Physica A 450 (2016) 670–681

Contents lists available at ScienceDirect

Physica A

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

# Cost effective campaigning in social networks

Bhushan Kotnis *, Joy Kuri

*Indian Institute of Science, Department of Electronic Systems Engineering, Bangalore, India*

## HIGHLIGHTS

- Optimal allocation of incentives can increase the size of the information campaign.
- Campaign size is an increasing function of the incentivized population.
- The non-trivial optimization problem can be solved using a simple linear program.

## ARTICLE INFO

*Article history:*
Received 5 April 2015
Received in revised form 7 November 2015
Available online 20 January 2016

*Keywords:*
Percolation theory
Optimization
Information spreading

## ABSTRACT

Campaigners are increasingly using online social networking platforms for promoting products, ideas and information. A popular method of promoting a product or even an idea is incentivizing individuals to evangelize the idea vigorously by providing them with referral rewards in the form of discounts, cash backs, or social recognition. Due to budget constraints on scarce resources such as money and manpower, it may not be possible to provide incentives for the entire population, and hence incentives need to be allocated judiciously to appropriate individuals for ensuring the highest possible outreach size. We aim to do the same by formulating and solving an optimization problem using percolation theory. In particular, we compute the set of individuals that are provided incentives for minimizing the expected cost while ensuring a given outreach size. We also solve the problem of computing the set of individuals to be incentivized for maximizing the outreach size for given cost budget. The optimization problem turns out to be non trivial; it involves quantities that need to be computed by numerically solving a fixed point equation. Our primary contribution is, that for a fairly general cost structure, we show that the optimization problems can be solved by solving a simple linear program. We believe that our approach of using percolation theory to formulate an optimization problem is the first of its kind.



## 1. Introduction

Online social networking platforms are being increasingly used by campaigners, activists and marketing managers for promoting ideas, brands and products. In particular, the ability to recommend news articles [1], videos, and even products [2] by friends and acquaintances through online social networking platforms is being increasingly recognized by marketing gurus as well as political campaigners and activists. Influencing the spread of content through social media enables campaigners to mold the opinions of a large group of individuals. In most cases, campaigners and advertisers aim to spread their message to as many individuals as possible while respecting budget constraints. This calls for a judicious allocation of limited resources, like money and manpower, for ensuring highest possible outreach, i.e., the proportion of individuals who receive the message.

* Corresponding author.
*E-mail address:* bkotnis@dese.iisc.ernet.in (B. Kotnis).

http://dx.doi.org/10.1016/j.physa.2015.12.127

Individuals share information with other individuals in their social network using Twitter tweets, Facebook posts or simply face to face meetings. These individuals may in turn pass the same to their friends and so on, leading to an information epidemic. However, individuals may also become bored or disillusioned with the message over time and decide to stop spreading it. Past research suggests that such social effects may lead to opinion polarization in social systems [3]. This can be exploited by a campaigner who desires to influence such spreading or opinion formation by incentivizing individuals to evangelize more vigorously by providing them with referral rewards in the form of discounts, cash back or other attractive offers. Due to budget constraints, it may not be feasible to incentivize all, or even a majority of the population. Individuals have varying amount of influence over others, e.g., ordinary individuals may have social connections extending to only close family and friends, while others may have a large number of social connections which can enable them to influence large groups [4]. Thus, it would seem that incentivizing highly influential individuals would be the obvious strategy. However, recruiting influential people can be very costly, which may result in the campaigner running out of funds after recruiting just a handful of celebrities, which in turn may result in suboptimal outreach size.

A resource constrained campaigner, for a given cost budget, may want to maximize the proportion of informed individuals, while other campaigners who care more about campaign outreach than resource costs, may desire to minimize costs for achieving a given number of informed individuals. We address both the resource allocation challenges by formulating and solving two optimization problems with the help of *bond percolation theory*.

A similar problem of preventing epidemics through vaccinations has received a lot of attention [5–9]. However, in these problems the cost of vaccination is uniform for all individuals, and hence it is sufficient to calculate the minimum number of vaccinations. Information diffusion can also be maximized by selecting an optimal set of seeds, i.e., individuals best suited to *start* an epidemic [10–12]. This is different from our strategy which involves incentivizing individuals to *spread* the message. It is possible to address the problem posed here using optimal control theory, which involves computing the optimal resource allocation in real time for ensuring maximum possible outreach size by a give deadline [13–17]. However, the optimal control solution is not only difficult to compute, but also very hard to implement as it requires a centralized real time controller. Furthermore, recent work, [13–17], on optimal campaigning in social networks does not address the problem of minimizing the cost while guaranteeing an outreach size. Our formulation allows us to solve both the problems.

Our model assumes two types of individuals viz. the '*ordinary*' and the '*selected*', and they are connected to one another through a social network. Before the campaign starts, the selected individuals are incentivized to spread the message more vigorously than the ordinary. We use the *Susceptible Infected Recovered* (SIR) model for modeling the information epidemic. For a given set of selected individuals, we first calculate the size of the information outbreak using network percolation theory, and then find the set of selected nodes which, 1. minimizes the cost for achieving a given proportion of informed individuals, and 2. maximize the fraction of informed individual for a given cost budget. We believe that our approach of using percolation theory to formulate an optimization problem is the first of its kind.

The detailed model description can be found in Section 2, percolation analysis in Section 3, the problem formulation in Section 4, numerical results in Section 5, and finally conclusions are discussed in Section 6.

## 2. Model

We divide the total population of $N$ individuals in two types: the ordinary (type 1) and the selected (type 2). Before the campaign starts selected individuals are provided incentives to spread the information more vigorously. These individuals are connected with one another through a social network, which is represented by an undirected graph (network). Nodes represent individuals while a link embodies the communication pathways between individuals.

Let $P(k)$ be the degree distribution of the social network. For analytical tractability, we assume that the network is uncorrelated [18]. We generate an uncorrelated network using the configuration model [19]. A sequence of $N$ integers, called the degree sequence, is obtained by sampling the degree distribution. Thus each node is associated with an integer which is assumed to be the number of half edges or stubs associated with the node. Assuming that the total number of stubs is even, each stub is chosen at random and joined with another randomly selected stub. The process continues until all stubs are exhausted. Self loops and multiple edges are possible, but the number of such self loops and multiple edges goes to zero as $N \to \infty$ with high probability. We assume that $N$ is large but finite. Let $\phi(k)$ be the proportion of individuals with $k$ degrees that are provided incentives for vigorously spreading the message, i.e., proportion of nodes with degree $k$ that are type 2 nodes. The goal is to find the optimum $\phi(k)$ for maximizing the epidemic size (or minimizing the cost). The actual individuals can be identified by sampling from a population of individuals with degree $k$ with probability $\phi(k)$.

We assume that the information campaign starts with a randomly chosen individual, who may pass the information to her neighbors, who in turn may pass the same to their neighbors and so on. However, as the initial enthusiasm wanes, individuals may start losing interest in spreading the information message. This is similar to the diffusion of infectious diseases in a population of susceptible individuals. Since, we account for individuals losing interest in spreading the message, we use a continuous time SIR process to model the information diffusion. The entire population can be divided into three classes, those who have not heard the message (susceptible class), those who have heard it and are actively spreading it (infected class) and those who have heard the message but have stopped spreading it (recovered class).

Let $\beta_1$ be the rate of information spread for an ordinary node (Type 1), while $\beta_2$ for a selected node (Type 2). In other words, the probability that a type $i$ individual 'infects' her susceptible neighbors in small time $\mathrm{d}t$ is $\beta_i \mathrm{d}t + o(\mathrm{d}t)$. Note that this is independent of the type of the susceptible node. Let $\mu_i$ be the rate at which type $i$ infected individuals move to the