# 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. Extensive experiments on real and synthetic datasets demonstrate that our solutions can effectively constrain the sensitive information diffusion, and enjoy a 40%less diffusion loss of non-sensitive information comparing with four baseline algorithms.

# Table of Contents

# List of Figures

# List Of Screens

# Chapter - 1
# 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's, 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]- [20]. 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] - [20]. 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] [21] [22]. 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]- [20], whose current strategies can mainly be classified into two aspects. The first is diffusing the truths over network to counteract rumors [12]- [14]. 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's, including personal information's, trade secrets, and etc. The second is temporarily blocking a number of users with high diffusion abilities [9] [10] [15] [16] or blocking a number of social links among users [17]- [20] 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's 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's, say information loss, which is not beneficial to the viral marketers that utilize information cascading to promote products.

## EXISTING SYSTEM

The existing attempts that share the closest correlation with prohibiting sensitive information diffusion belong to the rumor influence minimization, 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 informations, including personal informations, trade secrets, and etc.
- The second is temporarily blocking a number of users with high diffusion abilities or blocking a number of social links among users in hope of minimizing the diffusion of a rumor.

## PROPOSED SYSTEM

- 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.

- This enables us to determine the probability variations via social 1. https://www.douban.com/ links with high efficiency.
- Furthermore, for coping with the unknown diffusion abilities over the semi-known network, we propose to iteratively learn the unknown diffusion abilities through learning the reward distributions of the arms based on the rewards obtained from previously picked arms, and then determine the diffusion probability variations based on the learned reward distributions of arms.

**Chapter - 2**

# LITERATURE SURVEY

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 informations 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 informations, 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 some advertisements in viral marketing and some rumors 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 [25] [26], while the measures for increasing diffusion probabilities include sticking and adding pushes or deliveries of the posts reposted by given users

[16] [27]. 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 informations, 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]- [16]. 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, Wechat. 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 informations. 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. This enables us to determine the probability variations via social links with high efficiency. Furthermore, for coping with the unknown diffusion abilities over the semi-known network, we propose to iteratively learn the unknown diffusion abilities through learning the reward distributions of the arms based on the rewards obtained from previously picked arms, and then determine the diffusion probability variations based on the learned reward distributions of arms.

Our main contributions are summarized as follows: (1) We take the first look into minimizing the diffusion size of sensitive informations while preserving the diffusion of non-sensitive ones. We formulate the problem of interest into a constrained minimization problem where we characterize the intention of preserving non-sensitive information diffusions as the constraint. (2) We propose an efficient bandit based framework to jointly explore the solutions over the fully-known and semiknown networks within polynomial running time. Moreover, we design the distributed implementation scheme of our solutions for the further improvement of time efficiency. (3) We further extend our bandit based solution into a “learning- determining” manner for addressing the challenge of unknown diffusion abilities in semi-known networks. We theoretically prove that the regret bound of our solution is sub-linear to the diffusion time, indicating that the probability variations returned by our solution approximates to the optimal one with the increase of diffusion time. (4) We perform extensive experiments on both real and synthetic social network datasets. The results demonstrate that the proposed algorithms can effectively constrain the diffusion of sensitive informations, and more importantly, enjoy a superiority over four baselines in terms of 40% less information diffusion loss.