A Comprehensive Framework for Identifying Viral Marketing’s Influencers in Twitter

Faezeh Sadat Gohari¹, Shahriar Mohammadi²

Abstract

Viral marketing can lead to extensive knowledge of marketing campaigns across customers with lower costs. The important point of viral marketing is targeting the subset of customers that can influence on others. Such customers enhance the efficiency of a marketing campaign by maximizing propagation of viral message throughout the network. According to increasing the importance of the Twitter network for marketing efforts in recent years, the aim of this work is to identify the best influential individuals for the efficient performance of viral marketing campaigns in this network. Recent works on Twitter reveal the lack of a comprehensive framework for differentiation of influencers in viral marketing. Our qualitative research aims at the synthesis of results and theories from previous studies in a new comprehensive framework. The paper first provides a detailed review on previous works about the influence and diffusion of information on Twitter. Second, according to the important features of viral marketing’s influencers, it proposes a comprehensive framework for evaluating these features in terms of Twitter functions. This framework concentrates on all of the important factors for identifying viral marketing’s influencers. So, the most worthy twitterers with highest marketing value can be identified effectively based on our proposed framework.

Keywords: Viral Marketing (VM), Online Social Networks, Twitter, Influencers.

¹Master of Science, Department of Industrial Engineering, IT Group, K.N. Toosi University of Technology, Tehran, Iran. Tel: 0912-6088175 E-mail: fgohari@mail.kntu.ac.ir
²Assistant Professor, Department of Industrial Engineering, IT Group, K.N. Toosi University of Technology, Tehran, Iran. Tel: 0912-2978284 E-mail: mohammadi@kntu.ac.ir
1. Introduction

Online social networks (OSN) provide the right environment for social interaction among individuals. The growing popularity of OSN in recent years has led to the variety of marketing efforts from different campaign in these networks (Kim and Han, 2009). Network-based marketing is a collection of marketing efforts that take advantage of links between customers in the network to increase sales (Hill et al., 2006). One type of network-based marketing is called viral marketing (VM). The essence of VM is laid in the word-of-mouth (WOM) phenomenon. Viral marketing is electronic WOM that often uses social media applications for transmitting some kind of marketing message in an exponentially growing pattern. Viral approaches can lead to extensive knowledge of marketing campaigns across customers with lower costs. A successful VM epidemic entails three conditions: giving the right message to the right messengers in the right environment (Kaplan and Haenlein, 2011).

First, the right message is the message that is memorable, sufficiently interesting (Kaplan and Haenlein 2011) and has a unique power that encourages individuals to pass it on (Dobele et al., 2007).

Second, right messengers are subsets of customers who influence others (Flynn et al., 1994) and have a higher propensity to propagate the marketing message throughout the network based on their special personal and social characteristics (Gladwell, 2002). According to the classical laws of concentration, 20% of the messengers have potential to carry 80% of the propagation (Kaplan and Haenlein, 2011).

Finally, right environment is the communication channel that has to be chosen properly with considering the type of message and target customers (Kaplan and Haenlein, 2011).

One of the most popular OSN is Twitter network which is also a marketing platform based on its million users across the world (Milstein, 2008). The relationships between twitterers are “following” relationship. Every user can follow desired twitterers without the need of any permission. “Following” relationships are not necessarily bi-directional. That means followee may not follow his followers. Every twitterer can send tweets limited to 140 characters. Important actions that users do on twitter are: reply, retweet and mention – reply is a special kind of mention.
In this study, we concentrate on choosing right messengers (influencers) in VM process, and express our approach in the context of the Twitter network for the purpose of enhancement in efficiency of VM campaigns in Twitter.

The main contribution of this work is twofold. First, we provide a detailed state-of-the-art review on recent works specifically about the influence and diffusion of information on Twitter. Second, according to the important features of VM’s influencers, we propose a comprehensive framework for expressing and evaluating these features in the form of available metrics and algorithms of Twitter. Based on the proposed framework, we can investigate the overall marketing influence value for each twitterer with combining all of these features together.

The rest of this paper is organized as follows. The following section introduces our research methodology. Section 3 describes viral marketing, VM in online social networks, influencers in VM process and reviews the related works on Twitter. Section 4 demonstrates our proposed framework in details and Section 5 presents original findings of the work. We explain managerial implications in Section 6, and in Section 7 we present our conclusions and outline future lines of research.

2. Methodology

This paper uses a meta-analysis method and combines results from different studies, in the hope of constructing a comprehensive framework based on previous results. For this purpose, We studied related papers about influence and diffusion of information from year 1990 to 2013. We refined the related papers in terms of content, methodology and results. After the refinement process, we chose papers presenting new aspects of evaluating influence in a viral campaign. Then, we extracted the important features of VM’s influencers and presented a new analytical framework for evaluating these features in terms of Twitter functions.

3. Related Works

In this section, we introduce viral marketing, VM in online social networks and influencers in VM process. Also, we review the related works on Twitter.
3.1 Viral Marketing (VM)

The term viral marketing was coined by Rayport (1996). The concept of VM describes any strategy that encourages would-be advocates to pass on a marketing message to others. Like viruses, VM strategies take advantage of exponential growth to explode the message to thousands or millions.

With the introduction of electronic social media and its effect on facilitating interactions between companies and potential buyers, VM has gained huge popularity (Gruen et al., 2006) and several world-class companies and brands have inclined to it (Dobele et al., 2007).

3.1.1 Viral Marketing In Online Social Networks

Social networks play an essential role as a medium for the diffusion of information among people (Kempe et al., 2003). Information diffusion is a process whereby a new idea or action spreads widely through communication channels (Rogers and Rogers, 2003). Maximizing the spread of information depends on proper selection of initial target nodes (influencers). Importance of information diffusion circumstance in OSN has led to variety of empirical papers based on different networks such as Facebook, Flickr and so on (Aral et al., 2009; Cha et al., 2009; Sun et al., 2009).

In recent years, OSN became a new medium for VM. Viral marketing exploits the potential of social networks for affecting product adoption (Rogers and Rogers, 2003) by encouraging customers to share product information with each other (Leskovec et al., 2007).

Until now several researches have focused on information diffusion in OSN from the perspective of VM and product adoption: Palazuelos and Zorrilla (2012) used the FRINGE algorithm—an approach for detection of overlapping communities in graphs—in their work and measured the degree of influence of a node in Flickr network. Based on the result of this work, marketing campaigns can take advantage of the leadership structure of a social network for spreading a new product widely throughout the network. Guo et al., (2011) by studying the social interactions on Taobao network, verified that information passing exists on this network and its success rate depends on 4 variables: communication strength, time difference, product price, and product category. The results of this work can provide guidance to VM campaigns for promoting product
recommendation and adoption among online users. Stonedahl et al., (2010) in order to encourage faster adoption of product throughout the entire network defined a strategy space for customer seeding. They do this by weighting a combination of network features such as average path length, clustering coefficient and degree. They evaluate strategy effectiveness by simulating adoption on an agent-based model and empirically examined it on Twitter friendship data. Bhagat et al., (2012) studied the problem of maximizing product adoption which is the main goal in the VM. They proposed an intuitive model for product adoption maximization and evaluated it on two kinds of OSN in the context of movie and music. Their observations show that it is important to distinguish product adoption from influence because an influenced user may not adopt the product himself.

3.1.2 Identifying Influencers In VM

In previous works, different titles such as opinion leaders, influential people and key players are used to refer to an influential group in social networks (Momtaz et al., 2011). These are a small portion of people who have a lot of influence on the other people’s opinions and can act as filters or intermediaries to accelerate or stop the spread of information (Katz and Lazarsfeld, 2006).

In previous studies, a variety of influencer’s characteristics have been identified. But we can categorize the subset of them that have been more emphasized and have more effect on successful implementation of VM, in two different groups: personal and social characteristics. Personal characteristics are individual features. One of the most important personal characteristics is knowledge (Katz and Lazarsfeld, 2006; Li and Du, 2011). For example, a handphone manufacturer can engage those people who have knowledge about IT gadgets to potentially influence more people (Weng et al., 2010). Strengthmotivation for sharing information with others is another important feature of influencers (Gruen et al., 2006; Phelps et al., 2004). Beside this, influencers are more innovative(Li and Du 2011), and they would like to accept new things (Zhang and Dong, 2008).

Social characteristics are associated with the community that each person communicates with. One of the most important social characteristics is the social location—location of a person (node) in the network (Katz and Lazarsfeld, 2006; Zhang et al., 2010). This metric is used in the social network analysis (SNA). SNA has been the basis for many forms of marketing efforts (Lam and Wu, 2009). The most important and frequently used concepts for identifying influencers by SNA is network centrality (Borgatti, 2006) and one of the most frequently used centralities is network centrality – the number of direct connections of one node with others (Wasserman and Faust, 1994). Actually, the more connection with more persons, lead to more influences. The next
important feature is high social activities (Li et al., 2011; Rogers and Rogers, 2003). Two other considerable features are high social status and reputation (Rogers and Rogers 2003) and high similarity between influencer and follower (Li and Du, 2011; Zhang et al., 2010).

Table 1 shows the important factors for identifying influencers in VM that we discussed above.

3.2 Review Of Related Works On Twitter

Increasing the importance of information diffusion on OSN has led to several researches in the context of the Twitter network.

Huberman et al. (2008) explains that trying to rely on WOM to spread an idea in OSN entails finding the hidden network that consists of people who actually interact with each other. They introduced the concept of ‘friend’ and proved that two different networks exist: a dense and visible network consists of followers and followees, and a sparse and hidden network of actual friends. The latter is more influential because users with many actual friends tend to post more updates than users with few updates.

<table>
<thead>
<tr>
<th>Personal characteristics</th>
<th>Knowledge</th>
<th>(Katz and Lazarsfeld 2006; Li and Du 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Motivation</td>
<td>(Gruen et al. 2006; Phelps et al. 2004)</td>
</tr>
<tr>
<td></td>
<td>Innovative</td>
<td>(Li and Du 2011; Zhang and Dong 2008)</td>
</tr>
<tr>
<td></td>
<td>Tendency to accept new things</td>
<td>(Zhang and Dong 2008)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social characteristics</th>
<th>Social location</th>
<th>(Katz and Lazarsfeld 2006; Zhang et al. 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Social activities</td>
<td>(Li et al. 2011; Rogers and Rogers 2003)</td>
</tr>
<tr>
<td></td>
<td>Social status and reputation</td>
<td>(Rogers and Rogers 2003)</td>
</tr>
<tr>
<td></td>
<td>Similarity between influencer and follower</td>
<td>(Li and Du 2011; Zhang et al. 2010)</td>
</tr>
</tbody>
</table>

Table 1. Important Factors For Identifying Influencers In The VM
Leavitt et al. (2009) define influence on Twitter as the potential of a user’s action to initiate a further action by others. In this definition, actions mean tweet, reply and retweet; And according to the intended purpose of a user in the network are categorized into two different groups: content-based and conversation-based.

Work of Cha et al. (2010) includes comparison between three different measures of influence—indegree (i.e. the number of followers), retweets, and mentions as well as investigation of dynamics of user influence across topics and time. They showed that each metric leads to a different ranking and the most followed users are not necessarily influential on the other measures.

Kwak et al. (2010) compared three different ranking by: number of followers, page-rank, and retweets. They concluded while the first two rankings are similar, they are different from the latter. Also they focused on information diffusion through retweet trees and found that is no matter how many followers a user has, any retweeted post is likely to reach an average of 1000 users. This shows the power of retweeting in the diffusion of information in the network.

Weng et al. (2010) found high reciprocity (e.g., mutual following relationship) in a sample of twitterers and showed that its reason is homophily phenomenon. The homophily in social networks means that similarity between members is based on different characteristics (McPherson et al. 2001).

According to homophily in Twitter, the authors claim that topical similarity is an important criterion for measuring the influence. Therefore, they propose an extended PageRank algorithm, called TwitterRank to measure the influence of users on Twitter. TwitterRank measures the influence from two dimensions: the topical similarity between users and the link structure. The comparison between this new algorithm and previous related algorithms, namely PageRank and Topic-sensitive PageRank, verifies its superiority.

In a simple approach proposed by Anger and Kittl (2011), social networking potential (SNP) of each user is calculated by averaging the Retweet and Mention Ratio (i.e. the amount of user’s tweets that initiate a further action by others divided by the total amount of his tweets) and Interactor Ratio (i.e. the number of user’s followers that mention him or retweet his posts divided by the total amount of his followers). The results of evaluating this approach on Austria’s Top 10 twitterers indicate some similarities to the Klout (an online rating service) score, but for the most
followed users, the differences between Klout and SNP are noticeable. These differences are caused by the omission of the amount of followers in SNP, in spite of its large emphasis in Klout. Bakshy et al. (2011) analyzed the attributes and relative influences of twitterers by tracking diffusion events that occur on the Twitter follower graph. Since the focus of this work is on the predicting influences, they considered all users, not merely the most influential. They tested several hypothetical marketing strategies and concluded that individual-level predictions of influences are relatively unreliable. Therefore, they suggested that marketers should not seek for exceptional individuals. Indeed, in order to harness WOM diffusion reliably, they should target many potential influencers.

The summary of discussed researches on Twitter is presented in Table 2.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Descriptions</th>
<th>Results</th>
</tr>
</thead>
</table>
| (Huberman et al. 2008)  | Consideration of actual interactions, attempting to find hidden network between friends | -- Twitterers have a very small number of friends compared with declared contacts  
-- WOM outperforms in the sparse hidden network between friends |
| (Leavitt et al. 2009)   | Attention to the relationship between user and content on the platform, measuring the influence by analyzing a comprehensive set of conversation-based or content-based actions on Twitter | -- News outlets in Twitter are more influential in republishing content  
-- Celebrities in Twitter are better at making conversation rather than providing retweetable content |
| (Cha et al. 2010)       | Comparison between three measures of influence: indegree, retweets, and mentions, analysis of the dynamics individual’s influence across topics and time | -- Indegree alone can’t reveal the actual influence of a user.  
-- Most influential users can be influential over ranges of topics  
-- Retweet’s influence is laid in the content value of tweets; while, mention’s influence is laid in the name value of a user |
| (Kwak et al. 2010)      | Analysis of topological characteristics of Twitter, comparison between three measures of influence: indegree, PageRank, and retweets, study of information diffusion by constructing retweet trees | -- Existence of a gap in the influence inferred from indegree measure and that from the popularity of tweets  
-- Retweeting gives every user the high power to spread information widely |
| (Weng et al. 2010)      | Attention to high reciprocity in following relationships, proposition of a new TwitterRank algorithm for measuring the topic-sensitive influence of users | -- Reporting homophily on Twitter for the first time  
-- Superiority of new algorithm over related ones |
| (Anger and Kitzl 2011)  | Proposition of a simple and basic approach for measuring SNP for every twitterer, calculation of SNP as the average of Retweet and MentionRatio and Interactor Ratio | -- Similarity between this simple approach and Klout  
-- Differences between them are obvious for the most followed users |
| (Bakshy et al. 2011)    | Measuring influence based on the size of the entire diffusion tree associated with each event, predicting the influence as a function of attributes of individual seeds, evaluating the cost-effectiveness of a family of hypothetical marketing strategies | -- Predictions of which particular user will generate large cascades are unreliable.  
-- Under many circumstances, the most cost-effective performance can be satisfied through “ordinary influencers” |

Table 2. The Summary Of Related Works On Twitter
4. The Proposed Framework For Evaluating Total Marketing Influence Value

The concentration of previous works on Twitter is not specifically on the viral marketing phenomenon. Each work measures and compares the influence from different aspects. But, a successful viral marketing campaign needs to consider a combination of all important factors for identifying influencers. We improve the state-of-the-art by proposing a new framework for measuring all of the important social characteristics of influencers (section 3.1.2) in the form of available metrics and algorithms of Twitter. Using proposed framework, we can evaluate the influence of every twitterer from different dimensions and consequently investigate the total marketing influence value (TMIV) for every user with the combination of all of these dimensions together. Then, we can recognize the best influencers for successful implementation of viral marketing on Twitter. Fig1 illustrates our comprehensive proposed framework. In the next sections, we explain it in detail.

4.1 Evaluating Social Location And Similarity Between Influencer And Follower

One of the most famous metrics for evaluating social location in Twitter is indegree or the number of followers (equivalent to degree centrality). But previous studies show that it is not an accurate measure (Brin and Page, 1998). PageRank covers the shortcomings of indegree by taking the link structure of the whole network into account. Therefore, we also consider linking structure of the whole Twitter network for expressing social location of influencers in our framework. As we previously said, the work of Weng et al. (2010) improves over PageRank by considering both the link structure and topical similarity among twitterers. They proved that high degree of reciprocal relationship is a strong indicator of similarity between users. In the other words, two users follow each other because they have a similar topic interest. Since topical similarity is a good indicator of influence, we use it for expressing similarity between influencer and his followers in our VM framework.
## Important factors for identifying influencers in viral marketing

<table>
<thead>
<tr>
<th>Social factors</th>
<th>Related metrics in Twitter</th>
<th>Method of evaluation</th>
<th>TMIV for every twitterer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social location</td>
<td>Link structure of the whole Twitter network</td>
<td>TwitterRank algorithm</td>
<td>Nonlinear combination of factors based on BPNN</td>
</tr>
<tr>
<td>Similarity between influencer and follower</td>
<td>Topical similarity amongst twitterers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social activities</td>
<td>Tweet and reply</td>
<td>Activeness-based value formula</td>
<td></td>
</tr>
<tr>
<td>Social status and reputation</td>
<td>Retweet and mention</td>
<td>Retweet and Mention Ratio</td>
<td></td>
</tr>
<tr>
<td>Personal factors</td>
<td></td>
<td>Analysis of filled questionnaires by twitterers</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1. The Proposed Framework For Evaluating TMIV For Every Twitterer**

In our framework, we apply TwitterRank algorithm (Weng et al., 2010) to evaluate the user’s rank based on social location (link structure) as well as similarity among influencer and followers (topical similarity). This algorithm uses a formula for twitterers general influence which is measured as an aggregation of the topic-specific TwitterRank in different topics. In the calculation of Topic-specific TwitterRank, these notions are considered: First, the most influential user among the followees of a follower $i$, is the one who has the higher portion of tweets that are read by follower $i$. Second, the influence of a followee on follower $i$ is also related to the topical similarity between them. The similarity between two users in a specific topic can be evaluated as the difference between the probabilities that the two are interested in that topic. Therefore, in our framework we
use general influence formula in TwitterRank algorithm in order to calculate a general rank from the combination of both mentioned dimensions for twitterers.

4.2 Evaluating Social Activities

An influential user should have activeness in his community (Li et al., 2011). Li et al. (2011) for discovering influencers for marketing in the blogosphere, claimed that a more active blogger would show more interest to engage in conversation with other users. They used two metrics for evaluating activeness-based value of bloggers: “Content posts” and “Comment replies”. They calculated the activeness-based value of a blogger by summing of his content posts and comment replies during a period of time.

In our framework we can extend this definition in terms of twitter actions. The equivalent metrics for “content posts” and “comment replies” in Twitter are respectively “tweets” and “replies” from the active user to others. Therefore, we can also count the activity records of twitterers during a period of time in order to evaluate their influential rank from the social activities dimension.

4.3 Evaluating Social Status And Reputation

An influencer should be well respected and reputable in his community. This respect can be associated with the higher social status (i.e. income, occupation, and education and so on) (Rogers and Rogers 2003). In a virtual community, in contrast to traditional communities, factors such as income and occupation cannot be completely good indicators of social status of users. For this reason, in our framework we evaluate an influencer’s social status and reputation by the amount of attention from other twitterers to the content value of his tweets and his name value. Indeed, the more a twitterer publishes reliable tweets and engages others in conversation; he acquires the most attention from other twitterers. This leads to a higher reputation for him.

In the process of calculating market influence value of bloggers, Li et al. (2011) used two metrics: “Comments” and “citations”. Similarly, here we use two equivalent metrics on Twitter: “mention” and “retweet”. These metrics are good indicators of the amount of attention from the other twitterers to one user. Broader propagation of one’s tweets throughout the network by retweeting indicates the ability of that user to generate reliable content with pass-along value (Cha et al., 2010); and this leads to positive knowledge of that user in other’s minds. Also, the more a twitterer’s name is mentioned by others, indicates the ability of that user to engage others in a
conversation (Cha et al., 2010) and this likewise leads to a higher recognition of that user in the community.

Based on these metrics, we can use Retweet and Mention Ratio, in order to evaluate the influential rank of twitterers from the social status aspect.

4.4 Evaluating Personal Factors

For the purpose of evaluating personal characteristics of twitterers include knowledge, motivation, innovation and tendency to accept new things; we suggest the use of online questionnaires. We can adjust these questioners based on Likert scales and send them to the users. Then we analyze the filled questioners for twitterers in order to evaluate their influential rank from personal dimension.

4.5 Total Marketing Influence Value (TMIV)

Similar to Li et al. (2011), for calculating TMIV for every twitterer we can use the nonlinear combination of different influential ranks that have been calculated for different discussed dimensions. Due to the complexity of human behaviors in the network, Li et al. (2011) used a back-propagation neural network (BPNN) to deal with the uncertain weighting problem between different parameters for forecasting the final influence value.

In this final stage, after calculation of TMIV for twitterers, we can choose the subset of them that have the highest value. These are the best influential twitterers with the consideration of all important features for influencing in VM process.

By adopting this framework, marketing campaigns can differentiate influential twitterers in the viral marketing domain to reduce the marketing costs and promote a product/service with ease in this online social network.

5. Findings / Discussion

Recent works on Twitter reveal the lack of a comprehensive framework for differentiation of influencers in viral marketing. Previous works missed some important factors for identifying such individuals, because the main focus of most of them was not on viral marketing specifically. In our
comprehensive proposed framework, we concentrate on all of the important factors and combine them in terms of Twitter functions. The proposed framework can differentiate the best influencers for viral marketing campaigns in Twitter with consideration of the both important social and personal characteristics of influencers, which therefore will reduce the marketing cost and increase the knowledge of the campaign across customers.

6. Managerial Implications

This study helps viral marketing campaigns to effectively distinguish influencers in Twitter network; so they can reduce their marketing costs and promote a product/service with ease in this popular network. Our framework (Fig 2) has direct implications in design of viral marketing. First, it provides a clear relationship between important social features of influencers and different metrics of twitter which can help the viral campaign for evaluating such features. Second, marketers can use the TMIV model in order to well distinguish the first group of infected people with considering all of the social and personal factors together. Therefore, the viral message can spread widely throughout the network.

7. Conclusion And Future Work

The focus of this work was on identifying the best influencers in order to help succession in VM implementation in Twitter network. In this regard, we provided two major contributions. First, we reviewed the recent works about influence and diffusion on Twitter in detail. Second, according to the important features of VM’s influencers, we categorized them into personal and social groups and proposed a comprehensive framework. Based on this framework, firstly we can express and evaluate social features in terms of available metrics and algorithms of Twitter. And secondly we can investigate the total marketing influence value for each twitterer with combining all of the important social and personal factors together.

As a future research we suggest the empirical experiment of this framework on real data of Twitter network and compare it with similar related works. Also, it would be interesting to extend this framework to other popular online social networks such as Facebook with considering the metrics for evaluating TMIV for their users.
8. Authors Short Biographies Of Authors

[Faezeh S. Gohari is a postgraduate student of electronic commerce at K.N. Toosi University of Technology. She received her bachelor’s degree in computer engineering from Elm O Farhang University, Tehran-Iran. Her current primary research interests are in online social network analysis, e-marketing and recommendation systems in e-commerce]

[Dr. S. Mohammadi is a former senior lecturer at the University of Derby, UK. He also used to be a Network consultant in the UK for more than fifteen years. Now he serves as an assistant professor and chairman of the Information Technology department at K.N. Toosi University of Technology in Tehran, Iran. His main research interests are in the fields of Networking, Data Security, Network Security, e-commerce and e-commerce Security]

References


