MINIMIZING SENSITIVE INFORMATION DIFFUSION IN ONLINE SOCIAL NETWORKS

B. LAKSHMI PRIYA PG Scholar, *Department of Master of Computer Applications, Narayana Engineering College(Autonomous), Gudur.SPSR Nellore, AP*
P. VIJAY BHASKAR REDDY Assistant Professor, *Department of Master of Computer Applications, Narayana Engineering College(Autonomous), Gudur.SPSR Nellore, AP*

Abstract – The cascading of sensitive information such as private contents and rumors is a severe issue in online social networks. One approach for limiting the cascading of sensitive information is constraining the diffusion among social network users. However, the diffusion constraining measures limit the diffusion of non-sensitive information diffusion as well, resulting in the bad user experiences. To tackle this issue, in this paper, we study the problem of how to minimize the sensitive information diffusion while preserve the diffusion of non-sensitive information, and formulate it as a constrained minimization problem where we characterize the intention of preserving non-sensitive information diffusion as the constraint. We study the problem of interest over the fully-known network with known diffusion abilities of all users and the semi-known network where diffusion abilities of partial users remain unknown in advance. By modeling the sensitive information diffusion size as the reward of a bandit, we utilize the bandit framework to jointly design the solutions with polynomial complexity in the both scenarios. Moreover, the unknown diffusion abilities over the semi-known network induce it difficult to quantify the information diffusion size in algorithm design. For this issue, we propose to learn the unknown diffusion abilities from the diffusion process in real time and then adaptively conduct the diffusion constraining measures based on the learned diffusion abilities, relying on the bandit framework.
Index terms – Sensitive Information, Fully-known network, semi-known network.

I. INTRODUCTION

The prevalence of online social networks such as Facebook, Twitter and Wechat facilitates the information diffusion among users, and thus enables the efficient promotion of positive information, e.g., products, news, innovations [1]- [8]. Although such efficient diffusion can easily lead to large-scale diffusion called information cascading, the unconstrained cascading behavior could meanwhile cause the sensitive information to be incautiously diffused over the network [9]. Here the sensitive information refers to any kind of information that needs to be prohibited from cascading such as rumors, personal contents, and trade secrets.

The cascading of such sensitive information may cause the risk of leaking users' privacies or arising panics among publics [9]. With this concern, several social network medias (e.g., Facebook, Twitter) have claimed authorities to block accounts of users and delete some posts or tweets when they violate relevant rules about privacies or securities [9]. Thus network managers are able to take measures to prohibit the cascading of sensitive information.

The existing attempts that share the closest correlation with prohibiting sensitive information diffusion belong to the rumor influence minimization [9], whose current strategies can mainly be classified into two aspects. The first is diffusing the truths over network to counteract rumors. However, diffusing truths is only suitable for constraining the rumors, while is not suitable for constraining the diffusion of the other kinds of sensitive information, including personal information , trade secrets, and etc. The second is temporarily blocking a number of users with high diffusion abilities [9] [10] or blocking a number of social links among users in hope of minimizing the diffusion of a rumor. Although such strategy is effective for preventing rumors about some significant events like earthquakes, terrorist attacks and political elections, it is unrealistic for network managers to adopt this strategy on constraining the diffusion of sensitive information with various contents that widely exist in our daily lives. If network managers take such measure, it is required to block a much larger size of users or links. Then two critical problems arise.

Firstly, blocking too many users or social links will degrade user experiences and may arouse complaints for the right violation. Secondly, blocking users or social links for restraining rumors also brings the loss of the diffusion of positive information , say information loss, which is not beneficial to the viral marketers that utilize information cascading to promote products [1]- [6].
Regarding the limitations of existing solutions, in this paper, we take the first look into limiting the cascading of sensitive informations while preserving the diffusion of non-sensitive ones to lower the information loss. Considering the randomness of the users accepting information, diffused from their social neighbors, we adopt the widely used random diffusion model that each user diffuses information to his social neighbor successfully with a diffusion probability via the social link between them. Then our technical objective is adjusting the diffusion probabilities via social links to minimize the diffusion size of sensitive information under the constraint of keeping the value of the sum of diffusion probabilities via all social links. Corresponding to the reality, we consider a case where simultaneously diffuse over an online social network. In this case, decreasing diffusion probabilities models the measures such as deleting partial posts or fanpages reposted by users, while the measures for increasing diffusion probabilities include sticking and adding pushes or deliveries of the posts reposted by given users. Then, if network managers decrease the diffusion probability from a user holding rumors, the advertisements diffused from the user will inevitably be constrained as well. Thus, for lowering the diffusion loss of the advertisements and preserving the global diffusion ability of the whole network on diffusing non-sensitive information a natural approach is increasing the diffusion probabilities from one or more other users which hold the advertisements.
We study the problem of interest on both fully-known and semi-known networks which are the two main scenarios considered in current studies on information diffusion [1]. Over the fully-known network, we assume network managers know the diffusion abilities of all users. The examples for the fully-known network lie on the social networks for enterprises (e.g., Skype) or special interest groups (SIGs) (e.g., Douban1). As the full topology of a local social network, which consists of the staff of a same enterprise or the members in a same SIG, is available to network managers, it is feasible to quantify the diffusion abilities of all users.
On the contrast, the semi-known network here refers to the case that diffusion abilities of partial users remain unknown in advance. For example, the data of Facebook was reported to be utilized to influence the 2016 election in the US, which then led to a severe trust crisis for Facebook. Thus, due to the privacy concern and potential side effect, even for network managers, it is difficult to obtain the full topology of some global large scale social networks like Facebook, we chat. Unless the full network topology is known, we cannot evaluate the diffusion abilities of all users. Over the fully-known network, although we can determine the diffusion probability variations via social links through solving a constrained minimization problem, the huge size of social links in current large scale networks leads to the high complexity of the problem. Moreover, the unknown diffusion abilities of partial users over the semi-known network induce it infeasible to directly solve the constrained minimization problem for minimizing the diffusion size of sensitive information .
To tackle the above challenges, we utilize the constrained combinatorial multi-arm bandit framework to jointly design our solutions over the fully-known and semi-known networks, where we take the diffusion size of sensitive informations as the reward of a bandit and model the probability variations as the arms in bandit. With this mapping, we determine the probability variations through a constrained arms picking process with the aim of minimizing the obtained rewards. Through incorporating the constraint of diffusion probability variations into the construction of the arms of bandit, we relax the problem of interest into an unconstrained minimization problem when determining the diffusion probability variations based on the arms.

II. BACKGROUND WORK

For characterizing the information diffusion process in online social networks, Kempe et al. first propose two classic diffusion models: Independent Cascading (IC) model and linear threshold (LT) model. In the IC model, each user has a single chance to successfully diffuse the information to his neighbors with a given probability after this user having received the information. While in the LT

model, a user would get the information if a certain fraction of his neighbors have received the information. Since then, a great deal of works study the Influence Maximization (IM) problem, which focuses on efficiently selecting the optimal seed users to trigger a diffusion process in hope of maximizing the final information diffusion size [1]. Recently, due to the high cost of seeding influential users, Shi et al. [3] propose to let influential users repost the required information while seed the ordinary users for lowering the cost of IM campaign.

Similar to the multi-round setting in this paper, the seed selection for maximizing the information diffusion in multiple time rounds is considered in [2]. Moreover, considering the widespread interactions between the cyber (online) and physical (offline) worlds, offline events are utilized in [7] to further improve the performance of IM. On the contrast of the IM problem, there are also abundant researches focusing on minimizing the influence of rumors. One strategy for rumor influence minimization is diffusing the truths over network to counteract rumors. Specifically, the competitive linear threshold (CLT) model that characterizes the competing diffusion of truth and rumor is introduced in [12]. Then He et al. and Chen et al. propose to select a set of seed users to maximize the diffusion of truths under the CLT model. Chen et al. extends the IC model to describe the diffusion of positive information under the effect of negative information, and studies how to maximize the positive information diffusion. However, such clarifying measure cannot be used to constrain the diffusion of private sensitive information such as personal information , trade secrets.

Another class of rumor blocking measures focuses on blocking a certain number of influential users [9] or social links. On one hand, Song et al. propose to temporarily block a number of users with high diffusion abilities to reduce the diffusion of rumors before a deadline. With the consideration of user experiences, Wang et al. [9] study the online rumor blocking problem that periodically blocking a fraction of users during the rumor diffusion, and set a threshold to controls the blocking time of each user. Further, for coping with the unforeseen events in rumor diffusion, the adaptive blocking strategy is proposed. On the other hand, considering that straightforwardly blocking users is not desirable, propose to block a given number of social links for minimizing the diffusion of rumors. However, as we illustrated before, this kind of measures may incur much information diffusion loss, if being adopted to constrain the diffusion of the sensitive informations, considered in this paper. In addition, taking measures to constrain or promote information diffusion is also related to the studies about the effect of human behaviors on diffusion.

Besides the information diffusion, our work is also related to the combinatorial multi-arm bandit model. introduce the general multi-arm bandit model where only one arm is picked in each round. Recent studies utilize the combinatorial bandit in the IM problem over unknown or dynamic networks, where the diffusion probabilities in IC model are assumed to be unknown in advance. In each round, the solutions proposed first take the diffusion results in previous rounds as the feedback to learn the diffusion probability via each edge, and then conduct the seed selection based on the learned diffusion probabilities.

III. PROPOSED WORK

A. Diffusion Model

Over both the fully-known and semi-known networks, the sensitive information can only be diffused from the sensitive nodes. We assume there are T time rounds. A non-sensitive node will turn to sensitive once it receives sensitive information, and from then on until the end of the T rounds, as long as the sensitive information it holds are not out of date, it will have chance to diffuse sensitive information to its neighbors. The t-th round refers to the time from time stamp t to t+1. We use V t to denote the set of the sensitive nodes at time stamp t. We denote the edges whose source nodes are in sensitive as the target edges.

We study the problem of adaptively adjusting diffusion probabilities via edges at the beginning of each round, for taking measures in real time to minimize the sensitive information diffusion. For this end, we define the duration of each round as the time for two-hop diffusion. That is, the information diffuse two hops during a round.

B. Adaptive Diffusion In Fully-Known Networks

In the fully-known network, the problem of interest is a classical Linear Programming (LP) problem. However, the classical solutions (e.g., Simplex Algorithm, Ellipsoid Algorithm and Karmarkar Algorithm) for the LP problem cannot be efficiently applied to problem (1) in adaptive diffusion, due to the high dimension of the variable vector.

For the issue of the high complexity of classical solutions, we seek the solution for the adaptive diffusion based on the bandit framework. In particular, we model the probability variation vector in each round as the arm of a bandit, and model the diffusion size of sensitive information as the reward obtained from the bandit after picking such arm. By this, we explore the efficient solution for the adaptive diffusion through exploring efficient arm picking algorithm under the objective of minimizing obtained rewards. In addition, we adopt the bandit framework here also under the consideration that the bandit model will enable us to deal with the partial unknown diffusion abilities in the semi-known network (in Section 4). That is, we utilize the bandit framework to jointly design the solutions in both the fully-known and semi-known networks.

Algorithm in Fully-known Network

The aim of our algorithm for Adaptive Diffusion in Fully-known Network, named ADFN, is selecting a combination of base-arms with the minimum sum of mean rewards. We present the pseudo code of ADFN in Algorithm 1, where we use the combination to denote the set of the selected base-arms. In ADFN, we iteratively select the base-arm v with the minimum mean reward. Then, if the base-arm v is not conflicted with all the base-arms in the current combination and has negative mean reward, we add it into the combination. Furthermore, the super-arm in each round is determined by the sum of the probability variation vectors represented by the base-arms in the combination. Now, we present the complexity of the algorithm ADFN.

Algorithm 1: ADFN in the t-th diffusion round

Input: All the base-arms in the t-th round
Output: Variation Probability vector $\overrightarrow{\Delta\beta^t}$
$ActionPool \leftarrow$ All the base-arms, $combination \leftarrow \emptyset$;
while $ActionPool \neq \emptyset$ do
  v =MIN($ActionPool$);
    /* MIN(S) returns the item with the smallest reward in set S. */;
  if $\overrightarrow{D}^t \cdot \overrightarrow{\beta}_v > 0$ then
    End While ;
  end
  $ActionPool \leftarrow ActionPool \backslash \{v\}$;
  if $VALID(combination, v)$ then
    $combination \leftarrow combination \cup \{v\}$;
  end
end
for $\overrightarrow{\beta_i} \in combination$ do
  $\overrightarrow{\Delta\beta^t} = \overrightarrow{\Delta\beta^t} + \overrightarrow{\beta_i}$;
end
return $\overrightarrow{\Delta\beta^t}$

C. Adaptive Diffusion In Semi-Known Networks

Now, we proceed to explore the solutions for the adaptive diffusion over the semi-known network where the diffusion abilities of partial users remain unknown in advance. Over the semi-known network, besides the complexity issue, another major difficulty for solving the adaptive diffusion problem comes from the lack of exact diffusion abilities of partial target nodes.

Algorithm over Semi-known Network

We design the algorithm for minimizing the overall reward under the unknown mean rewards of base-arms as the “determining-learning” process. Initially, for the basearms whose two non-zero

elements are both on the target edges with informed destination nodes, we set the mean reward. Besides, for the base-arms associated with the uninformed destination nodes and without the exact mean rewards, we attach each of such base-arms an initial estimated mean reward which, obviously, is an unreliable estimated value. Then in each round, the algorithm mainly consists of two phases: (1) Determining the probability variation vector at the beginning of the round by determining the picked super-arm which consists of a set of base-arms; (2) At the end of the round, refining the estimated mean reward of each picked base-arm based on the rewards obtained from the picked base-arms in the current round.

In summary, we give in Algorithm 2 the framework for determining the probability variation vector over semi-known network. In the determining phase, we determine the super-arm as Exploration with probability, which decreases over time. Upon observing the diffusion size of sensitive informations during the current round, in the learning phase, we update the estimated mean reward of each base-arm combined into the picked super-arm.

Algorithm 2: Algorithm over semi-known network

```
for t = 1 to T do
    // Determining phase
    ε_t ← ε_0 / √t ;
    if ε_t then
        Super-arm ←Exploration;
    else
        Super-arm ←Exploitation;
    end
    Picking the super-arm;
    Observing the diffusion size of sensitive informations
     in current round;
    // Learning phase
    Updating the estimated mean reward of each
     base-arm in super-arm;
end
```

IV. RESULT

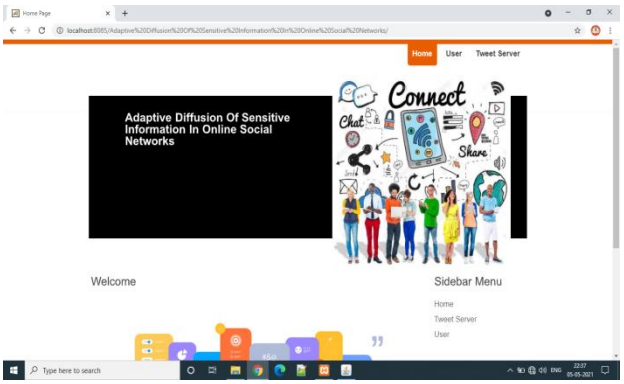

Fig. 1: Home Page

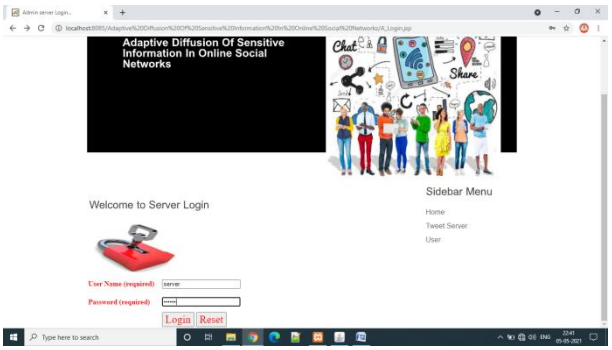

Fig. 2: Server Login Page

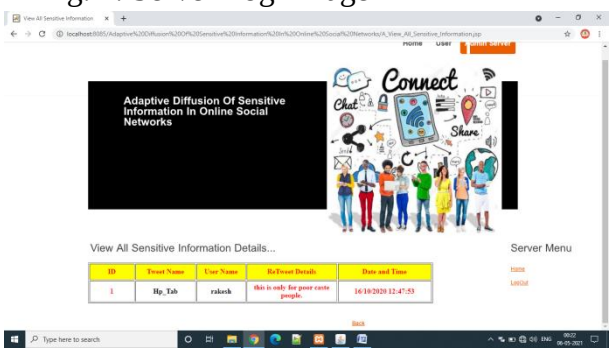

Fig. 3: Add Tweet Filter words

Fig. 4: Sensitive Information

V. CONCLUSION

In this paper, we study the problem of constraining the diffusion of sensitive informations in social networks while preserving the diffusion of non-sensitive informations. We model the diffusion constraining measures as the variations of diffusion probabilities via social links, and model the problem of interest as adaptively determining the probability variations through a constrained minimization problem in multiple rounds. We utilize the CCMAB framework to jointly design our solutions in the fully-known and semiknown networks. Over the fully-known network, we propose the CCMAB based algorithm ADFN to efficiently determine the probability variations via social links. Over the semi-known network, for tackling the challenge of unknown diffusion abilities of partial users, we propose the algorithm ADSN to iteratively learn the unknown diffusion abilities and determine the probability variations based on the learned diffusion abilities in each round.

REFERENCES

[1] Y. Li, J. Fan, Y.Wang, and K. L. Tan, “Influence maximization on social graphs: A survey”, in IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 30, no. 10, pp. 1852-1872, 2018.

[2] L. Sun, W. Huang, P. S. Yu, and W. Chen, “ Multi-round influence maximization”, in Proc. ACM SIGKDD, 2018.

[3] Q. Shi, C. Wang, J. Chen, Y. Feng, and C. Chen, “Post and repost: A holistic view of budgeted influence maximization”, in Neurocomputing, vol. 338, pp. 92-100, 2019.

[4] X.Wu, L. Fu, Y. Yao, X. Fu, X.Wang, and G. Chen, “GLP: a novel framework for group-level location promotion in Geo-social networks”, in IEEE/ACM Transactions on Networking (TON), vol. 26, no. 6, pp. 1-14, 2018.

[5] Y. Lin, W. Chen, and J. C. Lui, “Boosting information spread: An algorithmic approach”, in Proc. IEEE ICDE, 2017.

[6] Y. Zhang, and B. A Prakash, “Data-aware vaccine allocation over large networks”, in ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 10, no. 2, article 20, 2015.

[7] Q. Shi, C. Wang, J. Chen, Y. Feng, and C. Chen, “Location driven influence maximization: Online spread via offline deployment”, in Knowledge-Based Systems, vol. 166, pp. 30-41, 2019.

[8] H. T. Nguyen, T. P. Nguyen, T. N. Vu, and T. N. Dinh, “Outward influence and cascade size estimation in billion-scale networks, in Proc. ACM SIGMETRICS, 2017.

[9] B. Wang, G. Chen, L. Fu, L. Song, and X. Wang, “Drimux: Dynamic rumor influence minimization with user experience in social networks”, in IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 29, no. 10, pp. 2168-2181, 2017.

[10] Q. Shi, C. Wang, D. Ye, J. Chen, Y. Feng, and C. Chen, “Adaptive Influence Blocking: Minimizing the Negative Spread by Observation-based Policies”, in Proc. IEEE ICDE, 2019.