Influence Learning and Maximization

George Panagopoulos[1] and Fragkiskos D. Malliaros[2]

[1] École Polytechnique, Palaiseau, France
george.panagopoulos@polytechnique.edu

[2] Université Paris-Saclay, CentraleSupélec, Inria, Gif-Sur-Yvette, France
fragkiskos.malliaros@centralesupelec.fr

Abstract. The problem of maximizing or minimizing the spreading in a social network has become more timely than ever with the advent of fake news and the coronavirus epidemic. The solution to this problem pertains to influence maximization algorithms that identify the right nodes to lockdown for epidemic containment, hire for viral marketing campaigns, block for online political propaganda etc. Though these algorithms have been developed for many years, the majority of the literature focuses on scalability issues and relaxing the method's assumptions. In the recent years, the emergence of new complementary data and more advanced machine learning methods for decision have guided part of the literature towards learning-based approaches. These can range from learning how information spreads over a network, to learning how to solve the combinatorial optimization problem itself. In this tutorial, we aim to dissentangle and clearly define the different tasks around learning for influence applications in social networks. More specifically, we start from traditional influence maximization algorithms, describe the need of influence estimation and delineate the state-of-the-art on influence and diffusion learning. Subsequently, we delve into the problem of learning while optimizing the influence spreading which is based on online learning algorithms. Finally, we describe the latest approaches on learning influence maximization with graph neural networks and deep reinforcement learning.

Keywords: Influence Maximization · Machine Learning · Graph Mining · Social Network Analysis

1 Introduction and Objectives

Social influence governs multiple aspects of our lives. From deciding the product you will buy and the restaurant you will visit, to adapting political ideas and getting infected from viruses, peer pressure and the amount and quality of the interaction with other people can be a deciding factor for a person's life. In the real world it can be used from epidemic containment [7] to diminishing the misinformation in social networks [10,3]. To this end, social influence is a concept worth studying and the problem of influence maximization is one of the most challenging and timely in social network analysis. In its core, influence maximization is a combinatorial optimization problem that aims to find a bounded set of nodes in a network that can maximize spreading. This spreading might refer to political propaganda, product purchasing intent, a virus etc. Though the initial theoretical setting is rather well-studied, it suffers from some assumptions that restrict its effectiveness in the real world. For example, it has been observed that ignoring the structural impact of a node in influence relationships leads to inaccurate spreading prediction [1]. Moreover, the network topology alone is known to fail on predicting the spreading without temporal information [16,2] or content [4].

Recently, novel methodologies have emerged that either merge influence maximization with learning-based components from extraneous data, or fully transform it in a learning problem. In this tutorial, we are going to go through the literature connecting influence maximization with machine learning methodologies. These can be separated in the sections outlined below, which resemble solutions to the different problems pertaining to influence maximization. Methodologically this includes learning models ranging from recurrent neural networks and point processes, to multi-armed bandits, reinforcement learning and graph neural networks. From an algorithmic concepts, we delineate the basics of submodular maximization and performance guarantees, as well as heuristics and sketching. For each part of the tutorial, we aim to explain the most vital papers on the problem, discussing also some variants and extensions. The target audience of the tutorial includes (i) researchers in the area of machine learning, data mining, and web engineering with applications to social media and network analysis; (ii) graduate students interested in graph mining, algorithms, and machine learning; (iii) practitioners and members of industrial partners relevant to recommender systems, epidemiology, or marketing. The assumed background is sufficient knowledge of probabilities, graph concepts, and algorithm design.

Overall, we expect that the tutorial will be of great value for the ICWE community because of the aforementioned reasons on how timely is the problem. It could be argued that it is one of the most crucial problems in current social network analysis, with important implications in the real world. Its connection to the aforementioned fields as well as computational journalism renders it also rather interdisciplinary, hence its effect will be broad and lasting. The tutorial slides along with additional resources will be available online[3].

2 Outline of the Tutorial

In this section, we give a tentative outline of the tutorial. The proposed duration of the tutorial is half a day.

1. **Introduction**
   - What is influence
   - Exemplary applications
   - Metrics for influencer identification [22,28]
   - What is a diffusion cascade

[3] http://fragkiskosm.github.io/projects/influence_learning_tutorial/

- Influence evaluation [25]

2. **Traditional Influence Maximization**
   - Influence maximization [17]
   - Faster heuristics [5]
   - Faster algorithms [20,27]
3. **Influence and Diffusion Learning**
   - The need for influence estimation [1]
   - Learning influence [13,12,11]
   - Influence Maximization with influence estimation [24,26,14]
   - Diffusion prediction using neural networks [21,15]
   - Diffusion prediction using point-processes [30,8]
4. **Learning Influence Maximization**
   - Learning combinatorial optimization [18]
   - Graph reinforcement learning for IM [9,23]
5. **Online Influence Maximization**
   - Multi-armed bandits with edge feedback [6,29]
   - Multi-armed bandits with node feedback [19]
6. **Summary and Open Problems**
   - Realistic influence maximization
   - Pointer to other tutorials and data

Acknowledgements. Supported by ANR (French National Research Agency) under the JCJC project GraphIA (ANR-20-CE23-0009-01).

References

1. Aral, S., Dhillon, P.S.: Social influence maximization under empirical influence models. Nature Human Behaviour **2**(6), 375 (2018)
2. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: quantifying influence on twitter. In: WSDM. p. 65–74 (2011)
3. Budak, C., Agrawal, D., El Abbadi, A.: Limiting the spread of misinformation in social networks. In: The WebConf. pp. 665–674 (2011)
4. Chen, S., Fan, J., Li, G., Feng, J., Tan, K.l., Tang, J.: Online topic-aware influence maximization. Proceedings of the VLDB Endowment **8**(6), 666–677 (2015)
5. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: KDD. pp. 1029–1038 (2010)
6. Chen, W., Wang, Y., Yuan, Y., Wang, Q.: Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. The Journal of Machine Learning Research **17**(1), 1746–1778 (2016)
7. Drakopoulos, K., Ozdaglar, A., Tsitsiklis, J.N.: An efficient curing policy for epidemics on graphs. IEEE Trans. Netw. Sci. Eng. **1**(2), 67–75 (2014)
8. Du, N., Dai, H., Trivedi, R., Upadhyay, U., Gomez-Rodriguez, M., Song, L.: Recurrent marked temporal point processes: Embedding event history to vector. In: KDD. pp. 1555–1564 (2016)

9. Fan, C., Zeng, L., Sun, Y., Liu, Y.Y.: Finding key players in complex networks through deep reinforcement learning. Nature Machine Intelligence pp. 1–8 (2020)
10. Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., Zha, H.: Fake news mitigation via point process based intervention. arXiv preprint arXiv:1703.07823 (2017)
11. Feng, S., Cong, G., Khan, A., Li, X., Liu, Y., Chee, Y.M.: Inf2vec: Latent representation model for social influence embedding. In: ICDE (2018)
12. Gomez-Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. TKDD **5**(4), 1–37 (2012)
13. Goyal, A., Bonchi, F., Lakshmanan, L.V.: Learning influence probabilities in social networks. In: WSDM. p. 241–250 (2010)
14. Goyal, A., Bonchi, F., Lakshmanan, L.V.: A data-based approach to social influence maximization. VLDB (2011)
15. Islam, M.R., Muthiah, S., Adhikari, B., Prakash, B.A., Ramakrishnan, N.: Deepdiffuse: Predicting the 'who' and 'when' in cascades. In: ICDM (2018)
16. Karsai, M., Kivelä, M., Pan, R.K., Kaski, K., Kertész, J., Barabási, A.L., Saramäki, J.: Small but slow world: How network topology and burstiness slow down spreading. Physical Review E **83**(2), 025102 (2011)
17. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: KDD (2003)
18. Khalil, E., Dai, H., Zhang, Y., Dilkina, B., Song, L.: Learning combinatorial optimization algorithms over graphs. In: NeurIPS. pp. 6348–6358 (2017)
19. Lagrée, P., Cappé, O., Cautis, B., Maniu, S.: Algorithms for online influencer marketing. TKDD **13**(1), 3 (2018)
20. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: KDD. p. 420–429 (2007)
21. Li, C., Ma, J., Guo, X., Mei, Q.: DeepCas: An end-to-end predictor of information cascades. In: The WebConf. pp. 577–586 (2017)
22. Malliaros, F.D., Rossi, M.E.G., Vazirgiannis, M.: Locating influential nodes in complex networks. Scientific reports **6**, 19307 (2016)
23. Manchanda, S., Mittal, A., Dhawan, A., Medya, S., Ranu, S., Singh, A.: Gcomb: Learning budget-constrained combinatorial algorithms over billion-sized graphs. In: NeurIPS (2020)
24. Panagopoulos, G., Malliaros, F.D., Vazirgianis, M.: Influence maximization using influence and susceptibility embeddings. In: ICWSM. pp. 511–521 (2020)
25. Panagopoulos, G., Malliaros, F.D., Vazirgiannis, M.: Diffugreedy: An influence maximization algorithm based on diffusion cascades. In: Complex Networks (2018)
26. Panagopoulos, G., Malliaros, F.D., Vazirgiannis, M.: Multi-task learning for influence estimation and maximization. IEEE TKDE (2020)
27. Tang, Y., Shi, Y., Xiao, X.: Influence maximization in near-linear time: A martingale approach. In: SIGMOD. p. 1539–1554 (2015)
28. Tixier, A.J.P., Rossi, M.E.G., Malliaros, F.D., Read, J., Vazirgiannis, M.: Perturb and combine to identify influential spreaders in real-world networks. In: ASONAM. p. 73–80 (2019)
29. Wen, Z., Kveton, B., Valko, M., Vaswani, S.: Online influence maximization under independent cascade model with semi-bandit feedback. In: NeurIPS (2017)
30. Zhao, Q., Erdogdu, M.A., He, H.Y., Rajaraman, A., Leskovec, J.: Seismic: A self-exciting point process model for predicting tweet popularity. In: KDD. pp. 1513–1522 (2015)