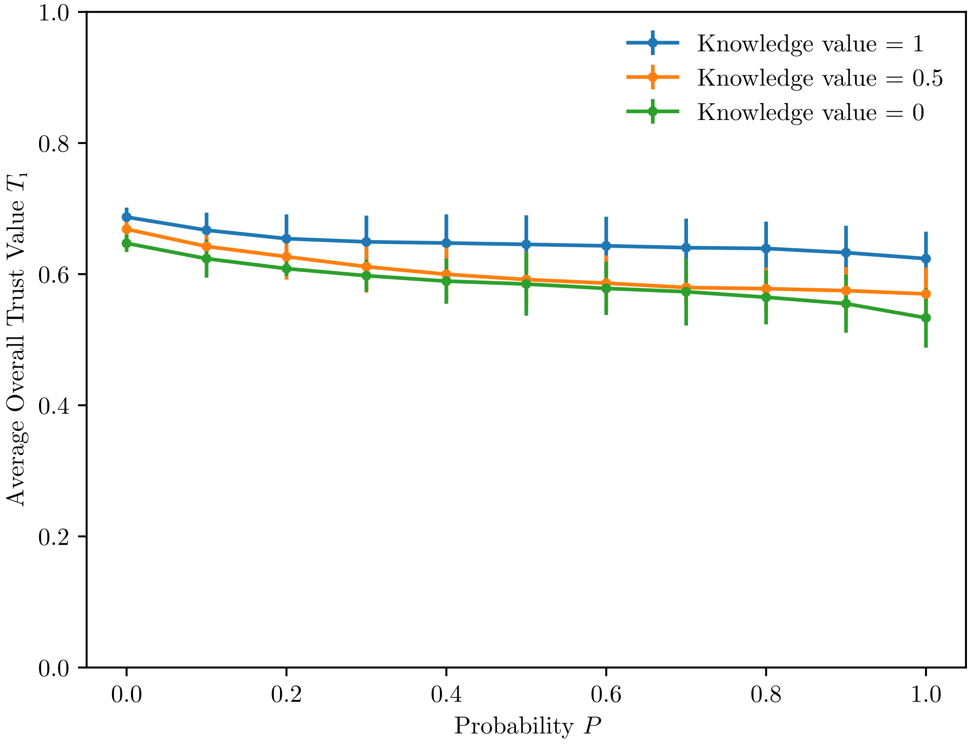


Fig. 8 The overall trust T_i against the probability of a malicious user sending fake news P

experiments in cases with the user i's knowledgeability Kn_i equal to 0, 0.5, and 1 for $\forall i \in U$. In this figure, the average overall trust $\overline{T}(n)$ decreases with increase in P for the three knowledgeable cases. This is because a large probability of sharing fake news P increases the appearance of fake news. This causes negative comments created by knowledgeable users, making the spreading speed of the tweet slow. A The higher number of negative comments $NC_i(n)$ affects a lower value of feedback factor $FF_i(n)$, while the lower number of

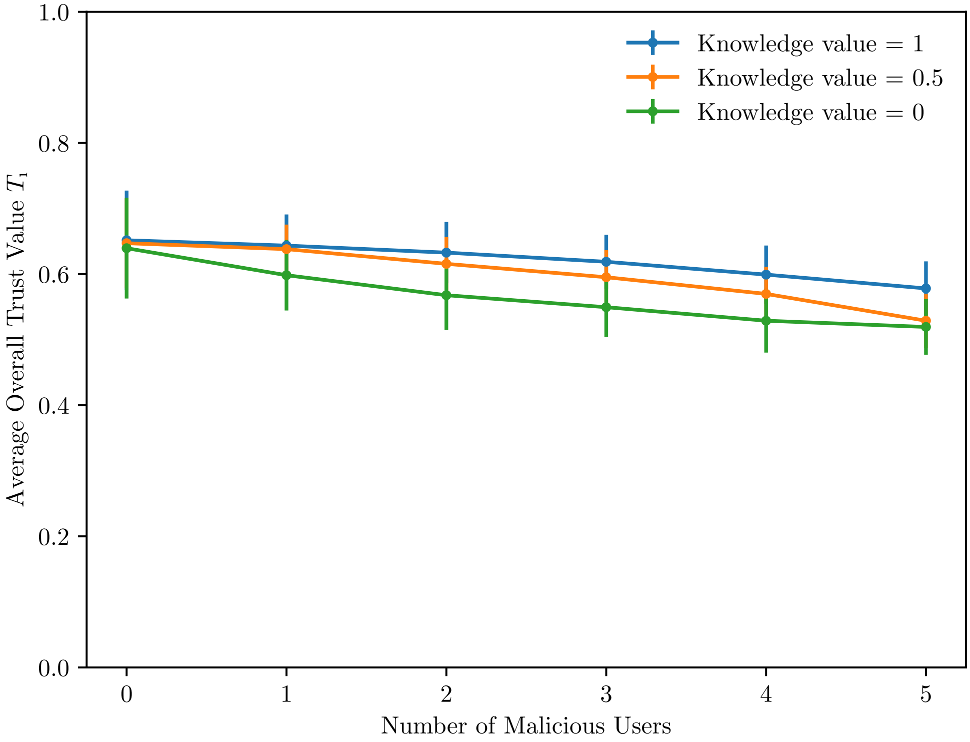


Fig. 9 The comparison between Overall Trust T_i and number of malicious users

tweet shared $sh(tw_k^{(i)}, n)$, and likes $lk(tw_k^{(i)}, n)$, affect a lower value of $BT_i(n)$.

Figure 9 shows how the number of malicious users affects the average overall trust $\overline{T}(n)$. In this figure, the overall trust decreases with increase in the number of malicious users, as expected. A remarkable point in the figure is that the overall trust values for three knowledgeability cases decrease similarly. This suggests that the knowledgeability itself is not effective in preventing a decline in the users' overall trust. This decreasing value matches overall trust value decreases in Gao et al. (2019) which shows how the trust degree react to malicious act.

7.3 Effect of big-five personality traits

In this subsection, we investigate how the Big-Five personality traits affect the overall trust. Since our trust model is heavily performed based on users interactions, the different on Big-Five personality traits will affect the overall-trust value. In this experiment, we set the values of the targeted personality traits between 0.1 and 0.9, keeping all the other parameters same as those in Schmitt et al. (2007). In openness scenarios, if an open-to-experience user is not generated based on the parameter settings, the system will select a user with the highest betweenness centrality $\sigma(i)$, as shown in Eq. (14), to ensure the dissemination of information to the majority of users in the network.

Figures 10, 11, 12, 13, 14 show the overall trust against the Big-Five personality traits of openness, conscientiousness, extroversion, agreeableness and neuroticism, respectively. The results shows that openness, conscientiousness and extroversion personality traits increase the overall trust value, while larger values of agreeableness and neuroticism make the overall trust value small. These results conform to Seidman (2020), which claims that neuroticism and agreeableness less correlate with trustworthiness, while the openness, conscientiousness and extroversion are highly related to trust.

Figure 10 shows the relation between the overall trust and openness. We conducted the experiments in cases of the knowledgeability $Kn_i = 0, 0.5$ and 1. We also investigate the cases where the first news created is normal and fake. In this figure, the overall trust values for three cases of knowledgeability are almost similar when the openness value is between 0.1 and 0.3. With the increase of openness value, however, the overall trust values with Kn_i =0.5 and 1 slightly increase, while in case of $Kn_i = 0$, the overall trust remains constant. This result is consistent with Pennycook and Rand (2021) which reported that users with more knowledge act carefully and have higher overall trust value. Figure 10 also shows that the overall trust is insensitive to the openness. This is because the openness does not take into account likes and comments. However, the openness has the role in information spreading and creating.

Figure 11 shows the overall trust against the conscientiousness. We observe the monotonic growth of the overall

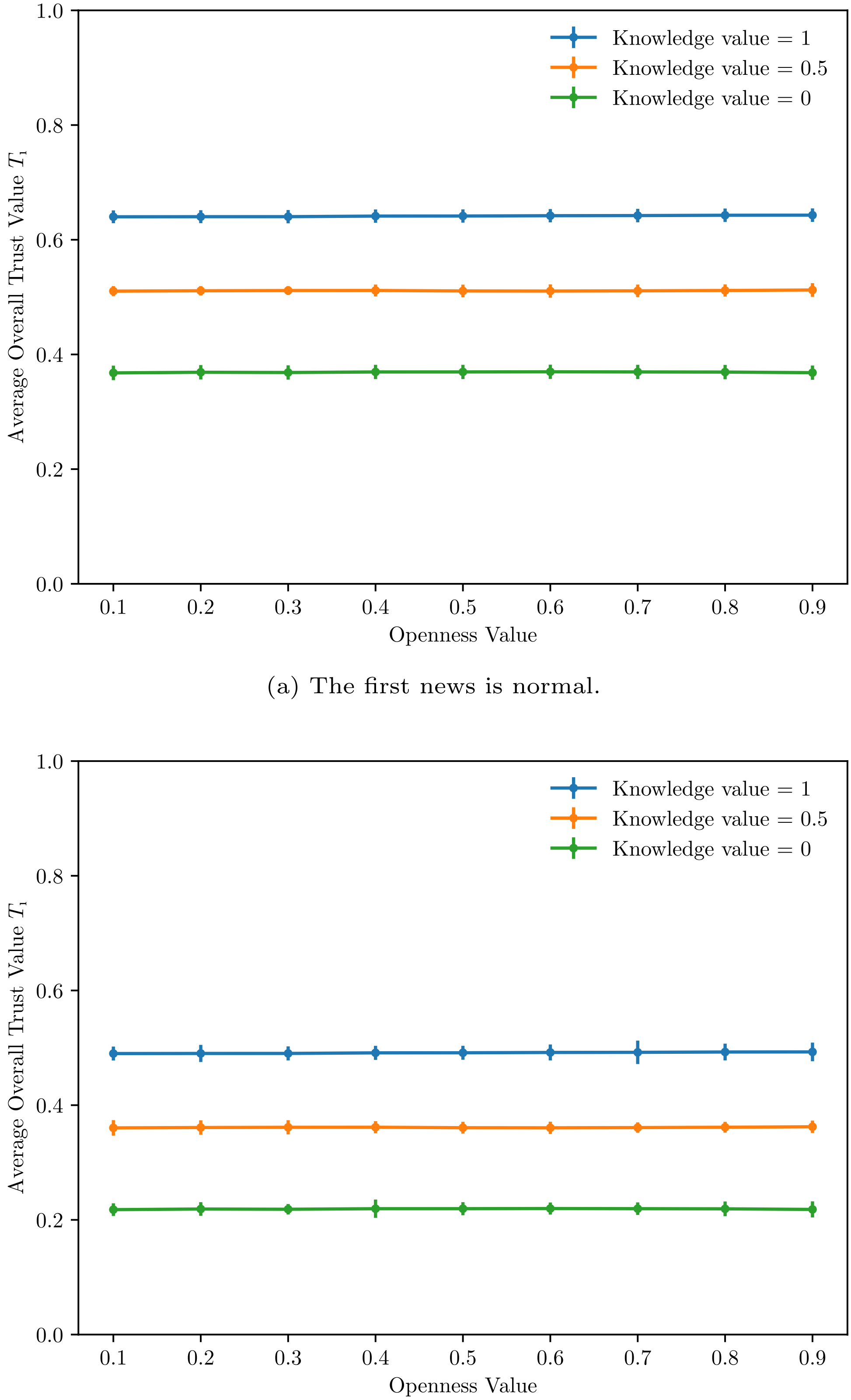
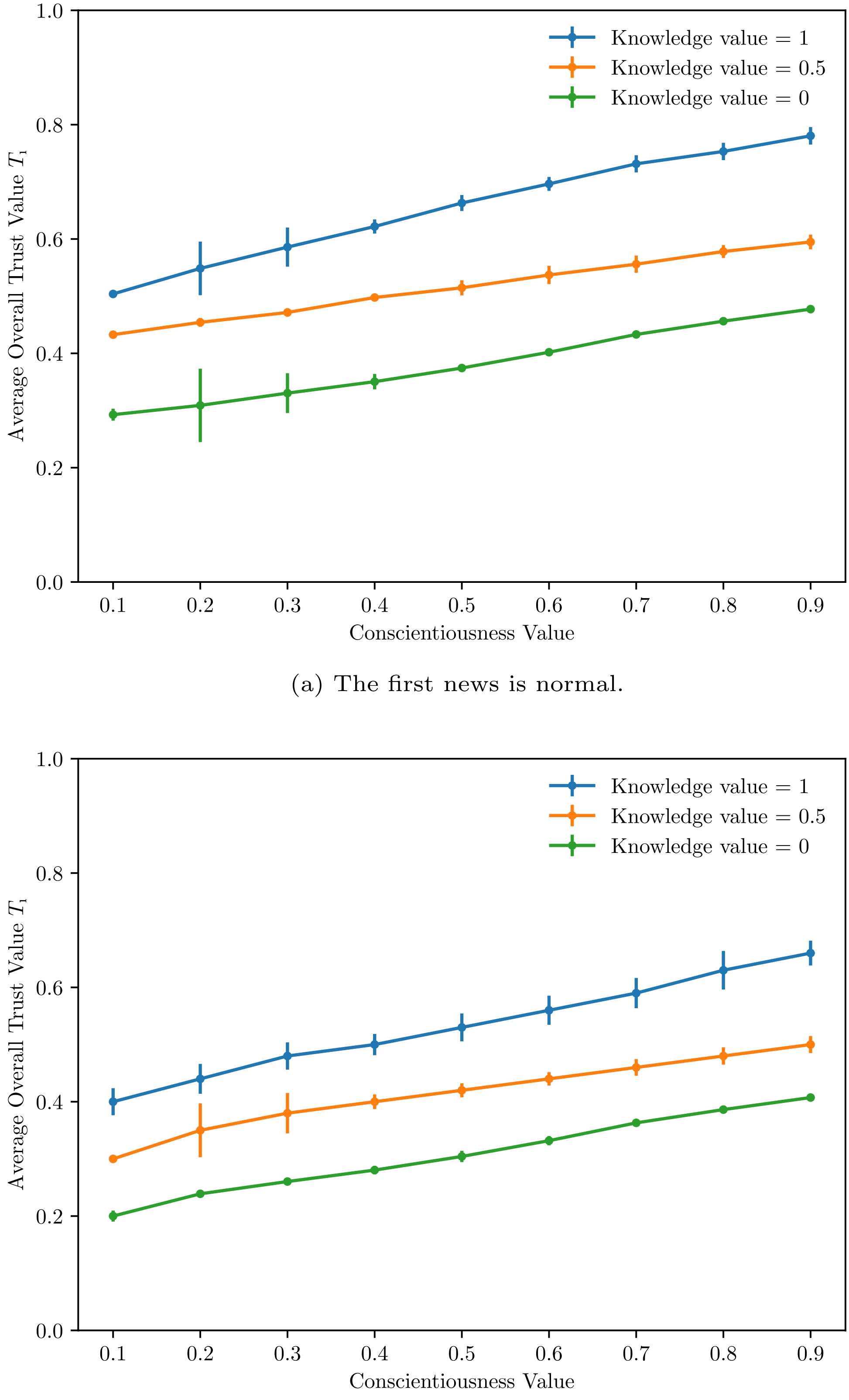


Fig. 10 Comparison of overall trust and openness personality rait

(a) The first news is normal.

(b) The first news is fake.

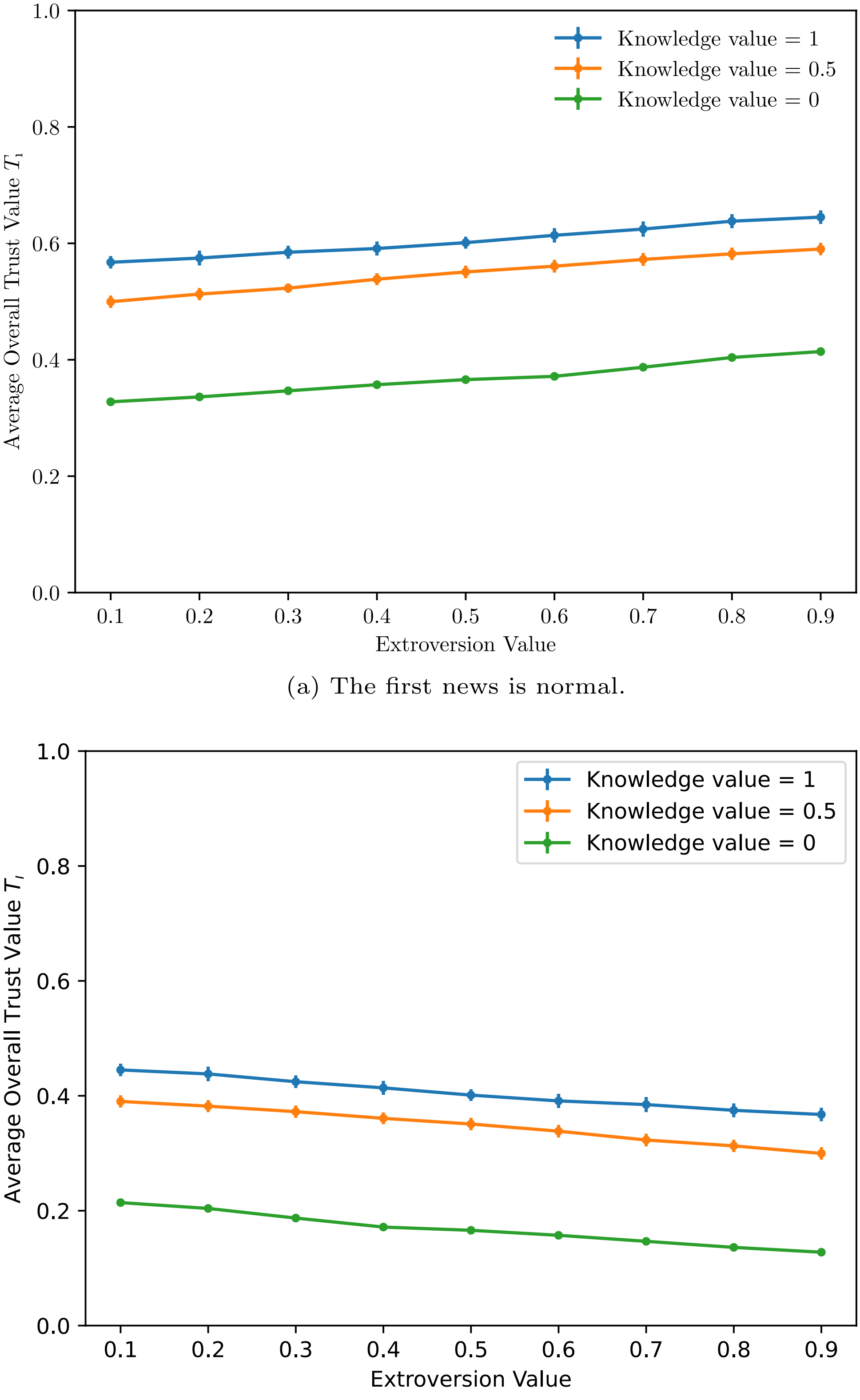
Fig. 11 Comparison of Overall Trust and Conscientiousness Personality Trait



(a) The first news is normal.

(b) The first news is fake.

Fig. 12 Comparison of Overall Trust and Extroversion Personality Trait

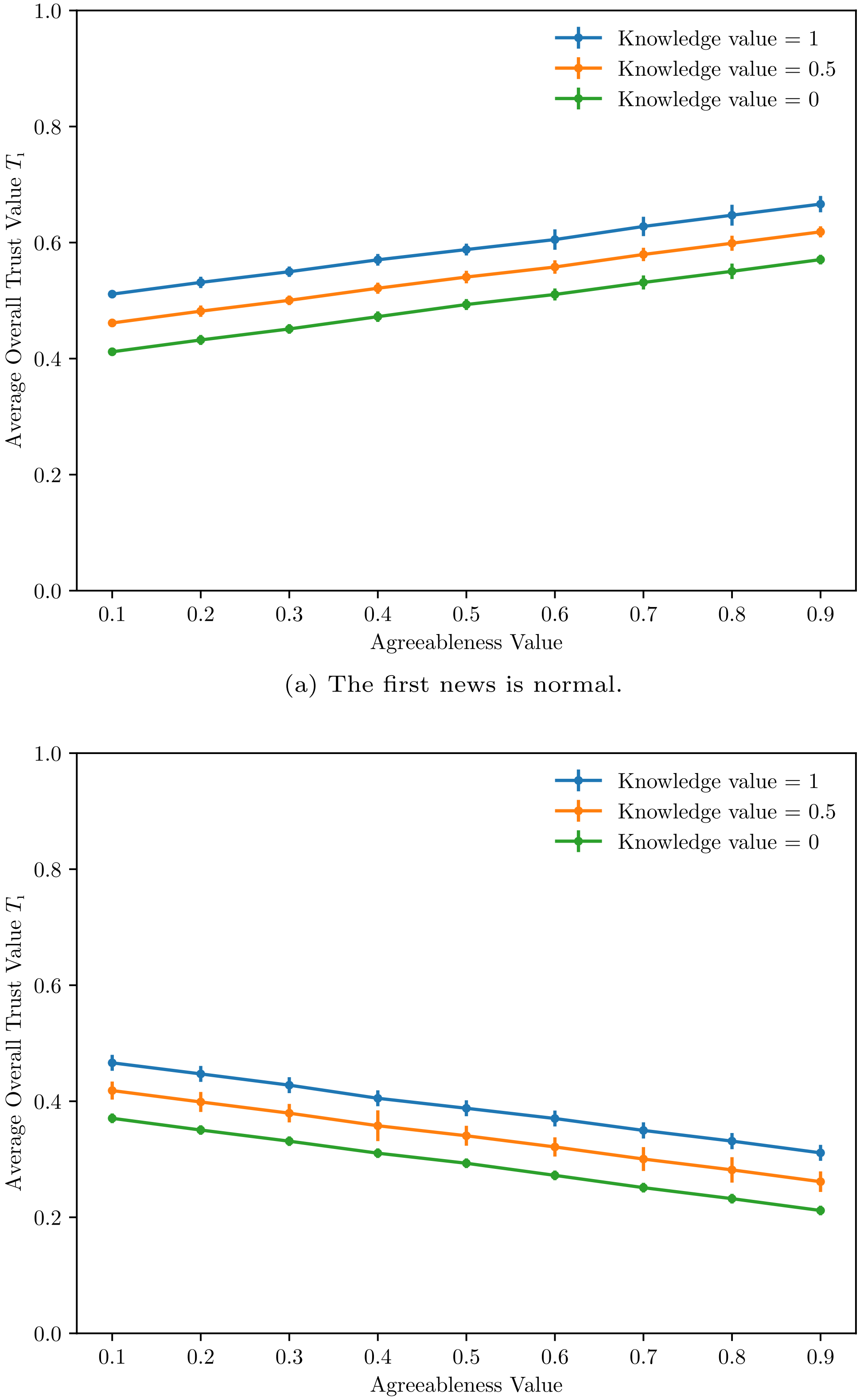


(a) The first news is normal.

(b) The first news is fake.

trust with increase in the conscientiousness value in different knowledge values. Users with high conscientiousness value is very cautious and not trapped with fake news while the news is disseminated in the initial phases, and tend to gather more opinions from the other users in the feedback section, then deciding to trust or not to trust the information. This

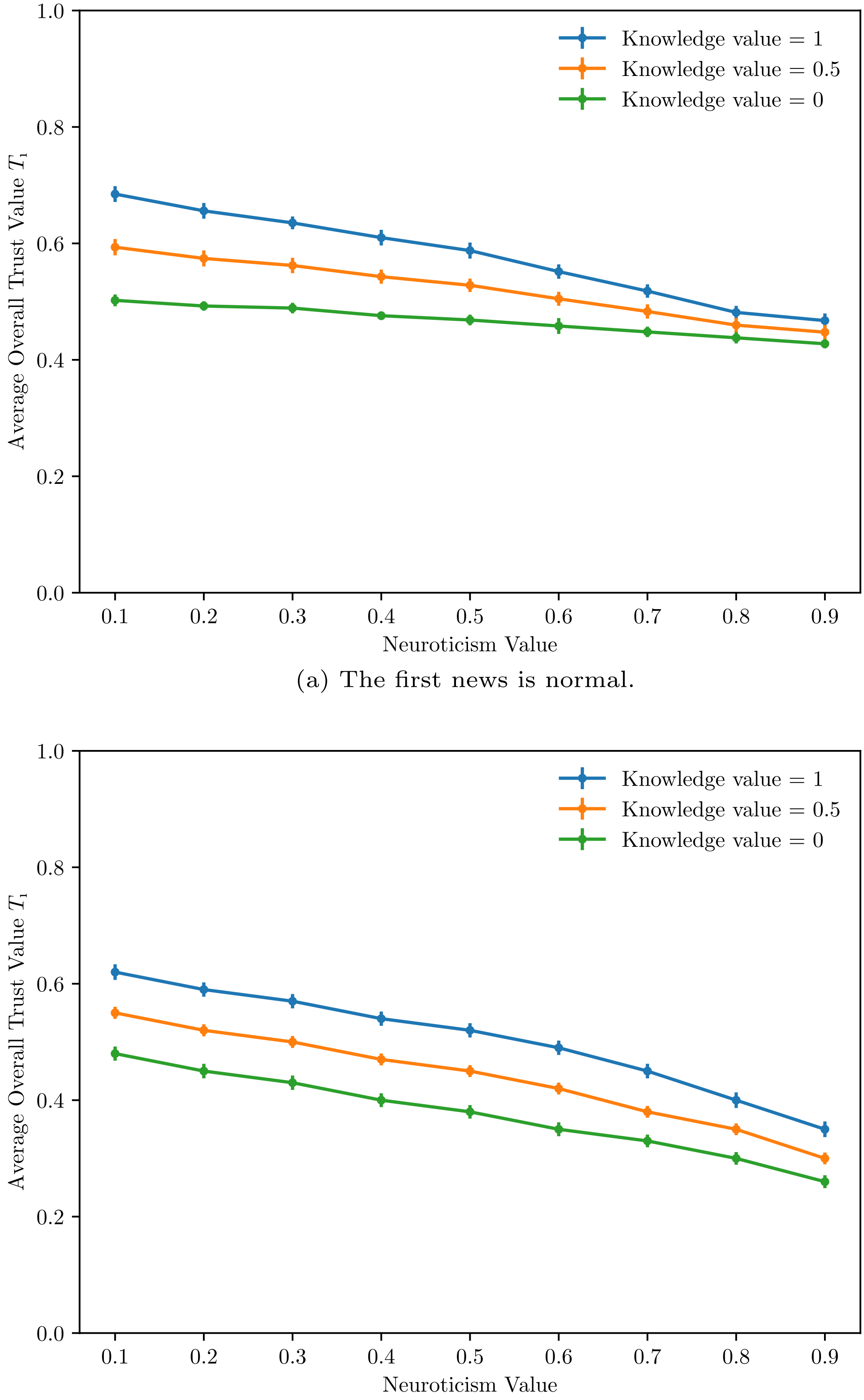
Fig. 13 Comparison of Overall Trust and Agreeableness Personality Trait



(a) The first news is normal.

(b) The first news is fake.

Fig. 14 Comparison of Overall Trust and Neuroticism Personality Trait



(a) The first news is normal.

(b) The first news is fake.

finding aligns with Zrnec et al. (2022), which shows conscientiousness users are more successful in detecting fake news. Therefore, our trust model can evaluate conscientiousness users with high overall trust values based on their cautious characteristics.

Figure 12 shows the overall trust against the extroversion. In Fig. 12a where the fist news is normal, the overall trust monotonically and gradually grows with increase in the extroversion value. Users with high extroversion are likely to share positive comment and likes in a single tweet, therefore it will increase the trust value when the fake news is few. Note also that high extroversion users are likely to attach a high number of pictures in a single tweet. This affects the included pictures value of the information-based trust (Seidman 2020). Furthermore, the overall trust will decrease while the fake news appears more, since the extroversion users tend to give high positive comment even to fake news. This finding is supported by Zrnec et al. (2022) showing that introverted users are reliable on detecting fake news.

Figure 13 illustrates the overall trust against the agreeableness. In terms of agreeableness, users with high agreeableness value are likely to share information without careful verification. In our simulation, the agreeableness users tend to have high level of tweet sharing and create positive comments. In Fig. 13a where the first news is normal, the overall trust values for three cases of knowledgeability are growing with increase in the agreeableness value. This result is supported by Wang et al. (2012), which reporting that with increase in the amount of shared information, the overall trust value will increase. When the first news created is fake, the overall trust monotonically decreases with increase in agreeableness. When user *i* has high agreeableness, user *i* is likely to share the news regardless of its trustworthiness, increasing the number of shares of fake news Fsh_i for the behavior-based trust $BT_i(n)$, as well as the number of positive comments $FPC_i(n)$ for the feedback factor $FF_i(n)$. This behavior resulted in a decrease in both $BT_i(n)$ and $FF_i(n)$.

Figure 14 shows the overall trust against neuroticism. Users with high neuroticism value tend to share fewer posts in SNS Wang et al. (2012). In Fig. 14, the overall trust values in three cases of knowledgeability are decreasing with increase in neuroticism value. This is because neuroticism is correlated with high fake news sharing and low logic value. Therefore this behavior decreases the information-based trust value. When the first news is normal, the differences among three curves is getting decreased with increase in the neuroticism value. Note that neuroticism users have higher likes probability and shares probability, but having less comment creation behavior. This makes the behavior-based trust $BT_i(n)$ large, while decreasing the feedback factor $FF_i(n)$ and information-based trust $IFT(n)$. When user *i*'s knowledgeability is $Kn_i = 1$, user *i* is likely to create more comments than users with small knowledgeability. When the initial news is fake, the spread and engagement with likes on the fake news negatively impact user feedback, decreasing both behavior-based trust and feedback factor of the users.

Based on the findings shown above, we can conclude that when the users' probability of sharing fake news is low, the high level conscientiousness and extroversion increase the overall trust value. On the contrary, the high level openness and neuroticism are correlated with decreasing the overall trust value.

In our simulation model, we introduce individuals into the social network who act maliciously independently. In the world of mass media communication, media plays a role in spreading information and shaping public opinions. In the context of social networking services, opinions from leaders or users with a high number of social links are highly influential compared to traditional media. As introduced by Zrnec et al. (2022), individual agendas have a significant correlation with leaders' agendas and are not dependent on media agendas. In this paper, we consider highly centralized users to be the most influential among the network. Although this is not directly correlated with leaders' agendas, we can see there is a possible relation among them. In this paper, we are only interested in how personality traits affect fake news dissemination, and addressing this issue will be an important point for future work.

8 Conclusion

In this paper, we proposed an analysis of how the information in SNS are disseminated using the factor of trust, Big-Five personality traits and agent-based modeling. In the proposed trust model, Big-Five personality traits were used for characterizing the personality of SNS users. The trust model was formed by five types of trust; identity-based, behavior-based, relation-based, feedback factor, and information-based. In order to evaluate the proposed trust model, we developed an agent-based simulation, conducting simulation experiments for the information dissemination in SNS. The overall trust value is low when the information-based trust is neglected while it is high when the behavior-based trust is neglected. The overall trust value is positively correlated with the trust level of the users. The overall trust value is decreasing when the probability of malicious users sharing fake news is high, while it is also decreasing when the number of malicious user increasing. Openness, conscientiousness, and extroversion are the attributes of the overall trust being increased, while the agreeableness and neuroticism decrease the overall trust of users.

Finally, our study addresses a valuable insight into how information is disseminated throughout the social network, and how personality can affect the reception of fake news. However, it is important to note the limitations of our research. The existing survey was conducted among a biased group, and the number of responses was relatively small. Therefore, there is a need for future work to conduct a survey that includes not only a large number of people but also a diverse range of individuals. Additionally, addressing the dissemination of fake news remains a significant issue. In future research, we aim to develop a system that effectively controls the dissemination of misinformation.

Author Contributions R.F.M designed this research, carried out the execution, and wrote the main manuscript. S.K. provided coordination and safety of this research. The authors read and approved the final manuscript.

Funding Funding This research was supported in part by Japan Society for the Promotion of Science under Grant-in-Aid for Challenging Research (Exploratory) No. 22K19776

Data availability Data is available upon reasonable request to the authors

Code Availability Code is available upon reasonable request to the authors

Declarations

Conflict of interest (check journal-specific guidelines for which heading to use) The authors declare that they have no Conflict of interest in this work.

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

References

Allcott H, Brennan M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31:211–236

Allport GW, Odbert HS (1936) Trait-names: a psycho-lexical study. Psychol Monogr 47(1):171

Antoci A, Bonelli L, Paglieri F, Reggiani T, Sabatini F (2019) Civility and trust in social media. J Econ Behav Organ 160:83–99. https://doi.org/10.1016/j.jebo.2019.02.026

Barabási A-L (2002) Linked: the new science of networks

Buchanan T (2021) Why do people spread false information online? the effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. PLoS ONE. https://doi.org/10.1371/journal.pone.0239666

Buchanan T, Kempley J (2021) Individual differences in sharing false political information on social media: direct and indirect effects of cognitive-perceptual schizotypy and psychopathy. Personal Individ Differ 182:111071. https://doi.org/10.1016/j.paid.2021.111071

Burbach L, Halbach P, Ziefle M, Calero Valdez A (2019) Who shares fake news in online social networks? In: Proceedings of the 27th ACM conference on user modeling, pp. 234–242

Burbach L, Nakayama J, Plettenberg N, Ziefle M, Valdez AC (2018) User preferences in recommendation algorithms: the influence of user diversity, trust, and product category on privacy perceptions in recommender algorithms. In: Proceedings of the 12th ACM conference on recommender systems. RecSys ’18, pp. 306–310. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3240323.3240393

Cheng X, Fu S, Vreede G-J (2017) Understanding trust influencing factors in social media communication: a qualitative study. Int J Inf Manage 37(2):25–35

Cigi-Ipsos: (2019) Cigi-Ipsos global survey internet security and trust 2019 part 3: social media, fake news and algorithms. Cigi-Ipsos. Accessed: 2020-03-15

Comfort LK, Ko K, Zagorecki A (2004) Coordination in rapidly evolving disaster response systems: the role of information. Am Behav Sci 48(3):295–313

Correa T, Hinsley AW, de Zúñiga HG (2010) Who interacts on the web?: the intersection of users’ personality and social media use. Comput Hum Behav 26(2):247–253. https://doi.org/10.1016/j.chb.2009.09.003

Costa P, McCrae RR (1999) A five-factor theory of personality. Five-Factor Model Personal Theor Perspect 2:51–87

Delone W, McLean E (1992) Information systems success: the quest for the dependent variable. Inf Syst Res 3:60–95. https://doi.org/10.1287/isre.3.1.60

DeYoung CG (2015) Cybernetic big five theory. J Res Pers 56:33–58

Erickson L (2011) Social media, social capital, and seniors: the impact of facebook on bonding and bridging social capital of individuals over 65. 17th Americas Conference on information systems 2011, AMCIS 2011 **1**

Esteves D, Reddy AJ, Chawla P, Lehmann J (2018) Belittling the source: trustworthiness indicators to obfuscate fake news on the web. CoRR **abs/1809.00494**arxiv:1809.00494

Gao Y, Li X, Li J, Gao Y, Philip SY (2019) Info-trust: a multi-criteria and adaptive trustworthiness calculation mechanism for information sources. IEEE Access 7:13999–14012

Hawkins JL, Weisberg C, Ray DW (1980) Spouse differences in communication style: preference, perception, behavior. J Marriage Fam 42(3):585–593

Jackson RC, Warren S, Abernethy B (2006) Anticipation skill and susceptibility to deceptive movement. Acta Physiol (Oxf) 123(3):355–371. https://doi.org/10.1016/j.actpsy.2006.02.002

Klimstra TA, Bleidorn W, Asendorpf JB, van Aken MAG, Denissen JJA (2013) Correlated change of big five personality traits across the lifespan: a search for determinants. J Res Pers 47(6):768–777. https://doi.org/10.1016/j.jrp.2013.08.004

Landrum AR, Eaves BS, Shafto P (2015) Learning to trust and trusting to learn: a theoretical framework. Trends Cogn Sci 19(3):109–111. https://doi.org/10.1016/j.tics.2014.12.007

Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild D, Schudson M, Sloman SA, Sunstein CR, Thorson EA, Watts DJ, Zittrain JL (2018) The science of fake news. Science 359(6380):1094–1096. https://doi.org/10.1126/science.aao2998

Liu F, Burton-Jones A, Xu D (2014) Rumors on social media in disasters: extending transmission to retransmission. In: PACIS 2014 Proceedings, p. 49

Lucassen T, Schraagen JM (2011) Factual accuracy and trust in information: the role of expertise. J Am Soc Inform Sci Technol 62(7):1232–1242

McCrae RR, Costa PT Jr (1997) Personality trait structure as a human universal. Am Psychol 52(5):509

Mohd Suki N (2014) Correlations of perceived flow, perceived system quality, perceived information quality, and perceived user trust on mobile social networking service (sns) users’ loyalty. J Inform Technol Res 5:1–14. https://doi.org/10.4018/jitr.2012040101

Murthy D (2018) Twitter: social communication in the Twitter Age, pp. 1–13. Cambridge, Polity Press, Cambridge

Neubaum G, Rösner L, Pütten AM, Krämer NC (2014) Psychosocial functions of social media usage in a disaster situation: a multimethodological approach. Comput Hum Behav 34:28–38

Nielsen RK, Fletcher R, Newman N, Brennen JS, Howard PN (2020) Navigating the 'Infodemic': how People in six countries access and rate news and information about coronavirus. Misinformation, science, and media. Reuters Institute, Oxford

Özer Ö, Zheng Y (2017) Establishing trust and trustworthiness for supply chain information sharing, pp. 287–312. Springer, Cham. https://doi.org/10.1007/978-3-319-32441-8_14

Pennycook G, Rand DG (2021) The psychology of fake news. Trends Cogn Sci 25(5):388–402. https://doi.org/10.1016/j.tics.2021.02.007

Preston S, Anderson A, Robertson DJ, Shephard MP, Huhe N (2021) Correction: detecting fake news on facebook: the role of emotional intelligence. PLoS ONE 16(10):0258719

Rand W, Herrmann J, Schein B, Vodopivec N (2015) An agent-based model of urgent diffusion in social media. J Artif Soc Soc Simul 18(2):1. https://doi.org/10.18564/jasss.2616

Riegelsberger J, Sasse MA, McCarthy JD (2003) Shiny happy people building trust? photos on e-commerce websites and consumer trust. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 121–128

Ross C, Orr ES, Sisic M, Arseneault JM, Simmering MG, Orr RR (2009) Personality and motivations associated with facebook use. Comput Human Behav 25(2):578–586. https://doi.org/10.1016/j.chb.2008.12.024

Rotenberg KJ (2018) The psychology of trust. Routledge, London

Rousseau DM, Sitkin SB, Burt RS, Camerer C (1998) Not so different after all: a cross-discipline view of trust. Acad Manag Rev 23(3):393–404

Ryu D, Abernethy B, Park SH, Mann L (2018) The perception of deceptive information can be enhanced by training that removes superficial visual information. Front Psychol. https://doi.org/10.3389/fpsyg.2018.01132

Schmitt DP, Allik J, McCrae RR, Benet-Martínez V (2007) The geographic distribution of big five personality traits: patterns and profiles of human self-description across 56 nations. J Cross Cult Psychol 38(2):173–212

Seidman G (2020) Personality traits and social media use. Int Encycl Media Psychol. https://doi.org/10.1002/9781119011071.iemp0295

Takahashi B, Tandoc EC Jr, Carmichael C (2015) Communicating on twitter during a disaster: an analysis of tweets during typhoon Haiyan in the Philippines. Comput Hum Behav 50:392–398

Twitter: the-algorithm. GitHub (2023)

Twomey C, O' Reilly G (2017) Associations of self-presentation on facebook with mental health and personality variables: a systematic review. Cyberpsychol Behav Soc Netw 20:587–595. https://doi.org/10.1089/cyber.2017.0247

Wang J-L, Jackson LA, Zhang D-J, Su Z-Q (2012) The relationships among the big five personality factors, self-esteem, narcissism, and sensation-seeking to chinese university students' uses of social networking sites (snss). Comput Hum Behav 28(6):2313–2319. https://doi.org/10.1016/j.chb.2012.07.001

Watson A (2022) News consumers who saw false or misleading information about selected topics in the last week worldwide as of february 2022, by region

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world'networks. Nature 393(6684):440–442

Zhou X, Zafarani R (2020) A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput Surv. https://doi.org/10.1145/3395046

Zrnec A, Poženel M, Lavbič D (2022) Users' ability to perceive misinformation: an information quality assessment approach. Inf Process Manage 59(1):102739. https://doi.org/10.1016/j.ipm.2021.102739