Hindering Influence Diffusion of Community

Jiadong Xie
xiejiadong0623@gmail.com
East China Normal University
China, Shanghai

ABSTRACT

Considering the rapid spread of incidents like rumours or epidemics, it is important to hinder their influence diffusion. However, none of the existing works can well control the influence diffusion of a community. Based on a novel metric named interaction frequency that can measure the influence diffusion of a community, we aim to remove b nodes from a given community such that the interaction frequency of the remaining nodes is minimized. We also design a polynomial-time algorithm for the problem. The experiments show our algorithm can efficiently hinder the influence diffusion.

CCS CONCEPTS

• **Networks**; • **Theory of computation** → **Graph algorithms analysis**; • **Information systems** → **Data mining**;

KEYWORDS

social network, influence diffusion, interaction frequency

ACM Reference Format:
Jiadong Xie. 2022. Hindering Influence Diffusion of Community. In *Proceedings of the 2022 International Conference on Management of Data (SIGMOD '22), June 12–17, 2022, Philadelphia, PA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3514221.3520250

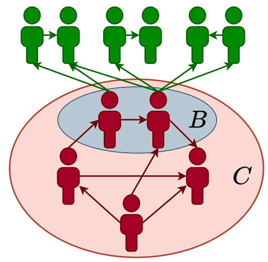

Figure 1: Motivation. Hinder the Influence of C by Removing B.

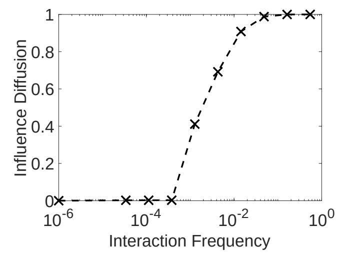

Figure 2: Motivation. Interaction Frequency Determines Influence Diffusion.

1 INTRODUCTION AND MOTIVATIONS

With the prevalence of social network platforms and the global COVID-19 epidemic, influence diffusion models [2] have attracted increasingly more attention than ever before. The infectious disease diffusion models usually consider human dynamics [5, 12] and complex contagion process [3]. In the area of social network, the previous works consider the network structure for defending network stability [17, 22] or directly minimize the influence by removing nodes [1, 20, 21] or edges [14, 15]. However, existing works have never considered community structure in diffusion models.

Our paper aims to hinder the influence diffusion of a community by removing nodes. Figure 1 presents a social network, where a rumour is diffusing in the community C and it may affect ignorant people (green) via directed edges. To hinder the rumour propagation of C, our problem propose to remove a set B from C such that the interaction frequency (a novel measure for influence diffusion) of the remaining nodes is minimized.

We formalize the influence diffusion of a community by interaction frequency. Our definition is inspired by a work on human incorporate network [7] that analytically finds (i) there is a transition that separates the local and global rumour spread; (ii) the transition point is highly related to the interactions between communities.

Figure 2 demonstrates that interaction frequency (IF) is effective in measuring influence diffusion. The experiment is conducted on the DBLP network and the SIGMOD community. For different data points in the figure, we remove different proportions of edges between SIGMOD and the outside. The x-axis is the IF of SIGMOD, and the y-axis reports the proportion of people that are influenced by a rumour simulation from SIGMOD. The simulation selects a random node in SIGMOD as rumour source to run Monte Carlo using the Maki–Thompson model ($\lambda = 0.1, \alpha = 1.0$) [7]. In Figure 2, influence diffusion is decreasing with the decrease of IF. When IF ≈ 0.001, we can completely prevent rumour spread.

2 RELATED WORK

Rumour Diffusion Model. The model is first proposed in [6] and the MT (Maki-Thompson) model further considers the formation of new edges [18]. In recent years, the models have included more details such as human dynamics [5, 12] and knowledge diffusion [19]. This work uses the MT model to investigate the effectiveness of our model.

Influence Minimization. The problem aims to minimize the expected probability of influence diffusion. The initial works minimize influence by removing edges [14, 15], while other works by removing nodes [1, 20, 21]. Our work has the same goal as influence minimization, but we use interaction frequency instead of probability to measure the diffusion.

Influence Maximization. The problem searches for a seed set of fixed size that can maximize the expected probability of influence

ACM ISBN 978-1-4503-9249-5/22/06.
https://doi.org/10.1145/3514221.3520250

diffusion, and it is first proposed in [8, 9]. It can also be formalized as a discrete optimization problem [13]. Influence maximization can be extended to a temporal setting where the timestamp of influence is considered [10, 11].

3 PRELIMINARIES

Let $G = (V, E)$ be a directed graph whose node set is V and edge set is E. If there exists an edge $\langle u, v\rangle \in E$, then v is an out-neighbor of u. The out-degree of a node u, denoted by d_u^+, is the number of out-neighbors of u, i.e., $d_u^+ = |\{\langle u, v\rangle \in E\}|$.

Given a community $C \subseteq V$ and a node $u \in C$, the cross-community out-degree of u is denoted by $d_u^{\notin C} = |\{\langle u, v\rangle \in E \mid u \in C \wedge v \notin C\}|$. The concept of interaction frequency is inspired by [7] and is a directed version of conductance [4].

Definition 3.1 (Interaction Frequency). Given a network G and a community C, the interaction frequency $\omega(C)$ of C is the sum of cross-community out-degree divided by the sum of out-degree, i.e.,

$$\omega(C) = \frac{(\sum_{u\in C} d_u^{\notin C})}{(\sum_{u\in C} d_u^+)}$$

.

According to [7] and Figure 2, rumour propagation can be largely controlled if the interaction frequency is close to 0. Consequently, we propose to remove a set B from C such that the interaction frequency of the remaining community is minimized. We formalize it as the LCIF problem.

Definition 3.2 (LCIF Problem). Given a directed graph $G = (V, E)$, a community $C \subseteq V$, and a budget b ($b < |C|$), *the least community interaction frequency problem* aims to remove a set B of at most b nodes ($B \subset C$) such that the interaction frequency of the remaining community $\omega(C \setminus B)$ is minimized, where

$$\omega(C \setminus B) = \frac{(\sum_{u\in C\setminus B} d_u^{\notin C})}{(\sum_{u\in C\setminus B} d_u^+)}$$

.

4 APPROACH

We use a binary search to find a optimal solution to LCIF problem, and the search lasts for T rounds. Let $\omega_{\text{low}} = 0$ and $\omega_{\text{high}} = \omega(C)$. For each round in the binary search, we (i) set $\omega' = (\omega_{\text{low}} + \omega_{\text{high}})/2$; (ii) test if ω' is a feasible interaction frequency, i.e., there exists a B' satisfying $\omega(C \setminus B') \le \omega'$ and $|B'| \le b$; (iii) set $\omega_{\text{high}} = \omega'$ if ω' is feasible and set $\omega_{\text{low}} = \omega'$ otherwise. After the binary search, we return ω_{high} as the least interaction frequency and return B_{ans} such that $\omega(C \setminus B_{\text{ans}}) \le \omega_{\text{high}}$ using Step (ii).

In the following, we detail Step (ii) that can both test if ω' is feasible and return the set B' that satisfies $\omega(C \setminus B') \le \omega'$.

Test the Feasibility of Interaction Frequency. Given a candidate interaction frequency ω', we compute a set $B' \subset C$ with the least budget $|B'|$ such that the interaction frequency is no more than ω', i.e., $\omega(C \setminus B') \le \omega'$. In case $|B'| \le b$, we can confirm that ω' is a feasible interaction frequency. Due to the equivalence below, we can greedily remove the node $u \in C$ with the highest $(d_u^{\notin C} - \omega' d_u^+)$.

$$\frac{\sum_{u\in C\setminus B'} d_u^{\notin C}}{\sum_{u\in C\setminus B'} d_u^+} \le \omega' \iff \sum_{u\in C\setminus B'} (d_u^{\notin C} - \omega' d_u^+) \le 0$$

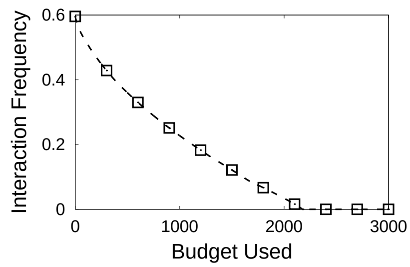

Figure 3: Our Method Decreases the Interaction Frequency of the SIGMOD Community.

Specifically, we (i) sort all nodes in C in decreasing $(d_u^{\notin C} - \omega' d_u^+)$ and store into C_{sorted}; (ii) gradually add the nodes in C_{sorted} to B' until we have $\sum_{u\in C\setminus B'} (d_u^{\notin C} - \omega' d_u^+) \le 0$; (iii) if $|B'| \le b$, then ω' is a feasible interaction frequency and the current B' satisfies $\omega(C \setminus B') \le \omega'$.

Time Complexity and Error Analysis. Let T be the number of rounds of the binary search and $n_C = |V(C)|$ be the number of vertices in C. Our binary search requires $O(n_C \log n_C)$ time each round, and $O(T \cdot n_C \log n_C)$ time overall. We also need to compute the degrees of each vertex in C using $O(\sum_{u\in V(C)} d_u^+)$ before the algorithm.

Assume B^* is the answer of the LCIF problem and B_{ans} is the output of our algorithm, then the error of our algorithm is bounded by $|\omega(C \setminus B_{\text{ans}}) - \omega(C \setminus B^*)| \le 2^{-T}$. Our algorithm returns the exact solution when T is large enough (e.g., 100).

5 RESULTS

We use the DBLP network from SNAP [16]. The network contains 317,080 nodes, 1,049,866 edges, and 8,734 ground-truth communities. We set $T = 500$ in the experiment.

In Figure 3, we apply our algorithm to SIGMOD community which contains 5172 nodes. Initially, the interaction frequency of SIGMOD equals 0.596, and the rumour can easily spread in the global network. With the increase of the budget, our algorithm can steadily decrease the interaction frequency. When the budget $b = 2177$ (removing 42.1% nodes in SIGMOD community), the interaction frequency of SIGMOD drops to 0, i.e., the rumour diffusion is completely removed according to Figure 2. The average runtime of different budgets is 2.95ms.

Contributions. Our principal contributions are as follows: (i) we define interaction frequency and identify its connection to influence diffusion; (ii) we propose a polynomial-time algorithm for minimizing interaction frequency; (iii) the experiments verify that the proposed algorithm can efficiently control the influence diffusion in real-world networks.

Acknowledgments. I would like to express my great appreciation to Mr. Deming Chu for his constructive suggestions and patient guidance. I also want to thank Mr. Xuliang Zhu, Prof. Fan Zhang, and Prof. Xuemin Lin for the valuable discussions and their assistance with this work.

REFERENCES

[1] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*. ACM, 665–674.

[2] Damon Centola. 2010. The Spread of Behavior in an Online Social Network Experiment. *Science* 329, 5996 (2010), 1194–1197.

[3] D. Centola. 2018. *How Behavior Spreads: The Science of Complex Contagions*. Princeton Univ. Press.

[4] Flavio Chierichetti, Silvio Lattanzi, and Alessandro Panconesi. 2010. Rumour spreading and graph conductance. In *SODA*. SIAM, 1657–1663.

[5] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences* 103, 7 (2006), 2015–2020.

[6] D. J. Daley and D. G. Kendall. 1964. Epidemics and Rumours. *Nature* 204 (1964), 1118.

[7] Jessica T. Davis, N. Perra, Q. Zhang, Y. Moreno, and A. Vespignani. 2020. Phase transitions in information spreading on structured populations. *Nature Physics* 16 (2020), 590–596.

[8] Pedro M. Domingos and Matthew Richardson. 2001. Mining the network value of customers. In *KDD '01*. ACM, 57–66.

[9] Pedro M. Domingos and Matthew Richardson. 2001. Mining the network value of customers. In *KDD '01*. ACM, 57–66.

[10] Nan Du, Le Song, Manuel Gomez-Rodriguez, and Hongyuan Zha. 2013. Scalable Influence Estimation in Continuous-Time Diffusion Networks. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013*. 3147–3155.

[11] Manuel Gomez-Rodriguez and Bernhard Schölkopf. 2012. Influence Maximization in Continuous Time Diffusion Networks. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*. icml.cc / Omnipress.

[12] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. 2004. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences* 101, 42 (2004), 15124–15129.

[13] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the Spread of Influence through a Social Network. In *KDD '03 (KDD '03)*. Association for Computing Machinery, 137–146.

[14] Masahiro Kimura, Kazumi Saito, and Hiroshi Motoda. 2008. Minimizing the Spread of Contamination by Blocking Links in a Network. In *AAAI '08*. AAAI Press, 1175–1180.

[15] Masahiro Kimura, Kazumi Saito, and Ryohei Nakano. 2007. Extracting Influential Nodes for Information Diffusion on a Social Network. In *AAAI '07*. AAAI Press, 1371–1376.

[16] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[17] Qingyuan Linghu, Fan Zhang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. 2020. Global Reinforcement of Social Networks: The Anchored Coreness Problem. In *SIGMOD*. ACM, 2211–2226.

[18] D. P. Maki and M. Thompson. 1973. *Mathematical Models and Applications: With Emphasis on the Social Life, and Management Sciences*. Prentice-Hall.

[19] Yamir Moreno, Maziar Nekovee, and Alessandro Vespignani. 2004. Efficiency and reliability of epidemic data dissemination in complex networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* 69 (06 2004), 055101.

[20] Senzhang Wang, Xiaojian Zhao, Yan Chen, Zhoujun Li, Kai Zhang, and Jiali Xia. 2013. Negative Influence Minimizing by Blocking Nodes in Social Networks. In *Late-Breaking Developments in the Field of Artificial Intelligence, Bellevue (AAAI Workshops, Vol. WS-13-17)*. AAAI.

[21] Ruidong Yan, Deying Li, Weili Wu, Ding-Zhu Du, and Yongcai Wang. 2020. Minimizing Influence of Rumors by Blockers on Social Networks: Algorithms and Analysis. *IEEE Trans. Netw. Sci. Eng.* 7, 3 (2020), 1067–1078.

[22] Fan Zhang, Jiadong Xie, Kai Wang, Shiyu Yang, and Jiang Yu. 2021. Discovering key users for defending network structural stability. *World Wide Web* (2021).