Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 1897–1907
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.22669.2398

An evolutionary game competitive model for increasing accuracy and maximizing influence in social networks

Saeedeh Noori[a], Hamidreza Navidi[a,*], Reza Bakhtiari[c]

[a]*Department of Mathematics and Computer Science, Shahed University, Tehran, Iran*
[b]*Department of Computer Science, Shahed University, Tehran, Iran*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Given the importance of maximizing influence in a social network, studies in this field often seek to find the nodes that have the most influence on the social network if designated as primary seeds. In this study, to reduce the complexity of computation algorithms, the problem is divided into several groups that aim to find a group of influential people among users of a social network. In this paper, a framework is introduced for solving the problem of influence maximization, which is based on the member clustering by the K means method, to improve the classification of network users, the data are weighted and the problem is modeled and analyzed as an evolutionary game. And finally, calculate its evolutionary stable strategy. This framework has been tested on real social network data for Abrar University students and we have achieved results such as increasing classification accuracy, reducing error function and finding a stable strategy in the community.

Keywords: Social network, Influence maximization, Distinction factor, k-means clustering, Evolutionary stable strategy.

1. Introduction

A social network consists of a series of people and relationships between them that these relationships can be information exchange, financial exchanges, interpersonal relationships including friendship, kinship and etc in social networks, individuals or organizations are represented as a node

*Corresponding author
Email addresses: Saeedehnoori00@gmail.com (Saeedeh Noori), Navidi@shahed.ac.ir (Hamidreza Navidi), Reza.Bakhtiari@shahed.ac.ir (Reza Bakhtiari)

Received: February 2021 *Accepted:* September 2021

or head and the social relations between them with the edge. Headers contain a user or set of users within networks, and edges are interdependencies between users that connect them. These heads and edges can include neighborhoods, job positions, or web pages.

Various researches have shown that by using the capacity and characteristics of social network, it is possible to identify some problems and find their solution, establishing relationships in social networks, managing these relationships and guiding them toward the target.

The importance of social network analysis is because relationships between individuals are more important than other attributes. Using this analysis can be a good scientific justification for many real-world phenomena. One of the old challenges in social networks is finding nodes that are studied for network analysis. A researcher may decide to study a group of people with a particular disease. However, it is difficult to know which of the patients in this study are appropriate. After determining appropriate nodes to analyze a social network, we need to determine the relationships between them. These relationships can be exchanged, exchanges, friendships, or any possible relationships between these members.

The identification of influential users in the social network has been studied from various angles and various models have been proposed to solve it. In [3], maximizing the influence of social networks using genetic algorithms is described by Jamal Kaksori. In [12], maximizing the influence on social networks with the help of a waterfall model is presented by Mehran Rostamnia. In [4], this problem is assumed to be a discontinuous optimization problem. For most of the models studied, including the model described in [2], the optimal solution is NP-hard. The framework presented in [11] is based on a simple linear model that is the optimal solution to the problem by solving a system of linear equations.

In this research, a new perspective has been proposed to find the most influential social network individuals. These people are selected with this goal that, if we provide them with a product or product, they introduce it to their relatives and relatives and cause the most spread of the idea or product in the network. These users are the social network of primary seeds.

1.1. Social Networks

The social network is expressed as a directed graph $g = (v, e)$ where v denotes the vertices and e is the edges between them. Effects can only propagate in one direction of edges. If we consider a undirected graph, such as friendship-based relations, each edge in this social network can be considered a bi-directional edge in both directions. For the vertices of the network one of these two states is possible. Active state, i.e., accepting a behavior by ross and passive state, i.e., not accepting behavior by ross. In most models, the tendency of each member to activation is an increasing function of activation of those around him. So, if the majority of friends accept a behavior, the individual 's desire to become active and accept it increases. Social networking has become a great effort to disseminate information and marketing because there has recently been a large number of social networking sites to influence people. Therefore, there is a possibility that the opinions and information of a large population can affect a short time. So many problems must be solved in this area including finding influential characters in the social network [4, 6].

1.2. Game Theory

Game theory is one of the tools for modeling multi-person strategic decisions and interpersonal interactions. Game theory tries to analyze the behavior and selection of individuals in game or strategic situations, in which one's success depends on the choice of others. Each player or player chooses the behavior that has the highest rewards. Game theory has been widely used in other sciences [7]. The main structure of this theory in most cases has a multidimensional matrix whose dimension reflects the outcome of a combination of different choices by players [1, 10].

1.2.1. Sustainable Evolutionary Strategy

An evolutionary stable strategy (ESS) is one of the topics of game theory. A sustainable evolutionary strategy can be stated that if there is a small mutation of this type of strategy, over time this strategy covers the whole society and will change society to its kind and will also dominate other strategies. Therefore, in order to select individuals as seeds to promote the product and to spread the idea, it is better to appoint people who have sustainable evolutionary strategy [1, 13]. Finally, formal definition of sustainable competitive strategy is discussed and it should be noted that within the framework of this research, we focus on symmetric games. The set of pure strategies in game theory is $K = \{1, 2, \cdots, k\}$ described as and related mixed strategies to form:

$$\Delta = \{x \in R^{+K} : \sum_{i \in k} x_i = 1\}$$

The outcome of the $x \in \Delta$ strategy when displayed in front of $y \in \Delta$ is displayed in the form $u(x, y) = x$, which will ultimately have a useful matrix for the interaction of all players. It should also be noted that the set of best answers $x \in \Delta$ against any $y \in \Delta$ strategy is $\beta^*(y) \subset \Delta$ It can be said that the $x \subset \Delta$ strategy is evolutionarily stable if it holds true:

1) $u(y, x) \leq u(x, x)\ \forall y$
2) $u(y, x) = u(x, x) \ \rightarrow \ u(y, y) < u(x, y)\ \forall y \neq x$

1.3. Clustering k-means

The most practical method for clustering data is k-means method. This method was proposed by McQueen in 1967. In this method, the number of clusters is predetermined and fixed. The method mentioned is suitable for data that is quantitative and have a mean value. The algorithm of this method is that the data first is partitioned into k groups and the distance between objects is computed from the centroid. The object is then moved closer to the nearest neighbor if the distance of each object is greater than the mean of the cluster. This process continues until members of the cluster are unchanged.

Algorithm 1 clustering algorithm k

Step 1: Divide the existing data into k clusters as desired
Repeat step:
1. Calculate the distance of each object from its center
2. Calculate the error function
Improvement step: Move the member that is less distant from the center of other clusters than the center of its own cluster to a closer cluster.
Stop step: Do not stop the members of the clusters or reduce the error function.

If we assume D is the set of objects with n objects and $C_1, C_2, \cdots, C_k$ represent K as a separate class D, the error function (EF) is defined as the sum of the distances of each object from its centroid.

$$EF = \sum_{i=1}^{k} \sum_{x \in C_i} d(X, \mu(C_i))$$

In this relation μ represents the center of the cluster and $d(X, \mu(C_i))$ represents the distance of the object from the center of its cluster. In this research, the distance is calculated on the basis of Euclidean and the aim is to minimize the error function [9].

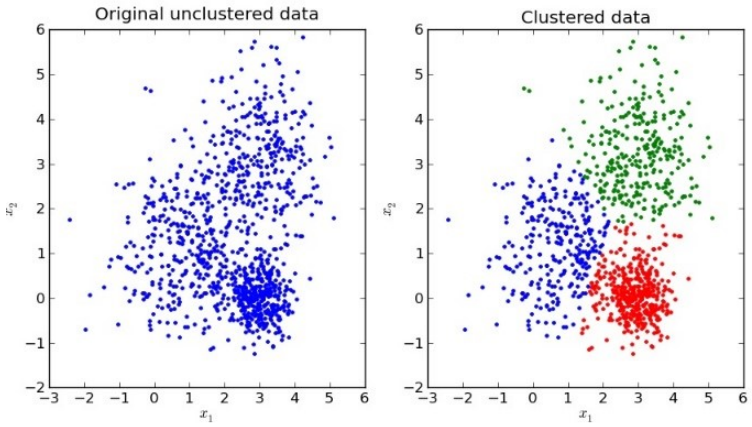

Figure 1: An image of the data of a social network before clustering and after clustering

1.4. Differentiation Factor

In this study, we intend to increase the separation accuracy by making changes in estimating the distance between data in the clustering method. The problem with previous methods is that the coefficient of all attributes in the corresponding attribute space is considered equal to 1, while in many real-world problems the importance of attributes in a set of data and thus in deciding on their type It will be different. To solve this problem, by calculating the weight of each attribute in the data set, we get a more realistic position of the data. In this way, the data of the same type are closer to each other, so they are separated from other data classes more carefully. There are several methods for weighting and prioritizing the characteristics, one of which is the differentiation factor method, which is described below.

In this method, first a specific and uniform interval for changing the weight of the characteristics is considered. Then, by assigning different weights in the range specified above to one feature and keeping the weight of other features constant in each step, the standard deviation of the data relative to the assigned weight is recorded. These steps are repeated for all features. The different weights assigned determine the importance of each attribute relative to the others by the following strategy:

An attribute that changes data more than others in exchange for a change in its weight by one standard deviation is more important, or in other words, the data is more sensitive to changes in that attribute. As a result, it will be more effective in determining the type of data.

It should be noted that the above strategy prioritizes only the characteristics and their significance coefficient is determined experimentally according to the problem conditions.

If the studied data have n characteristics, we do the following to calculate the differentiation factor:

$$Df_i = \sum_{j=1}^{n} w_j x_{ij} : \ i = 1, \cdots, m$$

And consider the weight vector $W = \{w_1, \cdots, w_n\}$ so that w_j for each $j = 1, \cdots, m$ is the optimal weight assigned to the i-th characteristic [8].

2. Provide a proposed algorithm

According to research on the solution of penetration maximization, most of them are looking for nodes that, if identified as the primary seeds, have the most influence on the desired social network. In

this study, in order to reduce the computation time and their complexity, instead of finding influential people in the network, a group of influential people in the social network has been found, and also to obtain this influential group of people, the average K clustering method The most appropriate method is to cluster large data, which we use, which increases the accuracy of classification. Then, by having the degree of influence of person to person in this society and using the calculation of the competency function, we obtained the degree of influence of each cluster towards itself and other clusters. Taking into account the calculated group influences, we will form an evolutionary game table and find a sustainable evolutionary strategy. In this case, it is initially assumed that not all social network users are aware of the existence of an idea or product to be introduced to the public, and their strategy is not to adopt that product or idea. When a company presents its idea to influential network users, there is in fact a leap in society, so it can be modeled using evolutionary game theory. According to the above, a framework for solving the problem of maximizing penetration in social networks with the approach of evolutionary game theory was presented. In this section, to test the functionality of this framework, the steps expressed on real data from a social network will be tested and the result will be fully explained.

These steps are expressed in the following algorithm:

Algorithm 2 The proposed algorithm for solving the penetration maximization problem

1: Extracting data from the studied social network.
2: Calculation of individual influence.
3: Calculation of data differentiation factor according to the contents of subsection 1-4.
4: Clustering of user differentiation factor by k-method based on algorithm 1.
5: Calculate the group influence or the same competency function according to the relation stated in section 2.
6: Create an evolutionary game table using the calculated values of the competency function.
7: Find a group of influential users by calculating a sustainable evolutionary strategy according to the conditions mentioned at the end of this section.

For further explanation, Figure 2 on the right shows a graph of the members of the social network, the values entered on each node indicate the difference between them, and also the clustering of users in this social network is shown on the left.

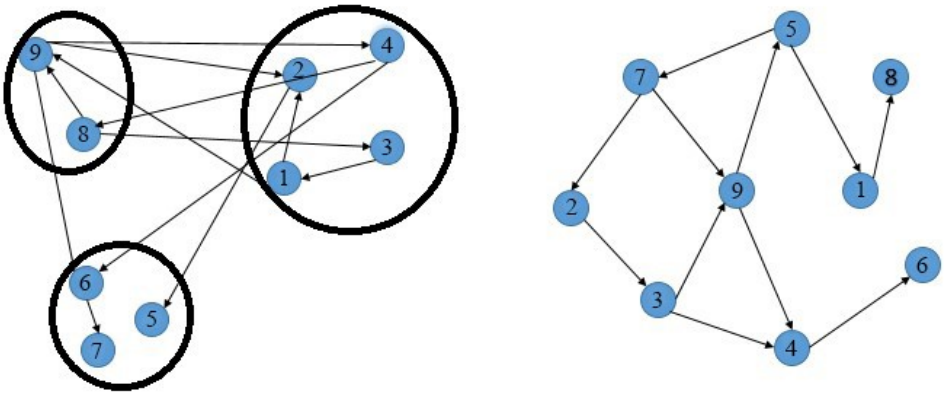

Figure 2: Graphics of a hypothetical social network

After clustering the members of the social network and determining the strategy of each cluster, we must examine the behavior of each cluster towards the other clusters. In fact, the competence

of clusters in relation to each other should be calculated in the theory of evolutionary games, which is the probability of influence that individuals of groups have on each other, and this is not the possibility of mutual influence. In evolutionary games, the behavior of each member in a cluster against its own groupmates and individuals in other groups must be analyzed. To calculate a single probability for each category, the average influence of the members of that category or competency function is used.

If it is assumed that the members of the community are in n categories, the calculation of the competency function of the categories with strategy A_i on the categories with strategy A_j will be according to the following formula:

$$u(A_i, A_j) = \frac{\sum_{a \in A_i, b \in A_j, a \neq b} P_{a,b}}{|\{(a,b) : (a,b) \in E, a \in A_i, b \in A_j, a \neq b\}|} \quad for\ i, j = 1, 2, \cdots, n$$

$P_{a,b}$ indicates the probability of user a's influence on user b. For example, if we have $n = 3, 9$ numbers are obtained. Using these numbers, the evolutionary game table is obtained in the form of table 1.

Table 1: Consequences of the evolutionary game

Outcome table	A_1	A_2	A_3
A_1	$U(A_1, A_1)$	$U(A_1, A_2)$	$U(A_1, A_3)$
A_2	$U(A_2, A_1)$	$U(A_2, A_2)$	$U(A_2, A_3)$
A_3	$U(A_3, A_1)$	$U(A_3, A_2)$	$U(A_3, A_3)$

Finally, using this table and considering the different modes of study of sustainable evolutionary strategy, an influential category in society is determined. Figure 3 illustrates this.

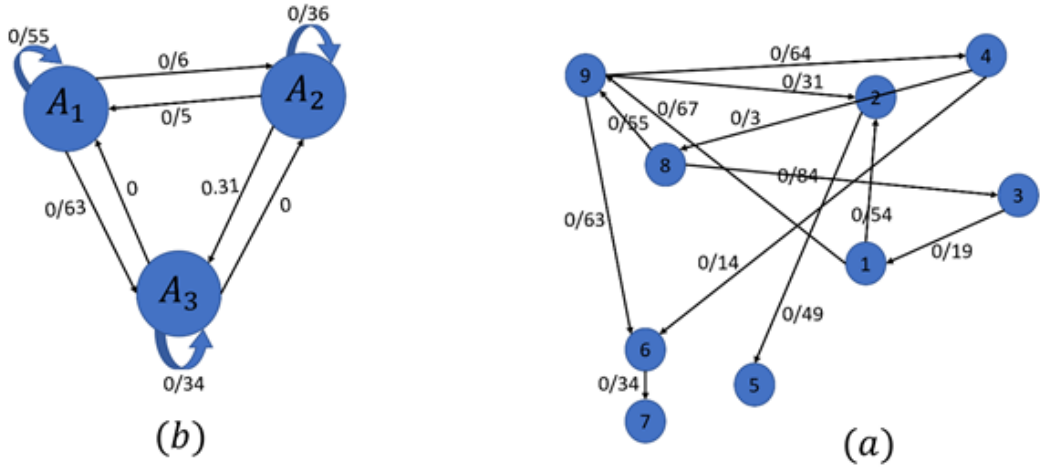

Figure 3: Calculation of the competency function for the sample graph

Part (a) The probability of penetration of social network members is shown as the weight of the edges and in part (b) the influence of clusters on each other is calculated according to the competency function.

In the graph of the above example, the table of the consequences of the evolutionary game, which in fact consists of the competency function numbers, is shown in table 2.

Table 2: Table related to the consequences of the evolutionary game

Outcome table	A_1	A_2	A_3
A_1	0.55	0.6	0.63
A_2	0.5	0.36	0.31
A_3	0	0	0.34

Once the game table is ready, we need to identify a sustainable evolutionary strategy, which will vary depending on how well the clusters penetrate each other:

To explain these conditions, we summarize the evolutionary game table as table 3:

Table 3: Summary table of the consequences of the evolutionary game

Outcome table	A_1	A_2	A_3
A_1	a	b	c
A_2	d	e	f
A_3	g	h	i

A) If a dominant strategy can be recognized as a sustainable strategy that:

1. If $a > d$ and $b > e$ and $c > f$, strategy A_1 defeats strategy A_2 .
2. If $a > g$ and $b > h$ and $c > i$, strategy A_1 defeats strategy A_3 .

In the same way, each of the remaining strategies is able to defeat the others and remain alone in society.

B) If one of the symmetric strategies (A_1, A_1) or (A_2, A_2) or (A_3, A_3) is out of balance, these strategies are evolutionarily stable.

1. If $a > d$ and $a > g$, the strategy (A_1, A_1) is a strict equilibrium.
2. If $e > b$ and $e > h$, the strategy (A_2, A_2) is a strict equilibrium.
3. If $i > f$ and $i > c$, the strategy (A_3, A_3) is a strict equilibrium.

C) If the game does not have a strict Nash equilibrium, it is necessary to pay attention to the mixed Nash equilibrium and check their evolutionary stability [8].

given the conditions expressed for evolutionary stable strategy, if we examine the evolutionary game table of the sample graph, we conclude that in this hypothetical social network strategy A_1 is an evolutionary sustainable strategy. Because the numbers in the first row of the table are larger than those of the other two rows. As A_1 result, strategy a beats up two more strategies and eventually embraces the whole society.

3. Case Study

The network studied in this chapter is a network of students of Abrar non-profit university compiled by Kermani et al [5]. This social network contains 163 students of Abrar Non-Profit University located in Tehran, who are studying industrial engineering and computer engineering. The relationship defined on this network is the messaging relationship. In this way, there is a

relationship from student a to student b if student a has the ability to send a message to student b. These students have three characteristics: the number of received messages, the number of answered messages and the number of forwarded messages. First, by weighting the features, we better display the differentiation of users and thus ease the categorization. In the process of implementing the differentiation factor, we will reach a matrix of standard deviation of the data in relation to the assigned weights to the characteristics, which is as follows:

$$\begin{bmatrix} 36.866 & 36.091 & 35.498 \\ 42.534 & 42.534 & 42.534 \\ 48.601 & 50.579 & 51.183 \end{bmatrix}$$

Therefore, considering the comparison of standard deviation changes in different characteristics and according to the data type, we define the optimal weight vector as follows:

$$\begin{bmatrix} 11.735 & 14.488 & 15.685 \end{bmatrix}$$

As mentioned in the previous section, standard deviation is a good tool for examining changes in coefficients, and the more a coefficient leads to standard deviation, the more that coefficient distinguishes social network members from each other. In this study, each of the coefficients is changed to $\{1, 2, 3\}$ and the other coefficients are considered constant and equal to the value of 1. Figure 4 shows the standard deviation changes based on the coefficient changes.

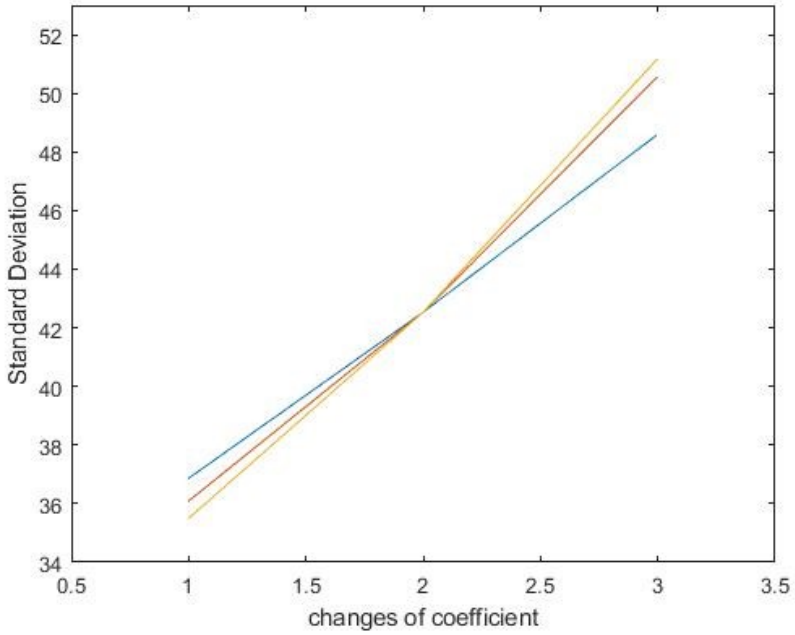

Figure 4: Diagram of changes in standard deviation based on changes in each of the coefficients

The lines of this diagram represent the changes in standard deviation based on one of the coefficients, and the line with the most slope means that it leads to the most changes and is related to the strongest coefficient.

Taking into account the above weights and after determining the differentiation factor of 163 students, we categorized the resulting numbers using the K method and assuming $K = 3$. Clustering is in 5 categories.

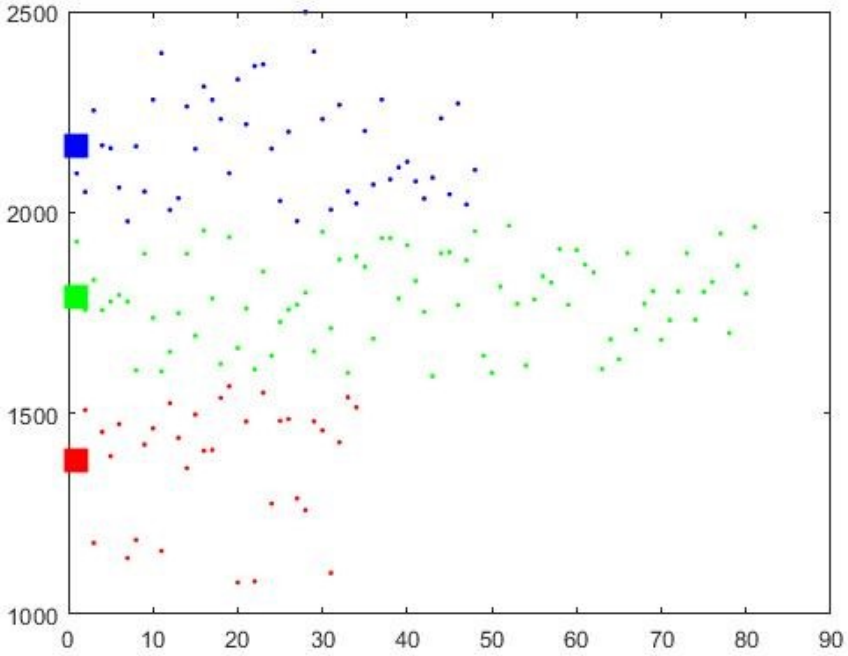

Figure 5: Examples of data classification of 163 students of Abrar University in MATLAB software

Also, the graph of the error function in K-clustering of the average is shown in Figure 6 and shows that this algorithm gradually achieves the lowest possible error, which means the error function here is the total distance of each user from his own cluster.

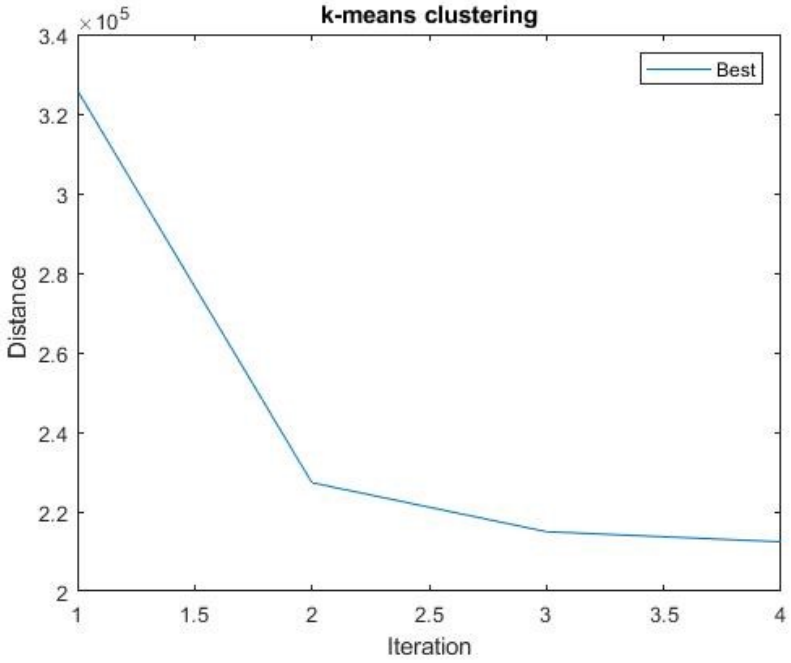

Figure 6: Diagram of the error function in K-clustering The average of the data of Abrar University students

After weighting, using the average K clustering method, we divide the data into 3 categories

and determine the strategy of each group, as well as the behavior of each cluster. In fact, the study of this behavior is the calculation of the competence of individuals in each category against other categories in the theory of evolutionary games.

For the given data, the table of sustainable evolutionary strategy is in the form of table 4:

Table 4: Evolutionary game implications for Abrar University data

Outcome table	A_1	A_2	A_3
A_1	0.06264	0.07082	0.08082
A_2	0.06845	0.07184	0.06844
A_3	0.08480	0.07114	0.07299

After preparing the table related to the game, we must determine the sustainable evolutionary strategy. According to the conditions mentioned in section 2, it can be concluded that the A_2 strategy is the sustainable evolutionary strategy. This means that in the end only group 2 can remain in the population and other groups will be excluded from society due to their lower qualifications than group 2 and will be replaced by people in the second group. Therefore, it can be said that the best option and choice to promote the product and spread the idea in this community, the users are in the second group and other groups in the community will be defeated. We can also note that if people in the community decide to change their strategy, the users of the second category will be stable in the community, and all users of the community tend to be present in this case and they will be more competent.

The program is written in the framework of this research for data related to 163 students of Abrar University was successfully implemented in 4.6031 seconds on a system with $AMD(A6-6310)$ processor and $6GB$ RAM and Windows 10 pro in MATLAB R2018b software and resulted in an error function equal to 212433.3051.

4. Conclusion

In this paper, an algorithm for improving network dissemination through data weighting and clustering is presented, which results in an accurate and practical answer to the problem of identifying the most influential people in the social network. In this context, using game theory, a game was considered between social network users and also by using the differentiation factor, the importance of users and their characteristics were displayed, based on which the importance of social network members was clustered and according to merit. Users derived from their influence on each other formed a sustainable the evolutionary game, which was eventually identified as a sustainable evolutionary strategy, the most influential members of the network were identified and considered as the primary seeds.

This framework was applied to data collected from the social network of Abrar Non-Profit University located in Tehran. According to information such as the influence of each user on other users and characteristics such as the number of messages received, the number of messages read, and the number of messages answered, the importance of the mentioned social network users was calculated and it was found that the second category users are more famous and influential than others. They are network users and have a sustainable evolutionary strategy, so they dominate other strategies and permeate the majority of society.

References

[1] H. Abdoli, *Game theory and its applications (incomplete, evolutionary and collaborative information games)*, Organization for the Study and Compilation of University Humanities Books (Samat), Center for Research and Development of Humanities, 2013.

[2] P. Domingos and M. Richardson, *Mining the network value of customers*, In: Proc. Seventh ACM SIGKDD Int. Conf. Knowledge Discovery and Data Min., ACM, 2001.

[3] J. Kaksury, *Maximize influence in social networks using genetic algorithms*, IranDoc, 2017.

[4] D. Kempe, J. M. Kleinberg and É. Tardos, *Maximizing the spread of influence through a social network*, In Proc. 9th ACM SIGKDD Conf. Knowledge Discovery and Data Min., 2003, pp. 137–146.

[5] M. A. A. Kermani, A. Aliahmadi and R. Hanneman, *Optimizing the choice of influential nodes for diffusion on a social network*, Int. J. Commun. Syst., 29(7) 2016 1235-1250.

[6] S. Jafari and H. Navidi, *A game-theoretic approach for modeling competitive diffusion over social networks*, Games, 9(1)(2018) 8 .

[7] E. Lotfi and H. Navidi. *A new model for determining the level of OPEC oil production based on price forecasting and game theory*, J. Model. Eng., 11 (35) .

[8] M. H. Manshei and P. Mousavi, *Sustainable evolutionary strategies in maximizing the influence of social networks*, Ministry of Science, Research and Technology - Isfahan University of Technology - Faculty of Electrical and Computer Science, 2014.

[9] M. Momeni, *Data clustering (cluster analysis)*, Mansour Momeni, 2011.

[10] H. Navidi, S. Ketabchi and M. Masibidgoli, *An Introduction to Game Theory*, Shahed University, Vice Chancellor for Research, Printing and Publishing Center, 2011.

[11] M. Richardson and P. Domingos, *Mining knowledge-sharing sites for viral marketing*, In: Proc. Eighth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Min., ACM, 2002.

[12] M. Rostamnia, *Maximizing the influence of social networks with the help of cascading model*, 2017.

[13] J. M. Smith, *The theory of games and the evolution of animal conflicts*, J. Theor. Biol. , 47 (1) (1974) 209-221.