論 文 の 内 容 の 要 旨
Abstract

論文題目　Efficient and Effective Identification of Influential Vertices in Social Networks
（ソーシャルネットワーク上の高影響力頂点集合を特定する効率的かつ効果的なアルゴリズム）

氏　　名　　大坂　直人

Social influence and information sharing occur in daily life, and social networks have been a place where such social interactions diffuse. The recent advancement of social networking services has significantly boosted the scale and speed of influence and information diffusion and enabled us to exploit influence diffusion for business use such as viral marketing. Moreover, we have been able to access a vast amount of trace of user actions at an individual level, which has encouraged a deep understanding of the mechanism of social influence at scale. Computational social influence is one of the research fields utilizing such data, aiming at analyzing, understanding, and optimizing social influence through modeling of the diffusion process, learning of model parameters, and optimization of the obtained networks. One of the most fundamental problems involving social influence optimization is influence maximization, which was formulated by Kempe, Kleinberg, and Tardos in 2003. Influence maximization is a graph optimization problem of finding a set of vertices that maximize the expected number of influenced vertices, i.e., the size of influence diffusion. Due to approximation algorithms with a theoretical guarantee and the potential application to marketing strategies and information dissemination, it has been actively studied in graph mining and graph database community for the last ten-odd years.

However, from an algorithmic point of view, the following challenges have remained unresolved. Firstly, influence maximization is still difficult to solve on real-world social networks even though there have been developed nearly-linear time approximation algorithms. This is due to the massive scale and dynamic nature of networks of the day and insufficient evaluation of algorithmic efficiency. Notably, the benchmarking study on existing influence maximization algorithms published by Arora, Galhotra, and Ranu in May 2017 has demonstrated that the setting of model parameters assigned to each edge has a significant impact on algorithmic efficiency, and there is no single state-of-the-art with the best trade-off between computation time and solution quality. Hence, boosting algorithmic efficiency even in an experimental sense is an urgent task. Secondly, influence maximization may result in ineffective strategies for influence diffusion. Since network diffusion is a probabilistic process, influence maximization has adopted expectation as a statistic to be optimized due to its simplicity and tractability. However, influence diffusion may end with a much smaller number of influenced vertices than the expectation. Expectation itself is not able to capture such a risk. Thus, it is unclear whether expectation maximization is able to produce low-risk strategies.

In this thesis, we address the above two challenges. In the first aspect, we explore efficient computation of influence maximization in practice. Our common tool for this purpose is the empirical observations of the diffusion process. There are two factors that may affect influence diffusion, i.e., the network structure and the setting of edge parameters. We conduct comprehensive experimental analysis using eighteen real-world networks and seven settings of edge parameters. We then discover the configurations of network and edge parameter setting for which existing algorithms become inefficient. We also find that existing algorithms incur redundant computation for such configurations. Based on the empirical analysis, we devise efficient algorithms under three situations below. First, we propose a fast algorithm for influence maximization. Our empirical observation tells us that for real-world networks, the difficult subproblem of influence maximization can be solved more quickly by using a simple linear time preprocessing technique. We experimentally compare the proposed algorithm with a number of existing algorithms. We show that heuristic algorithms often provide 10% less influential solutions while running faster than the proposed algorithm, and existing algorithms that have a theoretical guarantee of the solution quality demonstrate high-quality solutions; however, they cannot handle ten-million-edge networks for a certain setting of edge parameters. For such parameter settings, the proposed algorithm works and it provides comparable solutions to the existing algorithms. In particular, the proposed algorithm runs within two hours for a large network with hundreds of millions of edges. Further, we confirm the computation time reduction due to the proposed techniques by several orders of magnitude. Next, we develop a dynamic indexing algorithm for real-time influence maximization on evolving networks. We design a dynamic index structure, query algorithms for influence maximization, index update algorithms for graph changes. Then, we propose techniques for improving the algorithmic efficiency of naive update algorithms based on our empirical observation. We experimentally verify that our algorithm can update an index within one second on networks with tens of millions of edges for almost all configurations, which is several orders of magnitude smaller than that required to reconstruct an index from scratch. Then, we present a reduction algorithm for massive networks. In order to process billion-edge-scale networks, we consider reducing the size of an input at the expensive of solution quality. We propose a strategy for effectively identifying subgraphs that cause redundant computation and algorithms that produce a smaller graph that approximates an input graph. Throughout experimental evaluations using real-world networks with up to billions of edges, we confirm that an input graph is reduced to up to 4% and running influence maximization on the obtained graph achieves a few times speed-up without significant loss of solution quality.

In the second aspect, we address the risk of having a few influenced individuals. To this end, we employ portfolio optimization approach, which is a standard approach for risk management. Conceptually, in our context of influence diffusion, we virtually invest in the possible sets of vertices. Then, we adopt conditional value at risk as a statistic to be optimized instead of expectation. Conditional value at risk is one of the most popular risk measures in financial economics and actuarial science. Since we cannot use a standard approach for portfolio optimization because of exponentially many variables, we develop a new polynomial-time approximation algorithm. Our algorithm constructs a portfolio that approximates the maximum conditional value at risk within a constant additive error. Using relatively small network dataset, we experimentally demonstrate that the portfolios that our algorithm constructs achieve two times larger conditional value at risk than standard influence maximization, and the distribution of the number of influenced vertices is well concentrated on the expectation, which is desirable in terms of risk aversion.