Modeling Multi-source Information Diffusion: A Graphical Evolutionary Game Approach

Hong Hu*,Yuejiang Li*, H.V Zhao* and Yan Chen†
* Dept. of Automation and Inst. for Aritificial Intelligence, Tsinghua University,
State Key Lab for Intelligent Tech.&Sys., BNRist Center, Beijing, P.R.China
† School of Info.&Comm. Engr., Univ. of Electronic Science and Technology of China, Chengdu, P.R.China

***Abstract*—Modeling of information diffusion over social networks is of crucial importance to better understand how the avalanche of information overflow affects our social life and economy, thus preventing the detrimental consequences caused by rumors and motivating some beneficial information spreading. However, most model-based works on information diffusion either consider the spreading of one single message or assume different diffusion processes are independent of each other. In real-world scenarios, multi-source correlated information often spreads together, which jointly influences users' decisions. In this paper, we model the multi-source information diffusion process from a graphical evolutionary game perspective. Specifically, we model users' local interactions and strategic decision making, and analyze the evolutionary dynamics of the diffusion processes of correlated information, aiming to investigate the underlying principles dominating the complex multi-source information diffusion. Simulation results on synthetic and Facebook networks are consistent with our theoretical analysis. We also test our proposed model on Weibo user forwarding data and observe a good prediction performance on real-world information spreading process, which demonstrates the effectiveness of the proposed approach.**

I. Introduction

In recent years, with the popularity of the mobile Internet, online social networks have rapidly emerged, where information spreads in a sharing, editable and interactive pattern. Instead of only receiving the information in the past, many intelligent individuals have been enrolled in the creating and propagating process of information, which further results in the explosion of information. However, the convenience of spreading information can also be used in a malicious manner, where gossips and rumors propagate across online social platforms and sometimes lead to severe damage to the economy. Therefore, understanding the mechanisms of the information diffusion over social networks is critical not only to the personalized recommendation, holographic assistants platforms and other business applications, but also to national security and social stability.

In online social networks, information diffusion can be described as follows: when a piece of information is released by someone, his/her neighbors may be influenced and decide to forward the information or not. Thus, both how intelligent users of online social networks prefer the information and how they interact with each other will determine the diffusion process of the information.

There are many prior works on modeling and studying the information diffusion process in social networks [1][2][3],

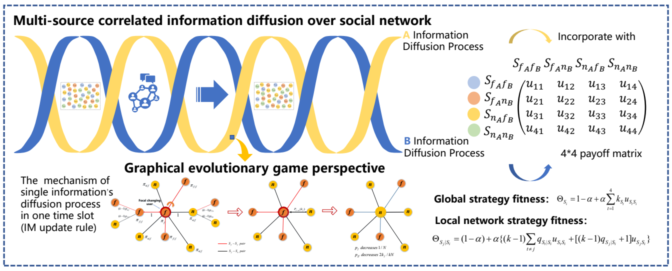


Fig. 1. An overview of mulit-source information diffusion process with graphical evolutionary game perspective

which mainly can be classified into two categories: empirical data-driven approaches and theoretical model-based analysis. Specifically, the data-driven approaches usually use a machine-learning algorithm to learn the parameters and features from large amounts of data. The statistical feature extraction mechanism has been extensively elaborated in [4]-[8]. In [4], the authors developed a data-driven framework which models the "forward probability" of a user's adoption of information and statistically learns how different contagions interact with each other through the Twitter dataset. However, users' interactions were ignored, and the learned "cluster" concept lacked corresponding real-world meanings. In [5], extensive mechanisms were designed to eliminate these shortcomings. Nevertheless, the approach did not has large improvement in prediction performance but has high sensitivity and strict requirement of the training dataset. Li [6] concluded the limitations of previous data-driven works and developed a micro-perspective statistical approach. The approach jointly exploited the global and local influence of users to calculate time-related social payoffs based on real data (Sina Weibo and Flickr). However, the approach did not consider multiple contagions and the extensibility of the approach is not satisfied.

In contrast to the data-driven approaches, model-based approaches focus on using different mathematical models to analyze the network's dynamic equations. Inspired by knowledge from psychology, sociology, and economics, these models usually have strong interpretability and extensibility. Some novel models include epidemic model [9], independent cascade (IC) model [10], linear threshold (LT) model [11] and their variants [12]-[15]. Using these models, many meaningful problems were investigated, e.g., influence maximization, accurate personalized marketing, and breaking news detection.

The work in [16] and [17] provided a new evolutionary game-theoretic approach of information diffusion. Considering all data are generated by users, it is not wise to only focus on optimizing a global objective function but totally ignore users' decisions, actions and their interactions [18]. Inspired by biology, evolutionary game theory (EGT) [19] was used to simulate the biological evolution originally. It comes naturally that we can take users' information exchanging, interaction and forwarding process as an analogy of creatures' complex natural selection to investigate the underlying principles in high-dynamic information diffusion processes. This promising framework can model users' microeconomic interactions while the information diffuses dynamically. It also has a strong ability of interpretability and extensibility, suggesting a suitable and tractable paradigm for the challenge of analyzing information diffusion [20].

In other areas of signal processing and special scenarios, the framework is also proved to be effective. In [20], the authors attempted to apply the framework in heterogeneous networks (each user has a diverse preference, thus, different payoff matrix). The work in [21] took irrational users (those who always intentionally forward fake news) into consideration. And in [22]-[24], the framework was extended to image processing and communication networking areas, suggesting the extensibility of this promising modeling tool.

In this paper, we study the multi-source information diffusion over social networks with a graphical evolutionary game perspective. Modeling users' decision making as an evolutionary game, we analyze the multi-source information diffusion dynamics. Through the study in this work, the main contributions of this paper can be summarized as follows:

1) Unlike the previous graphical EGT information diffusion models that consider the spreading of one single message and assume different diffusion processes are independent of each other, this paper studies multi-source correlated information, which is more consistent with the real-world scenarios of information diffusion.
2) The population dynamics, relationship dynamics and influence dynamics of multi-source information diffusion over social networks are derived specifically. To verify the model's effectiveness, we conduct the experiments not only in the synthetic network but also in real-world Facebook network. The simulation results are consistent with our theoretical derivation.
3) Since the competitive or cooperative relations of correlated information are hidden in the parameters of the 4*4 payoff matrix, we conduct a reverse experiment to learn the model-desired parameters from the crawled Sina Weibo users' forwarding data at correlated hashtags. We observe that with only 18 changeable parameters, our framework's prediction results can also fit well with the real-world diffusion data.
4) An effective prediction means that we can estimate future data from the past. Therefore, we limit the proportion of the training dataset and observe that only 60% ahead of the crawled data can fit the whole information-spreading process with less than 8% error rate, suggesting a good future prediction performance of our model.

The rest of this paper is organized as follows. In Section II, we first formulate the multi-source information diffusion process and define some important expressions. Then, in Section III, we theoretically extend the framework to multi-source information diffusion process by re-deriving the evolutionary dynamics in the 4*4 payoff matrix. The verification experiment and prediction test are in Section IV. The conclusions are drawn in Section VI.

II. Graphical Evolutionary Game Fomulation for Multi-source Diffusion

In this section, we will briefly introduce the basic concept of graphical evolutionary game theory and formulate the multi-source information diffusion process over social networks.

In the information diffusion process, graph structure refers to the topology of social network where the information is propagating. The nodes represent users and edges represent social relationships between users, e.g., friendship in Facebook. As for each user, he or she may have several strategies in the game, i.e forward the information or not. In this paper, we define users' strategy set $\{S_{f_Af_B}, S_{f_An_B}, S_{n_Af_B}, S_{n_An_B}\}$ as follows:

$$\begin{cases} S_{f_Af_B}, & \text{forward A and B,} \\ S_{f_An_B}, & \text{forward A not B ,} \\ S_{n_Af_B}, & \text{forward B not A,} \\ S_{n_An_B}, & \text{not forward A or B .} \end{cases} \tag{1}$$

which can be shorthand for $\{S_1, S_2, S_3, S_4\}$.

The utility function in the evolutionary game can be defined as "fitness". In the information diffusion scenario, users with larger fitness values have a bigger impact on their neighbors, and more users imitate their strategies. Specifically, the fitness is calculated by $\pi = (1-\alpha)B + \alpha U$ where B is the baseline fitness representing the users' personal interest and U is the fitness getting from interactions. The parameter α represents the selection intensity, i.e, the relative contribution of the game to fitness. Under usual circumstances, individual difference and preference have a large contribution to fitness, thus $\alpha \to 0$.

The payoff matrix in this paper is defined as $\{u_{s_is_j}\}_{4\times4}$ where $u_{s_is_j}$ denotes the interacting payoff of a user with strategy S_j meeting a user with strategy S_i, regardless of other strategies' influence. In the network, four strategies' population is donoted as $\{k_{S_1}, k_{S_2}, k_{S_3}, k_{S_4}\}$ and the average fitness of each strategy is $\{\Theta_{S_1}, \Theta_{S_2}, \Theta_{S_3}, \Theta_{S_4}\}$. Therefore, the fitness of strategy S_i can be calculated as:

$$\Theta_{S_i}(k_{S_1}, k_{S_2}, k_{S_3}, k) = 1 - \alpha + \alpha \sum_{t=1}^{4} k_{S_t} u_{s_ts_i} \tag{2}$$

With the time going forward, users may change their strategies according to their friends' fitness of adopting strategy randomly. Specifically, the strategy evolving process is divided into time slots. In each time slot, a user is selected randomly

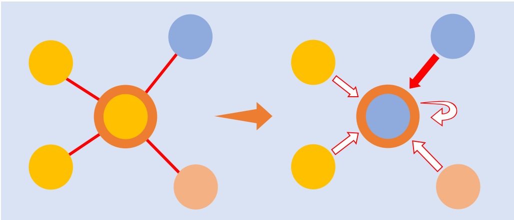


Fig. 2. IM update rule. The focal selected user will change or remain his/her strategy according to each strategy's fitness in his/her local neighbor network

as the focal user to update the strategy with a certain update rule which originated from biology evolving theory. Here, we adopt the imitation update rule (IM), where the selected focal user will either imitate the strategy of one neighbor or keep his/her current strategy unchanged. The probability of imitating a strategy(including himself/herself) is proportional to each strategy's fitness in the focal user's local neighbor network.

Since the strategy-updating process is happened in the local neighbor network of the selected focal user, rather than the global network, we define four kinds of local network to denote four kinds of focal user's neighbor networks: Q_1, Q_2, Q_3, Q_4, where $Q_i = \{q_{s_1|s_i}, q_{s_2|s_i}, q_{s_3|s_i}, q_{s_4|s_i}\}$. The $q_{s_j|s_i}$ donotes the percentage of strategy S_j users in the local network when focal user adopts strategy S_i. Therefore, ***local network state*** can be denoted as $Q = \{Q_1^T, Q_2^T, Q_3^T, Q_4^T\} = \{q_{s_j|s_i}\}_{4\times 4}$. Each strategy's global percentage can be defined as the ***global network state***, which is $X = (p_{s_1}, p_{s_2}, p_{s_3}, p_{s_4})^T$. Two defined states have several relations as follows:

$$\sum_{i=1}^{4} q_{s_i|s_j} p_{s_i} = p_{s_j}, QX = X \tag{3}$$

$$\sum_{i} p_{s_i} = 1, \Xi^T X = 1 \tag{4}$$

where $\Xi = (1,1,1,1)^T$.

III. Three Multi-Source Information Diffusion Dynamics' Derivation

In this section, with the definitions of the global network state and local network state, we will define and derive three dynamics of multi-source information diffusion over social networks as follows:

- Population dynamics: $\dot{p}_{s_i}, (i = 1,2,3,4)$
- Relationship dynamics: $\dot{p}_{s_j s_i}, (1 \le i,j \le 4)$
- Influence dynamics: $\dot{q}_{s_j|s_i}, (1 \le i,j \le 4)$

A. Population Dynamics Derivation

The population dynamics are the dynamics of each strategy's adopting percentage in global networks. In the following, we will analyze different scenarios with the objective of deriving the close-form expression of the population dynamics.

1) $\Delta p_{S_i} = -\frac{1}{N}$ scenario: We first analyze the scenario that a S_i user changes his/her strategy to other strategy. Specifically, according to IM update rule, the randomly selected changing user is with strategy S_i. Then, he/she will imitate a $S_j (j \neq i)$ user's strategy, proportional to the fitness of the corresponding strategy. In this case, the percentage of strategy S_i will decrease one unit ($\Delta p_{S_i} = -\frac{1}{N}$).

We suppose the selected user adopts strategy S_i and has k neighbors, among which k_{S_1}, k_{S_2}, k_{S_3}, k_{S_4} users adopt strategy S_1, S_2, S_3, S_4. Since this is a multinomial distribution, the probability of such a scenario is:

$$\Psi^{de}_{(S_i;k_{S_1},k_{S_2},k_{S_3},k_{S_4})} = \frac{k!}{k_{S_1}!k_{S_2}!k_{S_3}!k_{S_4}!} q_{s_1|s_i}^{k_{S_1}} \cdot q_{s_2|s_i}^{k_{S_2}} \cdot q_{s_3|s_i}^{k_{S_3}} \cdot q_{s_4|s_i}^{k_{S_4}} \tag{5}$$

Among the focal user's k nerghbors, for the neighbors adopting $S_j (j \neq i)$, each of them has $(k-1)q_{s_t|s_j}$ neighbors using strategy $S_t (t \neq i)$, as well as $1 + (k-1)q_{s_i|s_j}$ neighbors using strategy S_i averagely, where "1" means the focal user with the strategy S_i. Therefore, the average fitness of the focal S_i user's neighbors who adopt strategy S_j is

$$\begin{aligned}&\Theta_{s_j|s_i} = (1-\alpha) + \\ &\alpha\{(k-1)\sum_{t\neq j} q_{s_t|s_i} u_{s_t s_i} + [(k-1)q_{s_j|s_i} + 1]u_{s_j s_i}\}\end{aligned} \tag{6}$$

According to the IM update rule, the probability of a user updating his/her strategy is proportional to the fitness of each strategy. In such a case, the probability that the focal user updates his/her current strategy strategy from S_i to $S_j (j \neq i)$ is:

$$\mathbb{E}\left[P_{S_i \to S_j}(k_{S_1}, k_{S_2}, k_{S_3}, k)\right] = \frac{k_{S_j}\Theta_{S_j|S_i}}{\sum_t k_{S_t}\Theta_{S_t|S_i} + \Theta_{S_i}} \tag{7}$$

Therefore, the probability that the focal user updates his/her current strategy from S_i to any other strategies is:

$$\mathbb{E}\left[P_{S_i \to others}\right] = 1 - \frac{\Theta_{(S_i;k_{S_1},k_{S_2},k_{S_3},k)} + k_{S_i}\Theta_{S_i|S_i}}{\Theta_{(S_i;k_{S_1},k_{S_2},k_{S_3},k)} + \sum_t k_{S_t}\Theta_{S_t|S_i}} \tag{8}$$

Under the configuration of $\Psi^{de}_{(S_i;k_{S_1},k_{S_2},k_{S_3},k_{S_4})}$, the probability of $\Delta p_{S_i} = -\frac{1}{N}$ can be calculated as:

$$\begin{aligned}&\text{Prob}\left(\Delta p_{S_i} = -\tfrac{1}{N}\right) \\ &= \sum_{k_{S_3}=0}^{k-k_{S_1}-k_{S_2}} \sum_{k_{S_2}=0}^{k-k_{S_1}} \sum_{k_{S_1}=0}^{k} p_{S_i} \cdot \Psi^{de}_{S_i} \cdot \mathbb{E}\left[P_{S_i \to others}\right]\end{aligned} \tag{9}$$

2) $\Delta p_{S_i} = \frac{1}{N}$ scenario: Similarly, we analyze the reverse scenario that the selected focal user is with strategy $S_j (j \neq i)$ and imitates a S_i user's strategy. Therefore, the percentage of strategy S_i will increase one unit ($\Delta p_{S_i} = \frac{1}{N}$).

Since this process is a reverse process of the previous scenario, we can use (5),(6) to get configuration's probability $\Psi^{in}_{(S_j;k_{S_1},k_{S_2},k_{S_3},k_{S_4})}$ and the average fitness of the focal S_j user's neribors who adopt S_i strategy : $\Theta_{s_i|s_j}$

The probability that the focal user imitates his/her strategy from $S_j (j \neq i)$ to S_i is:

$$\mathbb{E}\left[P_{S_j \to S_i}(k_{S_1}, k_{S_2}, k_{S_3}, k)\right] = \frac{k_{S_i}\Theta_{S_i|S_j}}{\sum_t k_{S_t}\Theta_{S_t|S_j} + \Theta_{S_j}} \tag{10}$$

Accordingly, the probability that the focal user updates his/her current strategy from any other strategy to strategy S_i can be summarized as:

$$\mathbb{E}\left[P_{other\to S_i}\right] = \frac{k_{S_i}\Theta_{S_i|S_j}}{\Theta_{(S_j;k_{S_1},k_{S_2},k_{S_3},k)} + \sum\limits_t k_{S_t}\Theta_{S_t|S_j}} \tag{11}$$

Under the configuration of $\Psi^{in}_{(S_j;k_{S_1},k_{S_2},k_{S_3},k_{S_4})}$, the probability of $\Delta p_{S_i} = \frac{1}{N}$ can be calculated as:

$$\begin{aligned} &\text{Prob}\left(\Delta p_{S_i} = \tfrac{1}{N}\right) \\ &= \sum_{j,j\neq i} \sum_{k_{S_3}=0}^{k-k_{S_1}-k_{S_2}} \sum_{k_{S_2}=0}^{k-k_{S_1}} \sum_{k_{S_1}=0}^{k} p_{s_j}\cdot \Psi^{in}_{S_j}\cdot \mathbb{E}\left[P_{other\to S_i}\right] \end{aligned} \tag{12}$$

Combining (9) and (12), we have the probability of p_{s_i} increase $\frac{1}{N}$ and decrease $\frac{1}{N}$, respectively. Therefore, the expected variation of p_{s_i} can be calcultaed per unit time, which is defined as the dynamic of p_{s_i}:

$$\dot{p}_{s_i} = \frac{1}{N}\text{Prob}\left(\Delta p_{S_i} = \frac{1}{N}\right) - \frac{1}{N}\text{Prob}\left(\Delta p_{S_i} = -\frac{1}{N}\right) \tag{13}$$

As we have analyzed, the unkown variable in (13) is only local network state $q_{s_i|s_j}$, which will be caculated in the following contents.

B. Relationship Dynamics Derivation

The relationship dynamics $\dot{p}_{s_is_j}, (1\le i,j\le 4)$ are the dynamics of global edge states and suggest the dynamics of relationship among users, which have relations: $\sum\limits_{1\le i,j\le 4} \dot{p}_{s_is_j} = 0$. The $p_{s_is_j}$ denotes the percentage of the directed edge linking from a S_j user to a S_i user, which means, in our following analysis, $\dot{p}_{s_js_i}$ only include the scenario that the S_j node changes, rather than the S_i node changes, on account of avoiding a repeat caculation. And we have relations as follows:

$$q_{s_i|s_j}p_{s_j} = q_{s_j|s_i}p_{s_i}, p_{s_is_j} = p_{s_js_i} \tag{14}$$

After analyzing, there are two scenarios that change $p_{s_is_j}$.

- Scenario 1: $i \neq j$ condition

There are also two sub-scenarios in $i\neq j$ condition. When the randomly selected user adopts strategy S_j and then choose a $S_l(l\neq j)$ neighbor to imitate the strategy, the edge state $p_{s_is_j}$ decreases for one unit. The probability is:

$$P^{D_A}_{s_is_j} = p_{s_j}\cdot \sum_{l\neq j} q_{s_l|s_j}(k-1)q_{s_i|s_l} \tag{15}$$

On the contrary, when we first choose a $S_l(l\neq j)$ user, then choose a neighbor adopting strategy S_j, the edge state $p_{s_is_j}$ increases for one unit, whose probability is:

$$P^{I_A}_{s_is_j} = \sum_{l\neq j} p_{s_l}q_{s_j|s_l}(k-1)q_{s_i|s_j} + p_{s_i}q_{s_j|s_i} \tag{16}$$

In such case, with the help of above analysis and relations in (14), the relationship dynamic can be calculated as:

$$\begin{aligned} \dot{p}^A_{s_is_j} &= \frac{NP^{I_A}_{s_is_j} - NP^{D_A}_{s_is_j}}{Nk/2} \\ &= \frac{2}{k}\left[(k-1)\,p_{s_j}\sum_l q_{i|l}q_{l|j} - kq_{i|j}\right] \end{aligned} \tag{17}$$

- Scenario 2: $i = j$ condition

When we first choose a S_i user, then choose a neighbor adopting $S_l(l\neq i)$ for reproduction, the edge state $p_{s_is_j}$ decreases for one unit. The probability is:

$$P^{D_B}_{s_is_j} = p_{s_i}\sum_{l\neq i} q_{s_l|s_i}\left[(k-1)q_{s_i|s_l}+1\right] \tag{18}$$

In the reverse scenario, we first choose a $S_l(l\neq i)$ user as focal user, then choose a S_i neighbor for reproduction. Accordingly, the edge state $p_{s_is_j}$ increases for one unit, whose probability is:

$$P^{I_B}_{s_is_j} = p_{s_i}q_{s_i|s_i}(k-1)\sum_{l\neq i} q_{s_l|s_i} \tag{19}$$

Similarly, the relationship dynamic of this scenario can be calculated as:

$$\begin{aligned} \dot{p}^B_{s_is_j} &= \frac{NP^{I_B}_{s_is_j} - NP^{D_B}_{s_is_j}}{Nk/2} \\ &= \frac{2}{k}\left[(k-1)\,p_{s_j}\sum_l q_{i|l}q_{l|j} - kq_{i|j} + 1\right] \end{aligned} \tag{20}$$

As we can see, the relationship dynamics only have one unkown variable, local network state $q_{s_i|s_j}$, as same as the population dynanmic.

C. Influence Dynamics Derivation

The influence dynamics denote the dynamics of local network states and suggest the influence of one user on his/her neighbors. For instance, $\dot{q}_{s_i|s_i}\to 1$ means most neighbors of the focal users tend to choose the same strategy as the focal user S_i, indicating the focal user has a large influence.

However, due to the weak selection, global frequencies change at a rate of order α, greatly lower than the local frequencies, which change at a rate of 1. Therefore, we can regard the global network state as constant and derive the dynamics of the local network as follows:

$$\begin{aligned} \dot{q}_{i|j} &= \frac{\dot{p}_{s_is_j}}{\dot{p}_{s_j}} = \frac{\dot{p}_{s_is_j}}{p_{s_j}} + o(\alpha) \\ &= \frac{2}{k}\left[(k-1)\,p_i\sum_l q_{s_i|s_l}q_{s_l|s_i} - kq_{s_i|s_i} + \delta_{ij}\right] + o(\alpha) \end{aligned} \tag{21}$$

At the same time, when $\dot{q}_{s_i|s_j} = 0$, we can obtaion the temporarily steady local network state $q^*_{s_i|s_j}$, which is the only unkown varable $q_{s_i|s_j}$ in *Population Dynamics* and *Relationship Dynamics*. According to $\dot{q}_{s_i|s_j} = 0$ condition, we

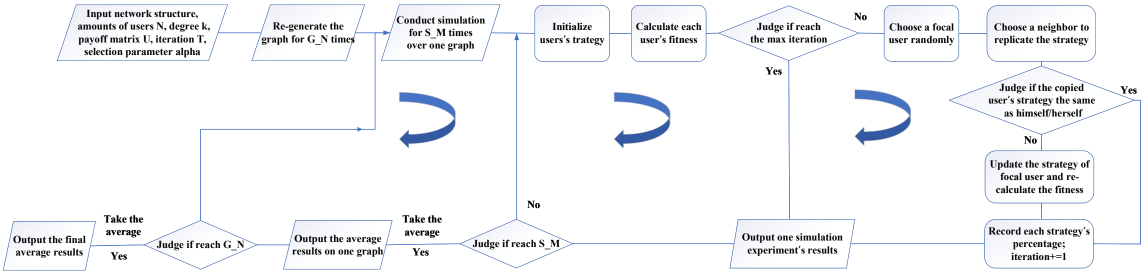


Fig. 3. Verification simulation flow chart

can have the equation as follows:

$$I + (k-1)\,Q^2 - kQ = 0 \tag{22}$$

where I is the identity matrix. When we take $QX = X$ and $\Xi^T X = 1$ into consideration, $(X\Xi^T)X = X, XI = X$. Therefore, local state $Q(i.e., \{q_{s_i|s_j}\}_{4\times4})$ can be a linear expression of $X\Xi^T$ and X. After substituting it into (22), we can obtain the unique non-trivial solution, which is as follows:

$$Q = \frac{k-2}{k-1}X\Xi^T + \frac{1}{k-1}I \tag{23}$$

Hence, after substituting the temporarily steady local network state $q_{s_i|s_j} = \frac{(k-2)p_{s_i}+\delta_{ij}}{k-1}$ into the population dynamic expression and make use of the weak selection($\alpha \ll 1$), we can simplify the close-form expression of population dynamics:

$$\dot{p}_{s_i} = \alpha' p_{s_i} \cdot \sum_j p_{s_j} \left[\Delta_{ij} - (k^2+k-6) \sum_i p_{s_i} u_{ji} \right] \tag{24}$$

where

$$\begin{cases} \Delta_{ij} = (k+3)u_{ii} + (k^2-k-2)u_{ij} - 3u_{ji} - (k+3)u_{jj}, \\ \alpha' = \alpha\frac{k(k-2)}{(k+1)^2(k-1)}. \end{cases} \tag{25}$$

IV. Experiments

In this section, several experiments are conducted to verify the theoretical derivation of the multi-source information diffusion dynamics and demonstrate the model's effectiveness of real-world's applications.

A. Synthetic Networks and Real-World Networks

The information diffusion is a process with large randomness but might be deterministic in statistical principle. Accordingly, we use the Monte-Carlo algorithm that relies on numerous repeated random sampling to investigate the underlying principles. In this paper, the experiments are conducted by simulating the fitness calculation and strategy update processes step-by-step. There are three loops ensuring that we can obtain statistically general conclusion and smooth curves. The simulation flow chart can be illustrated as follows:

Specifically, in each experiment, the network structure is re-generated for 20 times and the simulation experiments are conducted in the same network for 32 times. In each run, the fitness calculation and strategy update processes have 400 iterations, ensuring the experiment has a sufficient time span. The network size is set to 2000 users and the average degree is set to 20. The intensity of selection α is set as 0.01, i.e., weak selection. The initial percentage of each strategy is set as $\{0.1, 0.2, 0.3, 0.4\}$ and the payoff matrix is randomly set as:

$$U_{4\times4} = \begin{bmatrix} 0.4 & 0.48 & 0.48 & 0.64 \\ 0.48 & 0.24 & 0.64 & 0.32 \\ 0.48 & 0.64 & 0.24 & 0.32 \\ 0.64 & 0.32 & 0.32 & 0.2 \end{bmatrix} \tag{26}$$

Moreover, we extensively develop the experiments in the random regular network, Erdös Rényi (ER) random networks and Barabási-Albert (BA) scale-free networks (the specific methods are referred to [17]) to prevent the false results based on one particular realization of a specific network type. We compare the model-based predictions and the simulation results in these three synthetic networks, as we can see in Fig.4. In Fig.4 (c), the gap for Barabási-Albert (BA) scale-free network is due to the neglected dependence between the global network state and the network degree, which is analyzed in [17] specifically.

In the experiment of the real-world network, we choose the Facebook dataset in [25], which contains totally 4039 users and 88234 relations, and the average degree is about 40. Fig.5 shows the experiment results in theoretical prediction and Monte-Carlo simulation. It can be seen that the simulation results match well with the theoretical results, which verify the correctness and effectiveness of the model.

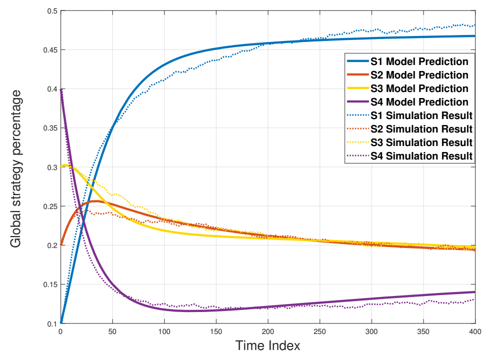
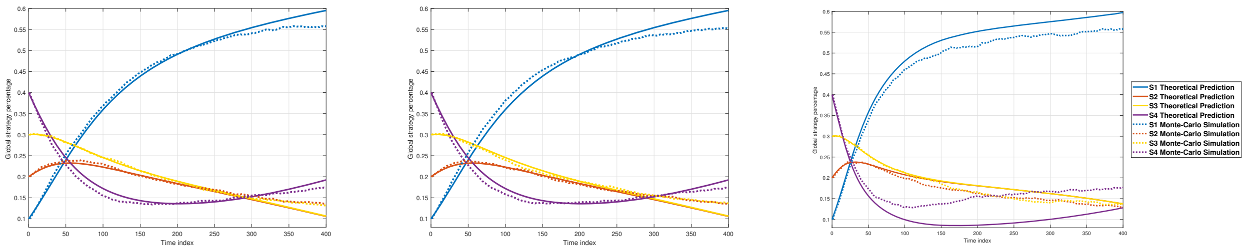


Fig. 5. Simulation results for real-world Facebook network



(a) Random regular network (b) ER random network (c) BA scale free network

Fig. 4. Simulation results for synthetic networks

B. Sina Weibo Real-World Data Experiment

In the above experiments, the users' payoff matrix is first set up, then the experiments of information diffusion over different networks are conducted. In this part, we consider a reverse process, which means that we use the crawled Sina Weibo (a popular Twitter-like social network) users' forwarding data to learn the parameters, e.g., payoff matrixes. Accordingly, we can use these learned parameters to predict the real-world diffusion process.

Specifically, in each iteration, we substitute the temporary parameters into the model to make a estimation, which is a 4*400 matrix (four strategies and 400 units of diffusion time). Therefore, the $F-norm$ of the absolute value matrix of the gap between estimated diffusion matrix and real-data diffusion matrix can be used as a reverse parametric adjustment quantized value for error function of learning algorithm. The parameters' training loss curve can be illustrated as follows:

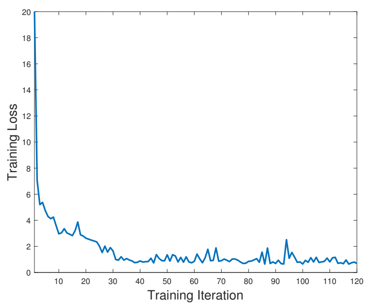
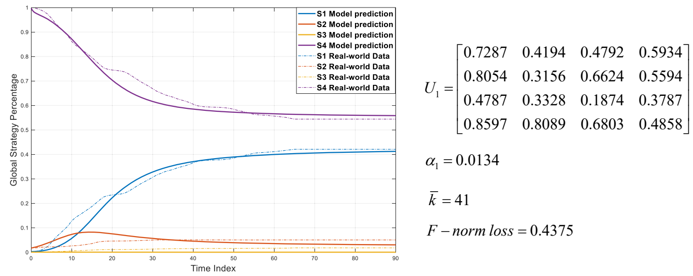


Fig. 6. The parameters' training loss curve

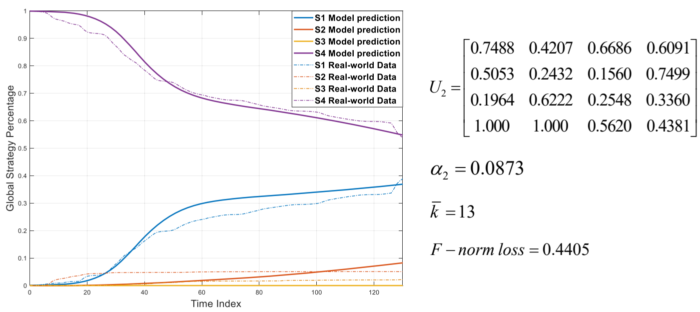
We choose two pairs of popular correlated Weibo hashtags for experiments. The first pair is "Canada arrests Huawei's global CFO Wanzhou Meng" and "Canada former diplomatist Michael Kovrig is in detention according to the Chinese government". They have arisen large attentions in Weibo social network and became Weibo's "hot search" breaking news at the same time since two hashtags are strongly relevant. Chinese people called it "It is impolite not to reciprocate" and most of them tend to forward these two hashtags together.

The second pair is "Notre-Dame de Paris caught fire" and "Famous French game company Ubisoft announced a €500000 donation and a free download for the popular game Assassin's Creed Unity". The donation is for rebuilding and the game appears proportionally 1:1 with the Cathedral's real-world counterpart, which may be the most sweeping and detailed buildings to ever appear in the game. They also mitivate the diffusion process each other, since different fileds of people (art & game) can communicate freely.

Four strategies are defined as the same as (1), and the learned parameters and final F-norm loss results are as follows:



(a) "Wangzhou Meng was arrested"correlated information

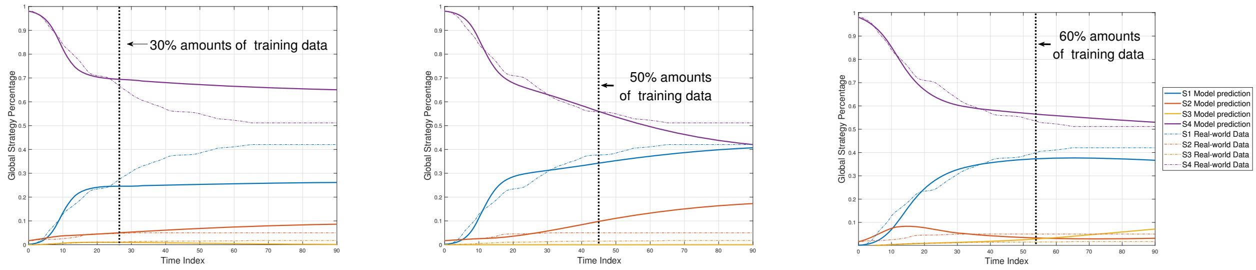


(b) "Notre-Dame de Paris caught fire"correlated information

Fig. 7. Comparisons between the model prediction and the real-world Weibo users' forwarding data

As we can see, the model's prediction curve matches well with the real-world information diffusion process. It means that by using only 18 parameters, our theoretical model can fit the real information diffusion process well. The mechanism of decision-making and users' interactions are modeled, and the relations of multi-source information are hidden in the payoff matrix, with which we can make the desired prediction.

Finally, we conduct another experiment to verify model's prediction performance. An effective prediction means that we can estimate future data from the past. Accordingly, we limit the proportion of the training dataset and use the



(a) 30% amounts (b) 50% amounts (c) 60% amounts

Fig. 8. Prediction results of limited proportion of training data

learned parameters to estimate the whole diffusion process. As we can see in Fig.8, when we only use 30% amounts of data for learning, the diffusion prediction does not seem reasonable. However, when only 60% ahead of training data are used, we can have a good prediction performance of the whole diffusion process at less than 8% error rate, which demonstrates the effectiveness of our theoretical model in real-world applications.

V. Conclusions

Overall, this paper proposes a new approach for multi-source information diffusion over social networks with a graphical evolutionary game perspective. Through the study, we formulate the problem, theoretically derive three dynamics of the networks and conduct different experiments to verify derivation's correction and model's effectiveness. The simulation's results are consistent with our analysis and the framework is proved to fit and predict the real-world information spreading process by learning from limited proportion of data well, which demonstrates the effectiveness of the model.

Acknowledgment

This work is supported by the National Key Research and Development Program of China (2017YFB1400100).

References

[1] H. V. Zhao, W. S. Lin, and K. R. Liu, *Behavior dynamics in media-sharing social networks*. Cambridge University Press, 2011.

[2] J. Zhang, Z. Fang, W. Chen, and J. Tang, "Diffusion of "following" links in microblogging networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2093–2106, 2015.

[3] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," in *2010 IEEE International Conference on Data Mining*. IEEE, 2010, pp. 599–608.

[4] S. A. Myers and J. Leskovec, "Clash of the contagions: Cooperation and competition in information diffusion," in *IEEE International Conference on Data Mining*, Conference Proceedings, pp. 539–548.

[5] X. Zhang, Y. Su, S. Qu, S. Xie, B. Fang, and P. S. Yu, "IAD: interaction-aware diffusion framework in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 7, pp. 1341–1354, 2018.

[6] D. Li, S. Zhang, X. Sun, H. Zhou, S. Li, and X. Li, "Modeling information diffusion over social networks for temporal dynamic prediction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1985–1997, 2017.

[7] F. Wang, H. Wang, and K. Xu, "Diffusive logistic model towards predicting information diffusion in online social networks," in *2012 32nd International Conference on Distributed Computing Systems Workshops*. IEEE, 2012, pp. 133–139.

[8] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs," *Proc. SAIM Int. Conf. Data Mining*, pp. 551–556, 2007.

[9] D. J. Daley and D. G. Kendall, "Epidemics and rumours," *Nature*, vol. 204, no. 4963, p. 1118, 1964.

[10] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos, "Interacting viruses in networks: can both survive?" in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 426–434.

[11] M. Granovetter, "Threshold models of collective behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[12] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," *Knowledge and information systems*, vol. 37, no. 3, pp. 555–584, 2013.

[13] A. Guille and H. Hacid, "A predictive model for the temporal dynamics of information diffusion in online social networks," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 1145–1152.

[14] M. Y. Li and J. S. Muldowney, "Global stability for the seir model in epidemiology," *Mathematical biosciences*, vol. 125, no. 2, pp. 155–164, 1995.

[15] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," in *2010 IEEE International Conference on Data Mining*. IEEE, 2010, pp. 599–608.

[16] C. Jiang, Y. Chen, and K. J. R. Liu, "Graphical evolutionary game for information diffusion over social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 524–536, 2014.

[17] C. Jiang, C. Yan, and K. J. R. Liu, "Evolutionary dynamics of information diffusion over social networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4573–4586, 2014.

[18] Y. Chen, C. Jiang, C.-Y. Wang, Y. Gao, and K. R. Liu, "Decision learning: Data analytic learning with strategic decision making," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 37–56, 2015.

[19] S. J. Maynard, "Evolution and the theory of games," *American Scientist*, vol. 64, no. 1, pp. 41–45, 1976.

[20] X. Cao, Y. Chen, C. Jiang, and K. J. R. Liu, "Evolutionary information diffusion over heterogeneous social networks," *IEEE Transactions on Signal & Information Processing Over Networks*, vol. 2, no. 4, pp. 595–610, 2016.

[21] Y. Li, B. Qiu, Y. Chen, and H. V. Zhao, "Analysis of information diffusion with irrational users: A graphical evolutionary game approach," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2527–2531.

[22] Y. Chen, Y. Gao, and K. R. Liu, "An evolutionary game-theoretic approach for image interpolation," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 989–992.

[23] C. Jiang, Y. Chen, and K. R. Liu, "Distributed adaptive networks: A graphical evolutionary game-theoretic view," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5675–5688, 2013.

[24] B. Wang, K. R. Liu, and T. C. Clancy, "Evolutionary cooperative spectrum sensing game: how to collaborate?" *IEEE transactions on communications*, vol. 58, no. 3, pp. 890–900, 2010.

[25] J. Leskovec, "Stanford large network dataset collection [online]." Available: https://snap.stanford.edu/data.