An epidemic model for correlated information diffusion in crowd intelligence networks

Yuejiang Li and H. Vicky Zhao
Department of Automation and Institute for Artificial Intelligence, Tsinghua University, Beijing National Research Center for Information Science and Technology (BNRist), Beijing, China, and

Yan Chen
School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

Abstract

Purpose – With the popularity of the internet and the increasing numbers of netizens, tremendous information flows are generated daily by the intelligently interconnected individuals. The diffusion processes of different information are not independent, and they interact with and influence each other. Modeling and analyzing the interaction between correlated information play an important role in the understanding of the characteristics of information dissemination and better control of the information flows. This paper aims to model the correlated information diffusion process over the crowd intelligence networks.

Design/methodology/approach – This study extends the classic epidemic susceptible–infectious–recovered (SIR) model and proposes the SIR mixture model to describe the diffusion process of two correlated pieces of information. The whole crowd is divided into different groups with respect to their forwarding state of the correlated information, and the transition rate between different groups shows the property of each piece of information and the influences between them.

Findings – The stable state of the SIR mixture model is analyzed through the linearization of the model, and the stable condition can be obtained. Real data are used to validate the SIR mixture model, and the detailed diffusion process of correlated information can be inferred by the analysis of the parameters learned through fitting the real data into the SIR mixture model.

Originality/value – The proposed SIR mixture model can be used to model the diffusion of correlated information and analyze the propagation process.

Keywords SIR model, Information diffusion, Crowd intelligence

Paper type Research paper
1. Introduction
With the recent development of social media, Internet of Things, big data, cloud computing and many other new technologies, we witness new industrial and social management patterns, and the emergence of a crowd cyber eco-system consisting of smart and deeply connected entities such as individuals, enterprises and government agencies (Chai et al., 2017; Shen et al., 2017; Nan et al., 2017; Huang et al., 2017). Such smart entities constantly interact with each other, influence each other’s decisions and have a significant impact on our society and economy. It is of crucial importance to study how such smart entities interact with each other, to understand their decision-making process, to analyze how they influence each other and the impact of such interactions on the entire crowd intelligence networks. Such investigation provides important guidelines on the design of efficient and effective mechanisms to manage such crowd intelligence networks.

In this paper, using information diffusion in social networks as an example, we study user behavior in such crowd intelligence networks. With deeply connected smart entities, information diffusion plays a critical role in the information sharing and network evolvement of such crowd eco-systems, sometimes detrimental to our society and economy. One example is the “salt panic” in China after the 2011 Tohoku Tsunami, where the news of nuclear leakage greatly stimulated the rumors like iodized salt can help ward off radiation poisoning, which lead to the long lines and mob scenes at stores and 10-fold jump of salt price throughout China (Pierson, 2011). Thus, it is important to study this complex information diffusion process over networks, to analyze its social and economic impact, and to prevent the propagation of such detrimental rumors.

Tremendous efforts have been dedicated to model the information diffusion process. The existing works on modeling information diffusion can be classified into two categories: graph and non-graph based approaches (Guille et al., 2013), which we will discuss in details below.

The graph-based models assumed that social networks could be described by graphs, where nodes represented users and an edge connecting two nodes represented a certain relationship between the two corresponding users. Two seminal models in this category are the Independent Cascades (IC) model (Goldenberg et al., 2001) and the Linear Threshold (LT) model (Granovetter, 1978). For the IC model, each edge between users could be classified as either the “weak tie”, i.e., the common relationship, or the “strong tie”, i.e., the closer and stronger relationship. Different probabilities were assigned to different types of edges, and the probability for a user to adopt a piece of certain information was determined by the probabilities of the edges connecting with him/her. To take user’s collective behavior into consideration, the LT model assumed that a user would accept a certain piece of information when the percentage of his/her neighbors who had adopted the information was above a threshold. Recently, the evolutionary dynamics of the natural ecological systems has been introduced to model the information diffusion over the social networks (Jiang et al., 2014a, 2014b; Cao et al., 2016). The authors modeled the information diffusion process with evolutionary game theory over the synthetic and real networks, and the evolutionary stable states were analyzed with respect to different types of information and different network structures.

The graph-based approaches could characterize the diffusion processes in the micro view, that is, how each intelligent individual was influenced by their neighbors and how he/she decided whether to adopt the information. However, these models were often limited by the difficulty to obtain the complete social network structure (De Choudhury et al., 2010).

The non-graph based approaches did not consider network structure in their analysis, and modeled the information diffusion process in a macro view. These models were mainly
based on the epidemic models (Daley and Kendall, 1964) from epidemiology. The canonical “Susceptible-Infectious-Recovered” (SIR) model and the “Susceptible-Infectious-Susceptible” (SIS) model from epidemiology were introduced to model the information and computer virus propagation over the online networks (Abdullah and Wu, 2011; Daley and Kendall, 1964; Lerman and Ghosh, 2010; Pastor-Satorras and Vespignani, 2001). These epidemic models classified users into different groups: those who had not heard and never spread the news (Susceptible), those who had received and forwarded the news (Infectious) and those who had stopped forwarding (Recovered). Using a few parameters to model the transition rates between groups, these models used the differential equations to describe and model the dynamics of the population in different groups.

Many works extended the classic epidemic models by introducing other groups to describe the diffusion process. Rui et al. proposed a Susceptible-Potential-Infective-Removed (SPIR) model that introduced the “Potential” groups into the classic SIR model (Rui et al., 2018). The “Potential” group was designed to describe the individuals who had heard of the information but did not become infectious and forward it. By introducing this new group, SPIR model matched the simulated information diffusion processes over the synthetic and real networks better than the classic SIR model. Considering the scenario where there were a few authorities that often clarified the fact and released the authoritative information to confirm or refute the content of network rumors, Xia et al. (2015) introduced a new group that represented these authorities into the classic SIR model, which is called the “SIAR” model. Through the simulation over synthetic networks, the authors showed that the “SIAR” model could realistically characterize the evolution of the rumor propagation. The works in Zhao et al. (2012), Zhao et al. (2013a) and Zhao et al. (2013b) introduced the forgetting and remembering mechanism, and designed a new group called “Hibernators” referring to the individuals who had transformed from the spreaders due to the forgetting mechanism and could be turned back to spreaders due to the remembering mechanism. Through analytical and simulations over homogeneous and heterogeneous networks, the authors found that this new group would reduce the maximum of rumor influence, and postpone the terminal time of the diffusion. Liu et al. considered the existence of the “super-spreaders” in the networks, whose spreading speed was much faster, and introduced a corresponding new group into the classic SIR model (Liu et al., 2016). The validation on real-world Weibo dataset of the proposed model was conducted and showed that this improved SIR model was much more promising than the classic SIR model in characterizing a super-spreading event of information propagation.

Most existing works focused on the diffusion process of only one information. However, the numerous smart entities in the crowd cyber system can interact with each other (Chai et al., 2017) due to the rapid development of the Internet. Thus, the diffusion process often exhibits more complex features, and different information often influences each other and spread together. The works in Beutel et al. (2012) and Prakash et al. (2012) modeled the interaction between different information propagation processes as the competition of users’ attention, while in reality, there may be complicated patterns on how different information propagation interact with each other, e.g., different information can also promote the propagation of each other.

Inspired by the works of Myers and Leskovec (2012), Sun et al. (2017) and Fu et al. (2019), in this work, we study how correlated information influence each other’s propagation over networks, and assume that intelligent individuals who have heard and spread the first information might have some pre-judgment and prior knowledge about the second information. Thus, the probability for him/her to spread the second information should be different from those who have not heard and never spread the first one.
Based on this assumption, this work extends the classic SIR model in Daley and Kendall (1964), Lerman and Ghosh (2010) and Abdullah and Wu (2011) and proposes the SIR mixture model to describe the diffusion process of two correlated pieces of information. The SIR mixture model includes two stages. The first stage represents the process where the first information spreads alone, while the two correlated information spread together in the second stage. By classifying the crowd into 8 groups according to their current states in the propagation of the two correlated information, the dynamics of the percentage of all groups can be described by 8 differential equations, respectively. Through the linearization of the differential equations, we can obtain the necessary conditions of the stable state of the information propagation. We also test our model with real data and infer the propagation process of two pieces of information.

The rest of the paper is organized as follows: Section 2 summarizes the classic SIR model and discusses its properties. Section 3 models the diffusion processes of correlated information, and provides the details of the proposed SIR mixture model. The stable state analysis is discussed at the end of this section. Section 4 validates the SIR mixture model with real data and infers the diffusion process. The conclusion is drawn in Section 5.

2. The classic susceptible–infectious–recovered model

There have been numerous works on the modeling and analysis of the dynamics of information diffusion over social networks. A class of models is inspired by the epidemic model from the epidemiology due to the similarity between the spread of infectious disease and the diffusion of information.

2.1 The classic epidemic model in epidemiology

The epidemic model was proposed by Kermack and McKendrick (1932) to model the disease spreading. The two basic and classic epidemiology models are the SIS model, also known as “Susceptible-Infectious-Susceptible” model, and the SIR model, also known as “Susceptible-Infectious-Recovered” model (Hethcote, 2000).

These two classic models both divide the whole population into different groups according to their state towards a single disease, which included “Susceptible” (S), “Infectious” (I) and “Recovered” (R). The susceptible (S) state refers to those who might be infected with the disease, the infectious (I) state refers to those who currently have the ability to spread the disease and the recovered (R) state refers to those who have recovered from the disease and get the immunity.

The state of each individual could change in the SIS and SIR model. The SIS model only includes S and I states. The susceptible individuals might be infected by the infectious individuals, so that their state could transit from the susceptible to the infectious. The infectious individuals might recover from the disease and their state transit back to the susceptible, that is, they do not get the immunity and might be infected again. However, in the SIR model, the infectious individuals might recover from disease and acquire the immunity, that is, they will never get the disease again and their states change to R, the “Recovered”. The state transition of the SIS model and the SIR model are shown in Figure 1(a) and (b), respectively.
In addition to the similar state description of the whole population, these two models both assume that the whole population is well mixed, and each person has the same probability to contact with all the other people. All individuals are homogenous, that is, the probabilities to transit between different states are the same for all individuals.

2.2 Epidemic models for information diffusion

The spread of disease share much common character with the diffusion of information. By analogizing the spread of disease in the population, Daley and Kendall (1964) first modeled the rumor propagation with the SIR model. In the past decade, the SIR model was also successfully used to model the information diffusion over the social networks (Lerman and Ghosh, 2010; Abdullah and Wu, 2011). The analogy between the spread of disease and the information diffusion with respect to the “SIR” model is listed in Table I.

The “SIS” model can also be used to model the information diffusion (Pastor-Satorras and Vespignani, 2001). However, the “SIS” model does not have the recovered groups. Thus, it cannot characterize the process that the population loses interest in the information and the infectious proportion decrease.

2.3 The dynamics of the classic SIR model

As illustrated in the previous subsection, the state transition of the SIR model is shown in Figure 1(b). To characterize the percentage of each group changing with time, the parameters \( \beta \) and \( \gamma \) are introduced to quantify the transition rates between S, I and R. The parameter \( \beta \) is the contact rate, and represents the average probability of adequate contacts (i.e., contacts sufficient for transmission) of an infectious person per unit time. The parameter \( \gamma \) is called the recovery rate which represents the average probability for an infectious individual to transit to the recovered (Hethcote, 2000). Let \( S(t) \), \( I(t) \) and \( R(t) \) denote the percentage of each group, and the dynamics of the SIR model can be described with the following differential equations:

\[
\frac{dS(t)}{dt} = -\beta I(t)S(t), \quad (1)
\]

\[
\frac{dI(t)}{dt} = \beta I(t)S(t) - \gamma I(t), \quad (2)
\]

\[
\frac{dR(t)}{dt} = \gamma I(t). \quad (3)
\]

The \( S(t), I(t) \) and \( R(t) \) should also satisfy the requirement \( S(t) + I(t) + R(t) = 1 \).

Although the above equations of the SIR model are non-linear, the stable state analysis can still be conducted. By setting equations (1)-(3) to zero, we can first obtain the equilibrium points denoted as \([s, i, r]^T\). According to the stability analysis at the equilibrium points in Hethcote (2000) and Piqueira (2010), the stable state should satisfy that \( i = 0 \) and \( s < \gamma / \beta \). The two conditions indicate that there does not exist any infectious individuals, and the percentage of the susceptible individuals is lower than a threshold which is determined by \( \beta \) and \( \gamma \) at the stable state.

The analysis of the SIR model’s dynamics in Hethcote (2000) also shows that the evolvement of the percentages of three groups in the SIR model can be obtained through
numerical calculations when the initial state \( \{S(0), I(0), R(0)\} \) and the parameters \( \{\beta, \gamma\} \) are given. The evolution process will finally reach one of the stable states.

3. Modeling correlated information diffusion based on the classic SIR model

Most prior works of SIR-based models only considered the diffusion of one single information. In this work, we extend the classic SIR model, and focus on the diffusion process of two pieces of correlated information, which takes the influences of correlated information on intelligent individuals into account.

3.1 The SIR mixture model

In this work, we consider two pieces of correlated information disseminated over the crowd intelligent networks and propose the SIR mixture model. Same as the classic SIR model (Abdullah and Wu, 2011; Daley and Kendall, 1964; Lerman and Ghosh, 2010), we assume that the whole population remains a constant with respect to time, and all individuals in the networks are well mixed and homogeneous.

Let \( E_1 \) and \( E_2 \) represent the correlated information respectively. According to the classic SIR model, each individual can have three possible states towards one information, denoted as \( \{S_1, I_1, R_1\} \) and \( \{S_2, I_2, R_2\} \), respectively. Combining the states with respect to \( E_1 \) and \( E_2 \), we will have 9 states, which are, \( S_1S_2, I_1S_2, R_1S_2, S_1I_2, I_1I_2, R_1I_2, S_1R_2, I_1R_2, \) and \( R_1R_2 \). However, in this work, we consider the simple scenario where individuals have limited attention and can only focus on one information at a time. This will also greatly simplify the theoretical analysis as demonstrated below. Therefore, the state \( I_1I_2 \) where the corresponding individuals spreading \( E_1 \) and \( E_2 \) simultaneously can be excluded. The total 8 possible states and their physical meanings are summarized in Table II. For notational simplicity, we also denote the percentage of individuals in each state at time \( t \) with the same annotation of the corresponding state when no confusions are made. For example, \( S_1S_2(t) \) stands for the percentage of individuals who are in the state of \( S_1S_2 \) at time \( t \), and other notations can be derived in the same way.

### Table I.
The analogy between disease spreading and information diffusion

| Analogy terms  | Disease spreading | Information diffusion       |
|----------------|------------------|----------------------------|
| Spreading object | Disease          | Information                |
| Susceptible   | People who can be infected | People who have not know the information |
| Infectious    | People who are infectious     | People who spread the information |
| Recovered     | People who are recovered     | People who do not spread the information anymore |

### Table II.
The states of the SIR mixture model and their physical meanings

| States  | Physical meaning                                      |
|---------|-------------------------------------------------------|
| \( S_1S_2 \) | People who do not know \( E_1 \) and \( E_2 \)          |
| \( I_1S_2 \) | People who spread \( E_1 \), but do not know \( E_2 \) |
| \( R_1S_2 \) | People who do not spread \( E_1 \) anymore, and do not know \( E_2 \) |
| \( S_1I_2 \) | People who spread \( E_2 \), but do not know \( E_1 \) |
| \( R_1I_2 \) | People who do not spread \( E_1 \) anymore, and spread \( E_2 \) |
| \( S_1R_2 \) | People who do not spread \( E_2 \) anymore, and do not know \( E_1 \) |
| \( I_1R_2 \) | People who do not spread \( E_2 \) anymore, and spread \( E_1 \) |
| \( R_1R_2 \) | People who do not spread \( E_1 \) and \( E_2 \) anymore |

Diffusion in crowd intelligence networks
To take the actual information diffusion scenario into consideration, two pieces of information may not begin to propagate at the same time, that is, one information might spread for a certain time \( t_0 \), and then, the other information begins to spread. Thus, our SIR mixture model can be correspondingly divided into two stages.

In the first stage where \( t \in (0, t_0) \), there is only the first information spreading across the networks. Without loss of generality, we assume that \( E_1 \) spreads first. In this stage, \( E_2 \) does not exist and the diffusion of \( E_1 \) will not be influenced by \( E_2 \). Thus, each individual can only change his/her own state in three possible states, \( \{ S_1S_2, I_1S_2, R_1S_2 \} \), while the percentages of the other five states are always 0 in this stage.

In the information diffusion scenario, we often consider there are only a few people begin to spread the information (Abdullah and Wu, 2011; Daley and Kendall, 1964; Lerman and Ghosh, 2010; Liu et al., 2016). Consequently, the initial state of the first stage should satisfy \( S_1S_2(0) \approx 1 \) and \( I_1S_2(0) \approx 0 \), and the percentages of people in other groups equal to 0 at time 0.

The state transition of the SIR mixture model in the first stage is shown in Figure 2, and we can show that this state transition diagram is exactly the same with only one information propagating as shown in Figure 1(b). The parameter \( \beta_1 \) and \( \gamma_1 \) are the contact rate and the recovery rate of the individuals who do not know \( E_2 \), and they are actually the parameters for \( E_1 \) to spread alone. Therefore, we can use the classic SIR model to describe this process, and the dynamics of the whole population in the first stage can be directly obtained from equations (1)-(3) as:

\[
\frac{dS_1S_2(t)}{dt} = -\beta_1 I_1S_2(t)S_1S_2(t), \tag{4}
\]

\[
\frac{dI_1S_2(t)}{dt} = \beta_1 I_1S_2(t)S_1S_2(t) - \gamma_1 I_1S_2(t), \tag{5}
\]

\[
\frac{dR_1S_2(t)}{dt} = \gamma_1 I_1S_2(t), \quad \text{and} \quad \gamma_1 \tag{6}
\]

\[
S_1I_2(t) = R_1I_2(t) = S_1R_2(t) = I_1R_2(t) = R_1R_2(t) = 0, \quad \text{for } t \in [0, t_0). \tag{7}
\]

At the end of the first stage at \( t_0 \), the percentages of individuals at state \( S_1S_2, I_1S_2 \) and \( R_1S_2 \) are \( S_1S_2(t_0), I_1S_2(t_0) \) and \( R_1S_2(t_0) \), respectively. The value of these three percentages are determined by the parameters \( \beta_1, \gamma_1 \) and \( t_0 \), that can be numerically solved and analyzed as in prior works of the classic SIR model (Hethcote, 2000). The percentages of the other five states still remain 0 as the time approaches \( t_0 \).

The second stage of the SIR mixture model begins at time \( t_0 \) where \( E_2 \) also begins to spread across the network together with \( E_1 \). The state transition of the SIR mixture model in the second stage is shown in Figure 3.

**Figure 2.**
The state transition of the SIR mixture model in the first stage.
The initial state of the second stage is determined by the end of the first stage. According to the previous analysis of the first stage, the percentage of individuals in state $S_1S_2$, $I_1S_2$ and $R_1S_2$ at $t = t_0$ are $S_1S_2(t_0)$, $I_1S_2(t_0)$ and $R_1S_2(t_0)$, respectively, while the percentages of other five states equal to 0. To characterize the beginning of propagation of $E_2$, we assume that a little perturbation of $S_1I_2$ exists. That is, at time $t_0$, some individuals whose state are $S_1S_2$ originally, become the “infectious” individuals of $E_2$ and change their current state to $S_1I_2$. Thus, this little perturbation will cause the percentage of $S_1I_2(t_0)$ to be slightly more than 0, and the percentage of $S_1S_2(t_0)$ to be slightly less than the original value calculated at the end of the first stage, while other percentages remain unchanged.

At the beginning of the second stage, most individuals are in states $S_1S_2$, $I_1S_2$ and $R_1S_2$. For the individuals in state $R_1S_2$ who had spread $E_1$, they can spread $E_2$, i.e., become the “infectious” individuals of $E_2$ with their states changing to $R_1I_2$, and finally lose interests in $E_2$ and become $R_1R_2$. This process is the “Part 2” shown in Figure 3. According to our assumption that each user can only spread one information at a time, the individuals in state $I_1S_2$ can only finish their spreading of $E_1$ first, and change their state to $R_1S_2$. After that, they can spread $E_2$, and the following state transition is the same as the individuals in state $R_1S_2$.

For the individuals in state $S_1S_2$, they can either firstly spread $E_1$ or $E_2$ in the second stage of the SIR mixture model, which are the “Part 1” and “Part 3” in Figure 3, respectively. Furthermore, for the individuals in state $S_1R_2$ who had spread $E_2$ first, they can still become a spreader of $E_1$ and transit their state in “Part 4” in Figure 3.

We can show from the state transition diagram in Figure 3 that both “Part 1” and “Part 4” correspond to the propagation of information $E_1$, while “Part 2” and “Part 3” correspond to the propagation of information $E_2$. “Part 1” is the propagation of $E_1$ for the individuals in $S_1S_2$ who do not know and never spread $E_2$, while “Part 4” is the propagation of $E_1$ for individuals who are in state $S_1R_2$ and have spread $E_2$. The difference between “Part 2” and “Part 3” can be derived in a similar way. Therefore, by comparing “Part 1” with “Part 4”, we can study how $E_2$ affects the propagation of $E_1$, and similarly, we can analyze how $E_1$ affects the propagation of $E_2$ by comparing “Part 2” with “Part 3”. 

To address the impact of correlated information on each other’s propagation, in this work, we assume that the individuals in state $S_1R_2$ who have spread $E_2$ have different contact rate of $E_1$ from those who have not spread and never know $E_2$ ($S_1S_2$), due to the prior knowledge of $E_1$ which is generated by spreading $E_2$. We denote the contact rate in the former case as $\alpha_1$, which is different from the latter case defined as $\beta_1$ previously. We also assume that the recovery rates of the individuals in state $I_1S_2$ and $I_1R_2$ are different to generalize the SIR mixture model, which are denoted as $\gamma_1$ and $\delta_1$, respectively. We can also denote $\beta_2$ and $\alpha_2$ as the contact rate of individuals in state $S_1S_2$ and $R_1S_2$. The recovery rate for individuals in state $S_1I_2$ and $R_1I_2$ are $\gamma_2$ and $\delta_2$, respectively. All parameters of the SIR mixture model are summarized in Table III.

![Figure 3](image-url)
According to the dynamics of the SIR model, we can use differential equations to describe the second stage of the SIR mixture model, where $E_1$ and $E_2$ spread together. Based on the assumptions and parameters’ physical meanings described previously, we can obtain that:

\[
\frac{dS_1S_2(t)}{dt} = -\beta_1 I_1(t) S_1(t) S_2(t) - \beta_2 I_2(t) S_1(t), \quad (8)
\]

\[
\frac{dI_1S_2(t)}{dt} = \beta_1 I_1(t) S_1(t) S_2(t) - \gamma_1 I_1(t) S_2(t), \quad (9)
\]

\[
\frac{dS_1I_2(t)}{dt} = \beta_2 I_2(t) S_1(t) S_2(t) - \gamma_2 S_1(t) I_2(t), \quad (10)
\]

\[
\frac{dR_1S_2(t)}{dt} = \gamma_1 I_1(t) S_2(t) - \alpha_2 I_2(t) R_1(t) S_2(t), \quad (11)
\]

\[
\frac{dS_1R_2(t)}{dt} = \gamma_2 S_1(t) I_2(t) - \alpha_1 I_1(t) S_2(t), \quad (12)
\]

\[
\frac{dR_1I_2(t)}{dt} = \alpha_2 I_2(t) R_1(t) S_2(t) - \delta_2 R_1(t) I_2(t), \quad (13)
\]

\[
\frac{dI_1R_2(t)}{dt} = \alpha_1 I_1(t) S_1(t) R_2(t) - \delta_1 I_2(t) R_2(t), \quad (14)
\]

\[
\frac{R_1R_2(t)}{dt} = \delta_2 R_1(t) I_2(t) + \delta_1 I_1(t) R_2(t), \quad (15)
\]

where $I_1(t) = I_1 S_2(t) + I_1 R_2(t)$ represents all individuals who are spreading $E_1$ at time $t$, and $I_2(t) = S_1 I_2(t) + R_1 I_2(t)$ represents all the individuals who are spreading $E_2$ at time $t$. equations (8), (9) and (11); equations (11), (13) and (15); equations (8), (10) and (12); and

| Parameter | Physical meaning |
|-----------|------------------|
| $\beta_1$ | Transition rate between $S_1S_2$ and $I_1S_2$ |
| $\beta_2$ | Transition rate between $S_1S_2$ and $S_1I_2$ |
| $\alpha_1$ | Transition rate between $S_1R_2$ and $I_1R_2$ |
| $\alpha_2$ | Transition rate between $R_1S_2$ and $S_1I_2$ |
| $\gamma_1$ | Transition rate between $I_1S_2$ and $R_1I_2$ |
| $\gamma_2$ | Transition rate between $S_1I_2$ and $S_1R_2$ |
| $\delta_1$ | Transition rate between $I_1R_2$ and $R_1I_2$ |
| $\delta_2$ | Transition rate between $R_1I_2$ and $R_1R_2$ |
| $t_0$ | Time for $E_1$ spreading alone |

**Table III.** The parameters of the SIR mixture model and their physical meanings.
equation (12), (14) and (15) describe the processes of “Part 1”, “Part 2”, “Part 3” and “Part 4” in Figure 3, respectively, based on the dynamics of SIR model.

3.2 Stable state analysis of the susceptible–infectious–recovered mixture model

Since we have derived the dynamical equations of the SIR mixture model in equations (8)-(15), we can perform the stability analysis. Let us denote the dynamical state as \( x(t) = [S_1 S_2(t), I_1 S_1(t), S_1 I_2(t), R_1 S_2(t), S_1 R_2(t), R_1 I_2(t), I_1 R_2(t), R_1 R_2(t)]^T \). By setting the dynamical equations (8)-(15) to zero, we can obtain the equilibrium points of the SIR mixture model, i.e., \( x_e = [S_1 s_2, I_1 s_2, s_1 s_2, r_1 s_2, r_1 r_2, r_1 r_2]^T \). Then by adopting the Lyapunov first method (Teschl, 2012), we can analyze the stability condition for the equilibrium points \( x_e \) (Liu et al., 2016; Piqueira, 2010).

We first show that the equilibrium points should satisfy the requirement that \( i_1 s_2, i_1 r_2, s_1 r_2 \) and \( r_1 r_2 \) should equal to zero. Mathematically, all parameters are both positive, and each entry of the equilibrium points \( x_e \) is non-negative due to their physical meanings. Then, by setting equation (15) to zero, we should have \( r_1 r_2 = 0 \) and \( i_1 r_2 = 0 \). Taking these two conditions into other equations, we will have \( i_1 s_2 = 0 \) and \( i_1 r_2 = 0 \). It is also reasonable in the physical aspect, because the equilibrium point means the percentage of each group should not change. Therefore, it should satisfy that no individuals spread \( E_1 \) or \( E_2 \) anymore at least, which means \( i_1 s_2, i_1 r_2, s_1 r_2 \) and \( r_1 r_2 \) should equal to zero. Thus, the equilibrium points can be simplified to \( x_e = [S_1 s_2, 0, 0, r_1 s_2, s_1 r_2, 0, 0, r_1 r_2]^T \). That is, at the equilibrium points, the state of each individual should be in one of \( S_1, S_2, I_1, S_1, S_1, R_1 \) and \( R_1 \).

Due to the non-linearity of the dynamical equations (8)-(15), we could not obtain the exact values of \( s_1 s_2, r_1 s_2, s_1 r_2 \) and \( r_1 r_2 \). However, we can adopt the Lyapunov first method to analyze the stability condition for \( s_1 s_2, r_1 s_2, s_1 r_2 \) and \( r_1 r_2 \). We can first take the linear approximation of the dynamical equations, which is the Jacobian matrix \( J \) of the right hand side of equations (8)-(15). Then, plugging the simplified equilibrium point \( x_e \) into the linear approximation of the dynamical system, we can have:

\[
\dot{x}(t) = J \cdot (x(t) - x_e)
\]

\[
= \begin{bmatrix}
0 & -\beta_1 s_1 s_2 & -\beta_2 s_1 s_2 & 0 & 0 & -\beta_2 s_1 s_2 & -\beta_1 s_1 s_2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\alpha_1 r_1 s_2 & 0 & 0 & -\alpha_2 r_1 s_2 & 0 & 0 \\
0 & 0 & -\alpha_1 r_1 s_2 & 0 & 0 & -\alpha_2 r_1 s_2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(16)

According to the criterion of Lyapunov stability in Teschl (2012) and Liu et al. (2016), the real part of the eigenvalue of \( J \) should be less than or equal to zero, and for the eigenvalues whose real part is equal to zero, the algebraic multiplicity should equal to the geometry multiplicity. We can first derive the characteristic equation of \( J \) from equation (16), which is:
The existence of the second peak may indicate that the corresponding online content has something new so that the volume would go up again. We attribute this to the online content variation in social media, the online contents exhibit rich temporal dynamics, that is, how popularity grows and fades over time. The dataset includes top-1000 online contents with blue solid lines. For the first kind of online contents, shown in Figure 4(a), their time sequences have only one peak. These diffusion processes can be described by the classic SIR model appropriately which is shown in Figure 4(a) with the green dot line. However, for the second kind of online contents, shown in Figure 4(b), there exist at least two peaks in the temporal sequences. The SIR model cannot fit the data properly, as shown in Figure 4(b) with the green dot line.

4. Real data validation with SIR mixture model

In this section, we use the dataset in Yang and Leskovec (2011) to validate our SIR mixture model. According to the study by Yang and Leskovec about the patterns of temporal variation in social media, the online contents exhibit rich temporal dynamics, that is, how contents' popularity grows and fades over time. The dataset includes top-1000 online contents with the largest overall volumes from Sept. 2008 to Aug. 2009. Each online content has a time sequence of 128 entries which indicates the numbers of news articles or the blog posts in each hour around the most popular period of the corresponding online contents.

Each online content can be regarded as the diffusing information, and the corresponding time sequence is the diffusion process of the information. Thus, the SIR model can be used to describe these time sequences, which correspond to the dynamics of the I groups.

To obtain the percentage of the news articles and blog posts with respect to each time unit, we adopt the method in Jiang et al. (2014b) and assume that the total number of news websites and bloggers is the maximal value of all time sequences in the dataset and divide all the time sequences by the maximal value, so that each entry of time sequences are normalized to [0,1].

Through the observation of the whole 1000 time sequences, they can be roughly divided into two kinds, whose representatives are shown in Figure 4 with blue solid lines. For the first kind of online contents, shown in Figure 4(a), their time sequences have only one peak. These diffusion processes can be described by the classic SIR model appropriately which is shown in Figure 4(a) with the green dot line. However, for the second kind of online contents, shown in Figure 4(b), there exist at least two peaks in the temporal sequences. The SIR model cannot fit the data properly, as shown in Figure 4(b) with the green dot line.
consisting of two pieces of correlated sub-information, and the first peak mainly results from the first sub-information, while the second sub-information lead to the second peak of the time sequences. Hence, the proposed SIR mixture model can be used to describe this kind of information diffusion process.

Because the dataset does not provide the detailed diffusion processes of each information, we only have the total volumes of each information over time. To fit the real data into our proposed SIR mixture model, we define the percentage of total infectious individuals of both $E_1$ and $E_2$ at time $t$, that is:

\[
I(t) = I_1(t) + I_2(t) = I_1S_2(t) + I_1R_2(t) + S_1I_2(t) + R_1I_2(t).
\]

Thus, $I(t)$ corresponds to the time sequences of the percentage of news articles and blog posts given by the SIR mixture model, and we also denote $\hat{I}(t)$ as the real time sequences of the percentage of the same news articles and blog posts.

Similar to the validation of the SIR model with real data, the aim of fitting time sequences with the SIR mixture model is to learn the parameters $\omega = [\beta_1, \beta_2, \gamma_1, \gamma_2, \alpha_1, \alpha_2, \delta_1, \delta_2, t_0]^T$, so that it minimizes the following MSE between the real data and the model fitting result, that is:

\[
\min_{\omega} \frac{1}{128} \sum_{t=0}^{127} \left| I(t) - \hat{I}(t) \right|^2.
\]

Because we cannot obtain the closed form expression of the $I(t)$ of SIR mixture model through such complex differential equations, we use particle swarm optimization (PSO) algorithm Kennedy and Eberhart (1995) to efficiently obtain the solution to equation (20). Due to the discreteness of $t_0$, we try each $t_0$ from 0 to 120 with a step size of 5. For a determined $t_0$, the PSO solver is used to obtain the optimal solution of the optimization problem equation (20) with respect to $[\beta_1, \beta_2, \gamma_1, \gamma_2, \alpha_1, \alpha_2, \delta_1, \delta_2]$. The result is shown in Figure 4(a) and (b) with orange dash lines.

**Notes:** (a) There is only one peak in the time sequence; (b) two peaks exist in the time sequence.
We conduct the real data fitting validation of the total 1000 time sequences in the dataset, and find that the average MSE given by the SIR mixture model is $5.19 \times 10^{-4}$. Compared to the average MSE given by the classic SIR model, $1.44 \times 10^{-3}$, the proposed model reduces the MSE by about 64%, which indicates that the SIR mixture model is much more promising than the classic SIR model in characterizing complex information propagation processes. Specifically, with the time sequence shown in Figure 4, we can show that the SIR mixture model performs as well as the classic SIR model in terms of the diffusion process with single peaks. However, our SIR mixture model outperforms the classic SIR model with respect to the diffusion process with double peaks.

Because no detailed information of each time sequence is provided in the dataset, we use the parameters learned by fitting into the corresponding real data to calculate the percentages of infectious individuals of $E_1$ and $E_2$ with respect to time $t$, which are $I_1(t) = I_1S_2(t) + I_1R_2(t)$ and $I_2(t) = S_1I_2(t) + R_1I_2(t)$, respectively. Then, we can use the detailed diffusion processes of $E_1$ and $E_2$ to analyze the evolvement processes of the correlated information diffusion and the influence between correlated information.

We take the time sequence in Figure 4(b) as an example, and the other time sequences can be analyzed in the same way. We first plot $I_1(t)$ and $I_2(t)$ in Figure 5(a) according to the SIR mixture model whose parameters are learned by the corresponding real data. It validates our assumption that the existence of the two peaks results from the spreading of $E_1$ and $E_2$, respectively. The interval between the two peaks is approximately 25 time units, which is also exactly the fitting results of $t_0$.

Through the fitting parameters, we observe that $\gamma_1 \approx \gamma_2 \approx 0.35$ and $\delta_1 \approx \delta_2 \approx 0.57$ which indicates that the crowd’s recovery rates are barely influenced by the other information. The reason might be that people pay more attention to the information that they are spreading now. Hence, the recovery rates are less affected by the correlated information.

The fitting results show that $\beta_2 \approx 0.54 < \alpha_2 \approx 1$, which indicates that in this particular example, people who have spread $E_1$ are easier to spread $E_2$. Thus, $E_1$ has a positive impact on $E_2$. We also plot $I_1S_2(t) = I_1R_2(t), S_1I_2(t)$ and $R_1I_2(t)$ in Figure 5(b), and the curve of $R_1I_2(t)$ grows faster and higher than one of $S_1I_2(t)$ which validates our analytical results.

The result in Figure 5(b) also indicates that $I_1R_2 \approx 0$ with respect to time. We also show that $\alpha_1 \approx 0.52 < \beta_1 \approx 0.63$, that is, the diffusion of $E_1$ is suppressed by $E_2$. We infer that it is mainly because people have more interests in the new information although this two information are correlated.

![Figure 5.](image-url)

**Notes:** (a) Percentages of $I_1$ and $I_2$; (b) percentages of $I_1S_2$, $S_1I_2$, $I_1S_2$ and $R_1I_2$
5. Conclusion
In this work, we extend the classic SIR model, and propose the SIR mixture model which formulates the process that two pieces of correlated information jointly propagates over the crowd intelligent networks.

To describe the influence between the correlated information, we exploit the character of the crowd intelligence, that intelligent individuals will have pre-judgment and prior knowledge over the second information for spreading the correlated first information. The influence will result in a different probability to spread the second information compared with those who have not heard and never spread the first one.

The crowd is divided into 8 groups according to their state towards two pieces of information. The SIR mixture model is a two-stage model including the process that the first information propagates alone, which can be modeled by the SIR model, and the process both information spread together whose dynamics can be described by eight differential equations. We also discuss the stable state of the SIR mixture model through linearization of the differential equations and obtain the condition for the stable state.

Finally, we validate our model with real data, and find that our model can be used to describe not only the information diffusion process with one peak, but also the more complex one with two peaks. We also use the parameters learned from the real data to reason about how correlated information interact with each other and propagate over the crowd intelligent networks.

References
Abdullah, S. and Wu, X. (2011), “An epidemic model for news spreading on twitter”, IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 163-169.
Beutel, A., Prakash, B.A., Rosenfeld, R. and Faloutsos, C. (2012), “Interacting viruses in networks: can both survive?”, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 426-434.
Cao, X., Chen, Y., Jiang, C. and Liu, K.J.R. (2016), “Evolutionary information diffusion over heterogeneous social networks”, IEEE Transactions on Signal and Information Processing over Networks, Vol. 2, pp. 595-610.
Chai, Y., Miao, C., Sun, B., Zheng, Y. and Li, Q. (2017), “Crowd science and engineering: concept and research framework”, International Journal of Crowd Science, Vol. 1 No. 1, pp. 2-8.
Daley, D.J. and Kendall, D.G. (1964), “Epidemics and rumours”, Nature, Vol. 204, p. 1118.
De Choudhury, M., Lin, Y.R., Sundaram, H., Candan, K.S., Xie, L. and Kelliher, A. (2010), “How does the data sampling strategy impact the discovery of information diffusion in social media?”, International AAAI Conference on Web and Social Media (ICWSM), pp. 34-41.
Fu, G., Chen, F., Liu, J. and Han, J. (2019), “Analysis of competitive information diffusion in a group-based population over social networks”, Physica A: Statistical Mechanics and Its Applications, Vol. 525, pp. 409-419.
Goldenberg, J., Libai, B. and Muller, E. (2001), “Talk of the network: a complex systems look at the underlying process of word-of-mouth”, Marketing Letters, Vol. 12 No. 3, pp. 211-223.
Granovetter, M. (1978), “Threshold models of collective behavior”, American Journal of Sociology, Vol. 83 No. 6, pp. 1420-1443.
Guille, A., Hacid, H., Favre, C. and Zighed, D.A. (2013), “Information diffusion in online social networks: a survey”, ACM Sigmod Record, Vol. 42 No. 1, pp. 17-28.
Hethcote, H.W. (2000), “The mathematics of infectious diseases”, SIAM Review, Vol. 42 No. 4, pp. 599-653.
Huang, Y., Chai, Y., Liu, Y. and Gu, X. (2017), “Intelligent interaction based on holographic personalized portal”, International Journal of Crowd Science, Vol. 1 No. 2, pp. 171-182.

Jiang, C., Chen, Y. and Liu, K.J.R. (2014), “Evolutionary dynamics of information diffusion over social networks”, IEEE Transactions on Signal Processing, Vol. 62 No. 17, pp. 4573-4586.

Jiang, C., Chen, Y. and Liu, K.J.R. (2014), “Graphical evolutionary game for information diffusion over social networks”, IEEE Journal of Selected Topics in Signal Processing, Vol. 8 No. 4, pp. 524-536.

Kennedy, J. and Eberhart, R.C. (1995), “Particle swarm optimization”, International Symposium on Neural Networks, pp. 1942-1948.

Kermack, W.O. and McKendrick, A.G. (1932), “Contributions to the mathematical theory of epidemics. II. The problem of endemicity”, Proc. R. Soc. Lond. A, Vol. 138 No. 834, pp. 55-83.

Lerman, K. and Ghosh, R. (2010), “Information contagion: an empirical study of the spread of news on digg and twitter social networks”, International AAAI Conference on Web and Social Media (ICWSM), pp. 90-97.

Liu, Y., Wang, B., Wu, B., Shang, S., Zhang, Y. and Shi, C. (2016), “Characterizing super-spreading in microblog: an epidemic-based information propagation model”, Physica A: Statistical Mechanics and Its Applications, Vol. 463, pp. 202-218.

Myers, S.A. and Leskovec, J. (2012), “Clash of the contagions: cooperation and competition in information diffusion”, IEEE 12th International Conference on Data Mining (ICDM), pp. 539-548.

Nan, Y., Liu, Y., Shen, J. and Chai, Y. (2017), “A study on MCIN model in intelligent clothing industry”, International Journal of Crowd Science, Vol. 1 No. 2, pp. 133-145.

Pastor-Satorras, R. and Vespignani, A. (2001), “Epidemic spreading in scale-free networks”, Physical Review Letters, Vol. 86 No. 14, p. 3200.

Pierson, D. (2011), “Japan radiation fears spark panic salt-buying in China”, Los Angeles Times, available at: http://articles.latimes.com/2011/mar/18/world/la-fg-china-iodine-salt-20110318

Piqueira, J.R.C. (2010), “Rumor propagation model: an equilibrium study”, Mathematical Problems in Engineering, pp. 1-7.

Prakash, B.A., Beutel, A., Rosenfeld, R. and Faloutsos, C. (2012), “Winner takes all: competing viruses or ideas on fair-play networks”, Proceedings of the 21st International Conference on World Wide Web, pp. 1037-1046.

Rui, X., Meng, F., Wang, Z., Yuan, G. and Du, C. (2018), “SPIR: the potential spreaders involved SIR model for information diffusion in social networks”, Physica A: Statistical Mechanics and Its Applications, Vol. 506, pp. 254-269.

Sun, L., Zhou, Y. and Guan, X. (2017), “Modelling multi-topic information propagation in online social networks based on resource competition”, Journal of Information Science, Vol. 43 No. 3, pp. 342-355.

Teschl, G. (2012), “Ordinary differential equations and dynamical systems”, American Mathematical Soc.

Xia, L., Jiang, G., Song, Y. and Song, B. (2015), “Modeling and analyzing the interaction between network rumors and authoritative information”, Entropy, Vol. 17 No. 1, pp. 471-482.

Yang, J. and Leskovec, J. (2011), “Patterns of temporal variation in online media”, Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 177-186.
Zhao, L., Wang, J., Chen, Y., Wang, Q., Cheng, J. and Cui, H. (2012), “SIHR rumor spreading model in social networks”, Physica A: Statistical Mechanics and Its Applications, Vol. 391 No. 7, pp. 2444-2453.

Zhao, L., Xie, W., Gao, H.O., Qiu, X., Wang, X. and Zhang, S. (2013a), “A rumor spreading model with variable forgetting rate”, Physica A: Statistical Mechanics and Its Applications, Vol. 392 No. 23, pp. 6146-6154.

Zhao, L., Qiu, X., Wang, X. and Wang, J. (2013b), “Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks”, Physica A: Statistical Mechanics and Its Applications, Vol. 392 No. 4, pp. 987-994.

Corresponding author
Yuejiang Li can be contacted at: lyj18@mails.tsinghua.edu.cn

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com