Effectiveness of Link Deletion Methods using History of Information Diffusion Cascades for Limiting their Future Spread on Social Media

Shiori Furukawa†, Sho Tsugawa‡, and Kohei Watabe⋆

† Graduate School of Engineering Information and Systems, University of Tsukuba
‡ Institute of Systems and Information Engineering, University of Tsukuba
1–1–1 Tsukuba, Ibaraki 305-8573, Japan
⋆ Graduate School of Engineering, Nagaoka University of Technology
Kamitomiokamachi 1603–1, Nagaoka, Niigata 940–2188, Japan

Email: s.furu@mibel.cs.tsukuba.ac.jp, s-tugawa@cs.tsukuba.ac.jp, k_watabe@vos.nagaokaut.ac.jp

Abstract— The dissemination of harmful information, such as fake news, on social media has become a serious issue. Therefore, many researchers have devoted considerable effort to limiting the diffusion of harmful information. A promising approach to limiting diffusion of such information is the application of link deletion methods to social networks. Link deletion methods have been shown to be effective in reducing the size of information diffusion cascades generated by synthetic models on a given social network. In this study, we evaluate the effectiveness of link deletion methods by using actual logs of retweet cascades, rather than by using synthetic diffusion models. Our findings reveal that even after deleting all outgoing links from users who have already retweeted a tweet, the final size of diffusion cascade of the tweet is estimated to be only 70% of the original size when the first 20% of the diffusion cascade is available. This suggests that despite the availability of the historical records of the initial phases of information diffusion, the effectiveness of link deletion methods to reduce the final sizes of the information diffusion cascades is limited.

1. Introduction

Although beneficial information abounds on social media, the dissemination of harmful information, such as fake news, has become a serious issue [1]. For instance, during the coronavirus disease 2019 (COVID-19) outbreak, there was widespread dissemination of misleading and false information on social media [2].

Thus, the link deletion problem has been studied for mitigating the spread of harmful information on social media [3, 4, 5]. The link deletion problem aims to identify links that trigger the spread of large-scale information cascades using a social network among social media users. Here, deleting a link between two nodes means (temporarily) blocking information diffusion between them. By blocking the spread of information through the identified links, the goal is to reduce the sizes of the information diffusion cascades. However, if information diffusion is blocked too heavily, there is a risk of blocking diffusion of useful information. Therefore, it is desirable to control the diffusion of harmful information by blocking the diffusion of information between as few users as possible. In previous studies, the effectiveness of several link deletion methods has been evaluated by using synthetic information diffusion models [4, 3, 6].

We have evaluated the effectiveness of link deletion methods to reduce the sizes of information diffusion cascades by using actual logs of retweet cascades rather than using synthetic information diffusion models [7]. This is because synthetic information diffusion models cannot fully reproduce the characteristics of information diffusion on real social media [8]. Our previous study [7] has shown that after deleting 10%–50% of links from a social network, the size of cascades after link deletion is estimated to be only 50% of the original size. This suggests that the effectiveness of the link deletion methods for reducing the sizes of information diffusion cascades is limited. Note that our previous study assumes that the links are removed before the information is disseminated [7]. Conversely, there exists another possible link deletion strategy that decides the links to be deleted by utilizing the history of the early stages of information diffusion. The effectiveness of link deletion methods when the history of the initial phases of information diffusion is utilized has yet to be determined.

In this paper, we evaluate the effectiveness of link deletion methods that integrate the history of the early stages of information diffusion and the social network structure to reduce the sizes of information diffusion cascades on social media. We assume that the information diffusion history of a given cascade that was observed during a certain timeframe is available. By exploiting the observed information diffusion history, a link deletion method determines which links to be deleted. Since it is not easy to conduct an experiment that actually blocks information diffusion, we use a method proposed in [7] to estimate the sizes of retweet diffusion cascades after link deletion. Using this method, we evaluate the effectiveness of link deletion methods that

ORCID iDs Shiori Furukawa: 0009-0000-5535-1383, Sho Tsugawa: 0000-0001-7837-2857, Kohei Watabe: 0000-0002-9246-9740

Table 1: Overview of the datasets

	Android	Christianity
Num. of users	9,958	2,897
Num. of links in social network	48,573	35,624
Num. of tweets	12	5
Ave. num. of retweets	33.3	22.9

use the initial history of information diffusion in addition to the social network structure for limiting the sizes of information diffusion cascades on social media.

2. Methodology

2.1. Datasets

We employ two datasets retrieved from Stack Exchanges, a social media platform designed for sharing questions and answers. The two datasets are the Android dataset [9] and the Christianity dataset [9], and both contain a social network among users, as well as diffusion cascades of tags. We regarded each tag as a tweet and the propagation of the tag as a retweet. Following [7], we only used tweets with 100 or more retweets. The basic statistics of the datasets are depicted in Tab. 1.

2.2. Estimating Cascade Size after Link Deletion

We introduce the method to estimate the size of a given information diffusion cascade when deleting several links in the social network, which is proposed in our previous study [7]. Let T be a set of tweets posted on social media during a certain period of time, and let L be a set of links to be deleted. Here, deleting link $(u, v) \in L$ means blocking information diffusion from user u to user v. We consider a method to estimate the size of each diffusion cascade regarding tweet $t \in T$ after deleting links in L. We denote the set of users who post or repost tweet t as R_t, and the set of users involved in tweet set T as $V = \bigcup_{t \in T} R_t$. The time at which user $u \in V$ posts t or repost t is denoted as $\tau(u, t)$, and the social network that expresses who-follows-whom relationships among users belonging to user set V is denoted as $G = (V, E)$. The link $(u, v) \in E$ represents user u following user v.

First, for each tweet $t \in T$, we construct a diffusion graph $H_t = (R_t, E_t)$ that represents the diffusion paths of tweet t. The set of nodes in H_t is equal to R_t, which is the set of users who tweeted or retweeted tweet t. The link $(u, v) \in E_t$ in H_t represents tweet t spreading from user u to user v. To obtain E_t is not obvious because each cascade does not contain information about the specific diffusion paths of the tweet. Here, we construct three diffusion graphs, one of which is non-tree diffusion graph (i.e., each node can have multiple parents), and the others are trees (i.e., each node has at most a single parent). In the non-tree diffusion graph, we consider a tweet to have spread from user u to user v when user v is following user u and the timing of user u's retweeted (or tweeted) tweet t is earlier than that of user v. In other words, if $(v, u) \in E$ and $\tau(u, t) < \tau(v, t)$, then $(u, v) \in E_t$. We also construct two-types of diffusion trees, which we call diffusion tree (first) and diffusion tree (last). In the diffusion tree (first), we consider a tweet to have spread from user u to user v when user v is following user u and the timing of user u's retweeted tweet t is earlier than that of user v and the earliest among user v's followees. In contrast, in the diffusion tree (last) we consider a tweet to have spread from user u to user v when user v is following user u and the timing of user u's retweeted tweet t is earlier than that of user v and the latest among user v's followees. We call the node that does not have any incoming links in H_t the seed user of tweet t, and denote the set of seed users of tweet t as S_t.

Next, we use the link deletion method to select a set of links L ($|L| = k$) to be deleted and construct a diffusion graph after link deletion $H'_t = (R_t, E'_t)$, where $E'_t = E_t \setminus L$. We use diffusion graph H'_t to find the number of users who receive tweet t after the link deletion. In graph H'_t, seed users can receive tweet t even after the link deletion. Moreover, we assume that all users within reach of the seed users in diffusion graph H'_t can receive tweet t. Then, we estimate the number of users who receive tweet t after link deletion as the number of nodes that are within the reach of the seed nodes $s \in S_t$ in diffusion graph H'_t

2.3. Link Deletion

For reducing the final cascade size of tweet $t \in T$, link deletion methods determine a set of links L to be deleted by using social network G and the history of the initial m retweets D_{t_m}. D_{t_m} is defined as

$$D_{t_m} = \{(u_1, \tau(u_1, t)), (u_2, \tau(u_2, t)), ..., (u_m, \tau(u_m, t))\}, \qquad (1)$$

where u_i is a user who posts i-th retweet of tweet t and $\tau(u_i, t)$ is the time of i-th retweet of tweet t. The basic idea of the link deletion methods is to remove outgoing links from users who have already retweeted tweet t and incoming links to users who are likely to retweet tweet t in the future. We denote the set of users who have already retweeted tweet t as $A_{t_m} = \{u_i \mid 1 \leq i \leq m\}$, and the set of users who are likely to retweet tweet t as P_{t_m}. To determine P_{t_m}, we use an existing information diffusion prediction method called Inf-VAE [9]. For each tweet t, the Inf-VAE estimates the probability that each user u will retweet tweet t in the future using social network G and the history of initial diffusion D_{t_m}. The set of users with high probability of retweeting a tweet t estimated by Inf-VAE is P_{t_m}. We set the probability threshold to be $|P_{t_m}| = 200$. The details of the link deletion methods are as follows.

1. **Influenced** selects k outgoing links from users in A_{t_m} in descending order of out-degree of nodes to which the links are connected.

2. **Predicted** selects k incoming links to users in P_{t_m} in descending order of out-degree of nodes to which the links are connected.

3. **Influenced-Random** randomly selects k outgoing links from users in A_{t_m}.

4. **Predicted-Random** randomly selects k incoming links to users in P_{t_m}

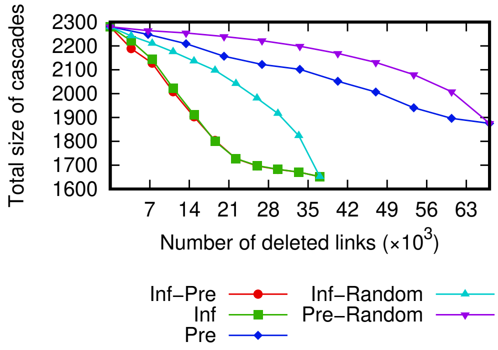
5. **Influenced-Predicted** selects k links from users in A_{t_m} to users in P_{t_m}. For link (u, v) ($u \in A_{t_m}$ and $v \in P_{t_m}$), out-degree of node v is calculated. Then, k links with high out-degree are selected as links to be deleted.

In what follows, for training the Inf-VAE model, 70% of the tweets in each dataset were used as the training data, and 10% of the tweets were used as the validation data. The rest of the 20% tweets were used for evaluating the effectiveness of the link deletion methods. For each tweet $t \in T$ the first 20% retweets were D_{t_m}.

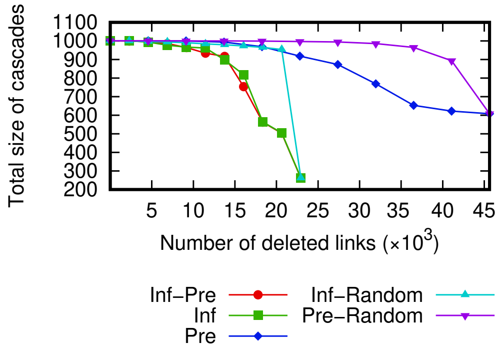
3. Results and Discussion

Figure 1 shows the estimated total size of retweet cascades against the number of deleted links in each dataset when using non-tree diffusion graph. From Fig. 1, we can see that in the Android dataset, even when all outgoing links from already retweeted users are deleted, the total estimated cascade size is about 70% that of the original cascade. For the Christianity dataset, although the total cascade size decreases by link deletion, the decrease is not large relative to the number of deleted links. For instance, after deleting 24,000 links from the the Christianity dataset, the total cascade size is about 25% that of the original cascade. Nonetheless, it should be noted that 24,000 links equate to nearly 70% of the links in the original social network. Therefore, these results suggest that the effects of link deletion methods to limit the size of tweet diffusion cascades are limited, despite having access to the initial history of diffusion cascades.

Upon examining the differences among the link deletion methods, we found that Influenced is more effective compared to Predicted. This finding indicates that preventing users who have already received the information from further dissemination is more effective than stopping the transfer of information to potential future recipients for curtailing the spread of information. In addition, the comparison between Influence and Influenced-Random demonstrates that selecting links to be deleted based on the out-degree of nodes to which the links are connected is an effective strategy. Moreover, the comparison between Influenced and Influenced-Predicted supports the usefulness of utilizing the diffusion prediction method to forecast the users who are likely to receive information, particularly when the number of links to be deleted is small.



(a) Android

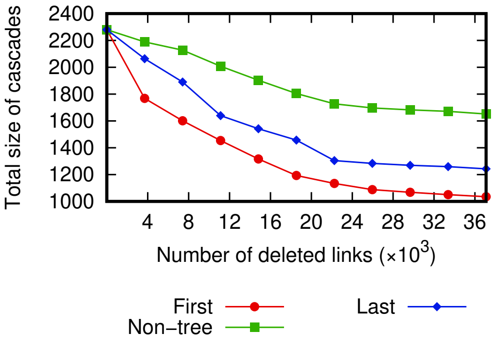


(b) Christianity

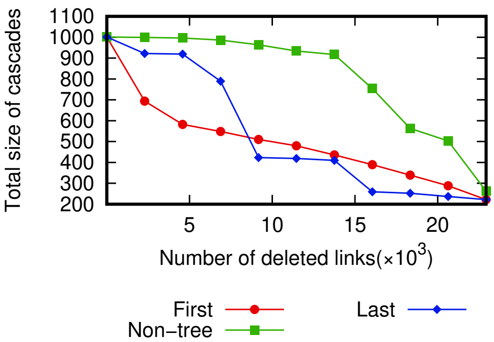
Figure 1: Total size of cascades after link deletion vs. the number of deleted links (non-tree diffusion graph)

Finally, we compare the total size of cascades when using non-tree diffusion graph, diffusion tree (first), and diffusion tree (last) when links are deleted using Influenced-Predicted (Fig. 2). Figure 2 shows that although the estimated sizes of cascades when using the diffusion trees are smaller than those when using the non-tree diffusion graph, the estimated sizes of cascades are still large. Even under the best case scenario, the total estimated cascade size is about 50% that of the original cascade in the Android dataset. This suggests that even under the optimistic estimation, the effects of link deletion are limited to reduce the cascade sizes.

In summary, our results suggest the challenges associated with reducing the size of information diffusion through the implementation of link deletion methods. Link deletion methods are not entirely effective in mitigating the dissemination of information through alternative media channels, beyond social relationships. Specifically, social media users can acquire and disseminate information through diverse sources, such as platforms' recommendations or other media. Consequently, it is suggested that to limit the transmission of information that is not mediated by social networks is important, as a means of constraining its diffusion on social media.



(a) Android



(b) Christianity

Figure 2: Comparison among different diffusion graphs (Influenced-Predicted)

4. Conclusion and Future Work

In this paper, we have evaluated the effectiveness of link deletion methods that integrate the history of the early stages of information diffusion and the social network structure to reduce the sizes of information diffusion cascades on social media. Our experimental results using social media datasets have demonstrated the limitations of the link deletion methods. The results have shown that even after deleting all outgoing links from users who have already retweeted a tweet, the final size of diffusion cascade of the tweet is estimated to be only 70% of the original size when the first 20% of the diffusion cascade is available.

In future work, we plan to examine methods to control the spread of information not via social networks. For instance, we expect that advertising counter campaigns [10, 11] is a promising approach.

Acknowledgments

This work was supported by JSPS KAKENHI Grant No. 20H04172, 22K11990, and 23H03379.

References

[1] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news: A survey on identification and mitigation techniques," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 3, pp. 1–42, 2019.

[2] D. Jolley and J. L. Paterson, "Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence," *British Journal of Social Psychology*, vol. 59, no. 3, pp. 628–640, 2020.

[3] R. Yan, Y. Li, W. Wu, D. Li, and Y. Wang, "Rumor blocking through online link deletion on social networks," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 13, no. 2, pp. 1–26, 2019.

[4] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos, "Gelling, and melting, large graphs by edge manipulation," in *Proc. 21st ACM International Conf. on Information and Knowledge Management (CIKM'12)*, 2012, pp. 245–254.

[5] Q. Y. C. Zhou, L. Xiang, Y. Cao, and L. Guo, "Minimizing the negative influence by blocking links in social networks," in *Proc. of Trustworthy Computing and Services*, 2015, pp. 65–73.

[6] A. Zareie and R. Sakellariou, "Rumour spread minimization in social networks: A source-ignorant approach," *Online Social Networks and Media*, vol. 29, p. 100206, 2022.

[7] S. Furukawa and S. Tsugawa, "Empirical evaluation of link deletion methods for limiting information diffusion on social media," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 169, 2022.

[8] R. M. Tripathy, A. Bagchi, and S. Mehta, "A study of rumor control strategies on social networks," in *Proc. 19th ACM International Conf. on Information and Knowledge Management (CIKM'10)*, 2010, pp. 1817–1820.

[9] A. Sankar, X. Zhang, A. Krishnan, and J. Han, "Inf-VAE: A variational autoencoder framework to integrate homophily and influence in diffusion prediction," in *Proc. 13th International Conf. on Web Search and Data Mining (WSDM '20)*, 2020, pp. 510–518.

[10] C. Budak, D. Agrawal, and A. E. Abbadi, "Limiting the spread of misinformation in social networks," in *Proc. 20th International Conf. on World Wide Web (WWW'10)*, 2011, pp. 665–674.

[11] Y. Fujita and S. Tsugawa, "Limiting the spread of misinformation on multiplex social networks," in *Proc. 2023 IEEE Annual Computers, Software, and Applications Conf. (COMPSAC'23)*, 2023.