Maximizing the Long-term Integral Influence in Social Networks Under the Voter Model

Chuan Zhou, Peng Zhang, Wenyu Zang, and Li Guo
Institute of Information Engineering, Chinese Academy of Sciences
Beijing, 100093, China
{zhouchuan,zhangpeng,zangwenyu,guoli}@iie.ac.cn

ABSTRACT

We address the problem of discovering the influential nodes in social networks under the voter model, which allows multiple activations to the same node, by defining an integral influence maximization problem in a long term. We analyze the problem formulation and present an exact solution to the maximization problem. We also provide a sufficient condition for the convergence of the integral influence. We experimentally compare the exact solution with other heuristic algorithms in the aspects of quality and efficiency.

Categories and Subject Descriptors H.2.8 [Database Management]: Database Applications - Data Mining
General Terms Theory, Algorithms, Performance
Keywords Influence Maximization, Voter Model.

1. INTRODUCTION

Influence maximization, defined as finding a small subset S of k nodes that maximizes spread of influence $\sigma_M(S)$ in social networks, has been extensively studied in the literature [2, 4, 6, 7, 9]. In these studies, several influence diffusion models were proposed to formulate the underlying influence propagation process, such as the linear threshold (LT) model, the independent cascade (IC) model and the voter model. To solve the influence maximization problem, considerable approximation algorithms and scalable heuristics were designed under these models.

In particular, the voter model, proposed in [3], is a fundamental probability model simulating opinion diffusion when people may switch opinions back and forth from time to time. Even-Dar and Shapira [4] studied the instant influence maximization problem in the voter model on simple undirected graphs, and they claimed that the best seeds for long-term instant influence maximization are the highest degree nodes. Li et al. [7] studied the influence maximization problem under the extended voter model, which incorporates negative influence in modeling the diffusion of opinion. Saito et al. [8] addressed the problem of discovering the influential nodes in a social network under an approximate vote model, by defining two influence maximization problems: instant and integral in short-term period.

However, the long-term integral influence maximization, as a counterpart, is still an untouched problem that worth exploring. This is because maximizing the integral sales of a new product in the long term is a question of great concern to marketing managers.

In this paper we discuss the maximization problem where the social network behaves like the voter model. Specifically, we analyze the problem formulation, propose an exact solution, and verify it experimentally.

2. PROBLEM FORMULATION

First, we provide a general definition of the voter model from the single-item view. For other definitions, refer to the work [4, 7].

Consider a directed graph $G = (V, E)$ with self loops and edge labels weight $w : E \to [0, 1]$. For convenience, let $w(u, v) = 0$ if $(u, v) \notin E$. For $v \in V$, the set of parents of v is denoted as $Par(v) := \{u \in V,\ (u, v) \in E\}$. For weight w, we assume that for each $v \in V$, $\sum_{u\in Par(v)} w(u, v) \leq 1$. Given a seed set $S \subseteq V$, the voter model works as follows. Let $S_t \subseteq V$ be the set of nodes that are activated at step $t \geq 0$ with $S_0 = S$. At step $t + 1$, every node $v \in V$ can be activated by its newly activated neighbors with probability $\sum_{u\in Par(v)\cap S_t} w(u, v)$. If v is activated successfully, then it is put into the set S_{t+1}. The process ends at a step τ with $S_\tau = \varnothing$. For simplicity, we still denote $S_t := \varnothing$ for $t > \tau$. The process $(S_t)_{t\geq 0}$ is Markovian.

The *long-term integral influence* triggered by S, i.e., the expected value of the total activation numbers from start to end, can be denoted as $\sigma_V(S)$, *i.e.*,

$$\sigma_V(S) := \mathbb{E}^S\Big[\sum_{t=0}^{\infty} |S_t|\Big].$$

The *long-term integral influence maximization problem* (**LIIM problem**) under the voter model aims to find a subset $S^* \subseteq V$, such that $|S^*| = k$ and $\sigma_V(S^*) = \max\{\sigma_V(S) \mid |S| = k, S \subseteq V\}$, *i.e.*,

$$S^* = \arg\max_{|S|=k, S\subseteq V} \sigma_V(S) \qquad (1)$$

where k is a given parameter.

3. ANALYSIS AND SOLUTION

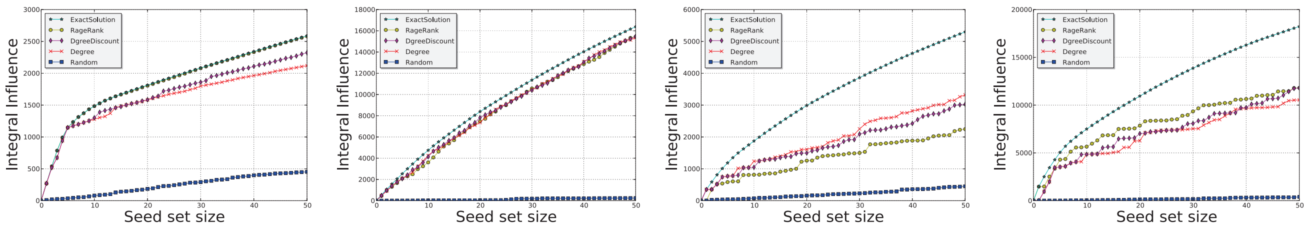
Theorem 1. *The long-term integral influence $\sigma_V(S)$ can be calculated as in Eq.(2)*

$$\sigma_V(S) = \sum_{t=0}^{\infty} \Pi_0^S \cdot W^t \cdot \mathbf{1} \qquad (2)$$

where $W = \{w(u, v)\}$ is the weight matrix in voter model.

Proof: By the Markov property of the voter model, we have $\Pi_t^S = \Pi_0^S \cdot W^t$, where Π_t^S is a row vector with element $\pi_t^S(v) := \mathbb{P}^S(v \in S_t)$. Hence, we obtain that, $\sigma_V(S) = \sum_{t=0}^{\infty} \Pi_t^S \cdot \mathbf{1} = \sum_{t=0}^{\infty} \Pi_0^S \cdot W^t \cdot \mathbf{1}$, where $\mathbf{1}$ is a column vector

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577376 .



(a) Facebook (b) Digger (c) Twitter (d) Epinions

Figure 1: Integral Influence *w.r.t.* seed size k on four data sets.

with all elements 1. Note that $\Pi_0^S \cdot W^t \cdot \mathbf{1}$ is a real number after matrix calculation. $\square$

Based on Eq.(2), one may raise a question, *under what conditions, the series* $\sum_{t=0}^{\infty} \Pi_0^S \cdot W^t \cdot \mathbf{1}$ *is convergent, and what the convergence limit is?* In the sequel, we derive Corollary 1 to answer these questions.

COROLLARY 1. *If the weight matrix W satisfies the condition $\max_v \sum_{u \in Par(v)} w(u,v) < 1$, then the series in Eq.(2) is convergent, and the limit of the convergence exists as in Eq.(3),*

$$\sigma_V(S) = \Pi_0^S \cdot (E - W)^{-1} \cdot \mathbf{1} \tag{3}$$

where E is a unit matrix and $(E-W)^{-1}$ is the inverse of $(E-W)$.

Proof: By matrix analysis [5], condition mentioned above implies $\| W \|_1 < 1$, which ensures the convergence of matrix series $\sum_{t=0}^{\infty} W^t$ with the limit $(E-W)^{-1}$. Hence we have $\sigma_V(S) = \Pi_0^S \cdot \left(\sum_{t=0}^{\infty} W^t\right) \cdot \mathbf{1} = \Pi_0^S \cdot (E-W)^{-1} \cdot \mathbf{1}$ where the first '=' is from Theorem 1. $\square$

Based on the analysis, we can provide an exact solution to the LIIM problem. Specifically, in order to maximize the long-term integral influence, we only need to choose an initial seed set S^* with the highest values in the column vector $\sum_{t=0}^{\infty} W^t \cdot \mathbf{1}$ or $(E-W)^{-1} \cdot \mathbf{1}$. Note that, although the LIIM problem can be addressed exactly in polynomial time [4], it is still time-consuming, especially when the network is large. Hence, it is necessary to take heuristic algorithms into comparison.

4. EXPERIMENTS

We conduct experiments on four real-world data sets (Facebook, Twitter, Digger, and Epinions [1]) to evaluate the long-term integral influence and solve the LIIM problem. We implement four heuristic algorithms, Degree [6], Pagerank [1], Degreediscount [2], and Random [6] for comparisons.

We measure the comparisons in two aspects: quality of the seed set (*i.e.* integral influence) and efficiency of the algorithm (*i.e.* running time). Moreover, to obtain the integral influence of heuristic methods for each seed set, we substitute them into Eq. (2) or Eq. (3). We assign weight of each directed link $(u,v) \in E$ in the network for the voter model as follows, $w(u,v) = \frac{1}{|Par(v)|+1}$ for $u \in Par(v)$, where $|Par(v)|$ denotes the number of parents of a node v. In our experiments, an undirected graph can be also regarded as a bidirectional graph.

Integral influence: We run tests on the four data sets to obtain integral influence $\sigma_V(S^*)$ *w.r.t.* parameter k from 1 to 50. From the results in Fig. 1, we can observe that no benchmark algorithms can beat the ExactSolution uniformly on the four data sets. The PageRank provides the nearest approximation in most cases, except on Twitter data. The results from Random show that maximizing the integral influence is far from trivial.

Time Cost: Fig. 2 shows time costs of selecting 50 seeds. From the results, we can observe that the benchmark algorithms, Random, Degree and DegreeDiscount, are very fast in selecting candidate nodes (which takes less than 1 second). The PageRank is slightly slower than the above three algorithms due to heavy iterations. The ExactSolution takes the longest time, but still acceptable (in a minute level).

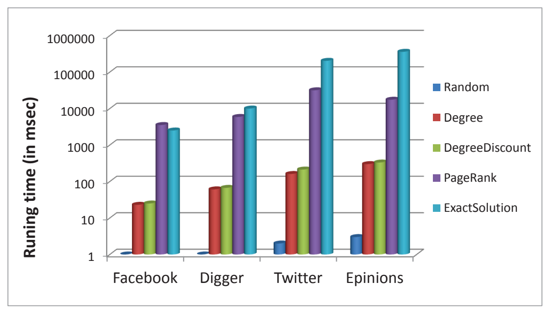


Figure 2: Runtime comparison.

Acknowledgements. This work was supported by the NSFC (No. 61370025) and the Strategic Leading Science and Technology Projects of CAS (No.XDA06030200).

5. REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[2] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD 2009*.

[3] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[4] E. Even-Dar and A. Shapira. A note on maximizing the spread of influence in social networks. In *Internet and Network Economics*, pages 281–286. Springer, 2007.

[5] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[6] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD 2003*.

[7] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *WSDM*, 2013.

[8] K. Saito, M. Kimura, K. Ohara, and H. Motoda. Efficient discovery of influential nodes for sis models in social networks. *Knowledge and information systems*, 30(3):613–635, 2012.

[9] C. Zhou, P. Zhang, J. Guo, X. Zhu, and L. Guo. Ublf: An upper bound based approach to discover influential nodes in social networks. In *ICDM 2013*.

[1] For details, visit *http://snap.stanford.edu/data/*