Electronic Commerce Research and Applications 13 (2014) 283–294

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

A targeted approach to viral marketing

Anastasia Mochalova *, Alexandros Nanopoulos

Katholische Universitaet Eichstaett-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany

ARTICLE INFO

Article history:
Received 24 September 2013
Received in revised form 30 May 2014
Accepted 1 June 2014
Available online 14 June 2014

Keywords:
Viral marketing
Social networks
Centrality scores

ABSTRACT

Viral marketing can be an effective marketing technique in social networks. Initiating from a set of influential seed users, it can activate a "chain-reaction" driven by word-of-mouth. The effectiveness of viral marketing lies in the fact that it conveys an implied endorsement from social ties. Existing approaches to selecting influential seeds depend on measures of global centrality within the structure of the social network – they select users that are central in the entire network according to some centrality measure (e.g., Eigenvector centrality). In this paper a new targeted approach to viral marketing is proposed that exploits prior knowledge about the potential market and uses local centrality scores to identify seeds that have high chances of reaching and activating many users in the potential market. The performance gained by the proposed approach is investigated with an experimental evaluation that uses data from real social networks. The results show that targeted approach outperforms existing, global centrality based methods. It is also shown that the relative performance of the targeted approach improves in the case where the majority of users are indifferent (or negative) to the viral marketing campaign.

1. Introduction

People are influenced by others. The social ties they have, strong or weak, influence their decisions, actions, and opinions. When choosing a product or service, people look for recommendations. They are more likely to trust a personal recommendation given by a friend or trusted acquaintance (Jurvetson 2000), a fact that increases the chances of them purchasing the product or service in question (Wang and Chang 2013). This is the reason why viral marketing in social media can become more powerful than third-party advertising, since it conveys an implied endorsement from social connections, that can be either friends, celebrities or any trusted person (Jurvetson 2000). Viral marketing relies on people to pass along a marketing message, which produces positive word-of-mouth and product awareness. With the fierce competition that e-commerce companies face today, a low-cost high-impact marketing strategy, such as viral marketing, can become valuable. The idea behind viral marketing (as defined by Bonchi et al. 2011) is that by targeting the influential users (called seeds) of a social network, it is possible to activate a "chain-reaction" of influence.A large percentage of the network can be reached driven by word-of-mouth, incurring only a small marketing cost for the initial seeds. Seed selection alone however, does not guarantee that the viral marketing campaign will spread – success is also dependant on the message being propagated and its viral properties (Dobele et al. 2005, Scharl et al. 2005).

Viral marketing has already proven to be an effective marketing technique in social networks (Leskovec et al. 2007) and has attracted a lot of attention in recent research (see this excellent survey Bonchi et al. 2011 and citations therein). Compared to more traditional ways, such as targeted marketing, viral marketing avoids the expense of contacting all members of a target group. In contrast, only a small number of influential seeds need to be contacted in order for the message to spread widely. Despite these advantages, existing approaches to viral marketing present some shortcomings which are described below (Section 1.1) and comprise the motivating factors of this study.

1.1. Need for a targeted approach

A common practice in viral marketing campaigns is to select the most influential users as initial seeds (Hinz et al. 2011). Such users are identified based on the *centrality* they have within the structure of the social network (Newman 2009). The widely used centrality scores, such as betweenness and Eigenvector centrality, select seed users that are *globally* central, i.e. their centrality is high with respect to the entire social network. The premise of this approach is that a few globally central users will be able to reach and activate most of the remaining users in the network.[1] However, this includes the implicit assumption that all remaining users have an equal interest in the product or message and therefore an equal

* Corresponding author. Tel.: +49 841 93721020.
E-mail address: anastasia.mochalova@ku.de (A. Mochalova).

[1] Later in the paper the term 'activation' refers generally to the adoption of the objective of the viral marketing campaign which can be expressed by buying a product, retrieving information about an item, etc.

http://dx.doi.org/10.1016/j.elerap.2014.06.002
1567-4223/

chance of being activated. In short, existing approaches (e.g., Hinz et al. 2011) select globally central seeds and assume that all remaining users have an equal activation probability.

The above approach is effective when no prior knowledge is available, however, in several cases information about past activities of users can be available. For instance, purchased products or products rated by users can be used to identify a set of users with demonstrated positive attitude towards a brand or similar product. Such users form a potential market because they are likely to become adopters in the current viral marketing campaign whereas the majority of remaining users are likely to stay indifferent or even be negative. Examples of potential markets are users who would consider buying the newly released album of a liked musical artist or sequel to a well played game.

Users in the potential market tend to be located together or in several clusters (Hogg 2010). Such self-clustering is described by the 'homophily' principle – the tendency for people to bond with others holding similar interests to themselves (Peres et al. 2010). Failing to use prior knowledge of the network can severely compromise the effectiveness of a viral marketing campaign as these clusters may be located far from the globally central users. Using knowledge of the network, these clusters can be targeted by a local centrality score.

Consider the example of the previous concept by observing a social network presented in Fig. 1. Each node represents a user and links between them represent their social ties. Chequered nodes represent users in the potential market. In this example, for simplicity of illustration, the potential market consists of only one group; in general, however, it might be split into several groups. The black node represents a user that is considered central according to a global centrality score (in this case, Eigenvector centrality (Borgatti 2005)) and will be selected as seed. If the viral marketing campaign is started from this globally central user, the viral process may stop before the message reaches any users in the potential market, because this seed is far away from it. This is supported by recent findings showing that influence propagation in viral processes in social networks often happens only within close proximity of the seeds (Cha et al. 2009). Therefore, globally central seeds can cause viral marketing to loose its focus and become sup-optimal. Required, therefore, is a targeted approach to viral marketing which selects seeds according to their *local* centrality with respect to the potential market. Such locally central seeds are more probable to reach and activate users in the potential market; e.g., the hatched node in the example in Fig. 1 is the seed selected using MFP (see Section 3.3 for mode details on MFP).

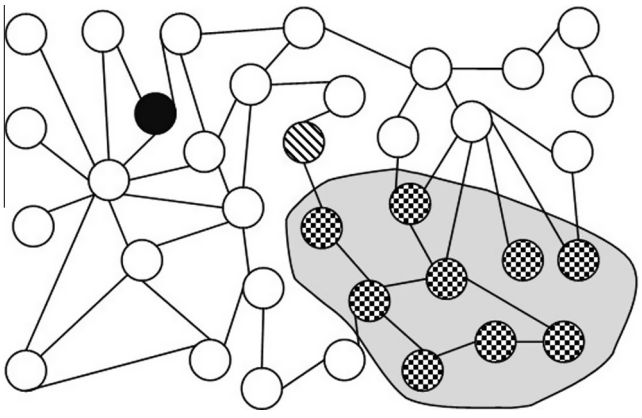

Fig. 1. Example of a social network. Black node denotes the globally central node; chequered nodes denote the potential market; hatched node denotes the node central w.r.t. the potential market.

1.2. Contribution

In this paper, a targeted approach to viral marketing is proposed. This approach combines the advantages of viral and traditional targeted marketing. Namely, it combines the efficiency that viral marketing has in social networks compared to traditional targeted marketing, by avoiding contacting each user in the potential market. However, similar to traditional targeted marketing, this approach exploits knowledge of the potential market to improve the effectiveness of the viral marketing process which otherwise does not exploit knowledge of a potential market and may become sub-optimal (see discussion in Section 1.1).

The main contributions of this paper are:

- Investigation is done into the local centrality scores and their application to viral marketing for the purpose of selecting seeds w.r.t. a given target in the form of users in a potential market.
- Performance of local centrality scores for the problem of seed selection in viral marketing is examined experimentally and compared to global centrality scores by using data from real social networks.
- The results demonstrate that the proposed targeted approach to viral marketing compares favorably to existing approaches that are based on global centrality scores.

The rest of the paper is organized as follows: In Section 2 the related work is presented. In Section 3 a formal description of the targeted approach to viral marketing is given and seed selection based on local centrality scores is introduced. The design of the experimental evaluation is presented in Section 4 which is followed by presentation of experimental results and their discussion in Section 5. The paper is concluded by Section 6.

2. Related work

Viral marketing can be stated as an influence maximization problem that is concerned with selecting the set of seed users in a social network who will initiate the spread of a piece of information and will cause the largest possible number of activations among remaining users (Kempe et al. 2003). The number of seed users is predefined and represents the cost to initiate the spread of information, i.e. the larger the seed set size, the higher the cost. According to this formulation, it has been shown that the influence maximization problem is NP-hard which means that finding the seed set that will achieve the maximal possible number of activated nodes requires execution time that grows exponentially with the size of the network (Kempe et al. 2003). Since real social networks are very large, finding the optimal seeds requires prohibitively large computational cost. To address this problem a greedy hill-climbing algorithm has been proposed that finds the seed set which activates at least 63% the number of nodes an optimal seed set would activate (Kempe et al. 2003). The popularity of the paradigm proposed in (Kempe et al. 2003) resulted in the development of several extensions that scale-up seed selection in very large social networks (Chen et al. 2009).

The aforementioned approaches exploit knowledge about pairwise influence between users called influence factors. A priori knowledge of influence factors is not possible in several real-world applications because influence is intangible and hard to measure. A data-driven approach has been proposed recently by several authors (e.g., Goyal et al. 2011, Li et al. 2010) which estimates influence factors based on actions previously performed by users, such as buying a product, writing a review, or joining a community. This approach can be used effectively in the case when the recorded actions are relevant to the current viral marketing