Modeling and Analysis of Uncertainty-based False Information Propagation in Social Networks

Jin-Hee Cho*, Trevor Cook*, Scott Rager†, John O'Donovan‡, and Sibel Adalı§
*US Army Research Laboratory, Adelphi, MD
Email: {jin-hee.cho.civ, trevor.j.cook.civ}@mail.mil
†Raytheon BBN Technologies, Cambridge, MA
Email: scott.rager@raytheon.com
‡University of California, Santa Barbara, CA
Email:jod@cs.ucsb.edu
§Rensselaer Polytechnic Institute, Troy, NY
Email: sibel@cs.rpi.edu

***Abstract*—To stop or mitigate the dissemination of false information in social networks, many studies have investigated the minimum number of seeding nodes required to significantly reduce the impact of false information. Although a person's confidence level, such as perceived certainty, and/or prior belief towards a given proposition can significantly affect their decision of whether to believe in true or false information, these topics have not been studied to date. In this work, we propose an opinion model based on *Subjective Logic* (SL), defining an opinion in belief, disbelief, and uncertainty, to study how to eradicate or mitigate the impact of false information by propagating true information to counter it. In the current form of an opinion in SL, when two agents interact with each other and update their opinions based on a consensus operator offered by SL, uncertainty continuously reduces whenever any new information, even conflicting evidence, is received. However, in reality, if a body of evidence is conflicting, with equal amounts supporting opposite positions, people often tend to be confused, leading to higher level of uncertainty. We enhance SL to deal with conflicting information associated with uncertainty. We map agents' opinion composition into each state in the SIR (Susceptible-Infected-Recovered) model to estimate the proportion of recovered agents who believe in true information. Our results show that agents' prior belief unfavoring false information can help guide their decisions towards a belief in true information even under high uncertainty. Further, having more true informers in a network can significantly increase the number of agents who believe in true information and the effect is more pronounced than having more frequent propagation of true information by fewer true informers.**

***Index Terms*—Subjective logic, opinion, false information, uncertainty, ambiguity.**

I. INTRODUCTION

The advent and proliferation of online social networks (OSNs) has been significantly contributing to propagating flooding amounts of information without any verification for its truthfulness. Many users in OSNs publish their own information or share others' information without checking its accuracy. Unfortunately, the unverified and potentially false information can drastically damage an individual's life or mislead the public for critical decision making such as voting for political leaders.

Two types of false information are *misinformation* and *disinformation*. While both contain false content, we distinguish between these two based on the intent of the person forwarding the information. Misinformation is wrong information disseminated by mistake or without bad intent, while disinformation is disseminated intentionally with the goal of misleading [13].

To study the propagation of false information and how to counter its effect, this work considers a social network where humans are modeled as agents interacting with other neighboring agents to form their own opinions. Agents' opinions evolve over time as they update their opinions based on their own perceived confidence level of their opinions and interactions with other agents. This work aims to model an agent's opinion and its evolution under uncertainty, which can be derived from lack of information or conflicting information. In addition, we study how to remove or mitigate the impact of false information by propagating true information. To this end, this work provides a model that answers the following research questions and study their impact on dissemination of false information:

- If conflicting evidence is received, how does it affect the agent's confidence (i.e., uncertainty) level of its opinion?
- How does an agent's prior belief affect its confidence and opinions used for a decision in whether to believe in true or false information?
- How does the frequency of true information propagating through the network or the number of true informers who originally disseminate true information affect the removal or mitigation of false information?

This work has the following key contributions:

- We model an agent's opinion based on Subjective Logic (SL) explicitly considering uncertainty in an opinion where the uncertainty can represent the agent's confidence level of its opinion. In particular, we consider the impact of conflicting evidence whose effect is not captured in the uncertainty dimension of an opinion in SL. The current SL only considers uncertainty in terms of lack of information by decreasing uncertainty continuously

upon receiving any new information although conflicting information may increase uncertainty.
- We associate an agent's expected opinion considering uncertainty of an opinion in SL with the agent's status in the SIR model [10] to capture agents that are recovered from false information and believe in true information.
- We conduct extensive simulation to investigate the key factors that impact the removal or mitigation of false information in terms of agents' opinions and the ratio of recovered agents.

II. Related Work

In this section, we discuss various approaches to correct or stop propagating false information in social networks and provide the background on SL which was used for formulating an agent's opinion in this work.

A. Countering False Information Propagation

We discuss the existing approaches to correct misinformation (or disinformation) in two directions: *network analysis-based approach* (NA) and *feature-based approach* (FA).

1) Network Analysis-based Approach (NA): Most NA aims to stop false information propagation by selecting a set of counter-misinformation nodes (i.e., nodes to disseminate true information against misinformation).

Influence Limitation (IL) algorithms have been used to counter misinformation based on Information Maximization (IM) approaches. Budak et al. [1] deal with information of competing campaigns in online social networks with the aim of limiting misinformation propagation by selecting a set of initial seed nodes propagating 'good' or 'bad' campaigns, which is an NP-hard problem.

Time constraints of misinformation propagation is also studied. Litou et al. [9] propose an information propagation model, called the *Dynamic Linear Threshold* (DLT) model, to distinguish credible information from misinformation. Nguyen et al. [11] study how to decontaminate the spread of misinformation in social networks with a small fraction of decontaminating nodes.

The process of misinformation is also modeled based on epidemic models. Zhao et al. [16] extends SIR (Susceptible-Infected-Recovered) model by incorporating the 'forgetting' factor to investigate the rumor propagation process based on the degree, infection rate, and forgetting rate. Zhao et al. [17] develop a Susceptible-Infected-Hibernator-Removed (SIHR) model that extends the SIR model by adding a direct link from ignorants (i.e., uninfected nodes by a rumor) to stiflers (i.e., stifling nodes against a rumor), which represents a type of people, hibernators.

2) Feature-based Approach (FA): FA is to identify the key features of false information or the sources (i.e., users) of the false information. Rumors can be spread out quickly based on a user's behavioral features or information features. Wang et al. [15] study the impact of uncertainty on information forwarding behavior for gossip diffusion in social networks. Liang et al. [8] take the rumor identification approach based on machine learning techniques in terms of feature design and selection.

Kumar and Geethakumari [5] develop an efficient misinformation detection algorithm based on the message attributes derived from cognitive psychology. Qazvinian et al. [12] propose rumor detection mechanisms by considering features based on content, network, and microblog platforms and validated the performance of the mechanisms using 10,000 tweets which are manually annotated.

Unlike the existing approach in both NA and FA, this work examines uncertainty of opinions in analyzing the propagation of false and true information considering agents' prior belief and their opinion updates upon receiving conflicting information.

B. Subjective Logic

SL is proposed to represent an opinion based on three dimensions, *belief*, *disbelief*, and *uncertainty*, where SL is derived by incorporating the perspectives from both probability models and logic [4]. SL has also been utilized in modeling uncertain belief under incomplete, unavailable information [2]. Recently, to deal with ambiguity derived from conflicting evidence, the trust revision algorithm is developed by discounting information based on the estimates of the information sources [4].

In this work, we adopt SL to explicitly deal with uncertainty an agent perceives in updating its opinion defined with belief, disbelief, and uncertainty. However, in SL, the degree of uncertainty is mainly considered in terms of the amount of information available to the agent. As more information becomes available, regardless of whether it is conflicting, uncertainty will decrease in SL [3]. The trust revision algorithm [4] can slow down the reduction of uncertainty; but uncertainty continues to decrease and reaches zero with the same amount of belief and disbelief. In such a situation, the model would not provide a useful support for the agent to make a decision under the same amount of information supporting the two extremes. To refine the limitation in SL, we propose a new operator that considers the increase of uncertainty upon receiving conflicting evidence, in addition to slowing down the reduction of uncertainty based on similarity-based opinion update. To the best of our knowledge, this work is the first that uses SL to solve a false information propagation problem where the opinion model deals with ambiguity derived from conflicting evidence.

III. Opinion Model

In SL, an *opinion* is represented by three dimensions: *belief* (b), *disbelief* (d), and *uncertainty* (u) [4]. A single opinion on a given proposition is represented by:

$$b, d, u \in [0, 1]^3, \quad b + d + u = 1 \tag{1}$$

A. Opinion Formation

The agent may have degrees of belief (i.e., agree) and/or disbelief (i.e., disagree) towards a given proposition with some degree of uncertainty. Agent i's opinion on proposition A is denoted as w_i^A. For simplicity, we omit A and use w_i to represent an agent i's opinion as:

$$w_i = \{b_i, d_i, u_i, a_i\} \tag{2}$$

where a_i is the base rate which normally represents general background knowledge [4]. a_i can be partially objective or partially subjective because individuals' observations are not the same. a_i can be interpreted as prior belief agent i has on a given proposition derived from past experiences (e.g., frequency-based probability) or mental reasoning process related to a bias or personal propensity in decision making process. In this work, either interpretation may be applicable. We examine its impact on decision making under uncertainty.

The base rate, a_i, affects expectation probability (i.e., a probability that an agent is expected to make a decision) in either belief or disbelief [4], denoted by E_{b_i} or E_{d_i}, respectively, where they are given by:

$$E_{b_i} = b_i + a_i u_i,\ E_{d_i} = d_i + (1 - a_i) u_i. \tag{3}$$

Note that $E_{b_i} + E_{d_i} = 1$ as $b_i + d_i + u_i = 1$.

In SL, an agent forms its opinion based on the amount of directly observed evidence. The following mapping rule is used to initialize agents' opinions:

$$b = \frac{r}{r+s+W}\ , d = \frac{s}{r+s+W}\ , u = \frac{W}{r+s+W}\ . \tag{4}$$

where r is positive evidence (i.e., agree) and s is negative evidence (i.e., disagree) for a particular proposition. For simplicity, we dropped the subscript i denoting the agent. W indicates the amount of uncertainty which can be affected by the inherent errors that can be introduced by the environment itself (e.g., imperfect observability).

B. Opinion Update

1) Like-Minded Opinion Consensus: Homophily in opinions, such as like-mindedness, significantly affects the way people update opinions [7]. In this work, the similarity of two agents' opinions, denoted by s_i^j, is computed based on *cosine similarity* [14] of the two opinion vectors i and j in terms of their belief and disbelief, (b_i, d_i) and (b_j, d_j), respectively. We omit the calculation of s_i^j due to the popularity of the cosine similarity. Note that agents' uncertainty levels are not considered to capture the similarity because having similar levels of uncertainty is not meaningful to update their opinions. The cosine similarity of the opinions of i and j is symmetric with $s_i^j = s_j^i$.

We use s_i^j, as a discounting operator [4], to determine the degree agent i accepts agent j's opinion. Given two vectors of opinions, $w_i = \{b_i, d_i, u_i\}$ and $w_j = \{b_j, d_j, u_j\}$, agent j's trust opinion in i's opinion, s_i^j, is given by $w_{i\otimes j} = \{b_{i\otimes j}, d_{i\otimes j}, u_{i\otimes j}\}$ where each element is estimated by:

$$b_{i\otimes j} = s_i^j b_j,\ d_{i\otimes j} = s_i^j d_j,\ u_{i\otimes j} = 1 - s_i^j (1 - u_j)\ . \tag{5}$$

Here, $u_{i\otimes j}$ is simply derived by $1 - b_{i\otimes j} - d_{i\otimes j}$ where $b_j + d_j + u_j = 1$. For simplicity, we omit the time step notation, but both sides of the equation refer to time step t.

We use SL's *consensus* operator [4] for an agent's opinion update upon receiving new information. The updated opinion of agent i after interaction with agent j is denoted as $w_i \oplus b_{i\otimes j} = \{b_i \oplus b_{i\otimes j}, d_i \oplus b_{i\otimes j}, u_i \oplus b_{i\otimes j}\}$ and each element is given by:

$$\begin{aligned} b_i \oplus b_{i\otimes j} &= \frac{b_i(1 - s_i^j(1-u_j)) + s_i^j b_j u_i}{\beta}\ , \\ d_i \oplus d_{i\otimes j} &= \frac{d_i(1 - s_i^j(1-u_j)) + s_i^j d_j u_i}{\beta}\ , \\ u_i \oplus u_{i\otimes j} &= 1 - \frac{(1 - u_i - s_i^j u_j)}{\beta}\ . \end{aligned} \tag{6}$$

where $\beta = u_i + 1 - s_i^j(1-u_j) - u_i(1 - s_i^j(1-u_j)) = 1 - s_i^j(1-u_i)(1-u_j)$ and $\beta \neq 0$ is assumed. $u_i \oplus u_{i\otimes j}$ is the same as $1 - (b_i \oplus b_{i\otimes j} + d_i \oplus d_{i\otimes j})$ and $u_i \oplus u_{i\otimes j}$ is simplified based on Eq. (5). Again, we omit the time step notation; here, the left side represents $w_i(t+1)$ while the right side uses the opinions at t such as $w_i(t)$ and $w_j(t) = w_{i\otimes j}(t)$.

2) Opinion Update to Deal with Ambiguity: In SL's original consensus operator [4], regardless of the differences between two opinions, agents' opinions are updated by increasing either belief or disbelief while continuously decreasing uncertainty. Even if Eq. (6) uses a discounter s_i^j to weight an agent's opinion more heavily from neighbors with similar opinions than those with dissimilar opinions, it still strictly decreases uncertainty, albeit slower than it would have without discounting. The uncertainty in SL mainly reduces when more evidence is received regardless of whether or not evidence is conflicting to previously received information. This effect is counter-intuitive to real life phenomena because receiving more information can create confusion when that information is supporting views of opposite extremes.

To capture uncertainty introduced by conflicting evidence, we propose an agent's opinion update that penalizes belief or disbelief values when conflicting information is received. In particular, we use the term *ambiguity* to refer to uncertainty derived from conflicting evidence. After agent i updates its opinion based on an interaction with agent j, as specified in Eq. (6), it then adjusts its opinion based on the distance between its own belief and disbelief. That is, if an agent has an opinion based on receiving the same amount of evidence supporting belief and disbelief ($b \approx d$), uncertainty should increase because conflicting evidence increases ambiguity. To this end, agent i with conflicting evidence adjusts its opinion on top of the updated opinion, $w_i = \{b_i, d_i, u_i\}$, based on Eq. (6). i's adjusted opinion, denoted by $w_i' = \{b_i', d_i', u_i'\}$, is formulated by weighting the distance between i's belief and i's disbelief with c_i, a conflict measure. We model c_i by:

$$c_i = \frac{|E_{b_i} - E_{d_i}|}{E_{b_i} + E_{d_i}} = |2b_i + 2a_i u_i - 1|\ . \tag{7}$$

where c_i is derived based on $b_i + d_i + u_i = 1$ and Eq. (3). Accordingly, $w_i' = \{b_i', d_i', u_i'\}$ is computed by:

$$b_i' = c_i b_i,\ d_i' = c_i d_i,\ u_i' = 1 - c_i(1 - u_i)\ . \tag{8}$$

u_i' can be derived based on $u_i' = 1 - (b_i' + d_i')$ where $b_i' + d_i' + u_i' = 1$. Note that we omitted the time step for simplicity. Both the left and right opinions are from the same time step (i.e., $t+1$ in Eq. (6)). In Eq. (8), c_i implies that when i has a similar level of expected belief (E_{b_i}) and expected disbelief (E_{d_i}), its belief (b_i') and disbelief (d_i'), as shown in Eq. (8), are significantly decreased. On the other hand, if E_{b_i} and E_{d_i} are significantly different, the original opinion based on the consensus operator, Eq. (6), is mostly kept intact. c_i can be equivalently represented by $|2b_i + 2a_i u_i - 1|$, as shown in Eq. (7). This form reflects that when b_i or d_i is relatively high, uncertainty, u_i, is low. On the other hand, when b_i and d_i are low, u_i must be high. Therefore, b_i and d_i are not heavily penalized simply when uncertainty is high; rather the penalty is higher when b_i and d_i are close to each other and u_i is relatively lower. This approach matches our desired real world scenario, because when u_i is high because of a lack of sufficient amount of information, new information can impact b_i and d_i. When uncertainty is low, however, and conflicting information is received, both b_i and d_i are reduced and replaced with rising u_i. Note that agent i's opinion can be updated by Eqs. (6) and (8) only when $u_i > 0$ (i.e., $\beta \neq 0$).

If agents do not have any prior belief towards belief (a_i) or disbelief ($1 - a_i$), an agent simply uses b_i and d_i to obtain c_i, instead of E_{b_i} and E_{d_i}, respectively. Thus, c_i is given by:

$$c_i = \frac{|b_i - d_i|}{b_i + d_i} = \frac{|2b_i + u_i - 1|}{1 - u_i} . \tag{9}$$

Eq. (9) clearly shows that when b_i and d_i are almost the same with small u_i, a higher penalty is applied than the case of b_i and d_i with high u_i. For simplicity, in Eqs. (8), (7), and (9), we omit notation for the time step. Here, both sides of the equations are at time step t.

An agent's opinion update can be summarized by two steps: (i) an agent updates its opinion based on Eq. (6) using a similarity-based discounting operator in Eq. (5); and (ii) the agent adjusts its opinion based on the distance between its belief and disbelief based on Eq. (7) (with prior belief/disbelief) or Eq. (9) (without prior belief/disbelief).

3) Opinion Decay over Time: Unless an agent receives new information by interacting with other agents, its opinion decays over time based on a decay factor, γ, over belief and disbelief while uncertainty increases in proportion to γ. For example, human cognition is limited by forgetting information over time. We model the decayed opinion by:

$$b_i = (1-\gamma)b_i,\ d_i = (1-\gamma)d_i\ , u_i = u_i + \gamma(1-u_i)\ . \tag{10}$$

Note that u_i is simply derived based on $1 - b_i - d_i$ where $b_i + d_i + u_i = 1$. Different from the opinion update by Eqs. (6) and (8) which allow the opinion update only for $\beta > 0$ and $u_i > 0$, respectively, the opinion decay based on Eq. (10) occurs at every time step. Therefore, even if u_i reaches 0, over time it can increase (i.e., $u_i > 0$) and accordingly agent i can update its opinion upon receiving new information from its neighbors. For simplicity, we omitted the time step notation, but the left side is at time $t+1$ while the right side is at t.

IV. Agent Model

This work considers an online social network as an undirected graph $\mathcal{G}$ where vertices, v_i's, are agents i's (e.g., users) in the set of $\mathcal{V}$ and the edges, e_{ij}'s (i.e., 1 for an edge and 0 for no edge), represent the relationships in the set $\mathcal{E}$. Agent i's neighbors refer to other agents directly connected to i. For information propagation, we set an initial number of seeding agents propagating false information, called *false informers* while setting an initial number of seeding agents disseminating true information, *true informers*, to counter the propagation of the false information.

A. Types of Agents

In an agent's opinion towards a given proposition, $w = \{b, d, u\}$, b is an agent's belief agreeing with false information, d is the agent's disbelief in false information, and u is the degree of uncertainty an agent has in its belief or disbelief (i.e., I don't know). An agent's initial opinion is set based on the mapping rule in Eq. (4) with different values of r, s, and W, depending on the type of agent. We consider the following three types of agents:

- *False Informer* (FI) disseminates false information with bad intention. The opinion of this agent is initialized with $(r, s, W) = (n, 1, 1)$ where $n >> 1$ (e.g., 1000), leading to $\{b, d, u\} = \{\frac{n}{n+2}, \frac{1}{n+2}, \frac{1}{n+2}\}$ implying that the agent highly agrees with false information (i.e., $b \to 1$) while it has low disbelief (i.e., $d \to 0$) and low uncertainty (i.e., $u \to 0$) towards the false information. FI does not change its opinion while influencing others' opinions.
- *Truth Informer* (TI) disseminates true information to counter false information. This agent starts its opinion with $(r, s, W) = (1, n, 1)$ where $n >> 1$, leading to $\{b, d, u\} = \{\frac{1}{n+2}, \frac{n}{n+2}, \frac{1}{n+2}\}$. This means that the agent disagrees with the false information while agreeing with true information. This is represented by low belief (i.e., $b \to 0$), high disbelief (i.e., $d \to 1$), and low uncertainty (i.e., $u \to 0$). Similar to FI, TI also does not change its opinion while influencing other agents' opinion.
- *Doubter* (D) has low confidence (i.e., $u \to 1$) in its own opinion by not initially agreeing or disagreeing with given false information (i.e., $b \to 0$ and $d \to 0$). This agent is initialized with its opinion with $(r, s, W) = (1, 1, n)$ where $n >> 1$, leading to $\{b, d, u\} = \{\frac{1}{n+2}, \frac{1}{n+2}, \frac{n}{n+2}\}$, implying low confidence in a given proposition due to lack of information (i.e., ignorance). Unlike FI and TI, D updates its opinion upon interacting with other agents as long as its uncertainty does not reach 0.

For our experimental analysis in Section V, we use an initial number of seeding nodes representing FIs, TIs, and Ds, denoted by s_f, s_t, and s_d, respectively.

B. Epidemic Status of Agents

We model the evolution of false information propagation using a variant of the SIR (Susceptible-Infected-Recovered) model [10]. The three states in the SIR model are defined based on the conditions associated with the expected belief or disbelief probabilities, E_{b_i} and E_{d_i}, as follows:

- *Susceptible* (S): An agent is not sure of its status in terms of whether it believes in false information or not. Agents in S have opinions with $E_b \leq 0.5$ and $E_d \leq 0.5$;
- *Infected* (I): An agent believes in false information to be true with $E_b > 0.5$; and
- *Recovered* (R): An agent believes in false information to be false (or believes in true information) with $E_d > 0.5$.

Among the three types of agents, FIs and TIs will not change their status and be fixed with I and R, respectively. On the other hand, Ds will change their status based on their updated opinion, with all Ds starting in state S. The epidemic state transitions possible for Ds are summarized in Table I. Note that when $u > 0$, an agent can move from its current status to any other status based on each transition condition.

The simulation will start with first all FIs propagating false information to their neighbors. Then, TIs propagate true information to their neighbors for m steps. Each propagation step continues until all Ds propagate to their neighbors. This approach reflects the nature of opinion propagation in social networks more accurately since in reality, people are not necessarily delivering what they exactly heard, but what they thought about it. The opinion updates are computed using Eqs. (6) and (8) with a perfect influence probability (i.e., 1). For simplicity, we adopt the high effectiveness (influence) property [1] assuming that when j propagates its opinion to its neighbor i, then i will update its opinion. Deriving the so-called different influence probabilities of one agent to another agent based on the informer's influence or receiver's characteristics will be considered in the future work and is out of the scope of this work.

V. Numerical Results and Analysis

In this section, we describe metrics used for our experiments and the details of our experimental setup. Further, we analyze the experimental results and discuss their overall trends.

A. Metrics

The following metrics are used for our experiments:

- *Agents' opinions*: Ds' opinions are captured in a histogram showing agents' belief, disbelief, and uncertainty.
- *SIR Ratio*: This refers to the ratio of agents in susceptible ($\mathcal{S}(t)$), infected ($\mathcal{I}(t)$), or recovered ($\mathcal{R}(t)$) status over the total number of doubters (s_d). We also use the aggregated metric for the recovered agents (R) across the T times of information propagation where all agents disseminate their opinions to their neighbors per time step (t) for $t = 1 - T$.

B. Experimental Setup

For the network topology, we use an ego-Facebook dataset [6] which gives a fully connected undirected network described by Table III. The key design parameters and their default values used for the environment setup are summarized in Table II. In Section V-C, we also vary parameters, including the number of true information propagation steps after the false information propagation (m), s_t and a_i to investigate their impact on the metrics in Section V-A.

In order for false information to be eradicated or mitigated from the network, we allow TIs to propagate true information more than FIs propagate false information. For example, at $t = 1$, FIs propagate false information over the network, and Ds receiving the false information propagate their opinions to their neighbors until each doubter has had a chance to propagate its opinion to its neighbors. For the next m time steps until $t = 1+m$, TIs propagate their opinions to their neighbors and then each doubter subsequently propagates their updated opinions to their own neighbors. By varying m, we examine the effect of increasing the frequency of true information propagation by a given set of TIs on opinions and SIR ratio.

C. Simulation Results

Fig 1 shows the evolution of opinions with respect to time steps for $m = 1$ or $m = 5$ where doubters (Ds) do not have prior belief or disbelief. Using $m = 1$ depicts the situation in which the same amounts of true and false information are propagated by FIs and TIs, one for each party, respectively. Using $m = 5$ means that after FIs propagate false information once, then TIs propagate true information five times. This pattern continues until t reaches T.

Findings from Fig. 1: (1) When agents do not have any prior belief and disbelief, they are vulnerable to adjust their opinion to a high belief with low uncertainty to the first information to which they are exposed. Since in this case, they are first exposed to false information, the majority of Ds transition to the Infected state and are unlikely to move from this state, even when the false information is followed by five instances of true information, as in Figs. 1b and 1d; and (2) in Figs. 1a and 1c, low uncertainty is observed because prior belief/disbelief is not used for computing the penalty based on the conflicting evidence based on Eq. (9).

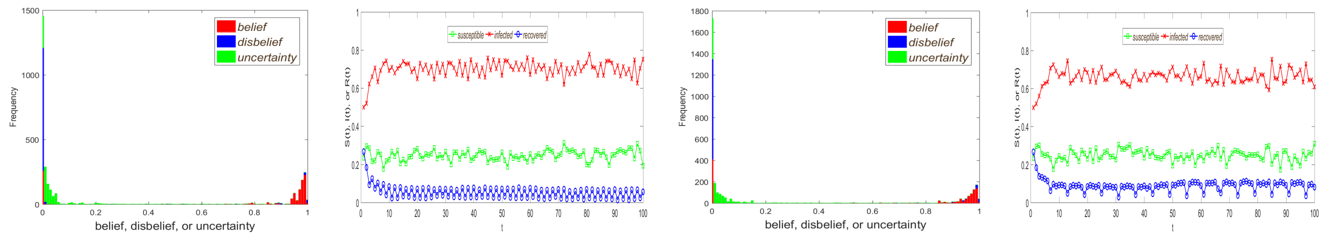
Fig. 2 uses the same setting of Fig. 1 except that agents have prior belief/disbelief.

Findings from Fig. 2: (1) Prior belief and disbelief with an equal weight ($a_i = 0.5$ and $1 - a_i = 0.5$) helps decisions to join true information as shown in Figs. 2b and 2d; more frequent true information propagation (i.e., $m = 5$) significantly increases the number of recovered agents by more Ds joining in believing in true information; and (2) compared to Fig. 1 with no prior belief, lesser uncertainty is observed particularly in 2c using higher m; opinion updates to deal with conflicting evidence based on Eqs. (7) and (8) allow true information to govern over false information because unnecessary increase of uncertainty due to ambiguity derived from conflicting evidence can be avoided.

Fig. 3 shows the effect of varying key design parameter values on the ratio of recovered agents (R). Fig. 3a examines the effect of m on R. In this case, we fix the % of FIs and TIs to 1% and 10%, respectively. All Ds use prior belief/disbelief with opinion decay factor, $\gamma = 0.05$.

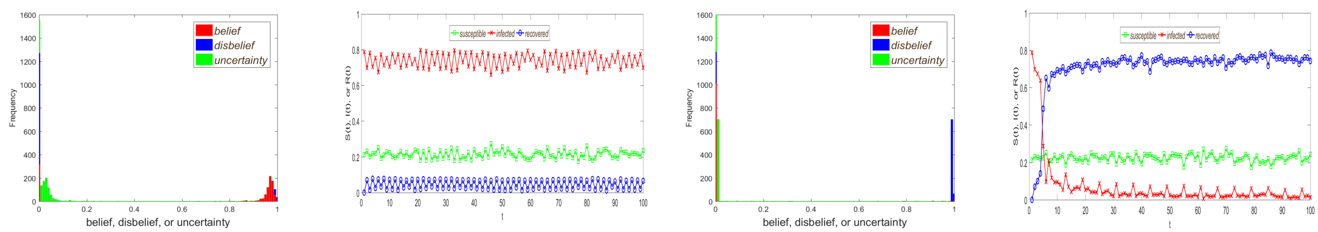
TABLE I: Transitions of SIR Epidemic Process by Doubters (Ds).

Status Change	Susceptible (S)	Infected (I)	Recovered (R)
SI	Current: $E_b \leq 0.5, E_d \leq 0.5, u > 0$	Next: $E_b > 0.5$	
IS	Next: $E_b \leq 0.5, E_d \leq 0.5$	Current: $E_b > 0.5, u > 0$	
IR		Current: $E_b > 0.5, u > 0$	Next: $E_d > 0.5$
RI		Next: $E_b > 0.5$	Current: $E_d > 0.5, u > 0$
RS	Next: $E_b \leq 0.5, E_d \leq 0.5$		Current: $E_d > 0.5, u > 0$
SR	Current: $E_b \leq 0.5, E_d \leq 0.5, u > 0$		Next: $E_d > 0.5$



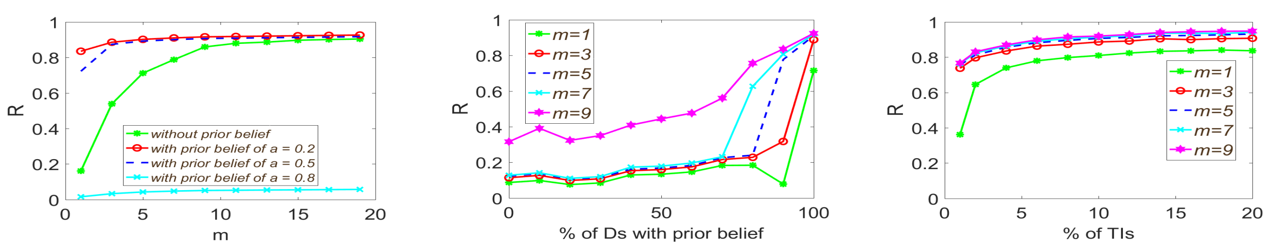
(a) Opinion histogram with $m = 1$ (b) SIR ratio with $m = 1$ (c) Opinion histogram with $m = 5$ (d) SIR ratio with $m = 5$

Fig. 1: Opinion histograms and SIR ratio without prior belief.



(a) Opinion histogram with $m = 1$ (b) SIR ratio with $m = 1$ (c) Opinion histogram with $m = 5$ (d) SIR ratio with $m = 5$

Fig. 2: Opinion histograms, opinion values, and SIR ratio with prior belief.



(a) R vs. m (b) R vs. % of Ds with prior belief (c) R vs. % of true informers

Fig. 3: Ratio of recovered agents (R) with respect to (a) varying the number of true information propagation per false information propagation (m); (b) the number of doubters with prior belief; and (c) the percentage of true informers.

TABLE II: Key parameters and their default values

param.	val.	param.	val.	param.	val.
n	1000	T	100	γ	0.05
$s_f = s_t$	11 ($\approx 1\%$)	s_d	1011 ($\approx 98\%$)	a_i	0.5

TABLE III: Network dataset statistics

N	1033	Ave. degree	51.785	Ave. path length	2.949
$\lvert\mathcal{E}\rvert$	26747	Modularity	0.54	Ave. clustering coeff.	0.534

Findings from Fig. 3a: (1) When Ds do not use prior belief/disbelief, a sufficiently large m is needed to reach the same R attained by Ds with prior belief; (2) when a D's prior belief is weighted against false information (i.e., $a \leq 0.5$), more TIs than FIs (i.e., 10% for TIs and 1% for FIs among N) cause most Ds to correctly disbelieve false information even with small m; and (3) if prior belief (a) is more weighted toward believing in false information (i.e., $a \geq 0.5$), the opposite effect occurs, in which almost all Ds believe in false information, no matter how high the ratio of true information to false information (i.e., $m : 1$) is.

Fig. 3b shows how varying the number of Ds with prior belief/disbelief and m affects R when all Ds' prior belief is set to $a = 0.5$.

Findings from Fig. 3b: (1) Larger m and more Ds with $a = 0.5$ increases R; in terms of m, the case with $m = 9$ outperforms the cases with $m = 1 - 7$ in R; and (2) when % of Ds with prior belief (s_d^*) is larger than 70, R significantly increases, compared to the case with $s_d^* < 70$.

Fig. 3c shows the impact of varying the % of TIs and m on R.

Findings from Fig. 3c: (1) Although more TIs (s_t) increases R, after a certain s_t (e.g., $s_t > 5\%$), the effect shows diminishing returns. This implies that to maximize the effect on R, a small number of s_t suffices; the effect of the set s_t on R can also be maximized by selecting TIs with high centrality, which we plan to investigate in future work; and (2) across all s_t's, for $m > 1$, larger m do not significantly increase R.

VI. Conclusion

We proposed an opinion model based on Subjective Logic [4] to deal with uncertainty derived from conflicting evidence and applied it to study false information propagation. In traditional SL, uncertainty is only considered in terms of lack of information, so it is strictly reduced upon receiving more evidence, regardless of receiving conflicting information. In this work, we enhanced the opinion model to deal with the so called ambiguity, considered as the uncertainty derived from conflicting evidence. The proposed opinion model uses expected probabilities in SL where an agent's prior belief or disbelief can significantly affect the expected belief and disbelief for an agent's decision on whether to believe in true or false information. In addition, we measured the recovered agents using the SIR model where each agent's expected belief and disbelief are mapped into the status (S, I, or R) to determine an agent's decision of believing in true or false information. We conducted sensitivity analysis by varying the amount of true information, compared to false information being propagated and agents' prior beliefs to investigate their impact on the opinions of the agents and the ratio of recovered agents. The key findings from our experiments include: (1) although prior belief can introduce higher uncertainty, it can lean the agent's decision towards a belief in true information because prior belief/disbelief can be used to support one side of an opinion, belief or disbelief, with a small amount of evidence; and (2) a small number of true informers can suffice to sway the majority of agents to believe in true information as long as agents do not have a high prior belief in false information.

Acknowledgment

This work was partially supported by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053; The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL, NSF, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] C. Budak, D. Agrawal, and A. E. Abbadi, "Limiting the spread of misinformation in social networks," in *ACM International World Wide Web Conference*, 2011.

[2] J. Cho and A. Swami, "Dynamics of uncertain opinions in social networks," in *IEEE Military Communications Conference (MILCOM)*, Baltimore, MD, 2014.

[3] A. Jøsang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 9, no. 3, June 2001.

[4] ——, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer, 2016.

[5] K. P. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," *Human-centric Computing and Information Sciences*, vol. 4, no. 1, p. 14, 2014.

[6] J. Leskovec. (2012) Social circles: Facebook. [Online]. Available: https://snap.stanford.edu/data/egonets-Facebook.html

[7] L. Li, A. Scaglione, A. Swami, and Q. Zhao, "Opinion dynamics and learning in social networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 3–49, March 2011.

[8] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, "Rumor identification in microblogging systems based on users' behavior," *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 99–108, Sept 2015.

[9] I. Litou, V. Kalogeraki, I. Katakis, and D. Gunopulos, "Real-time and cost-effective limitation of misinformation propagation," in *17th IEEE International Conference on Mobile Data Management*, 2016.

[10] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.

[11] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz, "Containment of misinformation spread in online social networks," in *Proceedings of the 4th Annual ACM Web Science Conference*, New York, NY, USA, 2012, pp. 213–222.

[12] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 1589–1599.

[13] B. C. Stahl, "On the difference or equality of information, misinformation, and disinformation: A critical research perspective," *Informing Science Journal*, vol. 9, 2006.

[14] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[15] Y. Wang, A. V. Vasilakos, J. Ma, and N. Xiong, "On studying the impact of uncertainty on behavior diffusion in social networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 2, pp. 185–197, Feb 2015.

[16] L. Zhao, Q. Wang, J. Cheng, Y. Chen, J. Wang, and W. Huang, "Rumor spreading model with consideration of forgetting mechanism: A case of online blogging livejournal," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 13, pp. 2619–2625, 2011.

[17] L. Zhao, J. Wang, Y. Chen, Q. Wang, J. Cheng, and H. Cui, "SIHR rumor spreading model in social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 7, pp. 2444–2453, 2012.