Omega 59 (2016) 263–278

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

A subjective evidence model for influence maximization in social networks ☆

Mohammadreza Samadi [a], Alexander Nikolaev [a,*], Rakesh Nagi [b]

[a] Department of Industrial and Systems Engineering, University at Buffalo (SUNY), Buffalo 14260, NY, United States
[b] Department of Industrial and Enterprise Systems Engineering, The University of Illinois at Urbana-Champaign, IL 61801, United States

ARTICLE INFO

Article history:
Received 6 October 2014
Accepted 30 June 2015
Available online 8 July 2015

Keywords:
Influence maximization
Social networks
Bayesian inference
Evidence
Seed selection

ABSTRACT

This paper introduces the notion of subjective evidence, which fuels a new parallel cascade influence propagation model. The model sheds light on the phenomena of belief reinforcement and viral spread of innovations, rumors, opinions, etc., in social networks. Network actors are assumed to be testing a Bayesian hypothesis, e.g., for making judgment about the superiority of some product(s) or service(s) over others, or (dis)utility of a given program/policy. The model-based influence maximization solutions inform the strategies for market niche selection and protection, and identification of susceptible groups in political campaigning. The NP-Hard problem of influential seed selection is first solved as a mixed-integer program. Second, an efficient Lagrangian Relaxation heuristic with guaranteed bounds is presented. In small, medium and large-scale computational investigations, we analyze: (1) how the success of an influence cascade triggered in a (sub)community, long exposed to an opposite belief, depends on the structural properties of the underlying social network, (2) to what extent growing (increasing the density of) a consumer network within a market niche helps a company protect the niche, (3) given a competitor's strength, when a company should counter the competitor on "their turf", and when and how it should look for limited-time opportunities to maximally profit before eventually surrendering the market.

1. Introduction and motivation

People tend to view product recommendations received from friends or through friends more favorably compared to advertisements offered by commercial mass media channels [15,41]. Social connections enable the propagation of ideas, judgments and opinions; the phenomenon where knowledge transfer between individuals significantly affects their decisions about purchasing a product is known as social influence/contagion [71,63,14]. Social influence and diffusion of innovations in social networks are mainly explored in managerial and sociological studies [75,1,2]. However, the need for simulating information diffusion/peer influence in social networks and solving optimization problems to algorithmically find potent success of cascade initiation strategies led to the introduction of the Influence Maximization (IM) problem. The objective of the IM problem is to find such early starters, termed seeds, for influence spread in a social network that will direct information transfer so as to achieve a desired impact on the expected product adoption, or people's decisions/judgments/opinions with respect to a query of interest [45,16].

Early mathematical formulations of the IM problem in social networks view social ties as indicators of dyadic dependence, where the random graph or Markov random field-based approach is a natural choice for model design [24,59]. More recent literature on the algorithmic analysis of influence spread has been dominated by diffusion-based models [45], in which ties are viewed as information flow channels. The Independent Cascade (IC) and Linear Threshold (LT) models are most notable ones, both allowing for elegant discrete optimization problem statements; these models also provided the basis for a streak of subsequent studies [46,36,72,23].

Application-wise, diffusion models have been found suitable for research studies in marketing [5,15] and health care [60]. However, algorithmic investigations up to date failed to culminate in significant managerial insights and strategies. This is in part due to the fact that existing models do not specify the medium and nature of influence flow through a network, i.e., fail to explain the diffusion of *what* leads to social influence, and *how* it does so.

This paper takes a previously unexplored approach to modeling the spread of competitive influence in social networks, rooted in Bayesian Inference theory and focused on propagation of *evidence*.

☆This manuscript was processed by Associate Editor Prokopyev.

* Corresponding author.
E-mail addresses: msamadi@buffalo.edu (M. Samadi), anikolae@buffalo.edu (A. Nikolaev), nagi@illinois.edu (R. Nagi).

http://dx.doi.org/10.1016/j.omega.2015.06.014
0305-0483/

Bayesian inference logic helps quantify social influence under the premise that people treat new information as evidence and update their beliefs in support of or against the null hypothesis. In this approach, network nodes represent intelligent agents (actors) who seek to form judgments about a product/query by testing a relevant hypothesis (e.g., that a particular claim is true), based on their prior beliefs as well as the knowledge acquired through friends. A node's decision to significantly favor the null hypothesis signals the node's "positive activation"; significantly favoring the alternative implies "negative activation"; finally, whenever the collected evidence is inconclusive, the node is labeled "inactive".

This paper presents a Parallel Cascade (PC) diffusion framework for modeling evidence spread through social networks. The flow of information in this PC model is classified as parallel duplication in the typology of flow processes on social networks, introduced by Borgatti [11], which supports the idea of belief reinforcement through subjective evidence duplication in social communication. The paper reports insightful observations, e.g., pertinent to the identification of penetrable market niches and convenient points of initial influence for conquering new market segments, obtained from solving basic instances of the PC model-based IM (PCIM) problem. The paper develops problem-specific optimization schemes for handling medium and large-scale instances of PCIM problem formulated as a Mixed-Integer program.

The rest of this paper is organized as follows. Section 2 reviews the literature on diffusion models for IM. Section 3 formally introduces the PC diffusion model, formulates PCIM problem and discusses its application to two empirical case studies. Offering a more computationally efficient approach to the problem, Section 4 presents a Lagrangian Relaxation heuristic tool suit, with solution quality guarantees achieved via two problem-specific heuristics for finding lower bounds for PCIM problem optima. Section 5 reports on the conducted experimental studies. Section 6 summarizes the findings, discusses the potential applications of the proposed methods and outlines future research directions. The paper contains two appendices: Appendix A presents the NP-hardness proof for the PCIM problem; Appendix B details the Subgradient Search algorithm for finding an upper bound for PCIM problem optima.

2. The landscape of the social influence research domain

The concept of word-of-mouth has received attention in the 1940s as an effective way for diffusion of information (e.g., about new products) and soon became a coined term in the experimental marketing research [53,76,44]. Models of information diffusion over networks, also first introduced in the marketing field, were developed more recently and found use in health care [55], sociology [50,69] and politics [20]. From an experimental point of view, the phenomenon of social contagion is known to be a significant factor affecting the strength of diffusion processes in social networks [52,1,58,3,62,4].

The investigations into the impact of influential people, or opinion leaders, on cascade formation comprise a large part of the literature. Opinion leaders are defined as the individuals that have the ability to strongly affect the opinions or decisions of their network peers [79]. While some studies degrade the value/power of opinion leaders for social cascade progression [7,74], most authors see opinion leader presence as a critical facilitator for cascade emergence [49,68,30,42,70]. Hinz et al. [41] experimentally showed that a wisely selected group of opinion leaders can increase the influence spread rate in a cascade up to eight times. Yet, two questions remain unanswered: *How can one select the appropriate opinion leaders for maximizing the spread of influence in a social network*? and *How does this selection depend the social network structure*? While the literature reviewed above is more concerned with exploring the mechanisms of successful cascade propagation, it does not provide a readily available method/solution/strategy (for a company or a political party) to artificially create a cascade in support of a product or opinion by recruiting the "best" opinion leaders. The latter objective, however, may be highly sought-after by research-aware practitioners.

The first organized efforts for identifying influential nodes in social networks relied on centrality-based heuristics [12]. The degree centrality heuristic assumes that any node with a large number of direct connections (called a hub) must be highly influential in a social network. The distance centrality heuristic, on the other hand, considers a node influential if it has short paths to other nodes in the network (called a bridge) [73,41]. The centrality-based heuristics, however, provide no quality guarantee for the solution of the IM problems with multiple required seeds. To formulate an algorithmic approach to finding influential node sets, the term "influence maximization" was coined by Domingos and Richardson [24]. While the first attempts to address the IM problem employed a Markov random field approach [24,59], Kempe et al. [45] were first to re-frame it as a discrete optimization problem.

The Independent Cascade (IC) model and the Linear Threshold (LT) model, proposed by Kempe et al. [45], are the most well-known diffusion models for IM; the optimization problems based on these models are NP-hard [72,17]. Kempe et al. [45] discovered a submodularity property of the IM objective function and presented a greedy seed selection algorithm with guaranteed, albeit loose, optimality bounds. The problem of assuming submodularity lies in the fact that under it, the marginal gain of adding new seeds should be decreasing, which does not support the idea of fixed threshold effects and reveals a manifest shortcoming in the respective diffusion models [37,50]. Furthermore, the original greedy algorithm and even its extensions were found overly demanding computationally [47,36,16,15,72].

A separate branch of literature has explored social influence from the empirical data mining perspective. As discovered by Aral et al. [2], a financially viable cascade initiation requires the selection (buy-in) of no more then 0.2% of the nodes in a network. This finding underlines the value of precise seed selection algorithms that can ensure a desirable cost/returns ratio in cascade seeding. However, one observes a gap between the literature based on data-driven studies and algorithmic research. The latter efforts, unfortunately, have often focused on computational investigations in impractical settings and failed to produce managerial insights. The present paper serves as a bridge between these two research thrusts. It designs a realistic diffusion model, strongly supported by mathematical sociology findings, and solves the seed selection problem optimally over real social network datasets, and hence, paves a way to rigorously explore strategic decision-making in social networks.

Note that the original IM problem formulation was concerned with maximizing the expected number of activated nodes *at the end of the diffusion process*, when the activation status of all the nodes becomes fixed, irrespective of the sequence and timing of node activation. However, in many practical IM applications, an influence campaign has a predefined time window, over which it has to achieve the maximized possible effect. Only recently, Goyal et al. [35] addressed the issue of unconstrained time horizon for IM and introduced MINTIME problem where the objective is to minimize the time until the activation of a predefined number of nodes.

Note also that in most IC and LT model-based seed selection problems have ignored the aspect of activation timing; furthermore, they assumed no competition. Meanwhile, the existing literature confirms the co-existence of (competing) opinions in