CrossMark

Modelling majority and expert influences on opinion formation in online social networks

Rajkumar Das[1] ⬤ · **Joarder Kamruzzaman[1,2]** ·
Gour Karmakar[1,2]

Abstract Two most important social influences that shape the opinion formation process are: (i) the majority influence caused by the existence of a large group of people sharing similar opinions and (ii) the expert influence originated from the presence of experts in a social group. When these two effects contradict each other in real life, they may pull the public opinions towards their respective directions. Existing models on opinion formation utilised the idea of expertise levels in conjunction with the expressed opinions of the agents to encapsulate the expert effect. However, they have disregarded the explicit consideration of the majority effect, and thereby failed to capture the concurrent and combined impact of these two influences on opinion evolution. To represent the majority and expert impacts, we explicitly use the concept of opinion consistency and expertise level consistency respectively in an innovative way by capitalizing the notion of entropy in measuring the homogeneity of a group. Consequently, our model successfully captures the opinion dynamics under the concomitant influence of majority and expert. We validate the efficacy of our model in capturing opinion dynamics in a real world scenario using the opinion evolution traces collected from a widely used online social network (OSN) platform. Moreover, simulation results reveal the impact of the aforementioned effects, and confirm that

✉ Rajkumar Das
  rajkumar.das@monash.edu

  Joarder Kamruzzaman
  joarder.kamruzzaman@federation.edu.au

  Gour Karmakar
  gour.karmakar@federation.edu.au

[1] Faculty of Information Technology, Monash University, Victoria, Australia

[2] School of Engineering and Information Technology, Federation University Australia,
  Victoria, Australia

⚫ Springer

our model can properly capture the consensus, polarization and fragmentation properties of public opinion. Our model is also compared with some recent models to evaluate its performance in both real world and simulated environments.

Keywords Opinion formation · Online social networks · Majority · Expert · Consistency

1 Introduction

Opinion formation dynamics captures the way individual opinions in a social group evolve under the presence of various social influences to form a collective opinion in society. Social influencs, such as conformation in a majority group [1], expertise in neighbour-hood and homophily, instigate repeated interactions among the group members and enforce them to revise their thoughts or adopt new information or refine their beliefs. Modelling such dynamics helps us to better understand the characteristics of these influences, and thereby helps to better understand the characteristics and composition of the final opinion. Moreover, capturing this opinion formation dynamics is of great significance as the developed models can be used in many cutting-edge applications in the World Wide Web (WWW) space. Examples of business related applications include analysing the evolution of the popularity of a product or business idea using customers' feedback, identification of experts and influential people and estimation of their influence for selecting target group for marketing, and investigating the success of viral marketing. Likewise, other applications include detection and monitoring of any socio-economic crisis or movement and then taking corresponding preventive steps, dissemination of information on new products, ideas and innovations, predicting the popularity of political parties and their agenda or strategies taken by government, and so on. If more realistic and accurate opinion estimation can be done, it will increase the likelihood of achieving the business and the operational objectives of the above mentioned applications. However, human behaviour is too complex to be abstracted, which makes the research in emulating the real world opinion evolution process more comprehensive and challenging.

The seminal works of DeGroot [12], Clifford and Sudbury [7], Hegselmann and Krause [16], Deffuant et al. [11], and Sznajd [32] have produced several opinion formation models in the literature. Based on the opinion representations and the corresponding opinion update rules, the existing models can be approximately categorized into two major streams. In the first modelling approach [11, 12, 16, 18, 34], opinions are represented using continuous values in a range and updated through repeated averaging of neighbours' opinions. Besides, the neighbours are selected using one or a combination of three neighbour selection strategies, namely: (i) all the neighbours [12], (ii) a selected subset of neighbours [16], and (iii) a randomly selected neighbour [11]. The other approach [7–9, 13, 17, 21] considers discrete valued opinions by representing them using some fixed discrete levels. In the latter approach, the update rule implies adoption of a neighbour's opinion using a selection strategy, e.g. a random neighbour [7], a probabilistically selected neighbour [9] and neighbour selected based on majority rule [8, 13].

The expert effect and the majority effect are two very important social influences that shape these neighbour selection and opinion formation process. Here, experts are agents with knowledge and experience, and whom other agents can rely on. They have the potential to draw other agents' opinions towards them. The majority effect is the other important driving force in opinion formation where opinions of a majority group from the neighbours

can drive the dynamics towards its direction. A strong community always attracts others to revise their opinions. It is noteworthy to mention here that sampling a population is the widely used way to predict the public opinion, and in most of the cases mining the majority opinion from the sample yields superior prediction of the outcome of opinion dynamics. The superiority of mining majority opinion for prediction of public opinion can be established through several examples, such as the poll prediction for 'Brexit', 'Scottish Referendum' and the 'US presidential election' (2008–2012). The importance of majority opinion in mining public opinion motivates us to integrate the majority rule in our modelling aspects. Since the majority rule does have reasonable fairness, it is universally adopted to make predictive analysis of many real-world influential events such as elections and referendums. Despite there exist a few debates about it (e.g., minority rights are neglected, failing to consider the intensity of preferences), the most popular discrete opinion formation process, namely 'Voter Model' [7] and its variants [8, 13, 17, 21] have embedded majority rule in their bedrocks as a fundamental principle for deriving opinions of agents participating in opinion dynamics. On the contrary, continuous opinion formation models lack such exploration of majority rule adequately. Nevertheless, in this paper, we study continuous valued opinion formation modelling as it can represent opinion in a desirable granular level to capture the dynamics more accurately.

To encode the expert effect, continuous opinion dynamics incorporates the concept of expertise level of an agent on its expressed opinion [6, 24]. Expertise level can be explicitly expressed by an agent as in the review system of technical articles where each review score is coupled with a level of expertise of the reviewer. If not explicitly expressed, it is possible to extract expertise level of an agent through text analysis of user activities and credentials using data such as user profile and posts in online social networks. Moussaid et al. [24] utilise an agent's confidence level in representing her level of expertise in their modelling. They distinguish experts as agents having very high confidence levels. However, the efficacy of their approach is limited due to (i) the use of a few intuitively defined opinion update rules considering a particular appication context and (ii) limited pair-wise interactions among the agents. Consequently, the model fails to capture the majority effect, i.e., the collective impact of a group of agents sharing similar opinions. On the other hand, uncertainty in beliefs is explored to determine the expertise level of an agent by Cho and Swami [6]. Likewise [24], work in [6] only employs pairwise interactions among the agents. As a result, both models fail to realize the conformity of an agent to a group with the majority of them sharing similar opinions (majority effect), while in Asch experiment [1] conformity to majority group is exhibited as a strong social influencing factor. Moreover, pairwise agent interactions can only capture the influence of expert agents on non-expert ones at individual level. This overlooks the influence due to the presence of a group of experts in an agent's neighbourhood.

In a nutshell, though majority effect has been exploited by a stream of discrete opinion models [32] and to some extent expert effect has been considered in some continuous models [6, 24], none of the models based on these two streams has considered the combined effect of majority and expert. It is clearly evident that in any real-world opinion formation process, the two effects occur concomitantly [24], thus both the majority supported opinions and experts' opinions coexist. In some occasions, the majority opinion is backed by experts or the expert opinion is supported by the majority people. On the contrary, in some situations, there might be disagreement between these two types of opinions. This necessitates the importance of combining majority effect with the impact of experts to study their joint influence on opinion formation, which is missing in [24] and in other works on opinion formation, but addressed in this paper.

While modelling the aforementioned combined influence, we emphasize on encoding the majority rule for continuous dynamics as it is not explored adequetly in the existing models. Complete random models [11, 19, 24] don't consider the majority effect at all, while the other existing models (averaging and bounded confidence models [12, 16]) partially capture the majority impacts. Averaging models take the average of all the neighbours' opinions of an agent without considering the fact that majority of them may or may not agree. Moreover, averaging models are adversely affected by outlier opinions. On contrary, in the bounded confidence models only a fraction of opinions conforming to the agent are considered. Those opinions are selected by a threshold of opinion distance from the agent. However, majority group can form beyond this opinion distance threshold. Even if the measure based on the opinion distance can properly capture the homophily principle present in social interaction, it overlooks the fact that a disagreeing majority opinion can have strong impact on agents and thereby shapes their opinion revision process. The value of consistency is widely used to represent the degree of homogeneity of a group. We use consistency to represent the impact of majority in an innovative way [10], and thus decouple it from other types of social influences (e.g., outliers, homophily). Here, we utilise the entropy of a set of opinions to measure such consistency. The majority rule suggests that, the closeness among majority supported opinions plays a vital role in an agent's opinion update process when it encounters with other agents' opinions. When the majority of neighbours share similar opinions, their influence as a group on an agent will be greater than their individual impact. Our definition of consistency using entropy [10] essentially captures the presence of such majority group in an agent's neighbourhood.

In brief, our research has the following contributions:

1. decouple majority influence from expert effect and formulate their individual and combined impacts,
2. utilise the concept of opinion consistency to measure the majority effect, while utilise both the value of expertise level and expertise level consistency to evaluate the expert effect, and use entropy in an innovative way to measure these consistencies,
3. detect the formation of a group of experts' opinions and its influence on nearby agents,
4. evaluate the efficacy of the proposed model using data collected from OSN.

The rest of the paper is organized as follows. In Section 2, we describe in details the formulation of the model's parameters (2.1) followed by the neighbour selection and opinion update strategies (2.2). Then in Section 3, we validate our proposed model using real world opinion formation data collected from Twitter. We investigate the characteristics of our model in more depth through simulation in Section 4.

2 Proposed opinion formation model

Consider a social network $G = (V, E)$ with $|V| = n$ agents forming a collective opinion on a particular topic. An agent $i \in V$ interacts to all its neighbours, $N_i = \{j | j \in V \land (i, j) \in E\}$ in each iteration. At time t, the opinion expressed by agent i is represented by $O_i(t) \in [0, 1]$. The range is called the Opinion Space (OS). Every agent is also associated with an expertise level $\mathcal{E}_i(t) \in [0, 1]$ besides its own expressed opinion $O_i(t)$. A agent with higher expertise level can be considered an expert whom other agents can generally rely on. In contrast, individuals with very low expertise levels represent lay people having poor topical knowledge and thus, very hard to be relied on for any information on that topic. To jointly integrate the majority and expert effects, our model considers four factors

that an agent is usually influenced by: (i) its own opinion $O_i(t)$, ii) its own expertise level $\mathcal{E}_i(t)$, (iii) neighbours' opinions $O_{N_i}(t) = \{O_j(t)|j \in N_i\}$, and (iv) neighbours' expertise levels $\mathcal{E}_{N_i}(t) = \{\mathcal{E}_j(t)|j \in N_i\}$. It is worth to mention here that although the opinions and expertise levels used in our models are numerical, people express their opinions in OSNs using text messages. To get opinions, we convert textual information into numerical values using a well-established method called sentiment analysis [22, 23, 27]. However, expertise level extraction is not straightforward as people will not necessarily express their expertise levels while expressing their opinions. We need to mine their activities and credentials to get their expertise levels. We will discuss more on opinion and expertise level extraction processes in Sections 3.2 and 3.3, respectively.

According to the literature [9, 24], in real life, opinion update process considers one of the three possible heuristics: (i) keep own opinion, (ii) make a compromise by averaging with neighbours' opinions and (iii) adopt neighbour's opinion; the choice is made based on the available information. To make opinion update as realistic as possible, our model considers all of them. Before describing how the opinion update process is implemented, the way our model captures the combined effect of majority and expert is described first in the following section.

2.1 Formulation of majority and expert effects

To measure the combined impact of the majority and expert effects, an agent computes the credibility of its neighbours' opinions considering the aforementioned four influencing factors (agent's and its neighbours' opinions and expertise levels). Here, the credibility score determines how much an agent can rely on its neighbours. To compute credibility, we need to individually measure the majority and expert effect present in the neighbourhood, and then combine them to get the overall credibility of neighbours' opinions. We use consistency for opinions and expertise levels to encode the majority and expert effect, respectively. Consistency ($\xi(A)$) of a set (A) of values indicates the overall similarity present in the set. The Shannon's entropy from information theory properly represents the consistency; a set with similar values results in a small entropy whereas diverge valued set has a large entropy. Consequently, a set's entropy normalized with the maximum possible entropy (e_m) [15] is a reasonable realization of its consistency.

Opinion consistency (majority effect) Opinion consistency of a neighbourhood is calculated from the entropy value as per (1).

$$\xi(O_{N_i}(t)) = 1 - \frac{e(O_{N_i}(t))}{e_m} \tag{1}$$

where, $e(O_{N_i}(t)) = \text{entropy}(O_{N_i}(t))$

$$= -\sum\nolimits_{O(t) \in O_{N_i}(t)} p(O(t)) \times \log(p(O(t)))$$

Here, $p(O(t))$ is the probability of opinion $O(t)$ in neighbours' opinion set $O_{N_i}(t)$. For simplicity in entropy computation, we assume that the Opinion Space (OS) comprises a set $B = \{b_1, b_2, \cdots, b_z\}$ of equal length blocks. Figure 1a illustrates an example distribution of an agent's neighbours' opinions using 10 blocks. For entropy computation, instead of evaluating individual opinion's probability $p(O(t))$, we calculate the probability of the
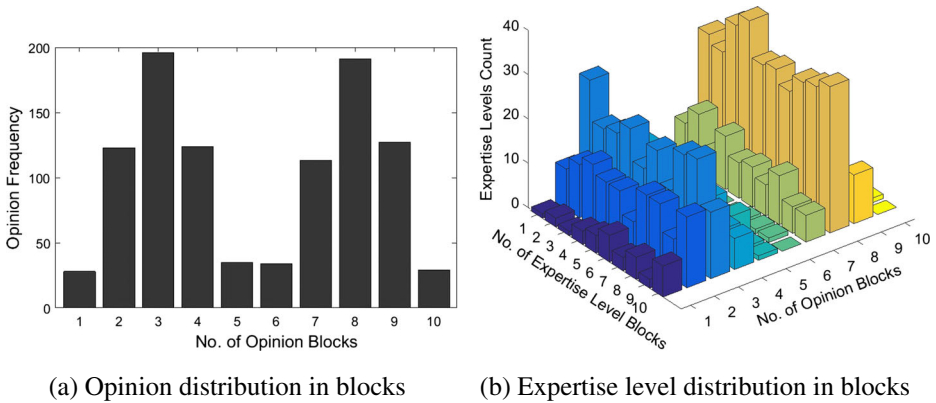
(a) Opinion distribution in blocks   (b) Expertise level distribution in blocks

Figure 1 Illustration of opinion and expertise level distributions in blocks to measure their consistencies

individual blocks. Here, a block's probability depends on the number of opinions belonging to that block. Thus, probability of a block b as represented by $p_{ib}(t)$ is computed as

$$p_{ib}(t) = \frac{|O_{ib}(t)|}{|O_{N_i}(t)|} \tag{2}$$

where, $O_{ib}(t)$ is the set of neighbours' opinions of agent i which belongs to block b and $|.|$ computes the cardinality of a set.

A high $\xi(O_{N_i}(t))$ indicates the presence of a large group of agents holding similar opinions, thus captures the majority effect. A low value designates no such majority gathering.

Expertise level consistency and expertise level credibility (expert effect) While computing expertise level consistency, we again take into account the possibility of forming different opinion groups and measure individual groups' expert effects. To compute the expert effect that a group exhibits, we need to find the consistency of the expertise levels of the agents whose opinions are member of that group (opinion block).

So, expertise levels inside an opinion block are further divided into expertise level blocks, as illustrated in Figure 1b, to compute consistency. As depicted in the figure, two same sized opinion blocks (3 and 8 of Figure 1a) with similar expertise level consistencies can have different expertise level values. Here, the agents of opinion block 8 have higher expertise levels than the agents in block 3. Thus, instead of having equal consistencies, block 8 will exhibit greater expert effect than its counterpart. Therefore, both the consistency and the values of the expertise levels determine the overall expert effect of a group of opinions. We incorporate the expected expertise level of the opinion groups to account for their values. This compound measure is named as expertise level credibility which is computed as per (3) for an opinion block $b \in B$ in an agent i's neighbourhood.

$$\Xi(\mathcal{E}_{ib}(t)) = \log(\xi(\mathcal{E}_{ib}(t)) \times \mathbb{E}(\mathcal{E}_{ib}(t)) + \eta)/\log(1 + \eta) \text{ where,} \tag{3}$$

$$\xi(\mathcal{E}_{ib}(t)) = 1 - \frac{e(\mathcal{E}_{ib}(t))}{e_m},$$

$$e(\mathcal{E}_{ib}(t)) = -\sum\nolimits_{\mathcal{E}(t) \in \mathcal{E}_{ib}(t)} p(\mathcal{E}(t)) \times \log(p(\mathcal{E}(t))),$$

$$\mathbb{E}(\mathcal{E}_{ib}(t)) = \sum\nolimits_{\mathcal{E}(t) \in \mathcal{E}_{ib}(t)} p(\mathcal{E}(t)) \times (\mathcal{E}(t))$$

Here, $\mathcal{E}_{ib}(t)$ is the set of expertise level values within the opinion block b in agent i's neighbourhood. $\xi(\mathcal{E}_{ib}(t))$ represents the consistency of block b and $\mathbb{E}(\mathcal{E}_{ib}(t))$ is its expected value. In the numerator of (3), $log()$ is used as it represents the diminishing property of a utility function in microeconomic theories [29]. η is used to keep the values of $\Xi(\mathcal{E}_{ib}(t))$ within a specific range suitable for specific applications. A high expertise level credibility $\Xi(\mathcal{E}_{ib}(t))$ indicates the presence of a large cluster of highly confident agents in block b. The expected expertise level credibility as evaluated in (4) further validates the presence of such expert groups in the neighbourhood of an agent.

$$\mathbb{E}(\Xi(\mathcal{E}_{ib}(t))) = \sum_{b \in B} \Xi(\mathcal{E}_{ib}(t)) \times p_{ib}((t)) \qquad (4)$$

So far, we have defined the theoretical underpinnings to capture majority and expert effects separately. However to realize their joint effects in opinion dynamics, (1), (3) and (4) can be used to represent the three major mutually exclusive real-world scenarios of opinion update process that are described below:

Scenario 1 (Joint consideration of majority and expert effects) This scenario captures the influence of a large group consisting of the majority neighbouring agents who are also experts. To represent the majority effect, the majority of the neighbouring agents have to possess similar opinions which yields high opinion consistency, while for the concurrent expert effect, their expertise levels are to be high and consistence. Therefore, their combined effect termed as the credibility score of neighbours ($\chi_{N_i}(t)$) at start of Section 2.1 can be calculated as the product of opinion consistency computed in (1) and expertise level credibility of neighbours computed in (4), which is defined as follows.

$$\chi_{N_i}(t) = \xi(O_{N_i}(t)) \times \mathbb{E}(\Xi(\mathcal{E}_{ib}(t))) \qquad (5)$$

Scenario 2 (Influence of small groups consisting of experts) In the absence of the joint effects, an emerging group of expert agents in the neighbourhood having similar opinions influences the agent to update its opinion. The group exhibits more influence as it grows larger, which is a characteristic reflecting majority phenomenon. The larger expertise level consistency within group ensures greater influence according to the expert effect. Finally, based on the homophily principle [16], the group closer in distance in opinion has more impact. Therefore, considering the relative size and consistency of a group, and an agent's opinion distance from that group, the influence of such a group on an agent's opinion is formulated as in (6).

$$\chi_{ib}(t) = \max_{b \in B} \left(\frac{\Xi(\mathcal{E}_{ib}(t)) \times p_{ib}(t)}{1 + abs(O_i(t) - \overline{O_{ib}(t)})} \right) \qquad (6)$$

Here, $abs()$ denotes absolute value and $\overline{O_{ib}(t)}$ is the average opinion of block b. $p_{ib}(t)$, $\Xi(\mathcal{E}_{ib}(t))$ and $abs(O_i(t) - \overline{O_{ib}(t)})$ represent the relative group size, its expert influence and the homophily principle, respectively.

Scenario 3 (Individual expert's influence) When there is no such group in the neighbours' opinions, it is only the individual experts that influence an agent. In this situation, the amount of influence depends on the opinion distance [16] and the relative expertise level [24] of an expert j to an agent i as captured by (7).

$$\zeta_{ij}(t) = \frac{\mathcal{E}_j(t) - \mathcal{E}_i(t)}{1 + abs(O_j(t) - O_i(t))} \qquad (7)$$

where, $j \in N_i$ and $\mathcal{E}_j(t) - \mathcal{E}_i(t) \geq \zeta_{th}(t)$

Here, $\zeta_{th}(t)$ determines the minimum level of expertise level that an expert should be higher than an agent to impact on its decision. The way of determining the suitable value of $\zeta_{th}(t)$ is to be discussed in the Section 2.4. The opinion update process in our proposed model is presented in the following section.

2.2 Model description

While revising an agent's opinion at time t, it looks for the presence of majority and expert effects in the neighbourhood as measured by the neighbourhood credibility score $\chi_{N_i}(t)$, small expert group influence $\chi_{ib}(t)$ and individual expert impact $\zeta_{ij}(t)$, and applies the opinion update process for the scenarios. The process is described latter in this section. However, the scenarios are mathematically defined using three bounding functions that are computed by applying bounding conditions on $\chi_{N_i}(t)$ and $\chi_{ib}(t)$ of (5) and (6), respectively. The bounding conditions are: (i) $\chi_{N_i}(t) \geq \mathcal{F}_1(B, \eta)$ for Scenario 1 with $\overline{\mathcal{E}_{N_i}(t)} \geq \mathcal{E}_i(t)$, (ii) $\chi_{N_i}(t) < \mathcal{F}_1(B, \eta)$ and $\chi_{ib}(t) \geq \mathcal{F}_2(B, \eta)$ for Scenario 2 with $\overline{\mathcal{E}_{ib}(t)} \geq \mathcal{E}_i(t)$, (iii) $\mathcal{F}_3(B, \eta) \leq \chi_{N_i}(t)$ for Scenario 3 with not satisfying the conditions for Scenario 1 and 2 and (iv) $\chi_{N_i}(t) < \mathcal{F}_3(B, \eta)$ for others. Therefore, $\mathcal{F}_1(B, \eta)$ is the bounding function for defining Scenario 1 which is computed as a lower bound of $\chi_{N_i}(t)$ under a particular representation of that scenario as described in Section 2.4. Similarly, defining appropriate representations for Scenario 2 and 3, bounding functions $\mathcal{F}_2(B, \eta)$ and $\mathcal{F}_3(B, \eta)$ are calculated respectively as the lower bounds of $\chi_{ib}(t)$ and $\chi_{N_i}(t)$ under those representations. The schematic diagram of the proposed model shown in Figure 2 exhibits the required information flows including neighbours' opinions ($O_{N_i}(t)$) and their expertise levels ($\mathcal{E}_{N_i}(t)$), scenario identification and the opinion and expertise level update process for each scenario.

We adopt weighted averaging as the compromise method since it is the widely accepted update process for dynamics with continuous valued opinions [11, 12, 16, 19, 30]. Based on the scenarios and bounding functions, an agent chooses one from the following three compromise options:

– **Compromising with whole neighbourhood (Scenario 1):** As with the condition (i) for Scenario 1 defined above, the collective influence of neighbourhood is very strong and its average expertise level $\overline{\mathcal{E}_{N_i}(t)}$ is higher than the agent's own expertise level $\mathcal{E}_i(t)$. Therefore, the neighbours are collectively as good as one single information source
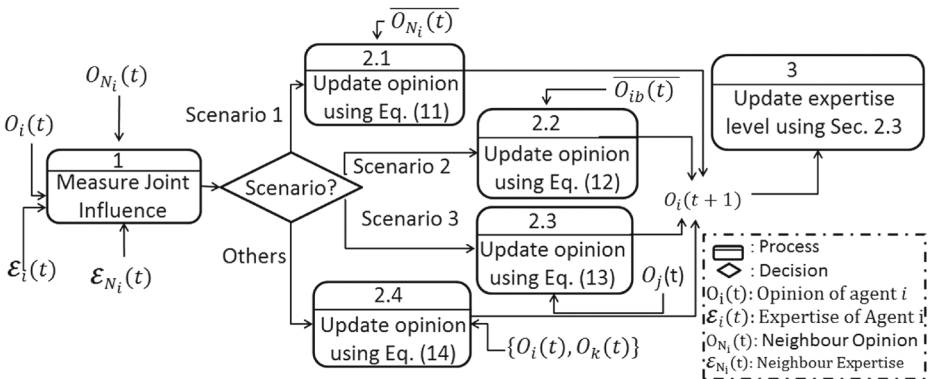

Figure 2 Schematic diagram of the proposed model

and agent i might compromise her opinion with the whole neighbourhood under this condition. However, it is observed in real life that an agent may not update her opinion even after being exposed to any information from experts and/or majority. To assimilate this observation in our modelling, we compute the probability P_1 of an agent to update her opinion under Scenario 1 as follows-

$$P_1 = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x^2}{2}} dx \tag{8}$$

where, Z is defined as,

$$Z = \frac{\overline{\mathcal{E}_{N_i}(t)} - \mathcal{E}_i(t)}{\sigma} \tag{9}$$

and the standard deviation σ is defined as

$$\sigma = \sqrt{\frac{1}{\| N_i \| - 1} \sum_{j=1}^{\|N_i\|} (\mathcal{E}_j(t) - \mathcal{E}_i(t))^2} \tag{10}$$

Here, P_1 considers the relative difference between an agent's and her neighbours' expertise levels. The higher this relative difference, the higher the value of P_1 and the greater the impact on the agent's opinion update is.

Therefore, with probability P_1, the agent i compromises her opinion as per (11).

$$O_i(t+1) = \frac{(\mathcal{E}_i(t) \times O_i(t) + \chi_{N_i}(t) \times \overline{O_{N_i}(t)})}{(\mathcal{E}_i(t) + \chi_{N_i}(t))} \tag{11}$$

Here, the neighbourhood opinion is represented as the average of their opinions, $\overline{O_{N_i}(t)}$ and the credibility score is as weight of averaging.

– **Compromising with a block (Scenario 2):** With $\chi_{N_i}(t) < \mathcal{F}_1(B, \eta)$, neighbours don't form one overall strong group to influence the agent. However, in absence of such a larger block defined for Scenario 1, in real life situation there may exist some relatively smaller blocks consisting of experts that can influence the opinion update process of an agent. $\chi_{ib}(t) \geq \mathcal{F}_2(B, \eta)$ and $\overline{\mathcal{E}_{ib}(t)} \geq \mathcal{E}_i(t)$ indicate the presence of such a group with average expertise level higher or equal to that of agent i. With the presence of such a block, it is highly likely that the agent would compromise her opinion with the representative opinion of that block. With the similar observation as in Scenario 1, we need to incorporate the possibility of an agent not changing its opinion in Scenario 2 as well. The probability P_2 of an agent to change its opinion under this scenario can also be computed using (8)–(10) using the expertise levels $\mathcal{E}_{ib}(t)$ of block b.

Therefore, the agent compromises with an expert block b as in (12) with probability P_2.

$$O_i(t+1) = \frac{(\mathcal{E}_i(t) \times O_i(t) + \chi_{ib}(t) \times \overline{O_{ib}(t)})}{(\mathcal{E}_i(t) + \chi_{ib}(t))} \tag{12}$$

where, $\overline{O_{ib}(t)}$ is the average opinion of block b in agent i's neighbours.

– **Compromising with an expert (Scenario 3):** Condition (iii) implies that there is no majority block consisting of experts in an agent's neighbourhood. In such a scenario, the agent may search for individual experts from its neighbourhood to compromise its opinion with one of their opinions using (13). The probability that an agent i will employ (13) to compromise with a particular expert j is $\zeta_{ij}(t)$ as defined in (7).

$$O_i(t+1) = \frac{(\mathcal{E}_i(t) \times O_i(t) + \zeta_{ij}(t) \times O_j(t))}{(\mathcal{E}_i(t) + \zeta_{ij}(t))} \tag{13}$$

As alluded in the beginning of Section 2, in addition to compromise, opinion update heuristics also allow an agent either to keep own opinion or to adopt one of the neighbours' opinions. The following scenarios embed these remaining two heuristics into our proposed model.

–   **Other Scenarios:** Condition (iv), i.e., $\chi_{N_i}(t) < \mathcal{F}_3(B, \eta)$ means neither a majority nor an expert effect is present in the neighbours. Therefore, an agent ignores them by keeping her opinion or adopting an opinion from her neighbours randomly. To implement the free-will present in all human decisions, we incorporate the free-will concept [28] in our model by making a probabilistic choice between the keep and adoption heuristic.

$$O_i(t+1) = \begin{cases} O_k(t), & \text{with probability } p \\ O_i(t), & \text{otherwise.} \end{cases} \tag{14}$$

where, $O_k(t) \in O_{N_i}(t)$.

Note, other scenarios are also considered in Scenario 1 and 2 when opinion update is unsuccessful with probabilities $(1 - P_1)$ and $(1 - P_2)$, respectively.

Algorithm 1 illustrates the above explained opinion update rules.

Algorithm 1 Opinion Formation Model

Input: $O_i(t), O_{N_i}(t), \mathcal{E}_i(t), \mathcal{E}_{N_i}(t)$
1:   Compute $\chi_{N_i}(t), \chi_{ib}(t), \zeta_{ij}(t)$ using (5), (6) and (7)
2:   **if** $\chi_{N_i}(t) \geq \mathcal{F}_1(B, \eta)$ and $\overline{\mathcal{E}_{N_i}(t)} \geq \mathcal{E}_i(t)$ **then**
3:   with probability P_1, update $O_i(t+1)$ using (11)
4:   with probability $1 - P_1$, update is unsuccessful
5:   **else if** $\chi_{ib}(t) \geq \mathcal{F}_2(B, \eta)$ and $\overline{\mathcal{E}_{ib}(t)} \geq \mathcal{E}_i(t)$ **then**
6:   with probability P_2, update $O_i(t+1)$ using (12)
7:   with probability $1 - P_2$, update is unsuccessful
8:   **else if** $\chi_{N_i}(t) \geq \mathcal{F}_3(B, \eta)$ **then**
9:   sort $\{\zeta_{ij}(t) | j \in N_i \wedge \|\mathcal{E}_j(t) - \mathcal{E}_i(t)\| \geq \xi_{th}\}$ in descending order
10: With probability $\zeta_{ij}(t)$, update $O_i(t+1)$ using (13) until successful
11: **if** update not successful **then**
12: with probability $1 - p$, $O_i(t+1) = O_i(t)$ and
13: with probability p, $O_i(t+1) = O_k(t)$ from (14)
14: Adjust expertise level using Section 2.3

2.3 Update of expertise level

Along with its opinion, like real world situation as described in [24], an agent level of expertise on a particular topic, especially her level of knowledge, might be changed depending on the information she receives from her neighbours. Seeing a large number of opinions (a sort of majority effect) close to an agent's own opinion certainly boosts up her level of expertise. On the contrary, a gathering of opinions at a distant opinion location in the OS amounts to knowledge that impacts negetively to shatter her expertise level. To update the expertise level of an agent in our model, we employ two thresholds from the literature: (i) homophily distance threshold (d_{th}) [11, 16], (ii) majority people threshold (λ_{th}) [25]. If the fraction of neighbours inside the homophily distance of an agent exceeds the majority proportion, expertise level is increased by a small amount, chosen randomly from the range

$[0, \triangle\mathcal{E}]$. According to the principle adopted in [4, 24] for updating agent's expertise level, a reasonable estimate for $\triangle\mathcal{E}$ which we use in our model is 0.15. Conversely, if the fraction of neighbours' opinions at the either end of the OS beyond the homophily distance of the agent is greater than the majority, expertise level is decreased by that small amount. Otherwise, expertise level is kept unchanged as there is no major opinion group in the neighbourhood to boost or shake the level of expertise.

2.4 Derivation of functions $\mathcal{F}_1(B, \eta)$, $\mathcal{F}_2(B, \eta)$ and $\mathcal{F}_3(B, \eta)$

We compute the aforementioned bounding functions by applying the majority rule in both opinion and expertise level domains. We consider the representation of Scenario 1 shown in Figure 3a to determine a lower bound of $\chi_{N_i}(t)$ as $\mathcal{F}_1(B, \eta)$. Here, the majority block contains at least 75% of neighbours. Their expertise levels are all high with the maximum possible consistency. Thus, an agent can't ignore the neighbours and they are as good as one unified information source. Consequently, the credibility score $\chi_{N_i}(t)$ of (5) computed under this assumption is a good approximation for $\mathcal{F}_1(B, \eta)$.

For the representation of Scenario 2 to compute $\mathcal{F}_2(B, \eta)$, we assume that an agent can't ignore a block of 30% majority having highly confident opinions and at a distant of 0.25 (the homophily threshold) as depicted in Figure 3b. This representation defines a lower bound of $\chi_{ib}(t)$ which can be used as $\mathcal{F}_2(B, \eta)$ and can be computed using (6).

Lastly, the representation of Scenario 3 shown in Figure 3c is considered to determine $\mathcal{F}_3(B, \eta)$. In this representation, just the majority opinions (50%) of the neighbours form a cluster around an expert and the rest are evenly distributed in other blocks. Moreover, their expertise levels are evenly distributed with zero consistency. In contrast to scenarios 1 and 2, this implies that neighbourhood has no consistent expert effect with a minimal of majority effect. Consequently, $\chi_{N_i}(t)$ of (5) under this representation can rationally be used as $\mathcal{F}_3(B, \eta)$ where agents look for potential individual experts to compromise with. Any other scenarios that do not meet the conditions for the above three scenarios are collectively represented as other Scenarios as described in Section 2.2.

It is noteworthy to mention here that the assumptions regarding the majority block size in computing the functions $\mathcal{F}_1(B, \eta)$, $\mathcal{F}_2(B, \eta)$ and $\mathcal{F}_3(B, \eta)$ for demarcating the scenarios are made intuitively by following majority rules. However, our model is flexible to adapt different majority block size heuristics for computing the bounding functions. Nevertheless, the choice is application specific and depends on how a particular application defines the majority rule. The variation of the bounding functions with respect to the majority block size assumptions and the corresponding effects on the opinion dynamics are observed in Section 4. Moreover, one can divide the other scenarios into multiple ones again as per the specific requirements of the particular applications.
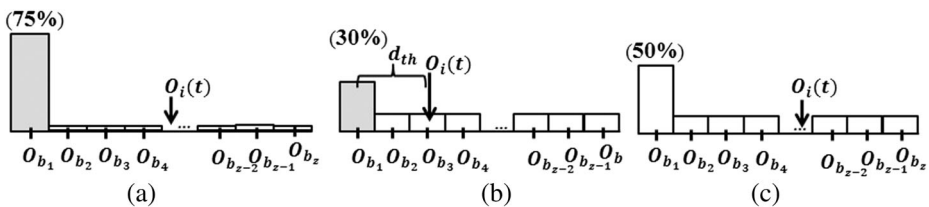

Figure 3 Schematic representation of Scenarios 1, 2 and 3 respectively for computing- **a**: $\mathcal{F}_1(B, \eta)$, **b**: $\mathcal{F}_2(B, \eta)$ and **c**: $\mathcal{F}_3(B, \eta)$. Highlighted blocks in (a), (b) denote high expertise level credibility

For the representation of Scenario 1 shown in Figure 3a, we obtain the opinion consistency $\xi(O_{N_i}(t))$ as

$$\xi(O_{N_i}(t)) = 1 - \frac{\left[\frac{3}{4}log(\frac{3}{4}) + \frac{1}{4}log\frac{1}{4(B-1)}\right]}{log(B)} \quad (15)$$

For majority block, $\xi(\mathcal{E}_{ib}(t)) = 1$ and $\mathbb{E}(\mathcal{E}_{ib}(t)) = 1$ that evaluates $\Xi(\mathcal{E}_{ib}(t)) = log(1 + \eta)/log(1+\eta) = 1$ using (3). In contrast, for all other blocks, $\xi(\mathcal{E}_{ib}(t)) = 0$ and $\mathbb{E}(\mathcal{E}_{ib}(t)) = 0.5$. Therefore, we have $\Xi(\mathcal{E}_{ib}(t)) = log(\eta)/log(1+\eta)$. Now, $p_{ib}(t) = \frac{3}{4}$ for majority block and $p_{ib}(t) = \frac{1}{4 \times (|B|-1)}$ otherwise. Therefore, $\mathbb{E}(\Xi(\mathcal{E}_{ib}(t))) = \frac{3}{4} \times 1 + \frac{1}{4} \times log(\eta)/log(1+\eta)$ from (4). Finally, putting these values in $\chi_{N_i}(t) = \xi(O_{N_i}(t)) \times \mathbb{E}(\Xi(\mathcal{E}_{ib}(t)))$, we can approximate $\mathcal{F}_1(B, \eta)$ as 0.4.

As discussed earlier for Scenario 2 in selecting $\mathcal{F}_2(B, \eta)$ using $\chi_{ib}(t)$, an agent can't overlook a group of neighbours having 30% majority with highly confident opinions and within its homophily distant threshold of 0.25. Consequently, $\mathcal{F}_2(B, \eta) = \chi_{ib}(t) = \max_{b \in B}(\frac{\xi(\mathcal{E}_{ib}(t)) \times p_{ib}(t)}{1+abs(O_i(t)-\overline{O_{ib}}(t))}) \approx 0.25$.

For Scenario 3 to calculate $\mathcal{F}_3(B, \eta)$, we obtain the opinion consistency, $\xi(O_{N_i}(t))$ as

$$\xi(O_{N_i}(t)) = \frac{e_m - e_{N_i}(t)}{e_m} = 1 - \frac{\frac{1}{2}[log(2) + log(2 \times (B-1))]}{log(B)} \quad (16)$$

Referring to Figure 3c, for evenly distributed expertise level $\xi(\mathcal{E}_{ib}(t)) = 0$ and $\mathbb{E}(\mathcal{E}_{ib}(t)) = 0.5$ for all $b \in B$. Thus, $\Xi(\mathcal{E}_{ib}(t)) = log(\eta)/log(1 + \eta)$. Now, $p_{ib}(t) = \frac{1}{2}$ for the majority block and $p_{ib}(t) = \frac{1}{2 \times (|B|-1)}$ otherwise, which evaluates $\mathbb{E}(\Xi(\mathcal{E}_{ib}(t))) = log(\eta)/log(1 + \eta)$ from (4). Finally, putting these values in $\chi_{N_i}(t) = \xi(O_{N_i}(t)) \times \mathbb{E}(\Xi(\mathcal{E}_{ib}(t)))$, we approximate $\mathcal{F}_3(B, \eta)$ as $\chi_{N_i}(t) \approx 0.15$.

Finally, we adopt the findings of the empirical studies done in [2, 24] regarding the level of expertise an expert has to determine the lower bound of the threshold $\zeta_{th}(t)$ used for (7). These studies have estimated the difference of expertise levels between two interacting agents which prompts the agent with lower expertise level to compromise with the one having higher level of expertise.

3 Model validation and performance evaluation

We have evaluated the strength of our model in capturing the evolution of opinion dynamics and analysed its various characteristics in two stages. First, we validate the efficacy of our model using real OSN data to see how closely our model is able to follow the opinion dynamics in real world, which is described in this section. The motivation is that OSNs are the most preferred platform for people to express their opinions in recent years. Most of the OSNs let their users to publish their ideas and thoughts using publicly accessible posts and to interact with others through different activities (e.g. comment, repost, like, share, mention). Moreover within some restrictions, it is easy to collect users' posts and interaction data from the OSNs through their publicly accessible APIs.

Second, we use simulated data to create various scenarios to analyse how our model behaves under those circumstances, which is elaborated in the next section. Especially through simulation, we observe the steady state characteristics of the dynamics. We also investigate whether the final opinion distribution of the agents yields consensus, polarization and fragmentation, and under which conditions these behaviours are observed.

3.1 Data collection and preprocessing

In the data collection process, we focused on Twitter (www.twitter.com), a micro-blogging site extremely popular for generating short textual contents. Twitter serves as a very effective platform of public opinion formation. Using its STREAMING API, we collected a continuous stream of tweets on the search keywords 'Vaccination' (accompanied with 'vaccine'), and 'Immunization'. The reason for choosing this topic is that there is an ongoing debate on the adverse effects of vaccination on babies (such as vaccination induced autism) and whether it is beneficial at all or it causes more harms to the kids when they grow up. Using REST API successively, we generated the underlying social network graphs for the involved users. For a snapshot of the ongoing debate on vaccination, we streamed twitter.com for consecutive three days on the aforementioned search keys and a substantial total of 21801 tweets were collected in the process. A manual preprocessing step to discard the irrelevant tweets yielded a dataset of size 14226. For statistically sound performance evaluation, we concentrated on users (85 such users) who have tweeted at least 10 times to express their opinions within the specified data collection period. All tweets from neighbours between the two consecutive tweets of a user have impacts on her expressed opinion for that time period. Opinion prediction is irreverent for any time step without such participating neighbours' tweets, because, in that case, one's opinion does not get influenced by others. Consequently, we ignored such time steps and also discarded the users who have 50% of their time steps with no participating neighbours. Eventually, we focused on 46 users (U) who had 10 or more tweets and more than 50% of their time steps had at least one participating neighbour.

3.2 Opinion extraction

To implement the proposed model in OSN environment, we have to extract individual users' opinions from their tweets. For mining these opinions, we need to convert the textual tweets into numerical values. Sentiment analysis [22, 23, 27] is the well-established technique for such transformation of text to a corresponding numerical value. Here, the numerical score of a user's sentiment represents her corresponding opinion value as mentioned in [27]. In our implementation, using sentiment analysis, we converted a user tweet text into numerical opinions in the continuous range [0,1]. For example, the sentiment score 0 expresses her opinion value 0, while the sentiment score 0.8 indicates her opinion value 0.8.

In our implementation, we adopted two different methods to extract the opinions, namely (i) a commercial sentiment analyser[1] and (ii) an open-source sentiment analysis tool named TextBlob[2]. Since sentiment analysis tools are not yet standardized or benchmarked, they may produce different sentiment values for the same text in some cases. As we also observed in our experiment, in some cases, the numerical values obtained from the two aforementioned tools differed. Therefore, to combine the best of the two tools, we took their average and used that value as the opinion expressed in a particular tweet text.

3.3 Expertise level extraction

The second important parameter that we need to extract from a user's tweet information is her expertise level. In our model, expertise level signifies the degree of expertise a user

---

[1]http://text-processing.com/

[2]https://textblob.readthedocs.io/en/dev/

have on a particular topic of discussion. The extraction of numerical expertise level is not that straightforward as extracting the opinions from a tweet text. An expert possesses more depth of knowledge on a topic than that of the general people which is reflected through her activities, affiliations and connections in Twitter. There are some existing works that rank Twitter users according to their expertise on a particular topic [5, 26, 31, 33]. To identify and rank experts, those works extract a number of features from Twitter data, namely, the content and originality of tweets, the number of followers and friends, re-tweets by others, user mentions, Twitter lists and user profile. However, it is not feasible to fully adopt those works as they focus on ranking among the experts rather than quantifying the expertise level of individual expert. On the other hand, to implement our model, expertise level of individual user needs to be estimated rather than ranking all the users. This expertise level is an estimation of an individual's perception of another user's degree of expertise on a topic when that individual encounters a tweet on that particular topic from the other user.

In our model, the expertise level is integrated as the extent of influence an expert exhibits on an individual. According to [3], the content of a user's tweets is an important measurement for her influence on others. On the other hand, similat to tweets, bio-information provides equivalently useful information for determining expertise of Twitter users, as found in the study in [20]. Consequently, the quality of a user's generated content in Twitter and the information provided in her bio-information are considered in our implementation to estimate a user's expertise level perceived by others. Here, by 'bio-information' we refer to the short description of a user in her Twitter profile, where she provides information such as professional and institutional affiliation, training and experiences, interests and qualifications to impress others. In our implementation, topic based keyword extraction and matching are used to evaluate the users' expertise levels in the range [0,1].

3.3.1 Keyword extraction

To estimate the expertise level of a Twitter user in our user set, we extract information from 50 Twitter accounts who are real experts in the field of vaccination. Accounts having sufficient number of tweets with quality information on vaccination and bio-information indicative of expertise were selected as real experts. We further validated them as experts by verifying the relevant information they provided in their bio-information, such as their professional affiliations. From their tweets and bio-information, two set of keywords ($\mathcal{K}$) were extracted, namely $\mathcal{K}^{\Theta}$ and $\mathcal{K}^{\beta}$, respectively. For keywords extraction, we used a rate limited free version of AlchemyAPI[3] from IBM Watson. Each extracted keyword $k \in \mathcal{K}$ is associated with a relevancy score τ_k that determines the importance of the keyword in representing the concept of the text.

3.3.2 Content quality, $\lambda_{\mathcal{C}}$:

For a user u having $\mathcal{N}$ number of tweets, consider $\Theta_u = \{\theta_1, \theta_2, \cdots, \theta_{\mathcal{N}}\}$ represents the set of tweets generated by her on our selected topic of vaccination. To measure the quality of the information generated through these tweets, they were checked whether they contain the keywords from $\mathcal{K}^{\Theta}$, where $\mathcal{K}^{\Theta}$ is the set of keywords extracted from experts' tweets. A matrix $M = [m_{jk}], 1 \leq j \leq \mathcal{N}, k \in \mathcal{K}^{\Theta}$ is constructed where $m_{jk} = \tau_k$ if tweet θ_j of user u contains the keyword $k \in \mathcal{K}^{\Theta}$ and 0 otherwise. Thus, M captures the quality of each

---

[3]http://www.alchemyapi.com/products/alchemylanguage/keyword-extraction

individual tweet of a user in terms of representing information as a real experts on the topic. The overall content quality $\lambda_\mathcal{C}$ is then computed as $\lambda_\mathcal{C} = \underset{1 \le j \le \mathcal{N}}{\text{median}}(\underset{k \in \mathcal{K}^\Theta}{\text{median}}(m_{jk}))$.

3.3.3 Bio-information quality, λ_β:

It refers to the similarity of a user bio-information with that of a real expert on the topic. We consider bio-information in estimating a user's expertise level since it presents the relevant information regarding her expertise on the topic. If $\mathcal{K}_u \subset \mathcal{K}^\beta$ represents the set of keywords extracted from the bio-information of user u that are also present in real experts bio-information, the quality λ_β was evaluated as $\lambda_\beta = \underset{k \in \mathcal{K}_u}{\text{median}}(\tau_k)$.

Finally, to determine the expertise level of a user using the two aforementioned quality measurements, we combined them by considering that they are equally important. Thus, $\mathcal{E}_u = \frac{1}{2}(\lambda_\mathcal{C} + \lambda_\beta)$.

3.4 Performance metrics

For any user $u \in U$, the opinions extracted from her tweets represent the true opinions O_R^u. Starting with the first element in O_R^u as the initial opinion, our proposed model estimates the subsequent opinions of a user using the model parameters to yield a predicted set O_S^u. Subsequently, we have validated our model by evaluating how close the estimated opinions of the users are to their true opinions. In this regard, we consider the following two well known metrics, namely: i) Root Mean Square Error (RMSE) and ii) Directional Symmetry (DS).

RMSE: It measures the overall discrepancies of our estimation process by statistically combining the deviation of the estimated opinions from the true opinions as shown in (17).

$$\text{RMSE} = \sqrt{\frac{\sum_{u \in U} \sum_{(o_r(t), o_s(t)) \in (O_R^u, O_S^u)} (o_r(t) - o_s(t))^2}{|U| \times |O_R^u|}} \qquad (17)$$

DS Since the opinions are regarded as a sequence of values as in time series analysis, we can use directional symmetry defined in (18) to statistically measure the performance in predicting the direction of change, either positive or negative, from one time step to the next.

$$\text{DS} = \frac{100}{|U| \times (|O_R^u| - 1)} \sum_{u \in U} \sum_{(o_r(t), o_s(t)) \in (O_R^u, O_S^u)} d_i \qquad (18)$$

$$\text{where, } d_i = \begin{cases} 1, & \text{if } (o_r(t) - o_r(t-1)) \times (o_s(t) - o_s(t-1)) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

3.5 Performance evaluation

We computed the RMSE and DS for our model based on the user set U. We also measured the same statistics for two recently developed and relevant opinion formation models , namely: i) Biased Voter Model [9] and ii) Moussaid Model [24]. Table 1 shows the evaluated performances of the three models.

From the table it is evident that our model outperforms the other two in terms of predicting the opinions and the directions of their changes. The reason for this better performance is that our model captures the real world opinion dynamics more effectively. We define multiple opinion update scenarios based on the majority and expert effects present in an agent's

Table 1 RMSE and DS for the Models

	Our Model	BVM	Moussaid
RMSE	0.18506	0.20097	0.21281
DS	0.45723	0.43427	0.19633

neighbourhood, which we believe more natural to predict the opinions. To be specific in the issues like vaccination, the general people are mostly guided by the expert opinions, and our model captures the expert effect more comprehensively than the other models.

We also investigated whether the performance improvements of our model depend on the number of participating neighbours in the prediction steps and the corresponding results are shown in Figure 4. From the figure it is apparent that, with a few exceptions, our model's performance is better than the existing models for prediction steps with different number of participating neighbours. It is important to note that most of the prediction steps in our data sample had participating neighbours in the range 1 to 40. Consequently, the first four results in Figure 4 are significant, where our model outperforms the other models in terms of prediction accuracy for both the metrics. The better performances validate the proposed model as a distinguished one to capture the real world opinion formation dynamics.

4 Simulation results and analysis

We adopted the Erdos-Renyi random graph with 1000 nodes to represent a social network in our simulation as it is one of the established way to represent such networks. Agents' initial opinions and their expertise levels were randomly distributed in the range [0, 1]. We experimented with two random distributions for the initial opinions: (i) uniform random and (ii) normal random with particular mean (μ) and standard deviation (σ). We considered 10 equal length opinion blocks as per the options in a survey in a scale from 1 to 10. Initial expertise levels were assigned in one of the two ways: (i) uniformly random across the agents and (ii) separately assigned for each opinion block to differentiate their expertise levels. To balance the randomness present in the free-will of human decisions, the probability
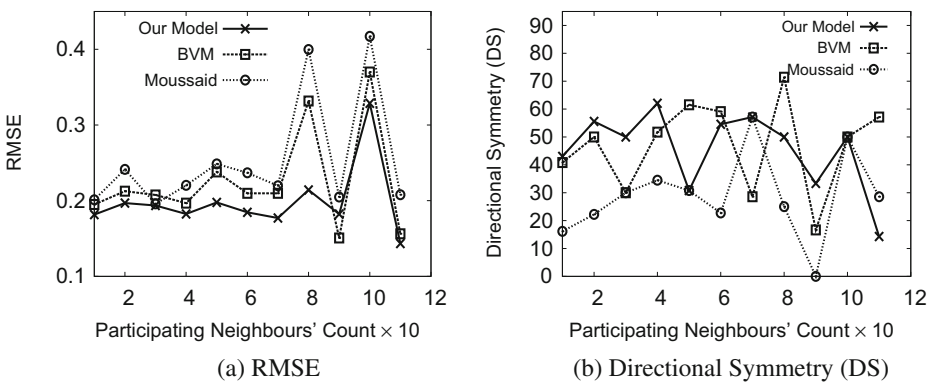

(a) RMSE

(b) Directional Symmetry (DS)

Figure 4 Measured performance of the models grouped by the the number of participating neighbours in the prediction steps

p as dictated in Steps (12-13) of Algorithm 1 for selecting between 'keep' and 'adoption' heuristics was set with 0.5. Finally, d_{th} and λ_{th} were assigned to 0.1 and 0.85 respectively to make the expertise level change conditions more stringent.

4.1 Majority (M) Vs. expert (E) effect

We examined the majority vs. expert effect through a controlled experiment. Agents started with a bimodal initial opinion distribution having two modes at $\mu_1 = 0.25$ (X) and $\mu_2 = 0.75$ (Y) with a small standard deviation ($\sigma_1 = \sigma_2 = 0.05$) for both. The majority effect was created by increasing the number of opinions in group X from 50% to 95%. However, their expertise levels were confined within [0-0.2] to make them non-experts. Moreover, inconsistent expertise levels in X decoupled them from any sort of expert effects. On the contrary, the agents in group Y were assigned to highly consistent expertise levels to reflect the expert effect. We increased their expertise levels in steps from 0.1(very low) to 1 (very high) in experiments with all agents having the same expertise level for each assignment to vary the extent of their expertise. Figure 5 illustrates the findings by showing the predominant effects in the converged opinions.

From our model, from Figure 5a it is clear that consistent agents with high expertise level (0.8 or higher) can be considered experts and drive the majority towards them. As the expertise level decreases, majority effect becomes more prominent. However, there is another region of parameters where both the effects are exhibited together and the region is identified as 'C' (combined). Figure 5b shows the results of Moussaid et al. [24] model for the same parameters. Their method fails to separate the majority and expert effects adequtely as indicated by the large number of grids with 'C'. It also cannot accurately capture the expert effect, as for very high expertise level of 0.8 $\sim$ 1, only 6 out of 18 cases i.e., (33.33% cases) show the expert influences. In contrast, for expertise level 0.8 $\sim$ 1, our model produces 100% expert effect. Moreover, our model more accurately captures the majority impacts as 75% agents having the same initial opinion have produced majority effect in the absence of highly expert agents. However, in such cases none of them has shown majority impact in the model by Moussaid et al. [24].
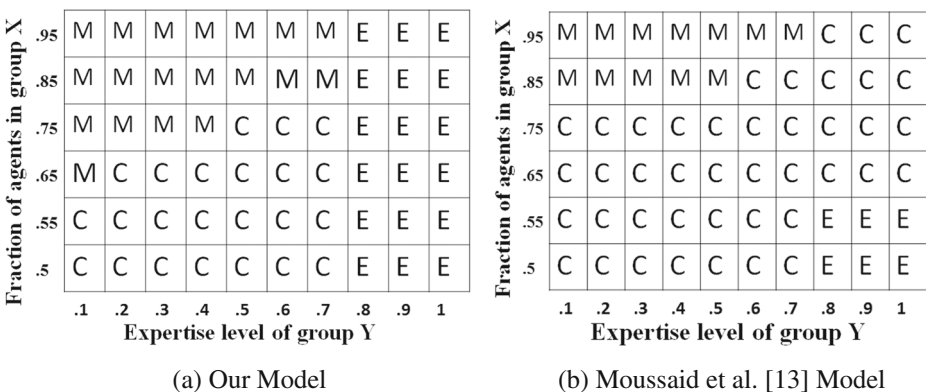

Figure 5 Majority vs. expert effects observed at the end of opinion evolution. Here, E: Expert, M: Majority, and C: Combined. Here, fraction of agents in group Y = (1 - fraction of agents in group X), and expertise levels of agents in group X ∈ [0 − 0.2]

4.2 Effects of bounding functions on opinion dynamics

As alluded in Section 2.2, our model defines four opinion update scenarios based on three bounding functions, namely $\mathcal{F}_1(B, \eta)$, $\mathcal{F}_2(B, \eta)$ and $\mathcal{F}_3(B, \eta)$. While computing values for these functions in Section 2.4, we have assumed some intuitive heuristics for the size of the majority block present in an agent's neighbourhood and its expertise level consistency. Moreover, they also depend on B, the number of blocks in computing the consistencies. For notational simplicity in the following discussions, we have omitted (B, η) from the functions' names, thus representing them by $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$.

Effects of $\mathcal{F}_1$ First, we illustrate how the values of $\mathcal{F}_1$ are varied based on these assumptions followed by the impacts of these variations on the opinion dynamics under the aforementioned experimental set-up in Section 4.1. Figure 6a shows the variations in $\mathcal{F}_1$ as a function of majority block sizes and the number of blocks B. In x-axis we vary the majority block size from 15% to 99%, while y-axis shows the number of blocks(B). The height of the surface in z-axis captures the variation in $\mathcal{F}_1$. Firstly, from the figure it is evident that, for a given value of B, $\mathcal{F}_1$ is monotonically increasing with the majority block size assumptions, except for 2. $B = 2$ yields irregularities in computation by producing same values of $\mathcal{F}_1$ for different majority block sizes. The observation is quite reasonable as our model is designed for continuous opinions, while $B = 2$ will produce a binary opinion dynamics. However, the higher the values of B, the higher the tendency of having linear variation of $\mathcal{F}_1$ over majority block sizes is. Here, specific applications might choose B differently based on the number of discrete levels it would allow the users to express their opinions. Secondly, $\mathcal{F}_1$ increases steadily for up to 60% of the majority block proportion whereas it attains a steep increase when the proportion exceeds 70%.

In Figure 6b, we illustrate the impacts of varying majority block sizes and the corresponding $\mathcal{F}_1$ values on the final opinion distribution. Here, we vary the majority block sizes from 25 to 90% and compute the corresponding $\mathcal{F}_1$ using $B = 10$. For each of the function values, we obtained the results like the one presented in Figure 5a. Then the proportions of the majority (M), expert (E), and combined (C) effects observed in the converged opinions are computed and plotted separately in Figure 6b. In our experimental set-up expert opinions (Group Y) never consist of a group greater than 50% as we varied the agent proportion in Y from 5 to 50%. Consequently, expert effect is more prominent for $\mathcal{F}_1$ values
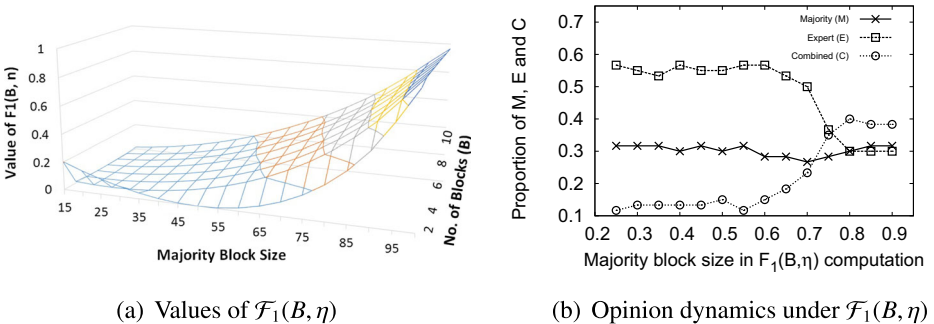

(a) Values of $\mathcal{F}_1(B, \eta)$ (b) Opinion dynamics under $\mathcal{F}_1(B, \eta)$

Figure 6 Analysis of $\mathcal{F}_1(B, \eta)$. **a** The values of $\mathcal{F}_1$ as a function of majority block size assumption and no. of blocks (B), **b** Proportion of majority (M), expert (E) and combined (C) effect at steady state as a function of the majority block size assumption in the computation of $\mathcal{F}_1$ with $B = 10$

with more relaxed majority block size assumptions (up to 0.55). With increased proportion of the majority block size in $\mathcal{F}_1$ computation from 0.6 to onwards, the expert group can't reach this majority in our experiments, which leads a decreasing expert effect.

On the other hand, for block size proportion below or equal to 60%, we observed that combined effect is relatively low. This is due to the fact that, with the more relaxed majority block size assumptions in our experimental set-up, an agent can distinguish between the majority and expert group effect and influenced by either of them. Based on the agent proportion in the group X and Y, both majority and expert group effects become strong enough to sweep agents from other groups to their respective directions. As the assumption in $\mathcal{F}_1$ becomes more stringent by increasing the proportion for majority block size, expert group Y increasingly fails to work in a group to act as a basin of attraction. Under these conditions, experts continue to influence agents at individual level. However, the impact of individual expert is far less than that of an expert group. Consequently, expert group Y fails to prevalently penetrate public attention which was observed by an increased fraction of combined effect in the dynamics. The majority effect remains steady throughout the parameter space. This is because, regardless of the block size assumptions, when majority group X attains high dominance from 75 to 95%, it has similar possibility to attract agents from the other group, especially when the expertise level of the latter group is less than 0.8. The most interesting observation from Figure 6b is that the three effects (M, E, C) become evenly poised after the majority block size assumption reaches 0.75 (75%). This also validates our consideration of 75% majority block size assumption while computing $\mathcal{F}_1(B, \eta)$ in Section 2.4.

Effects of $\mathcal{F}_2$ Figure 7a shows the impacts of $\mathcal{F}_2$ values on the final opinion distribution by varying majority block size assumptions from 15 to 75%, and then obtaining the results like the one presented in Figure 5a for each of such values. The proportions of the majority (M), expert (E), and combined (C) effects obtained for each $\mathcal{F}_2$ values are plotted

(a) Opinion Dynamics under $\mathcal{F}_2(B, \eta)$ (b) Opinion dynamics under $\mathcal{F}_3(B, \eta)$
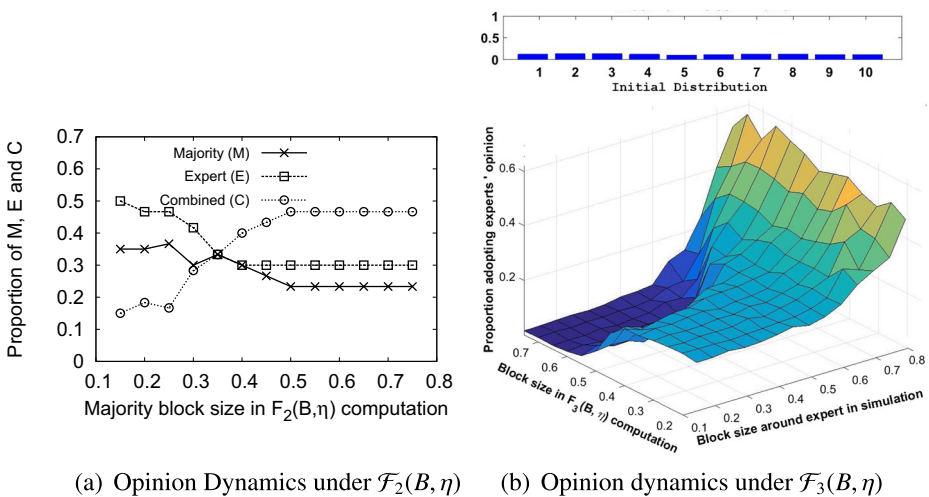
Figure 7 Analysis of $\mathcal{F}_2(B, \eta)$ and $\mathcal{F}_3(B, \eta)$. **a** Proportion of majority (M), expert (E) and combined (C) effect at steady state as a function of the majority block size assumption in the computation of $\mathcal{F}_2$, **b** Proportion of agents adopting experts' opinions by varying the majority block size assumption (x-axis) and the block size around experts in simulation (y-axis)

separately. From the figure we observe that the expert effect decreases with the increase in the majority block size assumptions till it reaches 45%. This is due to the fact that with more strict assumptions, it becomes harder to satisfy them by the expert groups. Moreover, the experimental set-up does not allow a group of expert with a proportion of agents greater than 50%. Consequently, fractions of E becomes steady after the majority block size reaches 50% (even starts from 40%).

With the early range of majority block size assumptions(15 to 35%), both the majority and expert effects are clearly distinguished as seen by a large fractions of M. Correspondingly, the combined effects (C) are lower than both M and E. However, with the increasing majority block size, the expert group fails to show more group effect. This leads to more cases where only individual experts can exhibit their influence. The experts individually can only attract a fraction of the agents in Group X rather than pulling the whole group, which results in more combined effects as observed by an upward trend in C. Consequently, the majority effect starts decreasing. The impacts of the individual experts' effects are discussed in the following paragraph. Nevertheless, the three proportions keep relative balance in the majority block size range of [0.3–0.45], that can be a suitable assumption to compute $\mathcal{F}_2$. This explains the rationale of using 30% majority block size assumption for computing $\mathcal{F}_2$ in Section 2.4.

Effects of $\mathcal{F}_3$ Figure 7b shows the impacts of $\mathcal{F}_3$ values on the final opinion distribution by varying the majority block size assumptions used in computing $\mathcal{F}_3$. Different from the two cases discussed above, here the experts do not form their own groups. Instead, they remain individuals with non-expert agents mixing with them. Therefore, we consider a uniform random initial opinion distribution as shown in the upper portion of Figure 7b, rather than starting with two distinct opinion groups (X and Y). Only 1% of the agent population are assigned with high expertise level of 1 to represent experts, and are placed in the last block of the OS ((0.9-1.0]). All other agents represent lay-people with an expertise level less than 0.2. To compute the values for $\mathcal{F}_3$, we varied the block size assumption around the individual experts in the range from 15 to 75% as shown in the x-axis of the lower part of Figure 7b. For each value of $\mathcal{F}_3$, we ran the simulation by increasing the actual block size around the expert ((0.9-1.0] of the OS) from 10 to 75% as shown in the y-axis, and obtained the proportion of agents adopting the experts' opinions.

It is eminent from the figure that, with the increase of the block size around experts in the simulation, the proportion of agents adopting experts increases. For all $\mathcal{F}_3$ values, block sizes beyond 60% (in y-axis) result in very high proportion of expert opinions' adoption as both the expert and majority effects become prominent in this range. However, for lower block sizes (10 to 45% in y-axis) around the experts in simulation, different $\mathcal{F}_3$ assumptions (majority block size in x-axis) yield different proportion of experts adoption. This proportion is large with more relaxed block size assumptions (15 to 50% in x-axis) as in simulation block sizes start from as low as 10% (in y-axis). On the other hand, the fraction is low for higher block size (55 to 75% in x-axis) assumptions as the smaller block sizes in simulation (10 to 45% in y-axis) cant satisfy those conditions. Looking from the y-axis, it is clear that the opinion dynamics make a transition in the block size range of 40 to 60%. Moreover looking from the x-axis, the $\mathcal{F}_3$ values computed using block size assumptions of 45 to 55% results in moderate transitions of experts' opinions adoption rate in comparison with the low (15 to 40%) or steep (60 to 75%) transitions of the adoption rate. Consequently, any value in these ranges can be suitable for $\mathcal{F}_3$. We chose 50% majority block size assumption in our simulation.

4.3 Consensus, polarization and fragmentation

Our model captures consensus, polarization, and fragmentation properties of opinion dynamics as shown in Figure 8. The first row of each of the figures shows agents' initial opinion distributions, while the second rows show the corresponding final opinion distribution after the dynamics converges. The first three figures (Figure 8a–c) illustrate the model behaviour in the presence of expert groups. The opinion blocks highlighted within rectangle represent the position of the expert groups. The other blocks contain only agents having low expertise levels. From the figures we observe that a single expert group leads others to consensus around it (Figure 8a). This is because the initial opinions are uniformly distributed in the OS, thus have no majority group. In absence of majority influence, expert group is the only source of strong influence on other non-expert agents and drives them towards it. On the other hand, the presence of more than one confident groups causes polarization (Figure 8b) or fragmentation (Figure 8c). Here, the reason behind the fragmentation of opinions is that the two confident groups reside at the two opposite ends of the OS. Therefore, due to their distances and opposite forces, agents in the middle are not convinced by them, thus form their own groups. However, due to a smaller distance between the expert groups in Figure 8b, they can sweep all the agents within their reach towards them, leaving no agent group in the middle. Consequently, two strong polarizations have been formed. Therefore, the distance between the expert groups determines the steady state behaviour of the dynamics.

The rest of the figures show the model behaviour in absence of an expert group. Without any expert, all agents converge in the middle which is an expected outcome of the dynamics (Figure 8d). However, a majority group attracts agents towards them in the absence of experts. In Figure 8e–f we show this impacts of majority in the absence of experts. For these two figures, simulation starts with two distinct opinion groups, where one group contains more agents than the other. Due to this majority, the group having more opinions attracts agents from the other group, thus strengthen the polarization in the majority group. Nevertheless, the polarization around the group depends on the majority proportion. A group of
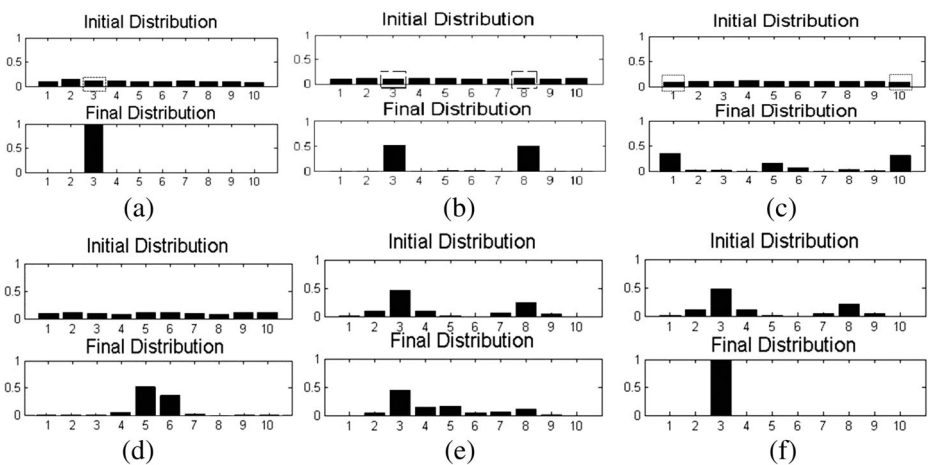

Figure 8 The initial and final opinion distributions of all agents. Expert blocks are highlighted within rectangle. (**a**) Consensus (**b**) Polarization (**c**) Fragmentation (**d**) Big cluster in the middle of OS, (**e**)–(**f**) Majority effect in absence of experts

60% majority attracts many agents to it (Figure 8e), whereas a group of 70% majority leads to consensus on its opinion (Figure 8f).

5 Conclusion

In this paper we introduce an opinion formation model by considering the combined impact of majority and expert effects. We for the first time introduce the concept of consistency in opinions and associated expertise levels of neighbouring agents and embed metrics derived from those to update an agent's own opinion as well as its expertise level. The theoretical concept of our model is validated using the data collected from Twitter, which shows the superiority of our model in estimating opinions in terms of RMSE and directional symmetry. The performance of our model is extensively analysed through simulation and compared with recent existing opinion formation models. Results show that our model captures the effect of majority and expert more accurately compared to other models. Our proposed model has also a far broader impact for instances, if an application of a specific domain requires to include a particular aspect of opinion dynamics, it can be easily done by adding a relevant scenario in it.

References

1. Asch, S.E.: Opinions and social pressure. Sci. Am. **193**(5), 31–35 (1955)
2. Bruza, B., Welsh, M., Navarro, D., Begg, S.: Does anchoring cause overconfidence only in experts? Cogn. Sci. Soc. **33**, 1947–1952 (2011)
3. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P.: Measuring user influence on twitter: The million follower fallacy. 4th International AAAI Conference on Weblogs and Social Media **10**, 10–17 (2010)
4. Chacoma, A., Zanette, D.H.: Opinion formation by social influence: from experiments to modeling. PLoS ONE **10**(10), e0140406 (2015)
5. Cheng, Z., Caverlee, J., Barthwal, H., Bachani, V.: Who is the Barbecue King of Texas?: A Geo-spatial approach to finding local experts on twitter. SIGIR 335–344 (2014)
6. Cho, J.-H., Swami, A.: Dynamics of uncertain opinions in social networks. IEEE Military Communications Conference 1627–1632 (2014)
7. Clifford, P., Sudbury, A.: A model for spatial conflict. Biometrika **60**(3), 581–588 (1973)
8. Crokidakis, N.: The influence of local majority opinions on the dynamics of the sznajd model. J. Phys.: Conf. Ser. **012016**, 487 (2014)
9. Das, A., Gollapudi, S., Munagala, K.: Modeling opinion dynamics in social networks. WSDM 403–412 (2014)
10. Das, R., Kamruzzaman, J., Karmakar, G.: Opinion formation dynamics under the combined influences of majority and experts. ICONIP 674–682 (2015)
11. Deffuant, G., Neau, N., Amblard, F., Weisbuch, G.: Mixing beliefs among interacting agents. Adv. Compl. Syst. **3**(1), 87–98 (2000)
12. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**(345), 118–121 (1974)
13. Galam, S.: Sociophysics: a review of galam models. Int. J. Mod. Phys. C **19**(03), 409–440 (2008)
14. Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N., Gummadi, K.: Cognos: Crowdsourcing search for topic experts in microblogs. SIGIR 575–590 (2012)
15. Hassan, R., Karmakar, G., Kamruzzaman, J.: Reputation and user requirement based price modelling for dynamic spectrum access. IEEE Trans. Mob. Comput. **13**(9), 2128–2140 (2014)
16. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence: models, Analysis and Simulation. J. Artif. Soc. Soc. Simul. **5**(3), 1–24 (2002)
17. Javarone, M.A.: Social influences in opinion dynamics: the role of conformity. Physica A: Statistical Mechanics and its Applications **414**, 19–30 (2014)
18. Kurmyshev, E., Juarez, H.A., Gonzalez-Silva, R.A.: Dynamics of Bounded Confidence Opinion in Heterogeneous Social Networks: Concord against Partial Antagonism. Physica A: Statistical Mechanics and its Applications **390**(16), 2945–2955 (2011)

19. Li, L., Scaglione, A., Swami, A., Zhao, Q.: Consensus, polarization and clustering of opinions in social networks. IEEE J. Sel. Areas Commun. **32**(6), 1072–1083 (2013)
20. Liao, Q., Wagner, C., Pirolli, P., Fu, W.-T.: Understanding experts' and novices' expertise judgement of twitter users. Computer-Human Interaction(CHI) 2461–2464 (2012)
21. Lima, F.W.S., Sousa, A.O., Sumuor, M.A.: Majority-vote model on directed Erdos-Renyi random graphs. Physica A: Statistical Mechanics and its Applications **387**(14), 3503–3510 (2008)
22. Liu, B.: Sentiment Analysis - Mining Opinions, Sentiments, and Emotions. Cambridge University Press, New York, pp. 1–367 (2015)
23. Martinez-Camara, E., Martin-Valdivia, M.T., Urena-Lopez, L.A., Montejo-Raez, A.: Sentiment Analysis in Twitter. Nat. Lang. Eng. **20**(01), 1–28 (2014)
24. Moussaid, M., Kaemmer, J.E., Analytis, P.P., Neth, H.: Social influence and the collective dynamics of opinion formation. PLoS ONE **8**(11), e78433 (2013)
25. Novak, S., Elster, J.: Majority Decisions: Principles and Practices. Cambridge University Press, New York, pp. 1–288 (2014)
26. Pal, A., Counts, S.: Identifying topical authorities in microblogs. WSDM 45–54 (2011)
27. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
28. Pineda, M., Toral, R., Hernandez-Garcia, E.: The noisy Hegselmann-Krause model for opinion dynamics. Eur. Phys. J. B **12**(86), 1–12 (2013)
29. Varian, H.R.: Microeconomic Analysis. Norton, New York (1992)
30. Varshney, K.R.: Bounded confidence opinion dynamics in a social network of bayesian decision makers. IEEE J. Sel. Top. Signal Process **8**(4), 576–590 (2014)
31. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. WSDM 261–270 (2010)
32. Xia, H., Wang, H., Xuan, Z.: Opinion dynamics: a multidisciplinary review and perspective on future research. Int. J. Knowl. Syst. Sci. **2**(4), 72–91 (2011)
33. Xu, Z., Ramanathan, J.: Thread-based Probabilistic Models for Expert Finding in Enterprise Microblogs. Expert Syst. Appl. **43**, 286–297 (2016)
34. Zhang, J., Hong, Y.: Opinion Evolution Analysis for Short-range and Long-range Deffuant-Weisbuch Models. Physica A: Statistical Mechanics and its Applications **392**(21), 5289–5297 (2013)