Understanding Information Diffusion under Interactions

Yuan Su† Xi Zhang‡ Philip S. Yu†§ Wen Hua† Xiaofang Zhou†§ Binxing Fang†

†Beijing University of Posts and Telecommunications, China
‡University of Illinois at Chicago, USA
§Institute for Data Science, Tsinghua University, China
¶The University of Queensland, Australia
¶Soochow University, China

{timsu,zhangx,fangbx}@bupt.edu.cn, psyu@cs.uic.edu, w.hua@uq.edu.au, zxf@itee.uq.edu.au

Abstract

Information diffusion in online social networks has attracted substantial research effort. Although recent models begin to incorporate interactions among contagions, they still don’t consider the comprehensive interactions involving users and contagions as a whole. Moreover, the interactions obtained in previous work are modeled as latent factors and thus are difficult to understand and interpret. In this paper, we investigate the contagion adoption behavior by incorporating various types of interactions into a coherent model, and propose a novel interaction-aware diffusion framework called IAD. IAD exploits the social network structures to distinguish user roles, and uses both structures and texts to categorize contagions. Experiments with large-scale Weibo dataset demonstrate that IAD outperforms the state-of-art baselines in terms of F1-score and accuracy, as well as the runtime for learning. In addition, the interactions obtained through learning reveal interesting findings, e.g., food-related contagions have the strongest capability to suppress other contagions’ propagation, while advertisement-related contagions have the weakest capability.

1 Introduction

During recent years online social networks have become ubiquitous in our life, and information diffusion in social networks has proved to play important and decisive roles in some situations, such as viral marketing. Specifically, a contagion is posted by some node in the network and exposed to its neighbors. If a neighbor forwards that contagion, infection occurs and the contagion begins to spread over the network.

To better understand the information dynamics in social networks, massive efforts have been devoted to this research area. However, most studies concentrate on the scenario that only a single piece of information spreads in the network at a time [Kempe et al., 2003; Goldenberg et al., 2001; Hethcote, 2000; Newman, 2003; Du et al., 2013; Cohen et al., 2014; Tang et al., 2015]. Recently approaches have started to consider interactions among contagions [Weng et al., 2012; Myers and Leskovec, 2012; Rong and Mei, 2013; Bi et al., 2013; Coscia, 2013; Valera and Gomez-Rodriguez, 2015; Pathak et al., 2010; Prakash et al., 2012; Karrer and Newman, 2011], whereas the interactions among explicit categories of contagions are rarely inferred. In [Myers and Leskovec, 2012] interactions among latent topics are considered which, however, are hard to understand and interpret. Actually, what is more interesting is the interactions among explicit categories, namely whether contagions belonging to one category (say food) would have some positive/negative effects on the spreading of contagions belonging to another category (say politics). However, to do this study, we would need to have the category of each contagion. Given the large number of contagions, it would be impossible to ask human to annotate all of them. How to find an efficient way to classify contagions with only minimum supervision is thus one of the key challenges in this work.

Besides, social roles and their interactions are also vital for information diffusion. For example, a contagion from a celebrity might have a higher possibility to get spread. Analogously, it is more likely for an ordinary user to forward a contagion posted by a celebrity than vice versa. Previous research has proved that the diffusion of contagions is affected by network structures [Yang et al., 2015]. Since social roles of users reflect network structures, it is intuitive to involve social roles according to their structure characteristics, as well as their interactions to build a more comprehensive model. After considering the interactions among contagions and the interactions among users, it is natural to ask whether there are interactions between users and contagions. The answer is obvious since each user has her own preference on contagions. So far, we have three kinds of interactions, and how to integrate them together into a coherent model is another challenge. Besides, once the model is built, fitting the model to get the interactions for each pair of contagions and users is prohibitive (quadratic in the number of contagions and users). Therefore, how to efficiently obtain the interactions poses a new challenge to our model.

Altogether, we illustrate a framework of information diffusion by incorporating three kinds of interactions: (1) User-Contagion Interaction (2) User-User Interaction (3) Contagion-Contagion Interaction. We study the scenario where a user needs to decide whether to forward a contagion given other simultaneously exposed contagions, and formulate the infection probability by incorporating the inherent
popularity of the contagion as well as three kinds of interactions. Since learning interactions for each pair of contagions and users is extremely time-consuming, some algorithms are proposed to reduce the cardinality. First, we apply a mixture of Gaussians model to explain the generation process of user network features, and use EM algorithm to extract social role distribution for each user. Then we propose a classification approach for contagions based on co-training [Blum and Mitchell, 1998], which uses a small number of labeled data and a large number of unlabeled data. After that, we achieve the category of each contagion and the social roles of each user. The proposed model statistically learns the interactions, and the resulting data assists to better comprehend the information diffusion process and provide a more accurate prediction for contagion adoption.

The contributions of this paper are threefold:

1) We propose an Interaction-Aware Diffusion (IAD) framework to model information diffusion process by incorporating three kinds of interactions, which provides new insights into how forwarding decisions are made.

2) To efficiently learn the interactions, a co-training based method is devised to classify the contagions, and a generative process is applied to obtain the social roles for users, which can significantly decrease the number of fitted parameters.

3) Experiments on a large-scale Weibo dataset [Zhang et al., 2013] not only prove the superiority of IAD framework to state-of-art works, but also reveal some interesting and useful findings. For example, contagions on food are more likely to suppress the propagation of other contagions, indicating strong possibility for food-related topics to attract people’s attention in Weibo.

2 Interaction-Aware Diffusion Framework

In this section, we first provide the statement and formulation of the problem, and then describe our approach and the learning process. Before going into details of IAD framework, we first define some important notations shown in Table 1.

### 2.1 Problem Statement

In a social network, when some new contagion is originated from one user, the information is exposed to its neighbors, and the exposed contagion is called an exposure. Since users have limited attentions [Weng et al., 2012], we make the assumption as [Myers and Leskovec, 2012] that a user reads through all the contagions her neighbors have forwarded, but only the most recent $K$ exposures that she can keep in mind. In social networks like Weibo and Twitter, tweets in a user’s reading screen are arranged in time descending order, i.e., users will first read the most recent contagions and then go backward. Therefore, there is a sliding window going back $K$ contagions that she keeps in mind.

The scenario we study here is when a user reads a contagion which is forwarded by one of her neighbors, given the sequence of contagions the user has previously read, what’s the probability of the user adopting this contagion. It is further described in Figure 1, where the set $\{m_1, m_2, ..., m_K\}$ is a sequence of $K$ contagions user $u_a$ has read and kept in mind, and $m_i$ ($i \neq 1, 2, ..., K$) is the contagion which is previously forwarded by $u_b$ and now examined by $u_a$. In this scenario, the forwarding decision made by $u_a$ is not only decided by the inherent characteristics of $m_i$, but also by three kinds of external interactions described as follows:

- **User-Contagion Interaction**: The interaction between the examining user and the examined contagion. As shown in Figure 1, it is $u_a$’s preference over $m_i$.
- **User-User Interaction**: The interaction between the examining user and the neighbor who has forwarded the examined contagion previously. In Figure 1, it is the effect $u_b$ has on $u_a$.
- **Contagion-Contagion Interaction**: The interaction among the examined contagion and other contagions the user has read recently. In Figure 1, it is the effect contagions $m_1$ and $m_2$ ($K = 2$) has on $m_i$.

Given the interacting scenarios, our task is to model the users’ adoption behaviour by incorporating the aforementioned interactions, and fitting the model to infer the interactions. The problem will be formulated in the next section.

### 2.2 Formulation

According to the interacting scenario (as shown in Figure 1), given $\{m_1, m_2, ..., m_K\}$ and $u_b$, the probability of infection by $m_1$ to $u_a$ is

$$P(I_{m_1(u_a)} | E_{m_1(u_b)}, E_{m_1, m_2, ..., m_K})$$  \hspace{1cm} (1)

Here $I_{m_1(u_a)}$ is the infection of $u_a$ by $m_i$, $E_{m_1(u_b)}$ is the exposure of $m_i$ which is forwarded by $u_b$, and $E_{m_1, m_2, ..., m_K}$ is the exposure set $\{m_1, m_2, ..., m_K\}$. We make the assumption as [Myers and Leskovec, 2012] that for any $k$ and $l$, $E_{m_k}$ is independent of $E_{m_l}$. Applying Bayes rule, we model Eq. (1) by

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>Users</td>
</tr>
<tr>
<td>$m$</td>
<td>Contagions</td>
</tr>
<tr>
<td>$r$</td>
<td>User roles</td>
</tr>
<tr>
<td>$t$</td>
<td>Contagion latent topics</td>
</tr>
<tr>
<td>$c$</td>
<td>Contagion categories</td>
</tr>
<tr>
<td>$\Delta \in R^{[u \times u]}$</td>
<td>User-user interaction matrix</td>
</tr>
<tr>
<td>$\Lambda \in R^{[m \times m]}$</td>
<td>Contagion-contagion interaction matrix</td>
</tr>
<tr>
<td>$\Omega \in R^{[u \times m]}$</td>
<td>User-contagion interaction matrix</td>
</tr>
<tr>
<td>$\Delta_{role} \in R^{[r \times r]}$</td>
<td>User-role interaction matrix</td>
</tr>
<tr>
<td>$\Lambda_{topic} \in R^{[t \times t]}$</td>
<td>Contagion-topic interaction matrix</td>
</tr>
<tr>
<td>$\Omega_{role}^{topic} \in R^{[r \times t]}$</td>
<td>User-role - contagion topic interaction matrix</td>
</tr>
<tr>
<td>$\Lambda_{category} \in R^{[c \times c]}$</td>
<td>Contagion category-category interaction matrix</td>
</tr>
<tr>
<td>$\Omega_{category}^{role} \in R^{[c \times t]}$</td>
<td>User-role - contagion category interaction matrix</td>
</tr>
</tbody>
</table>
Figure 1: An example of interacting scenario. User $u_a$ is exposed to contagions $\{m_1, \ldots, m_K\}$ (Here $K = 2$) and $m_i$ (forwarded by $u_a$’s neighbor $u_b$), and is examining whether to adopt $m_i$. $u_a$’s decision is influenced by: interaction between $u_a$ and $m_i$; interaction between $u_a$ and $u_b$; and interactions among $m_i$ and other exposing contagions ($m_1$ and $m_2$).

$$P(I_{m_i(u_a)}|E_{m_i(u_a)}, E_{\{m_1,m_2,\ldots,m_K\}}) = \frac{P(I_{m_i(u_a)})P(E_{m_i(u_a)}, E_{\{m_1,m_2,\ldots,m_K\}}|I_{m_i(u_a)})}{P(E_{m_i(u_a)}, E_{\{m_1,m_2,\ldots,m_K\}})}$$

$$= \frac{P(I_{m_i(u_a)})\prod_{k=1}^{K} P(E_{m_k}|I_{m_i(u_a)})}{P(E_{m_i(u_a)})\prod_{k=1}^{K} P(E_{m_k})} \times \prod_{k=1}^{K} \frac{P(I_{m_i(u_a)}|E_{m_k})P(E_{m_k})}{P(I_{m_i(u_a)})\prod_{k=1}^{K} P(I_{m_i(u_a)}|E_{m_k})}$$

$$= \frac{P(I_{m_i(u_a)}|E_{m_i(u_a)})\prod_{k=1}^{K} P(I_{m_i(u_a)}|E_{m_k})}{P(I_{m_i(u_a)})\prod_{k=1}^{K} P(I_{m_i(u_a)}|E_{m_k})}$$

Here we need to model $P(I_{m_i(u_a)})$, $P(I_{m_i(u_a)}|E_{m_i(u_a)})$ and $P(I_{m_i(u_a)}|E_{m_k})$ for each $k \in \{1, \ldots, K\}$, which are enforced between 0 and 1. Since each contagion has its inherent infectiousness, $P(I_{m_i})$ is defined as the prior infection probability of $m_i$, which can be obtained through dividing the number of its infections by the number of its exposures.

We define $\Omega(u_a, m_i)$ as the effect user $u_a$ has on contagion $m_i$ (User-Contagion Interaction), $\Delta(u_a, u_b)$ as the effect user $u_b$ has on user $u_a$ (User-User Interaction), and $\Lambda(m_i, m_k)$ as the effect contagion $m_k$ has on contagion $m_i$ (Contagion-Contagion Interaction). Then we model $P(I_{m_i(u_a)})$, $P(I_{m_i(u_a)}|E_{m_i(u_a)})$ and $P(I_{m_i(u_a)}|E_{m_k})$ as

$$P(I_{m_i(u_a)}) \approx P(I_{m_i}) + \Omega(u_a, m_i)$$

$$P(I_{m_i(u_a)}|E_{m_i(u_a)}) \approx P(I_{m_i}) + \Delta(u_a, u_b) \approx P(I_{m_i}) + \Omega(u_a, m_i) + \Delta(u_a, u_b)$$

$$P(I_{m_i(u_a)}|E_{m_k}) \approx P(I_{m_i}) + \Lambda(m_i, m_k) + \Omega(u_a, m_i)$$

Besides the proposed model adopting summations, we also conduct extensive experiments on the model adopting multiplications as well, but the model adopting summations performs better. Thus we apply the additive model in this paper.

Here we have connected the infection probability with three interaction matrices: (1) $\Omega \in R^{|u| \times |m|}$, (2) $\Delta \in R^{|u| \times |u|}$, and (3) $\Lambda \in R^{|m| \times |m|}$, where $|u|$ is the number of users and $|m|$ is the number of contagions. However, these matrices are impractical to learn, because $|u|$ and $|m|$ are extremely large in social networks. Thus, we model User Role-Contagion Topic interaction, User Role-Role Interaction and Contagion Topic-Topic Interaction instead, which will be illustrated in the next section.

2.3 The Proposed Approach

To decrease the fitted parameters, we utilize the network structures to infer users’ social roles, and use the contagion contexts to extract contagions’ topics. IAD framework is shown in Figure 2, which consists of five components:

- **User roles generation**: A generative process of user roles is proposed to distinguish different kinds of users.
- **Contagion latent topics extraction**: Latent topics are extracted as features for statistical model learning and contagion classification.
- **Statistical model learning**: Based on the outputs of the above two components, a statistical model is learned.
- **Contagion classification**: Based on latent topics, a co-training method of contagion classification is proposed. The categories derived here are explicit.
• **Interactions inference**: Given the results of contagion classification and the statistical model, interactions among contagions and users can be inferred.

Next we will introduce the process of user roles generation and contagion latent topic extraction in details, and then describe statistical model learning. The last two components will be illustrated in Section 3.

**User Role-Role Interaction.** User roles are defined as authority users, hub users and ordinary users in our work. Intuitively, an authority user has a large number of followers, while a hub user has lots of followers. A user may play multiple roles, for instance, an authority user may also be a hub user, and therefore we adopt a probability distribution over social roles for each user. Then we infer the interactions among different social roles. The results indicate how a user, with a specific roles distribution, influence other users’ probability of adopting a contagion.

We use PageRank score [Page et al., 1999], HITS authority and hub values [Kleinberg, 1999], in-degree and out-degree scores as features of users. A mixture of Gaussians model is proposed to explain the features generation process. Specifically, we assume the features of each user is sampled as a multivariate Gaussian distribution. Define \( \mathbf{r}(r_1, r_2, r_3) \) as user role vector, then for each role \( r_i \), we generate multivariate Gaussian distribution \( u|r_j \sim \mathcal{N}(\mu_j, \Sigma_j) \). EM algorithm is used to extract the role distribution for each user. After that, we assign each role \( r_j \) to the most relevant one of the three roles, according to that authority users have lots of followers and hub users have lots of followees.

Rather than modeling **User-User Interaction** denoted by \( \Delta \in \mathbb{R}^{[u]|x|[u]} \), we would model **User-Role-Role Interaction** instead, which is denoted by \( \Delta_{\text{role}} \in \mathbb{R}^{[r]|x|[r]} \). \( \Delta_{\text{role}}(r_i, r_j) \) is the affect role \( r_j \) has on role \( r_i \). Define \( \theta_{a,i} \) as the probability of user \( u_a \) belonging to role \( r_i \), and \( \sum_i \theta_{a,i} = 1 \). Now, \( \Delta(u_a, u_b) \) in Eq. (4) can be updated by

\[
\Delta(u_a, u_b) = \sum_i \sum_j \theta_{a,i} \Delta_{\text{role}}(r_i, r_j) \theta_{b,j} \quad (6)
\]

**Contagion Topic-Topic Interaction.** Each contagion is assumed to have a distribution on several topics, and \( t \) denotes the set of latent topics. LDA [Blei et al., 2003] is used to extract the latent topic distribution of each contagion. Then, instead of modeling \( \Lambda \in \mathbb{R}^{[m]|x|[m]} \) we would model a matrix \( \Lambda_{\text{topic}} \in \mathbb{R}^{[t]|x|[t]} \), which denotes the **Contagion Topic-Topic Interaction**. We define \( \theta_{t,a} \) as the probability of contagion \( m_i \) belonging to topic \( t_a \), and therefore \( \sum_a \theta_{t,a} = 1 \). Let \( \Lambda_{\text{topic}}(t_a, t_b) \) denote the impact of topic \( t_b \) has on topic \( t_a \). Now, \( \Lambda(m_i, m_k) \) in Eq. (5) can be updated by

\[
\Lambda(m_i, m_k) = \sum_a \sum_b \theta_{t,a} \Lambda_{\text{topic}}(t_a, t_b) \theta_{t,b} \quad (7)
\]

**User Role - Contagion Topic Interaction.** Instead of learning \( \Omega \), we build a matrix \( \Omega_{\text{role}} \in \mathbb{R}^{[r]|x|[t]} \) to denote the **User Role - Contagion Topic Interactions**. Then \( \Omega(u_a, m_i) \) in Eq. (3), Eq. (4) and Eq. (5) can be updated by

\[
\Omega(u_a, m_i) = \sum_j \sum_b \theta_{a,j} \Omega_{\text{topic}}(r_j, t_b) \theta_{b,i} \quad (8)
\]

**2.4 Model Learning**

The input of our model is a set of interaction scenarios. An example of interacting scenario is shown in Figure 1, which consists of the examining user \( u_a \), the examined contagion \( m_i \), user \( u_a \)’s neighbor \( u_b \) who has forwarded the examined contagion, and the exposing contagion set \( \{m_1, m_2, \ldots, m_K\} \) \( (i \neq 1, 2, \ldots, K) \). All the interacting scenarios comprise a set \( \{x_1, x_2, \ldots, x_n\} \), where \( x_i \) is the \( i \)th interacting scenario and \( n \) is the total number. For each interacting scenario, it can be observed whether the examining user has adopted the examined contagion or not, which can be denoted by \( y_i \in \{0, 1\} \) (1 for adoption and 0 for not). Then the training set \((\{x_1, y_1\}, (x_2, y_2), \ldots, (x_n, y_n)\) will be obtained. Let \( \pi(x_i) \) denote Eq. (1) for simplicity. Now, \( \pi(x_i) \) can be updated by \( \Omega_{\text{role}} \), \( \Delta_{\text{role}} \) and \( \Lambda_{\text{topic}} \) according to equations from Eq. (2) to Eq. (8), and the log-likelihood function is

\[
L(\Omega_{\text{role}}, \Delta_{\text{role}}, \Lambda_{\text{topic}}) = \sum_{i=1}^n (y_i \log(\pi(x_i)) + (1 - y_i) \log(1 - \pi(x_i))) \quad (9)
\]

Our goal is to estimate the parameters in \( \Omega_{\text{role}}, \Delta_{\text{role}} \) and \( \Lambda_{\text{topic}} \) to maximize the log-likelihood function. Stochastic gradient ascent is adopted to fit the model. In each iteration of parameters updating, if it will make any item with probability meaning lower than 0 or higher than 1, we won’t do any updating in this iteration, and goes to the next iteration.

**3 Classification of Contagions**

The interaction matrix \( \Lambda_{\text{topic}} \) and \( \Omega_{\text{role}} \) learned through our model are comprised of latent topics, which is difficult to interpret. In this section, we illustrate how to obtain interactions among explicit categories. We define \( |c| = 15 \) categories based on the Weibo dataset, involving advertisement, constellation, culture, economy, food, health, history, life, movie, music, news, politics, sports, technology and traffic.

To discover interactions among categories, contagions should be classified into categories first. However, contagions spreading in Weibo [Zhang et al., 2013] are not labeled to intrinsic categories. Labeled contagions are extremely expensive to obtain because large human efforts are required. Thus, only a few labeled contagions are available for learning. A classification approach based on co-training [Blum and Mitchell, 1998] is proposed. Specifically, each contagion in the dataset is described in two distinct views. One is the contagion itself, and the other is set of the other contagions posted by the same user. The intuition here is that contagions created from the same user are prone to have similar category. Then we build two classifiers for two views, and choose the latent topics as the features for each classifier. As described in section 2.3, contagion \( m_i \)’s latent topic distribution, denoted by \( \theta_{t,a}(a \in 1, \ldots, |t|) \), can be extracted using LDA. We define a contagion set \( M_i = \{m_1, m_2, \ldots, m_k\} \) to contain the other
contagions created by the same user. The latent topic distribution \( \Theta_{i,a}(a \in 1,..|t|) \) of \( M_i \) is obtained by \( \frac{\sum_k \theta_{i,a}}{k} \). Now, the two classifiers are listed as follows, and LIBSVM [Chang and Lin, 2011] is used for multi-class classification.

- **Classifier 1**: \( \theta_{i,a}(a \in 1,..|t|) \) as features for each contagion \( m_i \).
- **Classifier 2**: \( \Theta_{i,a}(a \in 1,..|t|) \) as features for each contagion set \( M_i \).

We labeled a minimum number of contagions for each category by hand for training in the beginning. After the initial training process, two classifiers go through the unlabeled contagions to make predictions. If the results from the two classifiers are the same for a contagion, this contagion is added to the labeled set and removed from the unlabeled set. Then a new set for training is obtained, and another iteration starts. In each iteration, there are some contagions moved from the unlabeled set to the labeled set. After enough contagions being labeled, we can derive the following two interactions.

\subsection*{Contagion Category-Category Interaction}

If the set of contagions \( \{m_1, m_2, ..., m_k\} \) belongs to category \( c_i \), the latent topic distribution \( \varphi_{i,a}(a \in 1,..|t|) \) of category \( c_i \) can be obtained through \( \frac{\sum_k \varphi_{i,a}}{k} \). We define \( \Lambda_{category} \in R^{c_i \times |c|} \) to denote Contagion Category-Category Interaction, where \( \Lambda_{category}(c_i, c_k) \) is the impact of category \( c_k \) on \( c_i \), that is

\begin{equation}
\Lambda_{category}(c_i, c_k) = \sum_a \varphi_{i,a} \Lambda_{topic}(t_a, t_b) \varphi_{k,b}
\end{equation}

\subsection*{User Role - Contagion Category Interaction}

Similarly, define \( \Omega_{category} \in R^{r \times |c|} \) to denote User Role - Contagion Category Interaction, where \( \Omega_{category}(r_i, c_j) \) is the interaction from user role \( r_i \) to category \( c_j \), that is

\begin{equation}
\Omega_{category}(r_i, c_k) = \sum_b \Omega_{topic}(r_i, t_b) \varphi_{k,b}
\end{equation}

\section*{4 Evaluation}

In this section, we conduct experiments based on a public Weibo dataset to evaluate IAD framework, and then discuss various qualitative insights.

\subsection*{4.1 Experimental Settings}

\textbf{Dataset} The Weibo dataset [Zhang et al., 2013] provides a list of Weibo users who have forwarded contagions, as well as the forwarding timestamp. Users’ friendship links are also recorded. Because of the crawling strategy, the distribution of retweet counts in different months is highly imbalanced. Thus, we select the diffusion data from July 2012 to December 2012, in which the retweet count per month is large enough and the distribution is more balanced. Consequently, we get 19,388,727 retweets on 140,400 popular microblogs. We delete the inactive users without any retweets in this period and obtain 1,077,021 distinct users for the experiment.

Then we do statistical analysis to extract interacting scenarios from the dataset. As illustrated in Sec. 2.1, it is assumed that the recent \( K \) exposures can be kept in the mind of a user, and here we set \( K = 1 \) and \( K = 2 \). For each user, when she examines a newly posted contagion, an interacting scenario occurs. If the examined contagion is adopted, the interacting scenario is a positive instance, otherwise it is a negative instance. We observe that the positive and negative instances are highly unbalanced in the dataset, so we sample a balanced dataset with equal number of positive and negative instances. In total, 38,777,454 interacting scenarios are got. We randomly use 90\% of the instances as the training set, and the remaining 10\% as the testing set. We set the number of latent topics set \( |t| = 20, 30 \) and 50 respectively.

\textbf{Baselines.} We compared our proposal with three baselines:

- **IP Model.** Infection Probability Model assigns the infection probability of a contagion to be the prior infection probability, which doesn’t consider the interactions among users and contagions.
- **IMM Model.** IMM Model [Myers and Leskovec, 2012] is a state-of-art work incorporating the interactions among contagions into its model. To make fair comparison, we use the same set of instances and the same setting of parameters as our work.
- **UI Model.** User Interaction Model is one component of IAD framework, which only considers the user-user interactions, specifically user role–role interactions.

In our proposal and the baselines, we set a predicting result to 0 if the predicting infection probability is less than 0.5, otherwise we set the predicting result to 1. Our model and the baselines are evaluated in terms of Precision, Recall, F1-score, as well as Accuracy. All experiments are performed on a dual-core Xeon E5-2690 v2 processor.

\subsection*{4.2 Results}

Table 2 shows the performance of our proposal and the baselines. It can also be observed our model constantly outperforms IP and IMM, which means only considering interactions among contagions in IMM is not sufficient. When \( K = 1 \), in terms of accuracy, the proposed IAD scheme outperforms IP by 4.14\% and outperforms IMM by 1.17\%. In terms of F1-score, the proposed IAD scheme outperforms IP and IMM by 6.38\% and 2.99\% respectively. When \( K = 2 \), our model performs better than IP and IMM by 3.89\% and
news, and especially politics. Ordinary users prefer contagions about economy, history, and ordinary users. Figure 3 (b) shows that authority users would like to adopt contagions from authority users there exists a status gradient on social roles seniority. Hub users are more likely to adopt contagions forwarded by users in role r_j, r_i as the ordinate and r_j as the abscissa, denotes the willingness of users in role r_i adopting contagions forwarded by users in role r_j; (b) User Role - Contagion Category Interactions \( \Omega_{role}^{category}(r_i, c_k) \), with \( r_i \) as the ordinate and \( c_k \) as the abscissa, denotes the willingness of users in role r_i adopting contagions of category \( c_k \); (c) Contagion Category-Category Interactions \( \Lambda_{category}(c_i, c_k) \), with \( c_i \) as the ordinate and \( c_k \) as the abscissa, denotes the influence of contagion in category \( c_k \) on contagion in category \( c_i \).

1.35% respectively in terms of accuracy, and achieves an improvement of 6.54% and 2.57% over IP and IMM in terms of F1-score. It can also be seen that our model constantly outperforms UI, which demonstrate that only consider interactions among users is also not sufficient. The results validate the effectiveness of our proposal, and demonstrate the interactions involved in the proposed model do play important roles in information diffusion process.

Taking the model complexity into consideration, IAD is much more efficient than IMM. Please note that the number of parameters to learn in IAD is 469, 999 and 2,659 respectively, whereas the number of parameters in IMM is 2,808,400 and 2,808,800 respectively. The difference in the time cost of the learning process between the two models is one order of magnitude, specifically about 6 hours in IAD vs. 76 hours in IMM under identical configuration (\( K = 2 \) and \( \vert t \vert = 20 \)).

4.3 Analysis of Interactions.

Throughout this section, we provide qualitative insights into the extent to which the interactions influence the adoption of contagions. After fitting the proposed model, \( \Delta_{role} \), \( \Omega_{role}^{category} \), and \( \Lambda_{topic} \) are obtained. Then the results are further processed by the classification process, and \( \Lambda_{category} \) and \( \Omega_{category} \) are derived. Here we show the results of \( \Delta_{role} \), \( \Lambda_{category} \) and \( \Omega_{category} \) when \( \vert t \vert = 50 \) and \( K = 2 \).

Figure 3 shows the contagion and user interactions. In Figure 3 (a), it can be seen that authority users are more likely to adopt contagions forwarded by other authority users, rather than those from hub users and ordinary users, which indicates there exists a status gradient on social roles seniority. Hub users would like to adopt contagions from authority users and ordinary users. Figure 3 (b) shows that authority users are more likely to adopt contagions about economy, history, news, and especially politics. Ordinary users prefer contagions on constellation, life, movie, music, sports and technology. Many hub users and ordinary users tend to adopt contagions about advertisement, and one possible reason is that they may be spam users. Figure 3 (c) reveals how different categories of contagions compete or cooperate to get propagated. It can be observed that on average, relationships between different categories are mainly competition, which validates the conclusion that attention is limited for individual users to adopt contagions [Weng et al., 2012]. It also shows that contagions belonging to food category are more likely to get adopted when simultaneously propagating with contagions belonging to other categories, i.e., the propagation of contagions on food are more likely to suppress the propagation of other contagions. In addition, contagions about constellation, culture, health and life also attract a lot of attention. On the contrary, contagions belonging to advertisement are least likely to suppress other contagions’ propagation, revealing that users are not interested in them.

5 Conclusion

A new information diffusion framework called IAD is proposed to analyze the users’ behaviors on adopting a contagion, in consideration of the interactions involving users and contagions as a whole. With this framework, we can quantitatively study how these interactions would influence the propagation process. To efficiently learn the interactions, we need to classify users and categorize contagions. Therefore, we use a generative process to infer user roles and a co-training method used to classify the contagions into explicit categories. Experimental results on large-scale Weibo dataset demonstrate that IAD can outperform the state-of-art baselines in terms of F1-score, accuracy and runtime in learning. Moreover, interesting findings are observed from the interactions, which are useful to various domains such as viral marketing.
Acknowledgments

This work was supported in part by State Key Development Program of Basic Research of China (No. 2013CB329605), the Natural Science Foundation of China (No. 61372191, 61472263), China Scholarship Council, the Australian Research Council (Grants No. DP150103008) and NSF through grants III-1526499.

References


