Adaptive Influence Maximization

Bogdan Cautis
LRI, Université Paris-Sud
Orsay, France
bogdan.cautis@u-psud.fr

Silviu Maniu
LRI, Université Paris-Sud
Orsay, France
silviu.maniu@lri.fr

Nikolaos Tziortziotis
Tradelab
LRI, Université Paris-Sud
Paris, France
ntziorzi@gmail.com

ABSTRACT

Information diffusion and social influence are more and more present in today's Web ecosystem. Having algorithms that optimize the presence and message diffusion on social media is indeed crucial to all actors (media companies, political parties, corporations, etc.) who advertise on the Web. Motivated by the need for effective viral marketing strategies, influence estimation and influence maximization have therefore become important research problems, leading to a plethora of methods. However, the majority of these methods are *non-adaptive*, and therefore not appropriate for scenarios in which influence campaigns may be ran and observed over multiple rounds, nor for scenarios which cannot assume full knowledge over the diffusion networks and the ways information spreads in them.

In this tutorial we intend to present the recent research on *adaptive influence maximization*, which aims to address these limitations. This can be seen as a particular case of the influence maximization problem (seeds in a social graph are selected to maximize information spread), one in which the decisions are taken as the influence campaign unfolds, over multiple rounds, and where knowledge about the graph topology and the influence process may be partial or even entirely missing. This setting, depending on the underlying assumptions, leads to variate and original approaches and algorithmic techniques, as we have witnessed in recent literature. We will review the most relevant research in this area, by organizing it along several key dimensions, and by discussing the methods' advantages and shortcomings, along with open research questions and the practical aspects of their implementation. Tutorial slides will become publicly available on https://sites.google.com/view/aim-tutorial/home.

ACM Reference Format:
Bogdan Cautis, Silviu Maniu, and Nikolaos Tziortziotis. 2019. Adaptive Influence Maximization. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3292500.3332285

ADAPTIVE INFLUENCE MAXIMIZATION IN ONLINE SOCIAL NETWORKS

Motivated by the need for effective viral marketing, Influence Maximization (IM) has been regarded as an important problem for CS researchers in the last two decades. There is a wealth of research on this topic, starting with the seminal paper of Kempe et al. [8, 9], which introduced stochastic, discrete-time diffusion models, such as Linear Threshold (LT) and Independent Cascade (IC). Parametric models such as IC and LT rely on diffusion graphs whose edges are weighted by influence probability. In [8, 9], authors show that selecting the set of seeds maximizing the expected spread is NP-hard in general, and they propose a greedy approximation algorithm. A rich literature followed, focusing on computationally efficient / scalable IM algorithms. The recent study of Arora et al. [1] summarizes state-of-the-art techniques; see also [2] for an overview. This research assumed an *offline* and *non-adaptive* problem setting. However, the need for *adaptive IM* algorithms arises naturally, for both efficiency and applicability in realistic scenarios, as discussed next.

First, given the stochastic nature of diffusion processes, even when fully explained via a parametric model like IC, such offline, one-shot decision making may often lead to sub-optimal spreads, as can be easily verified on even very basic examples [6]. Naturally, if the spread candidates could be selected in multiple rounds, during an influence campaign, an IM algorithm would be able to exploit at each round the knowledge on the already activated nodes, and thus re-focus the seed selection process accordingly at the next round. Therefore, the need for adaptive approaches arises naturally, and has been considered in recent research, e.g., in [5, 7, 13, 14, 16, 19].

Second, the large majority of studies in the IM state-of-the-art have as starting point a specific diffusion model, whose graph topology and influence probabilities are known in advance (*full-knowledge*). Yet knowing such parameters is widely regarded as unrealistic for many application scenarios. There are also situations where it is unreasonable to assume the existence of relevant historical data from which such influence parameters could be learned. Therefore, for such *partial knowledge* settings, several approaches have been proposed recently [10–12, 15, 17, 18], that learn the underlying diffusion parameters while running campaigns. To balance between exploration steps and exploitation ones, these approaches rely on *multi-armed bandits* techniques [3], and specifically the combinatorial structure [4].

Dimensions of Adaptive IM

The various adaptive IM studies exhibit significant differences in problem settings, which can steer the techniques and formal arguments in rather different directions.

Input Knowledge. The IM algorithm's input may be *complete* (or *full*), in the form of a diffusion graph G, known influence probabilities for G's edges, and a diffusion model that describes how spreads happen in G. When this information is *partial*, the IM algorithm's input may not include a diffusion model, or the influence probabilities, or both, as it may only include a coarse view of an unknown diffusion topology G [10, 11, 15].

KDD '19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08.
https://doi.org/10.1145/3292500.3332285

Feedback Model. After a spread process is initiated at a campaign round, an important facet is the observed *result* (or *feedback*). In certain application scenarios it may be reasonable to assume that only the activated nodes are observed (*node feedback*), while in others the edges over which the diffusion "travelled" may be fully observed (*edge feedback*). Moreover, feedback may be time-dependent (as in the case of *myopic feedback* [5, 13]), or may be as if instantaneous.

Optimization Objective. In the adaptive IM, the setting of utility function is even more important than in the offline case. Since spread seed selection is done repetitively, over multiple rounds, depending on the model assumptions, a node may be activated/re-activated multiple times. Accordingly, one may want to maximize either the number of activations or the number of distinct activated nodes.

Diffusion Model. While most studies adopt a setting whose underlying model is fixed, a few others take a more generic stand for adaptive IM and online learning, by assuming no diffusion model (*model-independence*) or by imposing mild assumptions thereof.

Budget-step. How the seed budget for the influence campaign is specified and can be spent may vary as well. In certain problem settings, a fixed number of seeds is to be chosen at each round, until the budget is consumed, while in others the set of seed nodes selected at a given round may be of arbitrary size (decided adaptively).

Re-seeding. Depending on all these key dimensions, the seed selection policy may proceed differently. For instance, it may or may not be able/willing to "re-seed" certain influential nodes. Similarly, it may deal with previously activated nodes by ignoring them later on or by still accounting for them in potential spread channels.

TUTORIAL ORGANIZATION

Next, we present the topics that will be covered in this tutorial.

Motivation. We will start by motivating the importance of studying social influence problems and their applications in real-world scenarios, and by presenting some examples on social influence and marketing on the Web.

Social Influence and Propagation Models. In this section, we will introduce the concept of social influence, by talking about how information propagates in social networks, with support from recent studies on word-of-mount effects. We will present the two main discrete-time propagation models, IC and LT. We will also discuss the differences between the social influence models and the well-known propagation models in network science (SI, SIS, SIR).

Brief Overview of Influence Maximization. We will review here the definition of influence maximization in the general (offline) case [8], and the theoretical properties of the spread functions resulting from the IC and LT propagation models. We will also review the complexity of the IM problem, along with its approximation guarantees. Finally, we will discuss the types of approaches used for accelerating the basic greedy algorithm.

Adaptive IM in the Full Knowledge Case. We will define the adaptive influence maximization problem, and we will discuss the most recent approaches in which full knowledge of the input is assumed. In particular, we will emphasize their formal properties, such as approximation ratio guarantees, under the full-feedback and myopic-feedback models. For this purpose, we will also provide a brief introduction to adaptive submodularity theory that constitutes a natural generalization of submodularity to adaptive policies.

Adaptive IM in the Partial Knowledge Case. In this part of the tutorial we will discuss the cases when partial or no knowledge is assumed along the following dimensions: (i) graph topology, and (ii) diffusion model. For all cases, we will discuss differences in objective functions, and the main types of feedback. We will emphasize that this problem setting leads to sequential learning instances with an *exploration-exploitation* trade-off. We will provide a concise introduction to multi-armed bandits algorithms and then discuss how such approaches are used in the adaptive IM. We will also review their main regret or waiting-time analysis results.

Complexity, Implementations, Datasets. The practical considerations are important even when dealing with the adaptive case of IM. We will cover the computational complexity of the presented research, and discuss how they can be implemented in practice. We will present and benchmark the methods currently implemented and available online. We will also discuss empirical evaluation strategies, along with the used datasets.

Comparison, Open Challenges, and Conclusions. We will end our tutorial with a comparison between the presented studies and, importantly, considerations pertaining to their applicability. Finally, we will outline the many challenges and open research questions in adaptive IM, for both sub-areas (*full* and *partial* knowledge), that are not addressed by the existing literature.

REFERENCES

[1] A. Arora, S. Galhotra, and S. Ranu. 2017. Debunking the Myths of Influence Maximization: An In-Depth Benchmarking Study. In *SIGMOD*. ACM.
[2] Çigdem Aslay, Laks V. S. Lakshmanan, Wei Lu, and Xiaokui Xiao. 2018. Influence Maximization in Online Social Networks – Tutorial. In *WSDM*.
[3] Nicolo Cesa-Bianchi and Gabor Lugosi. 2006. *Prediction, Learning, and Games*. Cambridge University Press.
[4] Wei Chen, Yajun Wang, and Yang Yuan. 2013. Combinatorial Multi-armed Bandit: General Framework, Results and Applications. In *ICML*.
[5] D. Golovin and A. Krause. 2011. Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization. *J. Artif. Int. Res.* 42, 1 (Sept. 2011), 427–486.
[6] Amit Goyal, Francesco Bonchi, and Laks V. S. Lakshmanan. 2011. A Data-based Approach to Social Influence Maximization. *Proc. VLDB Endow.* 5, 1 (2011).
[7] Kai Han, Keke Huang, Xiaokui Xiao, Jing Tang, Aixin Sun, and Xueyan Tang. 2018. Efficient Algorithms for Adaptive Influence Maximization. *PVLDB* 11, 9 (2018).
[8] D. Kempe, J. Kleinberg, and É. Tardos. 2003. Maximizing the Spread of Influence Through a Social Network. In *SIGKDD*. ACM, 137–146.
[9] David Kempe, Jon M. Kleinberg, and Éva Tardos. 2015. Maximizing the Spread of Influence through a Social Network. *Theory of Computing* 11 (2015).
[10] P. Lagrée, O. Cappé, B. Cautis, and S. Maniu. 2017. Effective Large-Scale Online Influence Maximization. In *ICDM*.
[11] P. Lagrée, O. Cappé, B. Cautis, and S.Maniu. 2019. Algorithms for Online Influencer Marketing. *ACM TKDD* (2019).
[12] S. Lei, S. Maniu, L. Mo, R. Cheng, and P. Senellart. 2015. Online Influence Maximization. In *SIGKDD*.
[13] Guillaume Salha, Nikolaos Tziortziotis, and Michalis Vazirgiannis. 2018. Adaptive Submodular Influence Maximization with Myopic Feedback. *ASONAM* (2018).
[14] Lichao Sun, Weiran Huang, Philip S. Yu, and Wei Chen. 2018. Multi-Round Influence Maximization. In *KDD*.
[15] S. Vaswani, B. Kveton, Z. Wen, M. Ghavamzadeh, L. Lakshmanan, and M. Schmidt. 2017. Diffusion Independent Semi-Bandit Influence Maximization. In *ICML*.
[16] S. Vaswani and L. Lakshmanan. 2016. Adaptive Influence Maximization in Social Networks: Why Commit when You can Adapt? *CoRR* (2016).
[17] S. Vaswani, V.S. Lakshmanan, and M. Schmidt. 2015. Influence Maximization with Bandits. In *Workshop NIPS*.
[18] Z. Wen, B. Kveton, and M. Valko. 2016. *Influence Maximization with Semi-Bandit Feedback*. Technical Report.
[19] Jing Yuan and Shaojie Tang. 2017. No Time to Observe: Adaptive Influence Maximization with Partial Feedback. In *IJCAI*.