On the Maximization of Influence Over an Unknown Social Network

Extended Abstract

Bo Yan
Beijing Institute of Technology
Beijing, China
yanbo@bit.edu.cn

Kexiu Song
Beijing Institute of Technology
Beijing, China
614361699@qq.com

Jiamou Liu*
The University of Auckland
Auckland, New Zealand
jiamou.liu@auckland.ac.nz

Fanku Meng
Beijing Institute of Technology
Beijing, China
782987923@qq.com

Yiping Liu
Beijing Institute of Technology
Beijing, China
2120161088@bit.edu.cn

Hongyi Su
Beijing Institute of Technology
Beijing, China
henrysu@bit.edu.cn

ABSTRACT

Influence maximization is a well-investigated problem which asks for key individuals who have significant influence in a given social network. This paper addresses this problem when the social network structure is hidden. We adopt the framework of influence learning from samples and build a neural network model to represent the information diffusion process. Based on the model, we propose two new algorithms NeuGreedy and NeuMax. NeuGreedy simulates the traditional greedy algorithm whilst NeuMax utilizes the weights of connections between neurons. We test the algorithms on both synthetic and real-world datasets. The results verify the effectiveness of the proposed methods as compared to existing algorithms with or without the network structure.

KEYWORDS

Social network; social influence; influence maximization; neural network; hidden network structure; machine learning

ACM Reference Format:
Bo Yan, Kexiu Song, Jiamou Liu, Fanku Meng, Yiping Liu, and Hongyi Su. 2019. On the Maximization of Influence Over an Unknown Social Network. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

Influence maximization (IM) is a crucial problem to many fields from viral marketing to crowd mobilization [3, 16]. The problem asks for a small set S of *seed nodes* in a given social network who would spread messages among adjacent nodes in the hope to trigger a a cascade of information propagation. The number of nodes eventually affected by this propagation process represents the influence $f(S)$ of the seed nodes S. The problem aims to find the seed nodes that have maximum influence [6, 17]. The problem apparently relies on the network structure which describes links between nodes [4, 13, 14, 20, 21]. In the real-world, however, this network structure is often unavailable due to, e.g., difficulty in collecting accurate social link data or privacy concerns. It is thus important to investigate the problem without the presence of the network structure.

Several studies address IM without a given network structure. Some works, e.g., [5, 12, 18, 19] assume that the network can be queried, while others, e.g., [7, 8, 15, 22] assume the presence of *cascade information* which makes the influence function PAC learnable. Instead, we adopt the paradigm of *influence maximization from samples* (IMFS), which aims to optimize the influence function f given sample pairs of the form $(S, f(S))$. Assuming the samples are taken from a specific distribution, the OPS algorithm solves IMFS and achieves an optimization rate of $\tilde{\Omega}(n^{-1/4})$ with polynomially many samples [2]. [1] then proposed the COPS algorithm for networks with a clear community structure [11]. The algorithm outperforms OPS for samples generated by the independent cascade (IC) model. A number of limitations exist with these methods as: Firstly, these methods require certain specific sample distributions; Then, OPS fails to capture other information diffusion models such as the linear threshold (LT) model.

This paper addresses IMFS with a learning-centered approach. (1) We use a neural network framework that trains a *influence prediction model* (IPM) to represent influence. (2) Two algorithms are proposed based on IPM: NeuGreedy which simulates the original greedy algorithm as presented in [10]; and NeuMax which utilizes the neural network structure. These algorithm exhibit superior performance as compared to the benchmark algorithms.

2 PROBLEM FORMULATION

Let V be a set of nodes denoting a set of n agents. A *diffusion instance* on V captures the initialization and outcome of an information diffusion process, i.e., it is a pair (S, R) where $S \subseteq R \subseteq V$; S is called the *seed set* and R is called the *activated set.* A *diffusion mechanism* $\mathcal{M}$ is a probabilistic distribution over all diffusion instances on V.

The *influence function* $f: 2^V \to [0,1]^V$ that corresponds to $\mathcal{M}$ is defined by $f(S) = (f(S)_v)_{v\in V}$ where $f(S)_v$ is the probability that v becomes activated in a diffusion instance in $\mathcal{M}$ with seed set S. Set $|f(S)| := \sum_{v\in V} f(s)_v$; it expresses the social influence of the set S. *Influence maximization from samples* (IMFS) assumes that we are given samples from the diffusion mechanism $\mathcal{M}$ and $k \in \mathbb{N}$. The goal is to find a set $S \subseteq V$ with $|S| \le k$ and the maximum $|f(S)|$.

*Corresponding author

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

3 METHODS

We train a feed-forward neural network to represent the influence function f given a collection H of diffusion instances. The problem can be seen as multi-label classification: A sample (S, R) is multi-hop encoded as a pair of $\{0,1\}$-vectors $\vec{S}$ and $\vec{R}$ where for any set $U \subseteq V$, the entry $\vec{U}_v = 1$ if $v \in U$ and $\vec{U}_v = 0$ otherwise. We use an m-dimensional *bias vector* $\vec{b} = (b_1, \ldots, b_m)$. Let $\varphi_{>0} : \mathbb{R}^m \to \mathbb{R}^m$ be the transformation that turns the negative entries of any vector into 0. Let $\vec{\sigma}$ denote the sigmoid squash function on a vector.Given matrices $\vec{K}, \vec{L}, \vec{b}$, the *influence prediction model* (IPM) is a vector $\hat{f}_{\vec{K},\vec{L},\vec{b}}(S)$ for any set $S \subseteq V$ by $\hat{f}_{\vec{K},\vec{L},\vec{b}}(S) = \vec{\sigma}(\varphi_{>0}(\vec{S}(\vec{K} + \vec{b}))\vec{L}^{\mathcal{T}})$. Using gradient descent, we find the parameters $\vec{K}, \vec{L}$ and $\vec{b}$ that minimize the cross-entropy loss $\mathcal{L}(f, \hat{f}_{\vec{K},\vec{L},\vec{b}}) = \sum_{v \in V} f(S)_v \log \left(\hat{f}_{\vec{k},\vec{L},\vec{b}}(S)_v\right)$.

The NeuGreedy **algorithm.** The *greedy algorithm* exploits submodularity of the influence function and solves the IM problem with a given social network structure with a guaranteed approximation ratio of $1-1/e$ [10]. The procedure builds the seed set S by iteratively adding to S the agent with the highest *marginal reward*, i.e., the net increase in S' influence. The NeuGreedy algorithm takes the same idea using the optimized IPM as a surrogate of the diffusion mechanism $\mathcal{M}$ to evaluate the influence function.

The NeuMax **algorithm.** Recall that the matrix $\vec{K}$ encodes the interplay between a set $S \subseteq V$ and m features, while $\vec{L}$ encodes the interplay between the same m features and the activated node set $R \subseteq V$. The "strength" of a node $v \in S$ to the ith feature is expressed by the ith row vector in $\vec{K}$, while the "strength" of the ith feature to a node $u \in R$ is expressed by the ith row vector in $\vec{L}$. Define the *influence matrix* as the product $\vec{W} = \vec{K}\vec{L}^{\mathcal{T}}$. The (i, j)-th entry in $\vec{W}$ represents in a certain sense how much activating node i may lead to the activation of node j. The core idea behind the NeuMax algorithm is to extract influential nodes by analyzing the influence matrix $\vec{W}$. Set importance$(v) := \sum_{u \in V} \vec{W}(v, u)$. For $i \in V$, normalize the column vector $\vec{W}(i)$ so that the lowest value having value 0 and the highest having value 1 to get $\tilde{\vec{W}}(i)$. $\tilde{\vec{W}}$ denotes the *normalized influence matrix* $\left[\tilde{\vec{W}}(1) \ \ldots \ \tilde{\vec{W}}(n)\right]$. For a seed set S and a threshold $\theta \in [0, 1]$, define the *expansion* as

$$\mathsf{Expand}_\theta(S) = \left\{ v \in V \;\middle|\; \sum_{u \in S} \tilde{\vec{W}}(u, v) > \theta \right\}.$$

The expanded set of S approximates the influence function f. To build a seed set, the NeuMax algorithm greedily selects nodes utilizing both views above. At every iteration, it evaluates the expansion A of the current seeds S. If $|A| < n$, it selects a node by the column view. Otherwise, it applies the row view.

4 EXPERIMENTS

Our experiments use three benchmark algorithms: •**Random.** Here the k seeds are selected at random. • COPS. [1] COPS is shown to have significantly better performance than other IMFS algorithms. • **Greedy.** [9] The algorithm has guaranteed approximation ratio but relies on the given network structure. We verify our methods over generated datasets from three real-world network data sets: A Twitter social network (soc-twitter-copen, 761 nodes, 1029 edges)[1], a power network (bcspwr06, 899 nodes, 2914 edges)[2], and an online voting network (rt-wiki-vote, 1454 nodes, 3377 edges [2]). To generate diffusion instances, we use LT and IC diffusion models. The generated instances are then used as training samples to produce our IPM. We also use the real-world diffusion instance dataset in [7] which contains 300 million blog posts and articles from 5,000 media sites. The data set contains 1000 agents and 29265 samples.

As shown in Fig. 1 (for real networks) and Fig. 2 (for the real-world samples), our algorithms clearly outperform the random benchmark by a large margin. They outperform COPS in most of the cases, and achieve comparable results as the greedy algorithm. For IC, NeuMax generally outperforms NeuGreedy while NeuGreedy is slightly better for LT. Over the real-world cascade data sets, NeuMax clearly outperforms the other algorithms, while COPS is much worse, matched even by the greedy benchmark with the number of seeds exceeds 20.

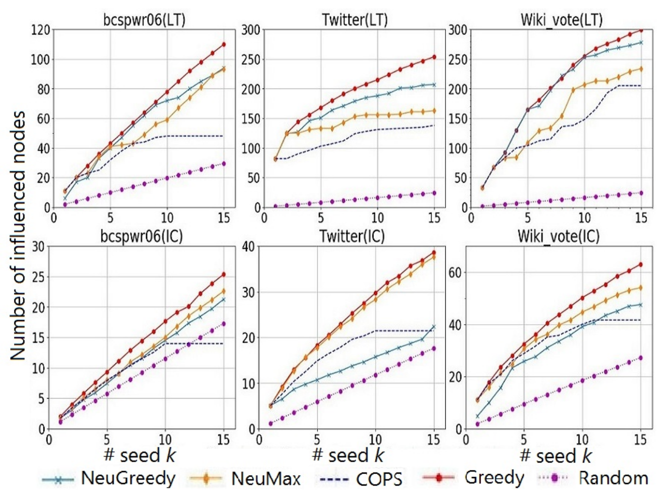

Figure 1: Comparison across the datasets: bcspwr06 (left), twitter(mid)) and wikivote (right). Samples are generated using LT (above) and IC (below) models.

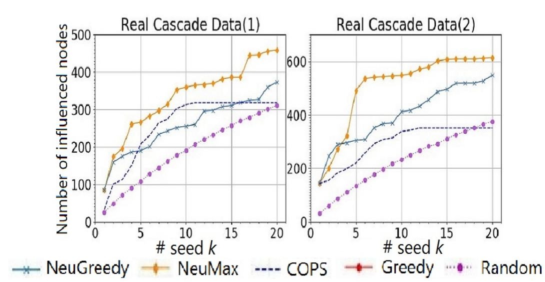

Figure 2: Results on real-world samples. The greedy algorithm cannot be applied due to lack of network data.

[1] http://konect.uni-koblenz.de/networks
[2] http://networkrepository.com

REFERENCES

[1] Eric Balkanski, Nicole Immorlica, and Yaron Singer. 2017. The Importance of Communities for Learning to Influence. In *Advances in Neural Information Processing Systems*. 5862–5871.

[2] Eric Balkanski, Aviad Rubinstein, and Yaron Singer. 2017. The limitations of optimization from samples. In *Proc. STOC*. 1016–1027.

[3] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 946–957.

[4] Yijin Cai, Hong Zheng, Jiamou Liu, Bo Yan, Hongyi Su, and Yiping Liu. 2018. Balancing the Pain and Gain of Hobnobbing: Utility-Based Network Building over Atributed Social Networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 193–201.

[5] Alexandra Carpentier and Michal Valko. 2016. Revealing graph bandits for maximizing local influence. In *International Conference on Artificial Intelligence and Statistics AISTATS*.

[6] Pedro Domingos and Matt Richardson. 2001. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 57–66.

[7] N. Du, Y. Liang, M. F. Balcan, and L. Song. 2014. Influence Function Learning in Information Diffusion Networks. In *International Conference on International Conference on Machine Learning*. 2016.

[8] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. 2010. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1019–1028.

[9] David Kempe, Jon Kleinberg, and Éva Tardos. 2010. Maximizing the spread of influence through a social network. *Progressive Research* (2010), 137–146.

[10] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 137–146.

[11] Jiamou Liu and Ziheng Wei. 2014. Community detection based on graph dynamical systems with asynchronous runs. In *2014 Second International Symposium on Computing and Networking*. IEEE, 463–469.

[12] Shodai Mihara, Sho Tsugawa, and Hiroyuki Ohsaki. 2016. Influence maximization problem for unknown social networks. In *Ieee/acm International Conference on Advances in Social Networks Analysis and Mining*. 1539–1546.

[13] Anastasia Moskvina and Jiamou Liu. 2016. Integrating networks of equipotent nodes. In *International Conference on Computational Social Networks*. Springer, 39–50.

[14] Anastasia Moskvina and Jiamou Liu. 2016. Togetherness: an algorithmic approach to network integration. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 223–230.

[15] Harikrishna Narasimhan, David C Parkes, and Yaron Singer. 2015. Learnability of Influence in Networks. (2015).

[16] Ory Okolloh. 2009. Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action* 59, 1 (2009), 65–70.

[17] Matthew Richardson and Pedro Domingos. 2002. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 61–70.

[18] Bryan Wilder, Nicole Immorlica, Eric Rice, and Milind Tambe. 2018. Maximizing Influence in an Unknown Social Network. (2018).

[19] Bryan Wilder, Laura Onasch-Vera, Juliana Hudson, Jose Luna, Nicole Wilson, Robin Petering, Darlene Woo, Milind Tambe, and Eric Rice. 2018. End-to-end influence maximization in the field. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1414–1422.

[20] Bo Yan, Yang Chen, and Jiamou Liu. 2017. Dynamic relationship building: exploitation versus exploration on a social network. In *International Conference on Web Information Systems Engineering*. Springer, 75–90.

[21] Bo Yan, Yiping Liu, Jiamou Liu, Yijin Cai, Hongyi Su, and Hong Zheng. 2018. From the periphery to the center: information brokerage in an evolving network. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. AAAI Press, 3912–3918.

[22] Jaewon Yang and Jure Leskovec. 2010. Modeling information diffusion in implicit networks. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 599–608.