WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM

e-ISSN: 2788-5879 p-ISSN: 2788-5879

Providing an Efficient Method to Identify Structural Balanced Social Network Charts using Data Mining Techniques

Razieh Asgarnezhad[1], Safaa Saad Abdull Majeed[1,2], Zainab Aqeel Abbas[1,2] and Sarah Sinan Salman[1,2,*]

[1]Department of Computer Engineering Isfahan (Khorasan) Branch Islamic Azad University Is- fahan , Iran
[2]Ministry of Education Educational Directorate of Wasit province Wasit Iraq

*Corresponding Author: Sarah Sinan Salman

DOI: https://doi.org/10.31185/wjcm.Vol1.Iss1.22
Received: January 2022; Accepted: March 2022; Available online: March 2022

ABSTRACT: As social communications become widespread, social networks are expanding day by day, and the number of members is increasing. In this regard, one of the most important issues on social networks is the prediction of the link or the friend's suggestion, which is usually done using similarities among users. In the meantime, clustering methods are very popular, but because of the high convergence velocity dimensions, clustering methods are usually low. In this research, using spectral clustering and diminishing dimensions, reducing the amount of information, reduces clustering time and reduces computational complexity and memory. In this regard, the spectroscopic clustering method, using a balanced index, determines the number of optimal clusters, and then performs clustering on the normal values of the normalized Laplace matrix. First, the clusters are divided into two parts and computed for each cluster of the harmonic distribution index. Each cluster whose index value for it is greater than 1 will be redistributed to two other clusters, and this will continue until the cluster has an index of less than 1. Finally, the similarity between the users within the cluster and between the clusters is calculated and the most similar people are introduced together. The best results for the Opinions, Google+ and Twitter data sets are 95.95, 86.44 and 95.45, respectively. The computational results of the proposed method and comparison with previous valid methods showed the superiority of the proposed approach.

Keywords: Efficient Method, Identify Structural, Balanced Social Network Charts, Data Mining Techniques

1. INTRODUCTION

With the expansion of social networks, the tendency of users to expand to these networks. In order to benefit from the benefits of these networks, users sought to identify and communicate with other users. Because what is supposed to be the main basis of social networks is the individual's friends and their relationship with them, and the amount of information they obtain through communication with them [1].

But the main issue in these networks is an increasing amount of information, which makes it difficult for users to identify and select others as friends. On the other hand, each user tends to prioritize communicating to users who have the most similarity (based on their interests, expertise, and activity). Based on the links that other friends have established, users who are reasonably likely to be similar to the user will be encouraged to be offered. For this purpose, it is necessary to have a system that, taking into account all the necessary information, offers the best person in each position as a friend's friend [2].

A lot of research has been done on friend suggestion systems in the social network. Many of these researches are used by common friends, and if they have two or more people in common, they will propose a link between the two [3]. Other

researches use their own characteristics, such as names, activities, and the number of friends at the moment and suggest them based on similarity between individuals. What is usually and mainly for research in this area is that many of them have tried to offer suitable friends to each user using the clustering method and putting the same users in a cluster. In short, the clustering approach is more widely used than other methods, which clustered users based on specific features, and clustered among similar users in a cluster, friendship is established.

In [4], they have proposed an algorithm based on the spectral clustering method for prediction of the bond; however, the related method is not balanced and can therefore be further improved. In the same vein, the current research uses a novel weights algorithm for the variables in the matrix used for clustering and determines the best weights in order to accurately predict the suggestion of friends and relatives Maximize This approach can improve the performance of the model relative to the standard (no weight).

In short, in this research, we try to identify and recommend to our users suitable spatial clustering techniques. In this way, by reducing the dimensions of the matrix, in addition to addressing the problems mentioned above, it is possible to store large volumes of social networking information and reduce computational complexity. In addition, the use of a balanced approach in performing spectral clustering and finding similarity of users can be superior to those in the literature on subject matter.

2. LITERATURE REVIEW AND RELATED WORKS

The research [5] attempts to provide a comprehensive model for offering friends on social networks. To this end, in addition to identifying appropriate friends based on concepts like common friends or friends of friends, the discussion of content publishing has also been considered. To this end, it is attempted to suggest suitable friends to the user in light of the social policy of the social networks based on the tendency of users to disseminate information, which, in addition to the higher probability of their acceptance by the users, tends to add more information to the user. Give in this model, two friend invariant factors and algebraic connection are used, which is calculated by calculating the Laplace conjugate matrix with the algebraic connection factor.

In the research [6], they first argued that most existing research and existing systems offer user friendliness based on their social graph, while this graph may not well reflect user preferences and interests. For this purpose, in this paper, a system is provided that the identification of the appropriate friends for each user does not limit their input information to network information only. In this research, we try to extract information about the lifestyle of users with the help of sensors. Then, using the techniques of the text, the similarity between the different styles of life of the users is measured and tried to identify the same users and suggest them as suitable friends. The method of operation of this system is to offer a list of similar people by asking the user and using the method described above. Finally, the system receives feedback from the user that it can use to improve its performance.

The paper presented in [8] states that all previous research and systems in the field of friend suggestion in the social network operate only on the basis of the connections between individuals and their strength or weakness; if the important factor to consider The characteristics of these people and the degree of similarity between the characteristics are taken. Therefore, in this paper, after reviewing and evaluating past work in this field, the proposed friend friendliness system has used the general information of individuals and FRUTAI has been named. This system can be applied to all social media because it is independent of the network structure. Also, with increasing user characteristics and features, it can be expanded with little cost.

In the research [9], they proposed a friendship suggestion system that uses the social impact of individuals for their suggestions. It is stated in this article that each person's network faces a huge amount of information that his friends share. Therefore, in order to avoid data overflow, the number and type of friends should be optimized. The most important thing about identifying people on the social network is the posts they bring to their friends. For this reason, the system uses the information distributed by each person for his ranking algorithm and the identification of his preferences. Now, if a person is offered to another as a friend, the preferences and interests obtained by this algorithm are identical to them.

In research [10], several factors have been taken into account in order to increase the accuracy of the model in proposing new friends to each user. These factors include the number of friends, the number of common member groups, the number of followers and followers. Also, in order to increase the speed of the offers to users and add the scalability feature, the proposed model is designed in an incremental way and therefore, for adding or removing a new link, there is no need to review all the existing links.

In the proposed system, a new, well-informed, location approach is also used. However, this cannot be the only criterion to suggest a friend to people. Hence, this system has the same level of self-interest in proposing to individuals. Also, in order to increase the accuracy and efficiency of the suggestions provided, the current location of the users is also considered, and in fact, it can be said that the system is somehow conscious of the context.

The proposed model [11] attempts to present a model for the suggestion of new friends by focusing on the network

topology and utilizing the genetic algorithm. To this end, we try to focus on the characteristics of the centrality as well as the application of the genetic algorithm in order to extract user preferences from friendship links, so that best friends can be offered to each user.

In a research [12] in his proposed model, he proposed new friends focusing on user interests. User preferences are modeled in two dimensions: field (time, location) and content. In order to extract user desires, it uses two categories of existing models: item-based models and user-driven models. In item-based models, the properties of each item are extracted and then extracted from the user's preferences based on user preferences. In user-based models, the interests of each user are extracted and analyzed.

3. PROPOSED METHOD

On each social network, users can select people as friends and can share their information with your friends, such as photos, news, personal information, and more. The existence of such links in a structured social network forms as a network, which is shown as a non-directional graph (Fig. 1)

FIGURE 1. **Graph structure in a social network**

.

The proposed method in this study will use spectral clustering to determine the similarity of the people and the suggestion of a friend to them. Then, based on an indicator, the similarity between clusters and clusters is calculated, and the most similar individuals are selected for suggestion as a friend. Note that in a graph derived from a social network, the nodes equivalent to individuals and arcs are equivalent to having friendship among individuals. The method uses the graph of the social network as an input and calculates a matrix called the Similarity matrix as output. Figure 2 is a demonstration of this fact.

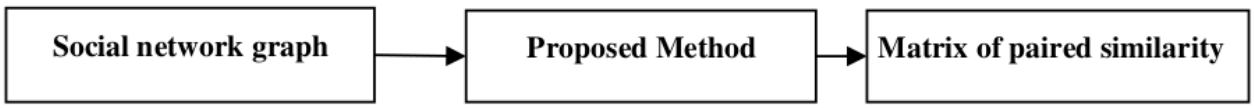

FIGURE 2. **Input and Output of the proposed method**

The similarity matrix is the proposed method of a square matrix, (i , j) in which the component reflects the similarity of i and j. accordingly, to a person like u, one can suggest the person with the most similarity on the output matrix.

Generally, Figure 3 illustrates the flowchart for the proposed algorithm.

To further illustrate the proposed method, further sub-sections will describe the steps involved.

3.1 WEIGHTED SPECTRAL DISTRIBUTION

Weighted spectral distribution is an indicator that can be used to determine the ability to perform clustering operations on data for a graph. This indicator, if larger than 1, indicates the fact that the existing graph is heterogeneous and the type

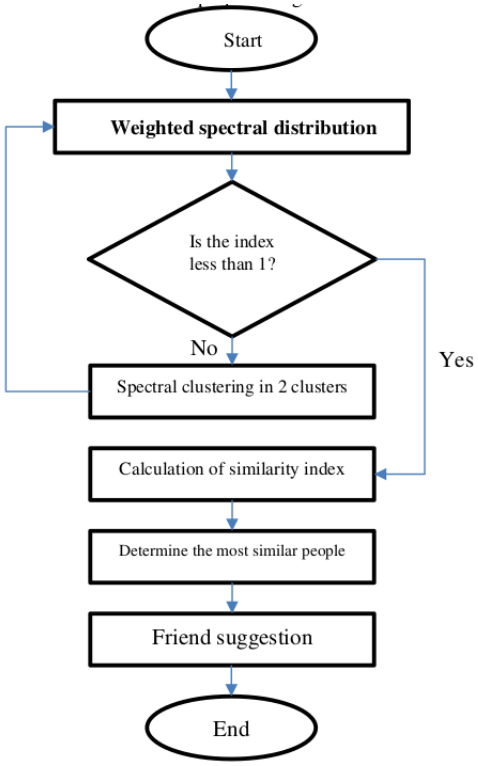

FIGURE 3. Flowchart for the proposed algorithm

of communication is not smooth. Therefore, it can be divided into two distinct clusters. On the other hand, if its value is less than 1, it means that the type of communication on the graph is homogeneous and there is no need for clustering. Accordingly, in order to implement the proposed method, it is necessary to calculate the spin-ray diffraction index for the observed graph. If the value of this index is greater than 1, then the graph is split into two clusters, otherwise the algorithm will stop. This index can be calculated using the following equation. This relationship is presented for classifying the operation, which is used in this research for data-clustering of social networking data.

$$\omega = \sum_{i-1}^{N} (1 - \lambda_i)^N \tag{1}$$

In which N is the number of nodes in the graph, and also i-th special value of the Laplace matrix is normalized. For the normalized Laplace matrix, further explanations will be given in the following sections.

3.2 CALCULATION OF SIMILARITY OF USERS

The final step of the algorithm will calculate the similarity between the elements. Researchers in [13] have introduced two functions for measuring the similarity of data, one for clustered data and one for data between two clusters. The research will also use the similarity method to calculate them, which are shown in accordance with the following relationships, respectively.

$$\text{sim}(i, j) = 1 - |D(i, I\ DX(i)) - D(j, I\ DX(j))| \tag{2}$$

$$\text{sim}(i, j) = \frac{1}{1 + D(i, I\ DX(j)) + D(j, I\ DX(i))} \tag{3}$$

Note that the value of the similarities will always be between 0 and 1, meaning zero means non-similarity, and 1 means complete similarity.

As a summation, the pseudocode of the proposed method is summarized in Fig. 3.

3.3 SIMULATIONS OF THE PROPOSED METHOD

In order to solve the proposed model, the data from Twitter, Google+ and Epinions social networks will be used. According to the proposed method, each set of data is divided into two clusters at each stage, and then the spectral

Input: Graph G
Output: Similarity matrix Sim

1. Compute weighted spectral distribution metric
2. **While** weighted metric is greater than 1
3. **Do**
4. Compute adjacency matrix A
5. Compute diagonal degree matrix D
6. Compute normalized laplacian matrix L
7. Find the eigenvalues of L
8. Compute eigenvectors corresponding to the selected eigenvalues (Matrix U)
9. Cluster data in matrix U using k-means with 2 clusters
10. **End while**
11. Compute similarity
12. Recommend friends base on maximum similarity

FIGURE 4. Pseudo-code of the proposed algorithm

harmonic distribution index for the clusters is calculated. The cluster (s) whose harmonic spectral index is less than 1 are homogeneous, and the rest are divided into 2 clusters. The clustering operation will continue to the point where the value of the harmonic spectral index for all clusters is less than 1. Upon completion of the clustering operation, 30% of the users are randomly selected and their connection with other users is eliminated. Then, the similarities are calculated with all other users, and similarities are found to exist with the predetermined value. In the final stage, the uniformity of the prior communications and communication is determined and the amount of the prediction accuracy of the link will be calculated. By analyzing the data set and using different values for the results, the final results are similar to those in Table 1.

Table 1. The results of solving the model with Euclidean interval function

Threshold	Dataset	Number of optimum clusters	precision (%)	Sensitivity (%)	Transparency (%)	Processing time (s)
8/0	Epinions	5	45/95	00/0	00/100	67/6
	Google+	2	84/80	003/0	59/92	59/3
	Twitter	7	45/95	00/0	00/100	89/4
6/0	Epinions	6	80/94	002/0	30/99	87/3
	Google+	2	44/86	00/0	05/99	66/8
	Twitter	5	56/92	002/0	97/96	26/3
4/0	Epinions	5	89/87	004/0	06/92	16/3
	Google+	2	53/85	00/0	06/98	61/3
	Twitter	5	89/87	004/0	06/92	43/3
2/0	Epinions	5	73/86	004/0	83/90	06/3
	Google+	2	84/80	003/0	59/92	59/3
	Twitter	5	73/86	004/0	84/90	48/3

Since the density of the proximity matrix is very low (i.e., the number of zero numbers on the adjacency matrix is much larger than the number of numbers 1), for comparison, one can use only precision. This fact is expressed in learning

issues. Accordingly, by comparing the obtained accuracy, it can be stated that the best accuracy obtained for the Epinions, Google+ and Twitter data sets is 95.95, 86.44 and 95.45, respectively. .

Using statistical tools, one can consider a lower bound and a high bound for the accuracy and processing time index. In this research, a box diagram for this work has been used. Data related to the accuracy and processing time are considered and based on which the corresponding box diagram for each of the data sets is drawn. Figures 5 and 6 illustrate the graphs for accuracy and processing time, respectively.

With respect to the accuracy graph, for example, it can be seen that the minimum value for the Epinions data set is 87% and the maximum value is approximately 95%. The same logic exists for other data sets. In addition, a similar interpretation can be used for processing time.

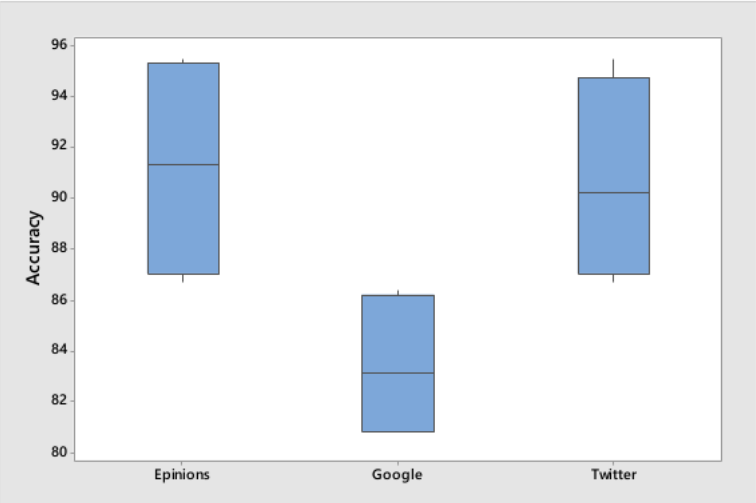

FIGURE 5. Accuracy box diagram by dividing the data set

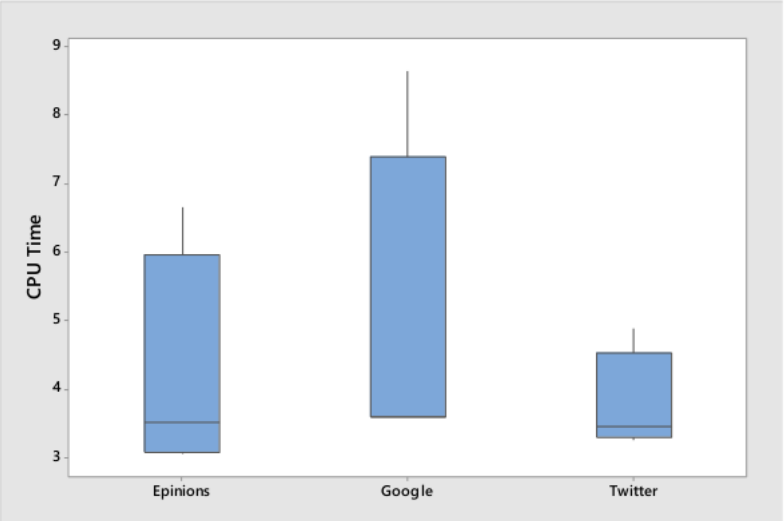

FIGURE 6. Box diagram of the processing time by dividing the data set

.

4. CONCLUSION

As social communications become widespread, social networks are expanding day by day, and the number of members is increasing. In this regard, one of the most important issues on social networks is the prediction of the link or the friend's

suggestion, which is usually done using similarities among users. In the meantime, clustering methods are very popular, but because of the high convergence velocity dimensions, clustering methods are usually low. In this research, by using spectral clustering and diminishing dimensions, the amount of information, the time of clustering, and the complexity of computing and memory were reduced. In order to further develop the proposed model, some issues are raised: the use of polar mapping can lead to improved performance of learning methods. Accordingly, it is recommended to map the polar coordinate axis after finding special vectors and then cluster the data. This can lead to improved performance indicators. The development of a proposed method for signaling networks, in which communications can be both positive and negative, can also be considered as a new approach, in which case the use of cluster Fuzzy fixing can be useful.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] D. Camacho, A. Panizo-Lledot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Information Fusion*, vol. 63, pp. 88–120, 2020.
[2] K. S. Krishnakumar and K. U. Maheswari, "A Comprehensive Review on Data Anonymization Techniques for Social Networks," *Webology*, no. 1, pp. 19–19, 2022.
[3] R. K. Langari, S. Sardar, S. A. A. Mousavi, and R. Radfar, "Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks," *Expert Systems with Applications*, vol. 141, pp. 112968–112968, 2020.
[4] W. Zou, X. Hu, Z. Pan, C. Li, Y. Cai, and M. Liu, "Exploring the relationship between social presence and learners' prestige in MOOC discussion forums using automated content analysis and social network analysis," *Computers in Human Behavior*, vol. 115, pp. 106582–106582, 2021.
[5] Z. Wu, Y. Lin, J. Wang, and S. Gregory, "Link prediction with node clustering coefficient," *Physica A: Statistical Mechanics and its Applications*, vol. 452, pp. 1–8, 2016.
[6] Z. Yu, C. Wang, J. Bu, X. Wang, Y. Wu, and C. Chen, "Friend recommendation with content spread enhancement in social networks," *Information Sciences*, vol. 309, pp. 102–118, 2019.
[7] Z. Wang and H. Qi, "Friendbook: A Semantic-based Friend Recommendation System for Social Networks", *IEEE Trans. Mob. Comput.*, p. 14.
[8] M. P. A. Kalaiyarasi, "Semantic-Based Friend Recommendation System for Social Networks," *Transactions on Mobile Computing*, vol. 14, no. 3, pp. 538–51, 2015.
[9] J. Chu, Y. Wang, X. Liu, and Y. Liu, 2020. "Social network community analysis based large-scale group decision making approach with incomplete fuzzy preference relations, Inf". *Fusion*, vol. 60, pp. 98120, Aug. 2020, doi: 10.1016/j.inffus.2020.02.005.
[10] Z. Zhang, Y. Liu, W. Ding, W. W. Huang, Q. Su, and P. Chen, 2019.
[11] F. Li, J. He, G. Huang, Y. Zhang, Y. Shi, and R. Zhou, 2015.
[12] P. Sun, S. Liu, N. Wu, B. Li, and J. Li, "Prediction and Clustering of User Relationship in Social Network," *Ubiquitous Intelligence and Computing and 2018 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2018 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2018 IEEE 12th Intl Conf on*, pp. 1155–1162, 2018.
[13] I. L. Hu, C. C. Chang, and Y. H. Lin, "Using big data and social network analysis for cultural tourism planning in Hakka villages," *Tourism and Hospitality Research*, vol. 21, no. 1, pp. 99–114, 2021.