A Pagerank-Inspired Heuristic Scheme for Influence Maximization in Social Networks

Bo Zhang, Nanjing University of Posts and Telecommunications, Nanjing, China

Yufeng Wang, Nanjing University of Posts and Telecommunications, Nanjing, China

Qun Jin, China Jiliang University, Hangzhou, China & Waseda University, Tokyo, Japan

Jianhua Ma, Hosei University, Tokyo, Japan

ABSTRACT

This article focused on seeking a new heuristic algorithm for the influence maximization problem in complex social networks, in which a small subset of individuals are intentionally selected as seeds to trigger a large cascade of further adoptions of a new behavior under certain influence cascade models. In literature, degree and other centrality-based heuristics are commonly used to estimate the influential power of individuals in social networks. The major issues with degree-based heuristics are twofold. First, those results are only derived for the uniform IC model, in which propagation probabilities on all social links are set as same, which is rarely the case in reality; Second, intuitively, an individual's influence power depends not only on the number of direct friends, but also relates to kinds of those friends, that is, the neighbors' influence should also be taken into account when measuring one's influential power. Based on the general weighted cascade model (WC), this article proposes Pagerank-inspired heuristic scheme, PRDiscount, which explicitly discounts the influence power of those individuals who have social relationships with chosen seeds, to alleviate the "overlapping effect" occurred in behavior diffusion. Then, the authors use both the artificially constructed social network graphs (with the features of power-law degree distribution and small-world characteristics) and the real-data traces of social networks to verify the performance of their proposal. Simulations illustrate that PRDiscount can advantage over the existing degree-based discount algorithm, DegreeDiscount, and achieve the comparable performance as greedy algorithm.

Keywords: Heuristic, Influence Maximization, Pagerank, Overlapping Effect, Word-of-Mouth

INTRODUCTION

This "word-of-mouth" (WOM) dissemination of information through social networks is of paramount importance in our everyday life. There is a wide range of situations (epidemiology, computer virus, marketing, political science, and agriculture etc.), in which users coordinate their decisions and form conventions through being influenced by behaviors of their friends

or neighbors, e.g., either to adopt a new behavior or product or not. Consider the following hypothetical scenario as a motivating example. A small company develops a new product and wants to market it through real social networks through the effect of WOM. Due to the limited budget for marketing, it can only select a small number of initial users in social networks to use it (by giving them gifts or payments). The company wishes that these initial users would start influencing their friends on the social network to use it, and their friends would influence their friends' friends and so on, and thus through so-called contagion process, a large population in the social network would finally adopt the application. The problem is whom to be selected as the initial users so that they eventually influence the largest number of people in the network.

The above problem, called influence maximization, was firstly stated by (Domingos, & Richardson, 2001): if we can try to convince a subset of individuals to adopt a new product and the goal is to trigger a large cascade of further adoptions, which set of individuals should we target in order to achieve a maximized influence? It was shown that, finding the influential set of nodes is NP-hard problem. Only for submodular function of diffusion model, the greedy algorithm can approximately work, with bounded threshold. (Kempe, Kleinberg, & Tardos, 2003) have proved that a simple greedy algorithm (choosing the nodes with maximal marginal gain) can approximate the optimal solution by a $(1-1/e)$, i.e., within 63% of optimal. However, the simple greedy-based approach has a heavy computation load. Specifically, the Greedy-based algorithm calculates the influence power precisely by enumeration. The more rounds the enumeration takes, the more accurate the result will get. However, when the network size increases, the computational time will increase dramatically, which prevents the greedy algorithm to become a feasible solution for the influence maximization problem in the real world. The research community has greatly studied the algorithmic aspects of maximizing influence in social networks, basically from two directions to solve the problem of efficiency: improving the greedy algorithm to further reduce its running time; the second is to propose new heuristics.

(Leskovec, Krause, Guestrin, Faloutsos, VanBriesen, & Glance, 2007) presented an optimized greed algorithm, which is referred as the "Cost-Effective Lazy Forward (CELF)" scheme. The CELF optimization uses the submodular property of the influence maximization objective to reduce the number of evaluations on the influence spread of nodes. Their experimental results demonstrate that, in comparison with simple greedy-based approach, CELF optimization could achieve as much as 700 times speedup in selecting seeds.

Considering the fundamental step of the greedy algorithm is to pick a node in each iteration from the remaining nodes, which attempts to make the maximum marginal contribution to the process of spread of information, (Narayanam, & Narahari, 2011) proposed a SPIN (ShaPley value based Influential Nodes) algorithms for computing the marginal contributions using the concept of Shapley (a well-known solution concept in cooperative game theory). Specifically, the Shapley value of a coalitional game provides the marginal contributions of the individual players to the overall value that can be achieved by the grand coalition of all the players.

Even with those improved greedy algorithms, their running time is still large and may not be suitable for large social network graphs. A possible alternative is to use heuristics. In sociology literature, degree and other centrality-based heuristics are commonly used to estimate the influence of nodes in social networks. Degree is frequently used for selecting seeds in influence maximization (so-called "degree centrality"): simply, the nodes with highest degree in remaining nodes are chosen. Surprisingly, if we choose all target users with high centrality, the resultant scheme only outperforms random selection for small target sets, due to overlapping effect (Han & Srinivasan, 2012). The main reason lies in that, the selection approach simply based on degree centrality does not take into account the neighborhood overlapping (so-called overlapping ef-

Related Content

Towards Event-Driven Context-Aware Web Services
Guadalupe Ortiz, Juan Boubeta-Puig, Alfonso García de Prado and Inmaculada Medina-Bulo (2013). *Adaptive Web Services for Modular and Reusable Software Development: Tactics and Solutions (pp. 148-159).*
www.igi-global.com/chapter/towards-event-driven-context-aware/69472?camid=4v1a

Development of a 4D Visualization Tool for Construction Planning
Takanori Terashma, Koji Makanae and Nashwan Dawood (2011). *E-Activity and Intelligent Web Construction: Effects of Social Design (pp. 13-21).*
www.igi-global.com/chapter/development-visualization-tool-construction-planning/53270?camid=4v1a

Organizing Thematic, Geographic, and Temporal Knowledge in a Well-Founded Navigation Space: Logical and Algorithmic Foundations for EFGT Nets
Levin Brunner, Klaus U. Schulz and Felix Weigel (2006). *International Journal of Web Services Research (pp. 1-31).*
www.igi-global.com/article/organizing-thematic-geographic-temporal-knowledge/3087?camid=4v1a

XWRAPComposer: A Multi-Page Data Extraction Service

Ling Liu, Jianjun Zhang, Wei Han, Calton Pu, James Caverlee, Sungkeun Park and Terence Critchlow (2008). *Web Services Research and Practices (pp. 189-223).*

www.igi-global.com/chapter/xwrapcomposer-multi-page-data-extraction/31214?camid=4v1a