Decision Science Letters 6 (2017) 207–220

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

Extracting new ideas from the behavior of social network users

Amir-Mohsen Karimi-Majd[a] and Mohammad Fathian[b*]

[a] Ph. D. candidate of Socio-economic Systems Engineering, School of Industrial Engineering, Iran
[b] Professor at School of Industrial Engineering, Iran University of Technology and Science, Tehran, Iran

C H R O N I C L E

Article history:
Received October 3, 2016
Received in revised format:
October 22, 2016
Accepted January 20, 2017
Available online
January 22 2017

Keywords:
Data mining
Graph theory
New product development
Idea generation
Behavior analysis

A B S T R A C T

Online social networks (OSNs) provide services targeting multifarious types of users in order to attract and retain them. For this purpose, developing new services according to user preferences has recently been under focused by various researchers. Most of present studies focus only on extracting the behavioral patterns of users, and neglect users' interactions, which is the main part of the social activities in OSNs. To cope with this issue, this paper proposes a new methodology to bring both dimensions of data, the extracted behavioral patterns of users and their social interactions, in order to reach a better analysis. Moreover, the idea provides a basis for considering other dimensions efficiently. In order to evaluate the performance of the methodology, this paper performs a case study, and conducts a set of experiments on the computer-generated datasets. The results indicates the great performance of the methodology.

1. Introduction

Nowadays, Online Social Networks (OSNs) compete with each other to attract more users. In order to attract users, deciding "which one of services should be developed for which group of users" and "how to present them to users" emerge as crucial issues (Kwon & Wen, 2010). Coping with these issues would lead to develop and renew services (such as publishing posts, creating communities, interacting between users and media, and so on), relating to the New Service Development (NSD) area, and extremely focusing on users' data. Hence, an OSN, stronger units in terns of NSD, will be more successful for the competition. In order to develop new services/products based on users' data, several approaches could be employed. A well-known approach is to analyze purchase users/customers' data, which could lead to develop new service/product packages using data mining methods (Liu & Shih, 2005; Liao et al., 2008; Liao et al., 2009; Liao et al., 2010; Zhu et al., 2015). Clustering and Association Rules (AR) mining are two famous methods for this purpose. As another data mining methodology, the

* Corresponding author. Tel: +98-21-73225048, Fax: +98-21-73021214
E-mail address: fathian@iust.ac.ir (M. Fathian)

doi: 10.5267/j.dsl.2017.1.002

researchers (Karimi-Majd & Mahootchi, 2015) claim that behavioral patterns of customers could express their unspoken *needs* and *wants* for developing “completely new” services. The idea behind this methodology is to determine how users use services (i.e., their behavioral pattern) and tell us which services they need. Using textual information of users for revealing new ideas has emerged as another wonderful approach in this line (Thorleuchter et al., 2010). In addition, users’ posts and comments would be analyzed in order to find new product ideas. Lee et al. (2015) investigasted the relationships between customer opinions and product attributes for developing saleable products. They developed a fuzzy association rule mining for that purpose.

Nevertheless, none of the above-mentioned methods would be sufficient for comprehensively analyzing OSNs. This is because they neglect the social interactions of users in their analysis. Social interactions together make up communities. Communities, since users influence on each other to reach similar preferences, trap opinions about goods, services, and brands (Sznajd-Weron & Sznajd, 2000), and define more real grouping among users (De Souza & Preece, 2004). These two characteristics would be employed to discover natural groups of network. Each of such groups owns linked members with rather the same opinions and preferences. Therefore, it is expected that the analysis of groups leads to more real results (Carmagnola et al., 2013). In order to reach such richer analysis, this paper proposes a new methodology by considering the two mentioned dimensions (i.e., the behavioral patterns of users and their social interactions) together. According to this idea, the proposed methodology aims to cluster a network by considering both behavioral and social dimensions, simultaneously. To do so, this methodology employs attributes in the form of a graph, as a common element between the dimensions in order to integrate them, followed by making an integrated network. Then, the clusters of attributes would be extracted from the integrated network using Spectral clustering algorithm. This methodology also creates a basis for efficiently considering more dimensions for the analysis.

The rest parts of this paper are as follows: the next section represents a brief review of the related works. Section 3 clearly describes our proposed methodology. The results of these experiments in order to compare our proposed methodology with another one would be found in Section 4. In this section, an interesting case study for showing the performance of the methodology was conducted. Sections 5 and 6 provide a discussion and some concluding remarks with possible future works, respectively.

2. Related works

New service development emerges as a vital issue in today’s business. Kumar and Phrommathed (2005) expressed that in order to achieve new products/ services, there are six steps as follows: 1) generating new ideas, 2) screening on hand ideas, 3) analyzing the akin businesses, 4) business development, 5) tests, and 6) commercialization. One can easily find that the success and the agility of passing such steps needs valuable and accurate ideas (Flint, 2002). To generate such ideas, many approaches and techniques have been proposed and employed. In this section, in order to review these methods, we categorize them in two groups, called the qualitative and the quantitative methods.

Brainstorming (Osborn, 1953) is known as an idea generation technique categorized as a qualitative method. This technique aims to provide a situation for creative thinking in order to generate many wild ideas in a short time. Scamper and forced relation techniques have been categorized in this group (i.e., qualitative methods). One can easily find that, in such methods, the experts’ knowledge, experience, and creativity play the main role in making management decisions. On the other side, the quantitative methods aims to generate new ideas based on the gathered information from customers or users. Quality Function Deployment (QFD) is a well-known technique in this group (Franceschini, 2016). Given the customers’ spoken requirements, as the qualitative input, QFD tries to map such information to particular plans, as the quantitative output. Unlike the advantages of QFD, gathering proper information as its inputs takes a long time while gathering information of customers’ transactions or profiles comes up as a rapid process in today’s web. Moreover, this technique often leads to a plan for improving present products, and does not guarantee to develop innovative ideas. In order to develop new ideas in a quantitative approach, we need to extract the latent knowledge from customers’ data. Data mining methods have been developed for such goal. Finding natural clusters among customers (i.e., clustering), and discovering interesting rules among their transactions (i.e., Association Rules mining) are two widely-used data mining approach in this area. For instance, Au and Chan (2003), and Niyagas et al. (2006) researched on the behavior of different bank's customers by means of mining and analyzing Association Rules (ARs) latent in their transactions. An example of such behavioral pattern (i.e., extracted ARs) is as follows: (Middle) ∧ (Salary>1000) → (Saving>300), which means if a customer is in *Middle* Age and his/her *Salary* is more than 1000, then his/her amount of *Saving* will be expectedly more than 300. The combination of AR mining and clustering have been proposed by Ahn and Sohn (2009) for the case of managing after-sales services. Developing new cosmetic packages (Liao et al, 2008) and designing new digital cameras (Bae & Kim, 2011) would be two other applications of using ARs found in the literature. In another interesting research, Cheng and Sun (2012) employed the extracted ARs for making strategic decisions in the field of Mobile commerce.

A drawback of the methods mentioned above is that they do not lead to completely new product or innovative service idea. As a good idea for tackling this issue, Karimi-Majd and Mahootchi (2015) proposed a methodology (called KM in this paper) to analyze the special graphs of the attributes (i.e., services and customers' attributes). This kind of graphs would be formed by extracting ARs and clustering them. Indeed, each cluster corresponds to a different graph. In order to extract ARs, they used Apriori method, which requires two thresholds for running: *Min-Support* and *Min-Confidence*. Moreover, for clustering ARs, they employed a variation of K-Medoid method (Park & Jun, 2009), which needs to determine the number of clusters, called *K*. To cope with these issues, they proposed a new index for evaluating the quality of the clustering when the parameters (i.e., *K*, *Min-Support*, and *Min-Confidence*) vary. After clustering ARs, the graphs would be found. Each graph can illustrate the gaps between attributes in its respective viewpoint. Analyzing these gaps helps us to generate innovative ideas. Fig. 1 schematically express the KM methodology.

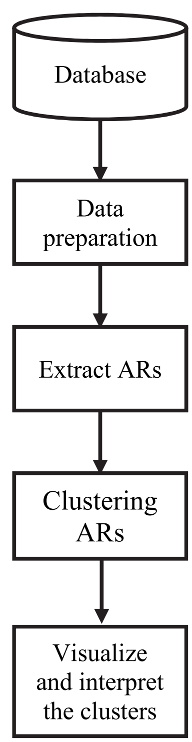

Fig. 1. KM methodology

They implemented their methodology on questionnaire data of a blogging websites' users. An interesting fact in their work is that they have focused on what happened between attributes to analyze the behavior of the customers instead of what occurred between customers. In order to make it clearer, the Fig. 2 demonstrates an extracted sub-graph of attributes, which is illustrated in (Karimi-Majd & Mahootchi, 2015).

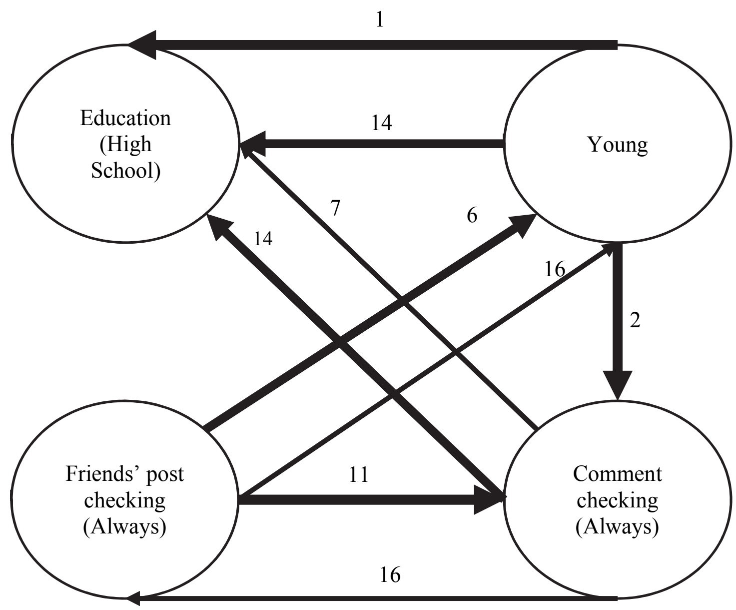

Fig. 2. The obtained sub-graph by the KM methodology

In Fig. 2, each rule has been depicted by means of arrow (s), and the rule number has been inserted in its respective arrow(s). The thickness of each arrow corresponds to its respective calculated customer satisfaction (CS). Since this sub-graph has been extracted using an optimization procedure, we could individually say about the gaps between the services. Based on the analysis present in Karimi-Majd and Mahootchi (2015), the users of that blogging website have two mainly behavior; a group of them prefer to social interactions who always check friends' posts and another group of them are not willing to communicate with others.

Another gap could be seen among the methods is that social aspect of users has been neglected. This gap has been encountered in other areas with an appropriate response. For instance, in the field of developing recommender systems, which focus only on in-hand products/services, social interactions has been considered in various new models (Deng et al., 2014; Yigit et al., 2015; Cena et al., 2016; Feng et al., 2016). Taking inspiration from such fields could be useful.

3. Proposed methodology

As stated before, the researchers in Karimi-Majd and Mahootchi (2015) reported that the innovative ideas lie at the heart of the groups of the interwoven behavioral patterns (i.e., clusters of ARs). The "behavioral patterns" refers to one dimension. The second dimension, which is not considered in their methodology, refers to "social patterns". Fig. 3 demonstrates different parts of the proposed methodology, called KF. In this figure, A, D^1, D^2, and D are matrices that are going to be described in the remaining parts of this section. Indeed, the remaining parts of this section are devoted to describe the dimensions, and to express how they can be integrated and clustered according to our methodology.

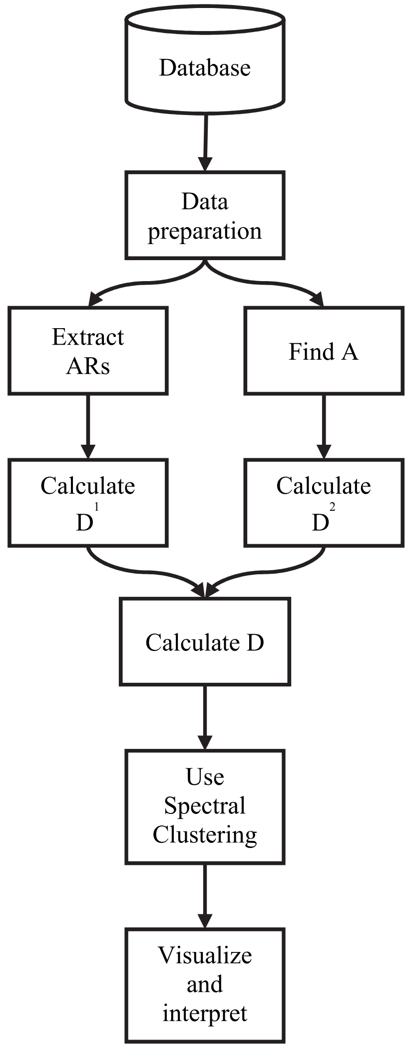

Fig. 3. Proposed methodology (called KF)

3.1. Extracting behavioral patterns: the first dimension

In this research, the behavioral patterns could be stated by ARs extracted from a given dataset, including *N* users and *F* attributes. *R* ARs (R indicates the number of ARs) could be demonstrated by a binary matrix, called *B*, including *R* rows and *F* columns. In this sense, ARs were supposed as the frequent patterns without any loss of generality. Clustering of the rows (i.e., ARs or the behavioral patterns) would be possible if an appropriate similarity measure were used. On the other hand, since the aim of the present research is to visualize the clusters of ARs according to the attributes engaging in their inclusive ARs, another approach is to focuse on the attributes instead of the ARs. Thus, the attributes could be clustered according to their similarities regarding the number of times they could simultaneously appear in ARs, i.e., rows of the matrix *B*. In other words, the intersection size of each pair of the attributes would be considered as their similarity for clustering them. However, a fairer similarity measure has been proposed by Jaccard (1901), which targets the proportion of the intersection size and the union size of two vectors (Strehl et al., 2000). The definitions of the size of intersection and the size of the union are as follows:

Definition 1. The size of the intersection of the two given binary vectors V1 and V2 with N elements is the number of the same elements appearing in both vectors.

$$I(V1,V2)=\sum_{i=1}^{N}V1_i\times V2_i\ . \tag{1}$$

Definition 2. The size of the union of the two given binary vectors V1 and V2, called U, is the number of all different elements appearing in the vectors.

$$U(V1,V2)=\sum_{i=1}^{N}V1_i+\sum_{i=1}^{N}V2_i-I(V1,V2)\ . \tag{2}$$

Then, by calculating the similarity between each pair of the attributes, the matrix D^1 could be extracted from the matrix B as follows:

$$d_{ij}=I(B_{.i},B_{.j})/U(B_{.i},B_{.j})\ , \tag{3}$$

where d_{ij} denotes the element of matrix D^1 appearing in the i^{th} row and j^{th} column, and $B_{.i}$ implies the i^{th} columns of matrix B.

The matrix D^1 has interesting properties. D^1 is square and also symmetric. With regarding to such properties, the well-known Spectral clustering method (Donath & Hoffman, 1973) would be proposed for our clustering purpose. This method calculates eigenvectors of the square matrix D^1 in order to clearly separate the data clusters from each other. Therefore, the quality of the obtained clusters would be higher. Once these clusters of the attributes were found, the graphs of ARs could be visualized as it has been performed in a previous study (Karimi-Majd & Mahootchi, 2015). Therefore, for each cluster, the visualization process would be performed by representing all ARs covered by all admissible combinations of the present attributes in that cluster. One can easily find the similar process in Fig. 1. In order to extract ARs, the well-known and highly-used AR mining algorithm, called Apriori (Han & Kamber, 2006) was employed. This algorithm requires two thresholds, called *Min-Support* and *Min-Confidence*, to be run. These thresholds could be found through a search procedure, as stated in (Karimi-Majd & Mahootchi, 2015).

3.2. Extracting social patterns: the second dimension

Calculating the elements of the matrix D^1, based on the behavioral patterns, could be easily and straightforwardly performed as stated before. The reason from the computational point of view is that the matrix form of the behavioral patterns, i.e., the matrix B, would be available, once ARs from the users' data are extracted. However, in the case of extracting the social patterns (i.e., the clusters of the users' interactions), there is no such matrix form. One may wonder that a similar matrix could be extracted from the matrix of the users' attributes, called *P* including *N* rows and *F* columns. This is incorrect, because the result of such calculations considers which user owns which attribute, and does not consider the adjacency matrix of users, called *A*, which shows which user interacts with which one. Another approach is the detection of communities among the users' interactions (Karimi-Majd et al., 2015). This approach is not also helpful because our methodology aimed to focus on finding the clusters among the present attributes.

To cope with this issue, here in this study, another approach has been proposed as follows. Similar to what was proposed to be performed on the behavioral patterns in the previous subsection, the attributes could be clustered based on the users' interactions (i.e., the social patterns). In other words, the aim of this research was to find the number of each pair of attributes which simultaneously appeared in both sides of the all users' interactions. Thus, the result matrix, called D^2, included *F* rows and *F* columns. The matrix D^2 could be calculated as follows:

$$D^2_{FxF}=(P^T_{FxN}\times A_{NxN}\times P_{NxF})/F^2, \tag{4}$$

where *T* implies to the transposed form of matrix. The first multiplication (i.e., $P^T\times A$) leads to a matrix representing the number of each attribute which appears at the end of the interactions of each user.

Then, this matrix would be multiplied by P in order to take the other side of each interaction into account. Thus, the numerator implied to the number of each pair of attributes simultaneously appeared in both sides of the all users' interactions. Since the calculated values might be too high in the case of large network (where a large number of attributes existed), the denominator was proposed to be considered. This is the square of the number of attributes, i.e., F, because F is considered twice in the numerator.

3.3. Integrating dimensions and clustering

Integrating such dimensions would be easily performed as stated previously by researchers (Tang & Liu, 2010). This approach provides a proper facility for finding the clusters (or even communities) in multi-dimensional networks similar to graphs of attributes (AGs). To do so, since the integration refers to union, the *Max* operator was employed, as the *Standard union* operator, for integrating D^1 and D^2. The result was called the matrix D.

$$D = max\ (D^1, D^2). \tag{5}$$

The clusters of the attributes could be found using the Spectral clustering method on the matrix D. The Spectral clustering method requires the number of clusters (called K) in order to be run, without knowing its value. On the other side, the Apriori method also needs to determine its respective thresholds, i.e., *Min-Support* and *Min-Confidence*, for finding ARs. As stated before, authors in (Karimi-Majd & Mahootchi, 2015) proposed a new index for evaluating the found clusters of ARs; however, our methodology aimed to do a classic clustering, not AR clustering. Thus, all the Cluster Validity Indices (CVI) in the literature could be used. One of them is called the Silhouette index (Rousseeuw, 1987) defined as follows:

$$S = \sum_{i=1}^{N} \left(\frac{(b_i - a_i)}{\max(a_i, b_i)}\right) / N, \tag{6}$$

where a_i and b_i imply the average distance of object i with all the other objects present in its respective cluster, and the lowest average distance of object i with all the other objects present in other clusters, respectively. The distance measure would be Euclidean distance.

Furthermore, as explained previously in a research (Karimi-Majd & Mahootchi, 2015), in order to find more proper clusters for effective visualization, the employed index should consider a factor for minimizing the number of placed attributes in the clusters. To do so, one would propose to use min-max operator as the denominator of the index, because this operator could control the maximum number of placed attributes in the clusters by minimizing it. Finally, the index is rewritten as follows:

$$S = \sum_{i=1}^{N} \left(\frac{(b_i - a_i)}{\max(a_i, b_i)}\right) / (N \times \min_k C_k), \tag{7}$$

where C_k is the number of objects (i.e., attributes) in the kth clusters. Using this CVI, our proposed methodology could be run over different combinations of K, *Min-Support*, and *Min-Confidence* to find the best one.

4. Experimental results

In order to evaluate and illustrate the performance of the proposed method, this section provides a case study. Before that, since only the KM methodology is related and may be a competitor for the proposed methodology, the results of some experiments on a number of artificial datasets will be presented in order to compare the methodology with the KM methodology. Note that the KM methodology only considers the behavioral patterns. All algorithms are coded and ran in MATLAB R2013a on a computer with Intel Core 2 Duo, 2.50 GHz, and 4.0 GB RAM.

4.1. Datasets and results

According to the authors' knowledge, there is no proper benchmark dataset on the web related to the main problem. Thus, we confined ourselves to a well-known artificial dataset, called ZAV dataset. ZAV benchmark dataset proposed by Zanghi et al. (2010) is a type of computer-generated dataset. This dataset includes nodes with their community labels, interactions, and integer attributes. The process of creating interactions and generating attributes are based on the random graph model (Erdös & Rényi, 1959) and random normal distribution, respectively. In order to generate such network by the means of simulation, the values of its five following parameters should be determined: 1) the size of the network *N*, 2) the number of attributes *F*, 3) the mixing parameter *dle*, 4) the difference between the mean of values of the attributes *dlf*, and 5) the number of clusters *Q*. Then, at the first step, the proposed algorithm generates *N* nodes considering *dle*, which are uniformly placed in *Q* clusters. At the second step, it simulates the values of *F* attributes for *N* nodes based on their cluster label *Q* and the average of the values of each attribute for each cluster considering *dlf*. After that, the required data will be prepared; therefore, no data preparation is required only for experiments on this kind of artificial datasets. In the first experiment, it was considered that *F* varied from 10 to 30, and *N*, *dle*, and *dlf* were constant, and equal to 1000, 0.2, and 5, respectively. As a pre-processing task, the integer attributes should have been transformed into the categorical attributes in order to extract ARs. The aim was to briefly study the behavior of the KF methodology, in comparison with another method, called KM, when the number of the attributes varied. Fig. 4 represents the achieved results.

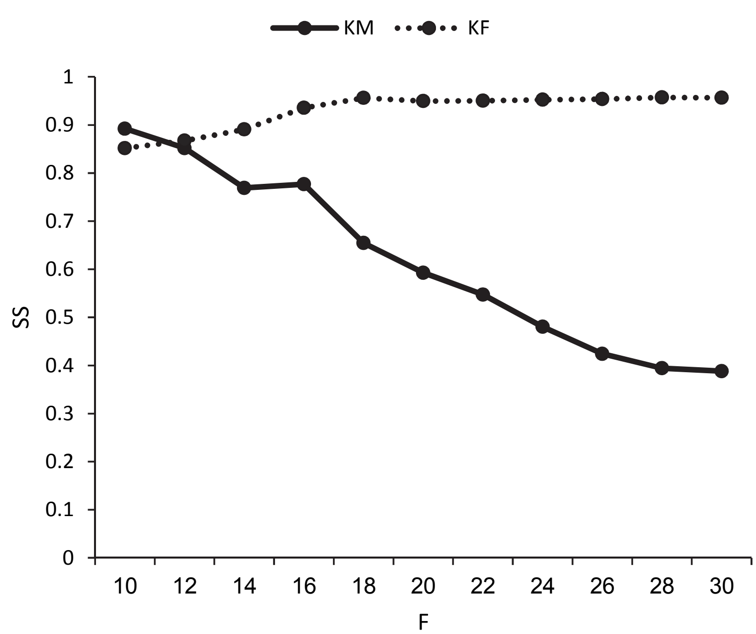

Fig. 4. The results achieved by the methods based on the SS when F varies

As illustrated in Fig. 4, although the KM methodology reaches a better result when *F* equals to 0.1, its respective results gradually decreases as F increases. However, in the case of the KF method, the results do not depend on the values of *F*. The reason of such behavior might be that when *F* increases, the diversity of the extracted ARs would be also increased. Thus, the relationships matrix of the ARs might be too sparse. In such situation, the KM methodology, which does not take the interactions into account, fails to find the highest quality clusters. On the other hand, the KF methodology trades off between the two present dimensions. Moreover, an interesting fact about the behavior of this method is that this method is stable when *F* varies. This might be due to the reason that as *F* varies, the structure of the respective network, i.e., interactions, does not significantly change. Thus, the stability of this method would be reached because the second dimension has more powerful impact on the KF's results. Other experiments corresponded to the situations when *dlf* varied from 5 to 15. In these experiments, the values of *N* and *dle* were similar to the previous experiments, but *F* was constant and equal to 10. Fig. 5 illustrates the results achieved by the methods based on the SS index.

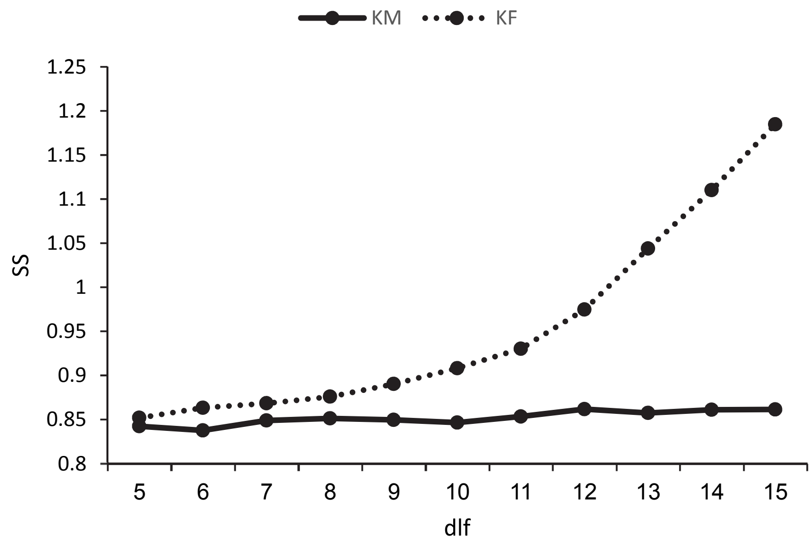

Fig. 5. Results achieved by the methods based on the SS when *dlf* varies

According to Fi. 5, when *dlf* increases, the quality of clustering obtained by the KF increases. This is due to the fact that as *dlf* increases, the separation between the clusters increases as well. In such situation, using eigenvectors of the respective matrix leads to clearer distinction of the latent clusters. Thus, the quality of the clusters extracted by the *KF* significantly increases as *dlf* increases. However, KM is almost indifferent when *dlf* changes. This is because the *KM* focuses on the clustering ARs, not on the present attributes. In other words, since the structure of the graph of ARs does not significantly change when *dlf* varies, the differences between the clusters of the extracted ARs may refer to the random events of the respective algorithm. Moreover, one can easily find that in most cases, the KF reaches to the greater results than the KM.

4.2. Case study

In the previous subsection, the obtained results of our methodology and another one were compared. It was demonstrated that the proposed methodology could reach better results, especially in the case of enormous number of attributes, similar to profiles of OSNs' users. Now, in order to indicate the performance of our methodology in analyzing the real-life problem, a case study is provided as follows.

4.2.1. Data preparation

In the first step of this study, as the database, a Persian OSN was chosen which was not permitted to disclose the network's name as requested by the administrator. The data has been collected by crawling among the users' friends of the network. For each unique user, the data of the respective profile was also extracted. The dataset included the profiles of 30711 users; each profile contained 23 attributes, and totally 31383 links (friendship) between the users. In the next step (i.e., the data preparation step), since many users had not yet completed their profiles, a serious missing data issue occurred. In order to solve this problem in the current study, it was decided to remove them from both the profiles database and the links database. After this cleansing, there were some users in the profile database who had no friendship in the links database. Thus, their profiles were also removed since the aim of this study was to provide a connected network in order to reach admissible results. The number of the remaining users and links were reduced to 676 and 1736, respectively. One could find that the density of the network (i.e., the ratio of the number of links to the number of users) has been increased from 1.02 to 2.57. Implicitly, the greater the density, the clearer social patterns would be expected to form.

In the case of attributes, the important ones were selected through a feature selection process. Age, gender, marital status, the number of months of membership, the number of friends, the number of sent friendship requests, the number of received friendship requests, the number of all posts, the number of all hot posts, the number of comments, the number of received comments, the number of likes, the number of received likes, the number of dislikes, the number of received dislikes, and the number of all re-shared posts were 17 remained attributes. Some of them were demographical attributes and other ones referred to the frequency of using the OSN services.

Since the next step could only work with the binary data in order to mine ARs, the remained attributes should have been transformed into the binary format. To do so, instead of each binary attribute (such as gender), two new attributes were considered (i.e., each one for each value), because both values had the chance for appearing in the ARs. In the case of categorical attributes (such as job), a new attribute for each value was presented instead of each main attribute. For instance, the attribute *Education* has been replaced with the attributes *Student*, *Freelancer*, *Teacher*, *Manager*, *Doctor*, *Worker*, *Employee*, *Expert*, *Artist*, *Housekeeper*, and *Other*. The ratio attributes (such as age) have also been transformed into the categorical attributes at first, and then into the binary attributes, similar to what was performed on the categorical attributes. Note that in the case of attributes referring to the frequency of using services, three levels were considered, as three categories for each attribute which were tried to find by clustering. These levels were: Low, Medium, and High. Finally, for each user, the dataset included 50 binary attributes.

4.2.2 Finding clusters

In the next step, the minimum values were determined for *Min-Support* and *Min-Confidence* thresholds which were equal to 0.1 and 0.5, respectively, leading to extract 23974 ARs. Then, these thresholds were increased to 1 by 0.1 in order to calculate different D^1 matrices, one for each combination of the thresholds. The values of K were also considered as integers between 2 and 20. The reason of this approach was explained (i.e., determining different values for the parameters) in the previous section. Finally, for each combination of the thresholds, the best number of clusters was found using the CVI described in the previous section. Fig. 6 illustrates the obtained results.

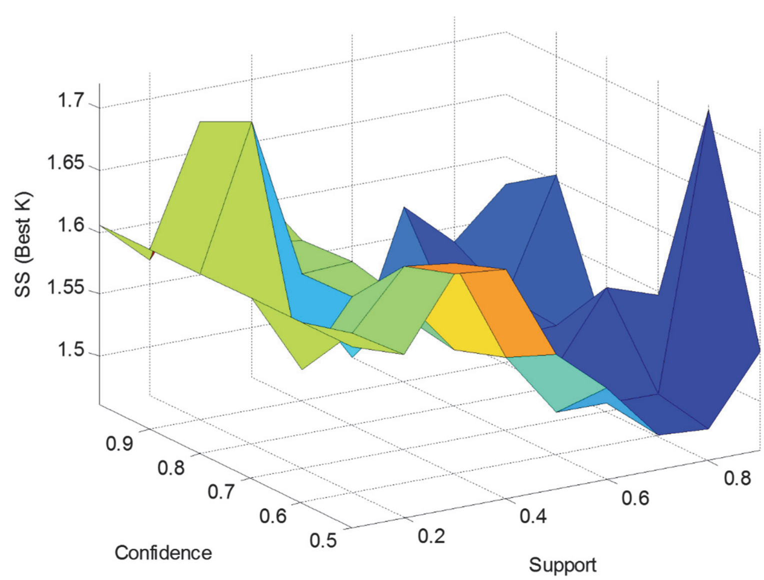

Fig. 6. The quality of clustering for different values of *Min-Support* and *Min-Confidence*

According to Fig. 6, the maximum refers to the best reached values as follows: K=20, *Min-Support*=0.2, *Min-Confidence*=0.8, and SS=1.72. The 20 extracted clusters consisted of 17 clusters each containing only one attribute, 2 clusters each containing 3 attributes, and 1 cluster containing 27 attributes. Each cluster, as stated before, should be visualized as a graph and interpreted for generating new service ideas. However, one could easily find out that interpreting the graphs containing one node or including many nodes is so hard. Thus, the clusters containing 1 or 27 attributes were withdrawn from the study. Consequently, two clusters containing 3 attributes would be remained for studying as follows.

One of these clusters is depicted in the form of a weighted graph (Fig. 7). The nodes of this graph refer to three attributes (i.e., Elderly, Married, and Employee), which are *co-occurred* among *friends* in the network (according to the idea of the methodology used in this study). Indeed, the elements of two matrices D^1 (refers to the co-occurring of attributes) and D^2 (refers to the friendship links) illustrate the weight (thickness) of the links between these nodes, which could be distinguished by red and blue lines for D^1 and D^2, respectively.

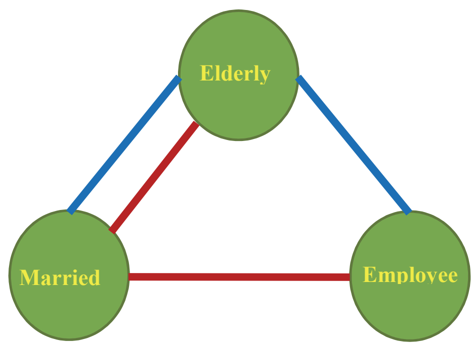

Fig. 7. Extracted graph 1

According to Fig. 7, an outstanding part of the users, among all types of them, is the group of elderly married employees (refers to the behavioral pattern) who are connected with each other (refers to the social patterns). They could be targeted for developing new services, which meet their common characteristics. For instance, a special service for making the charity or political communities with a proper advertising would be suggested. As an interesting fact of our methodology, a proper advertising or even an appropriate user interface could make all the users remain in the network and invite others, because of their social activities and this fact that a good opinion would be trapped in a community (1). One may claim that the classical clustering of users based on their attributes can reach such achievements. Note that such approach would fail to characterize between the extracted clusters, so that the most important is missing. Moreover, from a computational point of view, the classical clustering might lead to the wrong solutions because it has simultaneously involved all the attributes in the analysis, so the differences in some attributes would be compensated by the similarity in others.

Fig. 8 illustrates the other graph. According to this figure, another earnest group of users refers to the housekeepers who are strongly connected with each other. They write *comments* on their posts and give *likes* to each other both in medium level. On the other hand, the relationships between the nodes *comments* and *likes* are not significant; therefore, the node *Housekeeper* is the key node. Thus, if the manager aims to keep the housekeepers in the network (also absorb more housekeepers), he should develop services in order to facilitate these services, or make them more attractive. For instance, showing who gave *likes* to a post in a determined time could be a great idea or might develop a variation of internal chat system.

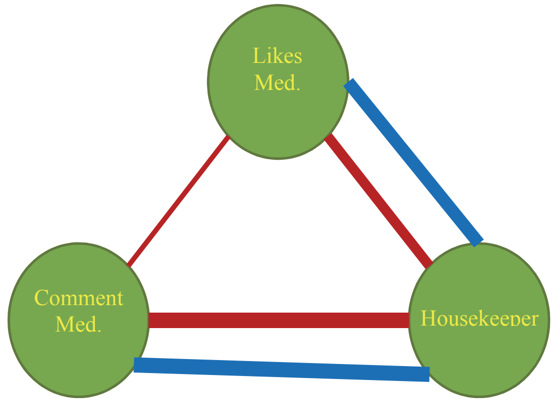

Fig.e 8. Extracted graph 2

5. Discussion

Authors in Karimi-Majd and Mahootchi (2015) claimed that each extracted graph of attributes has at least a unique message for generating new ideas. Such claim would be usual and acceptable in the area of pattern recognition, although, some results might actually be so hard to be interpreted. Indeed, without any opposition with the claim, we have found that the messages of the most of such graphs might not be heard for different reasons. The first one is that this characteristic might be inherited from AR mining, because some of the extracted ARs using the mining process would be redundant. The redundant ARs obviously could lead to obtain graphs without any message. Another reason might be that the messages should be discovered in the interpretation step, which significantly depends on the human characteristics. Some graphs would be so hard to be interpreted by some experts, but easy for the others. The lack of proper information might be another reason. As illustrated in the case study, the large amount of data would be removed in the data preparation step. Further research would be required in order to study the impacts of these reasons on the results and to develop the methods to cope with them. Our methodology could be also employed for the idea screening purpose, similar to the KM methodology. For instance, if the manager of our case study aims to develop a new service for adults, the first extracted graph tells that the new service should suit employees, and the second one insists on considering the main characteristics of the housekeepers in that service. These messages help managers to screen, and also to mature their ideas. Another idea, here, is to study the published contents by the elderly employees and housekeepers, as two important groups of the network's users of our case study. This approach helps managers to focus on important groups of users, and also important users (Zue et al., 2015), instead of all the users. The result might lead to screen the dissipated ideas.

There is an interesting aspect of the KM methodology. In this methodology, the impact of the calculated CS on the visualized ARs has been considered. Then, the authors (i.e., (Karimi-Majd & Mahootchi, 2015)) have shown the advantage of this idea in their case study as previously illustrated. In order to consider the impact of CS on the graph of all the ARs in the proposed methodology, one might propose to suppose a separate dimension for CS, similar to the KM methodology. Here, since the impact of CS should be visualized on the respective ARs, another suggestion has been made. Indeed, we proposed to multiply the CS of each ARs to its respective row in the matrix B. Then, since the newly calculated matrix B included real values, we proposed to employ the cosine similarity measure for calculating the matrix D^1. It is worth mentioning that the cosine similarity measure could be employed instead of the Jaccard measure even if the matrix B is binary. On the other hand, an underlying property of the matrix D^i (i.e., D^1 and D^2) was that D^i corresponded to AG, which had a potential to be the basis of integrating more dimensions with each other. Indeed, for each of the considered dimensions, a specific matrix D^i could be separately calculated Then, in the integration step, all the considered dimensions could be integrated together to make a unique matrix. Moreover, the obtained matrix could be clustered using all the clustering methods without any need to define a new one.

6. Conclusion

In order to attract more and more users, OSNs have to develop innovative services. To do so, they need to generate new ideas based on the requirements and preferences of users. The users' requirements would be latent in their behavior. If we accurately learn the behavioral patterns of the users, we can generate great ideas by analyzing the especial graphs of the present attributes. In this paper, a new methodology was proposed that not only considered the relationships between the present attributes for the analysis, but also took the users' interactions into account. Indeed, our methodology is able to find the cluster of attributes in multi-dimensional network.

Due to the lack of appropriate real-life data for the experiment, we employed a set of standard computer-generated datasets in order to evaluate our method. The results have shown the superiority of our proposed method in most of the cases. Moreover a case study was conducted in order to show the performance of our methodology. Our results criticized the claim that each extracted graph has at least one message for developing a new service idea. As a great characteristic of the proposed methodology, it can be noted that other dimensions such as liking, commenting, etc. could be added into consideration and integrated with other dimensions if it is needed. However, adding a new dimension did not always lead to the superior results. Therefore, a crucial issue, here, is finding the dimension which should be considered for reaching a better result. Coping with this issue has been left as a future work.

As another future work, the relationships between AGs, which were extracted from different dimensions of the users' data, would be performed on the other real-life dataset. Such analysis might make it possible to discover more facts about the users' behavior, leading to unhide the valuable gaps.

References

Ahn, J. S., & Sohn, S. Y. (2009). Customer pattern search for after-sales service in manufacturing. *Expert Systems with Applications*, *36*(3), 5371-5375.

Au, W. H., & Chan, K. C. (2003). Mining fuzzy association rules in a bank-account database. *IEEE Transactions on Fuzzy Systems*, *11*(2), 238-248.

Bae, J. K., & Kim, J. (2011). Product development with data mining techniques: A case on design of digital camera. *Expert Systems with Applications*, *38*(8), 9274-9280.

Carmagnola, F., Vernero, F., & Grillo, P. (2013). Advanced Social Recommendations with SoNARS++. *Interacting with Computers*, iwt028.

Cena, F., Likavec, S., Lombardi, I., & Picardi, C. (2016). Should I Stay or Should I Go? Improving Event Recommendation in the Social Web. *Interacting with Computers*, *28*(1), 55-72.

Cheng, L. C., & Sun, L. M. (2012). Exploring consumer adoption of new services by analyzing the behavior of 3G subscribers: An empirical case study. *Electronic Commerce Research and Applications*, *11*(2), 89-100.

De Souza, C. S., & Preece, J. (2004). A framework for analyzing and understanding online communities. *Interacting with Computers*, *16*(3), 579-610.

Deng, S., Huang, L., & Xu, G. (2014). Social network-based service recommendation with trust enhancement. *Expert Systems with Applications*, *41*(18), 8075-8084.

Donath, W. E., & Hoffman, A. J. (1973). Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, *17*(5), 420-425.

Erdös, P., & Rényi, A. (1959). On random graphs, I. *Publicationes Mathematicae (Debrecen)*, *6*, 290-297.

Feng, X., Sharma, A., Srivastava, J., Wu, S., & Tang, Z. (2016). Social network regularized Sparse Linear Model for Top-N recommendation. *Engineering Applications of Artificial Intelligence*, *51*, 5-15.

Flint, D. J. (2002). Compressing new product success-to-success cycle time: Deep customer value understanding and idea generation. *Industrial marketing management*, *31*(4), 305-315.

Franceschini, F. (2016). *Advanced quality function deployment*. CRC Press.

Han, I., & Kamber, M. (2006). Data mining concepts and techniques. *Morgan Kaufinann*.

Jaccard, P. (1901). Distribution de la florine alpine dans la bassin de danses. et dans quelques regions voisines. *Naturelles Bulletin de la Societe Vaudoise des Sciences*, 241–272.

Karimi-Majd, A. M., Fathian, M., & Amiri, B. (2015). A hybrid artificial immune network for detecting communities in complex networks. *Computing*, *97*(5), 483-507.

Karimi-Majd, A. M., & Mahootchi, M. (2015). A new data mining methodology for generating new service ideas. *Information Systems and e-Business Management*, *13*(3), 421-443.

Kumar, S., & Phrommathed P. (2005). *New product development: an empirical study of the effects of innovation strategy, organization learning and market conditions*. Springer, New York.

Kwon, O., & Wen, Y. (2010). An empirical study of the factors affecting social network service use. *Computers in human behavior*, *26*(2), 254-263.

Lee, C. K. H., Tse, Y. K., Ho, G. T., & Choy, K. L. (2015). Fuzzy association rule mining for fashion product development. *Industrial Management & Data Systems*, *115*(2), 383-399.

Liao, S. H., Hsieh, C. L., & Huang, S. P. (2008). Mining product maps for new product development. *Expert Systems with Applications*, *34*(1), 50-62.

Liao, S. H., Chen, Y. N., & Tseng, Y. Y. (2009). Mining demand chain knowledge of life insurance market for new product development. *Expert Systems with Applications*, *36*(5), 9422-9437.

Liao, S. H., Chen, Y. J., & Deng, M. Y. (2010). Mining customer knowledge for tourism new product development and customer relationship management. *Expert Systems with Applications*, *37*(6), 4212-4223.

Liu, D. R., & Shih, Y. Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, *42*(3), 387-400.

Niyagas, W., Srivihok, A., & Kitisin, S. (2006). Clustering e-banking customer using data mining and marketing segmentation. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *2*(1), 63-69.

Osborn, A. F. (1953). Applied imagination, principles and procedures of creative thinking.

Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, *36*(2), 3336-3341.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Strehl, A., Ghosh, J., & Mooney, R. (2000, July). Impact of similarity measures on web-page clustering. In *Workshop on Artificial Intelligence for Web Search (AAAI 2000)* (pp. 58-64).

Sznajd-Weron, K., & Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, *11*(06), 1157-1165.

Tang, L., & Liu, H. (2010). Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, *2*(1), 1-137.

Thorleuchter, D., Van den Poel, D., & Prinzie, A. (2010, January). Extracting consumers' needs for new products-A web mining approach. In *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on* (pp. 440-443). IEEE.

Yigit, M., Bilgin, B. E., & Karahoca, A. (2015). Extended topology based recommendation system for unidirectional social networks. *Expert Systems with Applications*, *42*(7), 3653-3661.

Zanghi, H., Volant, S., & Ambroise, C. (2010). Clustering based on random graph model embedding vertex features. *Pattern Recognition Letters*, *31*(9), 830-836.

Zhu, Z., Su, J., & Kong, L. (2015). Measuring influence in online social network based on the user-content bipartite graph. *Computers in Human Behavior*, *52*, 184-189.