Opinion leaders selection in the social networks based on trust relationships propagation

Samad Mohammad Aghdam, Nima Jafari Navimipour*

Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Received 26 November 2015; revised 13 February 2016; accepted 14 February 2016
Available online 6 April 2016

Abstract

Today, social networks become very popular and include a wide range of users. In these networks, some users have a great influence ratio to other users who are called opinion leaders. They can use their influence on many issues, such as political, economic, education, social, etc. In this paper, we propose a new framework to select the opinion leaders in online communities. The framework uses the trust relationship between the users and evaluates the total trust value (TTV) of primary opinion leaders between other users to select the highest of them. According to the obtained results, the proposed framework in comparison of top in-degree method, top out-degree method, top centrality method and hybrid IO-degree method provides better results in the social network marketing (SNM) campaigning.

Keywords: Opinion leader; Social networks; Trust; Similarity; Sociality

1. Introduction

Online communities are becoming more and more popular as it is shown by the growth on the number of their members [8]. Also, the popularity of social networks has led to new ideas and caused to use that in different cases, like combining with Cloud/Grid computing [17,18,20,22,24,28], Peer-to-Peer networks [19], Expert Cloud [2,17,18,21,23], cooperative games [3] or in therapeutic cases [1]. All the registered users in these communities are not enough honestly or reliable. When information is to be shared or a user is interested in knowing the opinions of others, uncertainty might be a problem. It would be desirable to have a tool that aids users in overcoming this uncertainty, where the trust plays an essential role [8].

Online public opinions closely are connected with the various contradictions and sensitive issues in the social transformation period. In the process of public opinion transmission, any Internet users expressed their ideas who can be familiar with computers operating in a particular way, such as participating in the topic discussion or social networks Ref. [32]. Opinion leaders play an important role in improving communication and encouraging group members in order to have a greater level of information exchange Refs. [4,33]. Their superior status, education, and social prestige enabled them to influence followers, which

* Corresponding author. Tel.: +98 9144021694; fax: +98 4134203292.
E-mail address: jafari@iaut.ac.ir (N. Jafari Navimipour).
Peer review under responsibility of University of Kerbala.

http://dx.doi.org/10.1016/j.kijoms.2016.02.002
2405-609X/© 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of University of Kerbala. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
are element key to making the connected community and make the best group performance. With attention to prominent and effective diffusion strategy, using opinion leaders has been studied in many fields such as anthropology, sociology, and business Refs. [6,34]. Also, opinion leaders have demonstrated significant impact in political participations and discussed to sustain any prolonged political action Ref. [35]. Opinion leaders have identified a prerequisite for guiding and interfering public opinion on the Internet due to the identification of opinion leaders is very important and meaningful [14].

In this paper, we introduce a framework to increase the accuracy of opinion leaders' selection method in the social networks. At the first step, we have removed the duplicate comments, self-trust statements, and troll's opinions, then try to select the opinion leaders that have a maximum amount of trust between users.

The rest of this paper is organized as follows. Section 2 discusses the related works; Section 3 presents the research framework; Section 4 presents discussions and comparison of the proposed mechanism with the existing methods; in Section 5, concludes the paper with some suggestion for future research directions.

2. Related works

In recent years, the identification of opinion leaders is a subject which has been a great interest. So far, some researchers were studied on opinion leaders where first of them is situated within the field of sociology. Numerous studies such as [5,10,16,26,29,30] have been conducted in order to understand the concept of opinion and characteristics of leaders distinguishing them of their followers. The studies on opinion leader identification can be divided into two parts: (1) link-based opinion leader identification methods that consider the social interactive structure of the network; and (2) mixture of opinion leader identification methods that combine the social link information with semantic-based information embodied in documents [9,12,15,27].

As an example of link-based opinion leader selection, Carson, Tesluk et al. [4] have identified opinion leaders with an excellent edge rank algorithm based on excellent network theory, applied it to rank excellent edges and used the ranking result to identify opinion leaders in opinion excellent network model. Alternatively, Li, Ma et al. [13] have proposed an improved combined framework identifying the opinion leader in online learning communities, which ranked opinion leaders based on four distinguishing features: expertise, novelty, influence, and activity. Furthermore, the performances of opinion leaders were further investigated in terms of longevity and centrality.

In another study, Cho, Hwang et al. [6] have found that those opinion leaders are the optimum marketing choice in terms of diffusion speed and maximum cumulative number of adopters, used a social network method and threshold model and conclude that opinion leaders with a high sociality and centrality (users who have high degree of connections to other people) are the best ones for fast diffusion.

In case of SNM, about 78% of the trust of customers to the social network communities are based on opinion leaders' recommendations for products and services [25]. So, how to identify the opinion leaders effectively is the key to raise sales and brand awareness. In sales and marketing field, some studies focus on developing various indexes such as in-degree, out-degree, betweenness and closeness in social network analysis (SNA) to identify opinion leaders. These studies are interested in identifying opinion leaders who will forward be marketing messages to other users via their trust and distrust networks. Moreover [11] Table 1 provides side by side comparison of the reviewed methods.

Trust can assist entities making decisions before establishing collaborations. It is desirable to simulate the behavior of users as in social environments where they tend to trust users who have common interests or share some of their opinions, i.e., users similar to them. In this direction Ref. [36] in a research introduce the concept of context similarity among entities and derive a similarity network. Then, it defines a trust model that allows to establish trust along a path of entities. Also, many other methods for evaluating trust in the social community are proposed, for example, in Ref. [11] three methods (Knowledge Score, Matching coefficient, Jaccard coefficient) which obtain trust degree between users in social networks are introduced. Which knowledge score reflect both the evaluation score and the intensity of a relationship between two users and two other methods are acting based on the structural and social similarity between two users.

3. Research framework

For selecting opinion leaders in the social network, at first, we must have access to relationships between
users and data sets of the social network. However, it is possible that some of these data may not be accurate. So, we should first remove incorrect data and then obtain trust relationships to select opinion leaders. The research framework is illustrated as a graphical form in Fig. 1.

At the first step, preliminary data filtered to obtain accurate results; then, at the second step, opinion leaders identified, which step one divided into three parts and step two divided into two parts. The rest of this section comprehensively defined these steps.

3.1. Data filtering

To avoid of error causes and losing of accuracy for the results, the data were filtered. Filtering of data removes self-trust statements, duplicate comments and troll's comments by employment three components as follows.

3.1.1. Removing self-trust statements

The order of self-trust statements are those comments which a person gave to itself (e.g., user “X” issues “Y” time’s self-trust statements), so must remove this cases. Algorithm 1 shows how to remove the self-trust statements. For understanding the algorithm, we must know a comment structure. The structure of a comment is represented in Table 2.

<table>
<thead>
<tr>
<th>Algorithm 1. Removing the self-trust statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: ( P_x ) is an array list, include ( X ) part of comments.</td>
</tr>
<tr>
<td>2: ( P_y ) is an array list, include ( Y ) part of comments.</td>
</tr>
<tr>
<td>3: ( P_z ) is an array list, include ( Z ) part of comments.</td>
</tr>
<tr>
<td>4: STSC includes a number of self-trust statements.</td>
</tr>
<tr>
<td>5: ( \text{for} \ ( N = 0 ; \ N &lt; = \text{length of comments} ; N++) )</td>
</tr>
<tr>
<td>6: ( \text{if} \ (P_x N == P_y N) \ \text{then} )</td>
</tr>
<tr>
<td>7: remove ( P_x, P_y, P_z N )</td>
</tr>
<tr>
<td>8: ( N-- )</td>
</tr>
<tr>
<td>9: STSC ++</td>
</tr>
<tr>
<td>10: \text{end if}</td>
</tr>
<tr>
<td>11: \text{end}</td>
</tr>
</tbody>
</table>
3.1.2. Removing duplicate comments

Another irrelevant data duplicate comments, which includes duplicate comments for a user (e.g., User “x” issues a trust statement “y” times (More than one time) for user “z”), so this will be correct. Algorithm 2 shows that how the duplicate comments have been deleted.

3.1.3. Removing trolls comments

In Internet slang, trolls are persons that sow discord on the Internet by starting arguments or upsetting people, posting inflammatory, extraneous, or off-topic messages in an online community with the deliberate intent of provoking readers into an emotional response. Moreover, they disrupt normal on-topic discussion. Here, we are going to delete trolls comments. To counteract the effects of troll comments, we practice in the following way.

So for trolls recognition Eq (1) is defined.

\[
\text{Troll} = \alpha \times \left( \text{in} - \text{degree}^{-}(i) \right) + \left(1 - \alpha\right) \times \left( \text{out} - \text{degree}^{-}(i) \right)
\]

where \((\text{in} - \text{degree}^{-}(i))\) and \((\text{out} - \text{degree}^{-}(i))\) represent negative in and out comments of user \(i\) and \(\alpha = 0.7\), which increases the value of negative input comments toward negative output comments and all in and out comments \((i)\) represent all negative and positive comments of user \(i\) which taken and gave to another user. The result of Eq (1) is a digit between 0 and 1 and the user who this digit for it more than a threshold (in this paper 0.33), is considered to be a troll. Therefore, the comments by this user to other users are removed. Algorithm 3 shows how to remove the trolls’ comments.

3.2. Opinion leaders’ selection

In this paper, opinion leaders are obtained using a trust metric and measuring the strength of the trust relationships among users. Namely, with obtaining the total degree of trust for each user, we identified which is that a leader or not; this action takes place in two stages evaluating trust and opinion leaders selecting.

3.2.1. Trust evaluating

Our method uses the trust relationship between users and evaluates the total trust value between users to select best opinion leaders. For opinion leaders’ selection, in the first step, we adopt a trust metric to estimate the strength of trust relationships among users on social networks.

The adopted metric to measure the strength of trust relationships are Jaccard coefficient (JC) \([11]\) based on the structural and social similarity between two users. According to \([31]\), there exist a positive relationship between the similarity among users and the strength of trust established among users. For analysis, we use the number of users whom both users, \(i\) and \(j\) trust to estimate social similarity (and the strength of trust) between users, \(i\) and \(j\). Then, we transform the value of JC metric between 0 and 1 with computing a ratio of the number of users who are trusted by both users, \(i\) and \(j\) to the number of users that are trusted by either a user \(i\) or a user \(j\), but not both. We define this metric in Eq (2).

\[
(\text{JC}) = \frac{|\text{out} - \text{degree}(i) \cap \text{out} - \text{degree}(j)|}{|\text{out} - \text{degree}(i) \cup \text{out} - \text{degree}(j)|}
\]

where \(|\text{out} - \text{degree}(i) \cap \text{out} - \text{degree}(j)|\) represents the number of users who are trusted by both users \(i\) and \(j\) and \(|\text{out} - \text{degree}(i) \cup \text{out} - \text{degree}(j)|\) represents the
number of users who are trusted by either a user $i$ or a user $j$, but not both \cite{7,11}.

3.2.2. Opinion leaders selecting

To select $N$ opinion leaders among $M$ users, in the first step, to reduce the computation of trust relationships and obtain better results, primary selection of opinion leaders takes place using four methods, which introduced in the Section 2, (Hybrid IO-degree, top out-degree, top in-degree and top centrality methods). So $N \times 3$ top users (three times greater than requested opinion leaders, the number 3 are obtained from the experiments) elected with this four methods, which finally we choose the method that gives the best results as a suggested method. Top in-degree, top out-degree, hybrid IO-degree and top centrality methods respectively obtained from the Eqs. (3)–(5) \cite{11} and (6) \cite{6}.

**Top in – degree** = $\text{Max} \ (\text{in comments}) \quad (3)$

Where in comments include positive and negative comments which a user is taken from the other and max mean is, that user which have a maximum in a comment.

**Top out – degree** = $\text{Max} \ (\text{out comments}) \quad (4)$

Where out comments include positive and negative comments which a user gave to other and max mean is, that user which have a maximum out comment.

**Hybrid IO – degree** = $\alpha \times \text{(in – degree)} + (1 - \alpha) \times \text{(out – degree)}$  \quad (5)

Where $\alpha = 0.7$, where $\alpha$ increases the value of input comments toward output comments.

**Top centrality** = $\text{Max} \ (\text{in degree} + \text{out degree}) \quad (6)$

Where it means that user has high sum of in and out comments.

In the second step, total trust value for each $N \times 3$ user with other users is calculated (For primary opinion leaders of four methods). Total trust value for user $x$ is obtained by Eq. (7)

**Algorithm 2. Removing the duplicate comments**

1: $P_x$ is an array list, include $X$ part of comments.
2: $P_y$ is an array list, include $Y$ part of comments.
3: $P_z$ is an array list, include $Z$ part of comments.
4: DCC is including number of duplicate comments.
5: $\text{for} \ (N = 0 \ ; \ N < = \text{length of comments} ; N++)$
6: $\text{for} \ (M = N+1 \ ; \ M < = \text{length of comments} ; M++)$
7: if $(P_x \ N == P_x \ M \text{ and } P_y \ N == P_y \ M \text{ and } P_z \ N == P_z \ M)$
then
8: remove $P_x, P_y, P_z \ N$
9: $M --$
10: DCC ++
11: end if
12: end
13: end

**Algorithm 3. Removing the trolls comments**

1: TLIS is an array list, include user id of trolls.
2: $P_x$ is an array list, include $X$ part of comments.
3: $P_y$ is an array list, include $Y$ part of comments.
4: $P_z$ is an array list, include $Z$ part of comments.
5: TCC is including number of trolls' comments.
6: $\text{for} \ (N = 0 \ ; \ N < = \text{length of TLIS} ; N++)$
7: $\text{for} \ (M = 0 \ ; \ M < = \text{length of comments} ; M++)$
8: if $(\text{TLIS} \ N == P_x \ M)$
then
9: remove $P_x, P_y, P_z \ M$
10: TCC ++
11: end if
12: end
13: end
Out-degree includes both positive and negative comments, so the total trust value for each \( N \times 3 \) top user is calculated by using positive and negative out-degree using (7) and then divide to the number of all users for obtaining a digit to total trust value between 0 and 1.

In the last step, results are sorted in descending order to determine the order of \( N \times 3 \) users. Then, \( N \) top opinion leaders are elected.

4. Discussion and comparison

According to research framework, data filtering is applied to data set, then the opinion leaders are selected with all introduced methods, and in the last, numbers of returned confiding users are calculated for all methods, and the results are compared with each other. A notable point is that, since the trust metric acting based on the social similarity, so when we select the opinion leaders with top total trust value, it means that selected opinion leaders in addition to the primarily selected opinion leaders have highest similarity and sociality degree among other users.

4.1. Dataset

Used data sets are collected from opinions website which is a social network site where each user can rate reviews written by other users and generally this data sets included the relationships of trust between epinions website users. Data sets are publicly available at http://snap.stanford.edu/data/soc-Epinions1.html. ~100,000 opinion between ~15,000 users of the data sets used throughout this paper, the structure of a comment represented in Table 2.

4.2. Simulation environment

The used environment to simulate this investigation is C#. An important reason for choosing it is the structure of arrays in this language, which are defining as one piece in main memory and causing to increase the processing speed.

\[
\text{Total trust value (TTV)} = \sum_{j=1}^{n} \left| \frac{\text{out-degree}(x) \setminus \text{out-degree}(j)}{\text{out-degree}(x) \cup \text{out-degree}(j)} \right|
\]  
(7)

4.3. Results of data filtering

To increase the accuracy of the results, the dataset is filtered. Fig. 2 represents the number of removed comments in the cases of self-trust statements, duplicate comments or trolls comments.

As shown in Fig. 2, through all comments, trolls comments included 6123, duplicate comments included 1730 and self-trust statements included 1250 comments.

4.4. Extracting total trust value

After clearing data, we identify 100 opinion leaders with highest trust values according to the proposed framework that described in Section 3. Table 3 shows the obtained total trust value for 10 users.

The trust value for each user, as shown in Table 3, represents the same total trust value for that user.

4.5. Comparison

After opinion leaders selection with all methods returned real opinion leaders for each method obtained, and results displayed in Fig. 3.

Fig. 3 represents returned percent of real opinion leaders for all methods from 100 selected opinion leaders by each method. As shown in Fig. 3, TTV method (hybrid IO-degree) returned best results.

In the SNM, the manager uses an opinion leader for responding the campaigns. Furthermore, the opinion leader...

---

Fig. 2. The numbers of filtered comments.
leaders forward marketing messages to other users. To determine whom users are attention to this messages and they become a real customer or confiding users, accordingly which people tend to trust other people that behave like them. We calculate the similarity between the opinion leaders and all other users and if the result of similarity degree for each user is more than a threshold, we have considered that user as an actual customer. We use Eq. (2) for extracting similarity between tow users. Similarity threshold in our study were (0.05).

For comparing the proposed method, we review four other ways for identifying the opinion leader, which include selecting users with top in-degree, top out-degree, hybrid IO-degree [11], top centrality [6] and TTV methods. As shown in Fig. 3, TTV (hybrid IO-degree) is the best method among other TTV methods, so we choose it as a representative of TTV method. Fig. 4 represents the percent of returned confiding users using eight methods.

After calculating the trust of degree between selected opinion leaders using the eight methods, the result shows that the TTV (hybrid IO-degree) method with returning 29% of all users as confiding user (real customer) is the best method among top in-degree with 20%, top out-degree with 19%, hybrid IO-degree with 22%, top centrality with 15%, TTV (top in-degree) with 24%, TTV (top centrality) with 18% and TTV (top out-degree) with 23%. Furthermore, Fig. 5 represents the similarity degree sum of confiding users for selected opinion leaders by eight methods.

Table 3
Obtained total trust value for 10 users (data sets from (http://www.epinions.com)).

<table>
<thead>
<tr>
<th>User id</th>
<th>Trust value</th>
<th>User id</th>
<th>Trust value</th>
<th>User id</th>
<th>Trust value</th>
<th>User id</th>
<th>Trust value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26259984260</td>
<td>0.0521</td>
<td>19761434500</td>
<td>0.0514</td>
<td>19761434500</td>
<td>0.0514</td>
<td>19761434500</td>
</tr>
<tr>
<td>2</td>
<td>19761434500</td>
<td>0.0514</td>
<td>01397086596</td>
<td>0.0464</td>
<td>01397086596</td>
<td>0.0464</td>
<td>01397086596</td>
</tr>
<tr>
<td>3</td>
<td>29600878468</td>
<td>0.0456</td>
<td>29600878468</td>
<td>0.0456</td>
<td>29600878468</td>
<td>0.0456</td>
<td>29600878468</td>
</tr>
<tr>
<td>4</td>
<td>03267664020</td>
<td>0.0441</td>
<td>32676614020</td>
<td>0.0442</td>
<td>32676614020</td>
<td>0.0442</td>
<td>32676614020</td>
</tr>
<tr>
<td>5</td>
<td>00000312291</td>
<td>0.0427</td>
<td>00000312291</td>
<td>0.0424</td>
<td>00000312291</td>
<td>0.0424</td>
<td>00000312291</td>
</tr>
<tr>
<td>6</td>
<td>10286567300</td>
<td>0.0412</td>
<td>10286567300</td>
<td>0.0412</td>
<td>10286567300</td>
<td>0.0412</td>
<td>10286567300</td>
</tr>
<tr>
<td>7</td>
<td>04936863620</td>
<td>0.0402</td>
<td>04936863620</td>
<td>0.0402</td>
<td>04936863620</td>
<td>0.0402</td>
<td>04936863620</td>
</tr>
<tr>
<td>8</td>
<td>07340396420</td>
<td>0.0387</td>
<td>02983235620</td>
<td>0.0401</td>
<td>02983235620</td>
<td>0.0401</td>
<td>02983235620</td>
</tr>
<tr>
<td>9</td>
<td>24187867012</td>
<td>0.0375</td>
<td>24187867012</td>
<td>0.0375</td>
<td>24187867012</td>
<td>0.0375</td>
<td>24187867012</td>
</tr>
<tr>
<td>10</td>
<td>00000340335</td>
<td>0.0363</td>
<td>00755601284</td>
<td>0.0399</td>
<td>00755601284</td>
<td>0.0399</td>
<td>00755601284</td>
</tr>
</tbody>
</table>

Fig. 3. Real opinion leader for each method.
of opinion leaders to more investigations. Fig. 6 represents the percent of returned confiding users to each method in the reference size of opinion leaders.

As shown in Fig. 5, in all methods, after increasing the size of opinion leaders to a certain range a small change is occurred. In addition, Table 4 shows the side-by-side comparisons between TTV and seven important methods in this domain in terms of using social relationships, using network topology analysis, using textual content, using removing trolls’ effects, and trust relationships.

Table 4 represents the features used in the 8 methods through the 5 features. As the table shows, none of the methods except TTV used the removing trolls effects which increase the accuracy of the primary data and trust relationships which improves the results. It should be noted, however, the TTV method uses these two features, in comparison with the super edged rank algorithm method which requires network topology analysis and combined framework method, is simpler and less costly.

5. Conclusions and future studies

In this paper, a new framework for identifying opinion leaders based on trust relationship among users is proposed. Identified opinion leaders can be used on many issues, such as political, economic, education, social and etc. In the proposed method, after filtering the values of data sets on three parts, consists of
removing self-trust statements, duplicate comments, and trolls comments, our method identifies those users as opinion leaders which have a top total trust value by using two parts: trust evaluating and opinion leaders selection. The adopted metric to measure the strength of trust relationships is JC [11] based on the structural and social similarity between two users. So, in this research with calculating total trust value from users’ comments, opinion leaders are identified. The results demonstrated that the returned percent of real opinion leaders (73% of real opinion leaders) and confiding users (29% of all users) using the proposed method is more than of those users selected by using the top in-degree method (20% of all users), top out-degree method (19% of all users), top centrality method (18% of all users), hybrid IO-degree (22% of all users) methods. As an extension of the current study, we like to use other trust metrics in the proposed method, for example, knowledge score (KS) or matching coefficient (MC) [11]. Also, we have the interest to increase the accuracy of selecting the best opinion leaders by considering items like user rating and network topology.

References


[27] L. Page, et al., The PageRank Citation Ranking: Bringing Order to the Web, 1999.


