Improved Leach algorithm for detecting opinion leaders in social networks

Afsaneh Razghandi^a, Yasser Elmi^a

^a, *Department of Computer Engineering, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran*

Abstract

Today, the development of human communication through social networks has increased the importance of social network analysis. In social networks, the role of nodes is not the same and important nodes play a key role in network performance. The nodes which have more influence in a social network are called "opinion leaders".

Measuring the importance of each node is needed in order to finding the opinion leaders of social networks. This parameter is measured using the centrality criteria in the network's graph. In this paper, the well-known Leach algorithm that used to find the cluster heads in wireless sensor network has been modified to determine opinion leaders in social networks. The evaluation of the proposed algorithm is performed using centrality (degree, closeness, betweenness and eigenvector) indicators. Simulation results show that the proposed algorithm has a significant relationship with the standard centrality indices with a correlation value greater than 0.7.

Key Words: Social Network Analysis, Opinion Leaders, Wireless Sensor Networks, Leach Algorithm, Centrality.

1- Introduction

Due to the increasing development of social communication through electronic tools, the analysis of social networks resulting from these communications has also gained special importance. Social network analysis has emerged as a key technique in modern sociology. The importance of this field of research is in humanities, biology, economics, geography, information sciences, and it is also studied as a popular topic in computer science[1,2].

Social networks are usually modeled as a graph. In this graph, the vertices represent the actors and the edges represent the relationship between them. The status of nodes in social networks is not the same and there is a big difference between their importance. Nodes in different situations have different levels of influence on network survival. In general, important nodes play a key role in social network performance. Different letters have been referred to these influential nodes in social networks. In most studies, these people are called "opinion leaders" and in other studies, they are referred to as "influential people", "market professionals" and "key actors" [3,4].

Measuring the importance of nodes have been done in order to find the influence of that nodes in network. In the context of social networks, the importance of nodes is calculated by centrality criteria.

1-1 Leach algorithm

Leach is a cluster-based routing protocol in wireless sensor networks [5]. The purpose of this protocol is to reduce the energy consumption of the nodes in order to improve the life time of the wireless sensor network. In standard Leach algorithm, the nodes are organized as clusters and in each cluster one node is selected as the cluster head[6]. In each round of this algorithm, $p\%$ of the nodes determined as the cluster heads. Each node is assigned a random value between 0 and 1. If this number is less than the threshold specified in Eq. 1, during current round the node will be determined as the cluster head, otherwise it will continue to operate as a normal node[7].

$$T(n) = \begin{cases} \frac{p}{1 - p * (r \bmod \frac{1}{p})} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

In Eq. (1), p is the desired percentage of cluster heads for each round that has been predetermined, r represents the current period number, and G is a set of nodes that were not cluster head in the previous $(r \bmod \frac{1}{p})$ step. This equation is designed so that at each $(\frac{1}{p})$ round, all of nodes are become cluster head only once, so the energy consumption is distributed throughout the network. Various versions of the Leach algorithm such as E-LEACH, TL-LEACH, M-LEACH, LEACH-C and V-LEACH have been proposed by researchers that more details in this regard can be studied in [8-12].

1-2 Centrality criteria

In the 1950s, attempts were made to define the graph centrality criteria. Different centrality criteria have been introduced to measure the influence of actors in the social network. The most common centrality that have been introduced are :1) degree centrality 2) closeness centrality 3) betweenness centrality and 4) eigenvector centrality. These centrality criteria are briefly discribed as follows:

1) *Degree Centrality* of each vertex is defined by the number of edges adjacent to that vertex.

$$C_D(j) = \sum_{j=1}^n A_{ij} \quad (2)$$

where A represents the graph adjacent matrix.

2) *Closeness Centrality* measures the distance of one vertex from all other vertices in the network's graph. This centrality is measured by the length of paths or steps required by one person to reach other people on a social network.

$$C(v) = \frac{1}{\sum_{t \in v} d(v, t)} \quad (3)$$

where $d(v, t)$ represents the distance between vertices v and t .

3) *Betweenness centrality* of vertex v is the number of times the other vertices of the graph pass through this vertex to reach each other. The betweenness centrality of vertex v is calculated as follows:

$$g(v) = \sum_i \sum_j \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (4)$$

where σ_{st} represents the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ indicates the number of paths that passes through v .

4) *Eigenvector Centrality* (also called eigen centrality or prestige score) is a measure of the influence of a node in a network. The eigenvector centrality score of node i , denoted by x_i , is calculated as follows:

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^N A_{ij} x_j \quad (5)$$

where $M(i)$ indicates the set of neighbors of nodes i and λ is a constant called the eigenvalue.

In this paper, using the concept of centrality and applying the improved Leach algorithm, opinion leaders of social networks are determined. By modifying the standard Leach algorithm and evaluating the results using the centrality criteria, a new criterion for determining centrality and finding opinion leaders has been proposed.

2- Related Works

Social networks analysis is the one of most important field of sociology, economics and computer science. Okamoto et al. [13], considered all the characteristics provided for opinion leaders, influential people, market experts and key actors, and presented a complete classification including structural, relational and individual characteristics for opinion leaders. Numerous researches have examined the types of centrality criteria which can be studied in [14]. In [14], using the concept of centrality (degree centrality and closeness centrality), the performance of social networks has been analyzed [15,16].

In [17], using graph mining algorithms, a solution for clustering social networks (based on the centrality of nodes) is proposed. In paper presented by Jensen et al., by calculating the centrality metrics for network nodes, important nodes are identified and placed in the center of the cluster. Nettleton et al. [18], using graph mining algorithms, offer a solution to improve social network clustering (based on edge betweenness algorithm). In their research, a big dataset is generated by modeling a social network using huge graphs (where nodes are the same as individuals, organizations, and groups), and big data analysis algorithms have been used to find opinion leaders of social network.

In paper introduced by Abdulsalam et al. [19], a distributed data aggregation method with a clustering structure was presented. In their method, the amount of energy consumption in the data aggregation step was compared and the step of cluster construction is considered for both algorithms (i.e. the proposed CDDA and the same Leach algorithm). This method has an effective improvement in energy consumption in environments with high data correlation compared to the Leach algorithm, while in environments with little correlation between data, the two algorithms work almost identically. In [20], a clustering method to reduce energy consumption in wireless sensor networks is presented, which is the most effective method for scalability and reducing energy consumption in wireless sensor networks. The simulation results show that algorithm

presented by Sasirekha et al. can reduce the energy consumption of the wireless sensor network and significantly increase the life time of these networks. In [21], a hybrid multi-step routing algorithm is proposed, which combines hierarchical and planar multi-path routing algorithms. Their primary goal is to minimize energy consumption in the wireless sensor networks. Approach to improve the network clustering for efficient use of energy in wireless sensor networks is presented by Baghouri et al.[22].

3- Proposed Method

The main idea of the proposed algorithm is to make changes in the standard Leach algorithm. These changes have been made with the aim of determining cluster head as the opinion leader of social network. Based on the standard Leach algorithm, a random number (between 0 and 1) is assigned to each node. If this number is less than a certain threshold, the node is selected as the cluster head. To analyze this method, the types of centrality criteria introduced in the previous sections were calculated. These results showed that the standard Leach algorithm has very little correlation with the calculated centralities. Therefore, changes were made in standard Leach algorithm to improve these results (in terms of the degree of correlation). In order to achieve better load balancing, proper distribution of cluster heads and reduce the random effect of the standard Leach algorithm, the parameter of physical distance of nodes was added to the Leach algorithm. These changes lead to meaningful relationship and correlation between the selected cluster heads of the modified Leach algorithm and the standard centralities. The pseudo-code of the proposed algorithm is as follows:

Algorithm1 : Pseudocode of proposed algorithm

N: node numbers, **S:** node characteristics, **M:** percentage of advanced nodes, **R_{max} :** maximum number of rounds, **G:** set of nodes that have not been supervised, **C1:** number of clusters

Part I: Operation for epoch...

```

for  $r = 0$  to  $R_{max}$ 
    if  $\left( \text{mod} \left( r, \left\lfloor \frac{1}{p} \right\rfloor \right) \right) = 0$  then
        for  $i = 1$  to  $n$ 
             $S(i).j \leftarrow 0$  ;  $S(i).c_l \leftarrow 0$ 
        end
    end
    for  $i = 1$  to  $n$ 
        if  $(S(i).E \leq 0)$  then
             $dead \leftarrow dead + 1$ 
            if  $(S(i).E = 0)$  then
                 $Advance \leftarrow Advance + 1$ 
            end
            if  $(S(i).E \neq 0)$  then
                 $Normal \leftarrow Normal + 1$ 
            end
        end
        if  $(S(i).E > 0)$  then
             $S(i).type \leftarrow 'N'$ 
        end
    end
    for  $i = 1$  to  $n$ 
        if  $(S(i).E > 0)$  then

```

```

        S(i).type ← 'N'
    end
    Ti ← random(0,1), i = 1..N
    if (S(i).E ≤ 0) then
        D ←  $\frac{p}{1-p \cdot \text{mod}(r, \lfloor \frac{1}{p} \rfloor)}$  D
    end
    if Ti ≤ D then
        C1 ← C1 + 1
    end
    S(i).type ← 'C' ; S(i).G ←  $\lfloor \frac{1}{p} \rfloor - 1$ 
    Di ←  $\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$ 
    Ti ← Ri3 + Di ; Tmax ← maxi∈N Ti
end
for i = 1 to n
    Ti ← Ti - Tmax
end
for i = 1 to n
    if S(i).type = 'N' AND S(i).E > 0 then
        if (C1 ≥ 1) then
            Dmin ←  $\sqrt{(x_i - x_{c1})^2 + (y_i - y_{c1})^2}$ 
        end
        for c = 1 to C1
            temp ← mini∈N (Dmin,  $\sqrt{(x_i - x_{ch})^2 + (y_i - y_{ch})^2}$ )
            if (Ti < Dmin) then
                Dmin ← Ti
            end
        end
    end
end
end

```

In original version of Leach protocol, all nodes with a fixed probability are selected as cluster heads, so some unsuitable nodes may be selected as cluster head. Therefore, some choices may cause rapid energy consumption in important nodes (which plays main role in the network because these nodes connect two subnets of social network graph) and this problem causes the network to be partitioned. Some changes have been made to improves standard Leach algorithm. These changes are expressed in Eq. (6)–(9).

$$D_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad (6)$$

$$T_i = R_i^3 + D_i \quad (7)$$

$$T_{max} = \max_{i \in N} T_i \quad (8)$$

$$T_i = T_i - T_{max}, i \in N \quad (9)$$

Using the distance criterion, the location of each node was determined, so it is expected that the possibility of selecting the cluster head will no longer be completely random. In addition, no changes have been made to the time complexity of the modified algorithm compared to the original Leach algorithm, because no new loops or similar commands have been added to the proposed algorithm in the modified version. The correlation coefficient can also be used to test the revised version.

4- Experimental results

After performing the simulations using the values and methods described, different results were obtained which are presented in this section. Correlation calculations are one of the most basic methods of "statistical analysis". Its purpose is to measure the type of relationship and the degree of similarity between different attributes of objects and phenomena that are being studied. "Correlation coefficient" is a statistical index that determines the different degrees of relationship between two dependent variables on a fixed and finite scale.

4-1 Correlation of standard datasets

In this section, the data correlation was calculated on 6 standard datasets. These datasets include social networks with a size of 34 to 1490 nodes. Table.1 shows the degree of correlation between the centrality criteria of standard datasets.

Table 1) The degree of correlation in standard datasets

Dataset Name	Karate	Celegansneural	Political-Blogs	Dolphins.Sex	Football	Lesmiserables
Size	34	297	1490	112	123	150
Betweenness	0.80	0.70	0.70	0.70	0.50	0.80
Closeness	0.06	0.60	0.01	0.10	0.80	0.06
Eigenvector	0.70	0.80	0.80	0.70	0.90	0.70
Betweenness-Closeness	0.03	0.40	0.007	0.10	0.50	0.03
Betweenness-Eigenvector	0.50	0.50	0.50	0.40	0.50	0.50
Closeness-Eigenvector	0.10	0.70	0.02	0.10	0.70	0.10

4-2 Solidarity on artificial social networks

In Table. 2, the analysis results of the correlation degree between the proposed algorithm and standard centralities was presented. Datasets are generated randomly and the number of nodes is between 10 to 80 nodes. Similarity of results with standard centralities were computed. Finally, correlation of 0.7 is reached. The proposed algorithm with the standard centrality criteria has a significant logical relationship and the correlation of their output values was more than 0.7.

Centrality criteria are calculated using NodeXL network analysis software. Then the correlation of the output of the proposed algorithm with the outputs of the NodeKL software is compared. 10 to 80 nodes have been compared. Table. 2 shows the degree of correlation between different centrality criteria compared to each other and compared to the modified Leach algorithm. It can be seen in Table. 2, the Eigenvector-Closeness criteria have the highest level of correlation and the Eigenvector-Leach criteria have the lowest level of correlation.

Table 2) The degree of correlation in artificial datasets

Size (nodes)/ Correlation	10	20	30	40	50	60	70	80
--------------------------------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

Betweenness	0.974	0.687	0.804	0.895	0.964	0.919	0.941	0.936
Closeness	0.997	0.994	0.988	0.995	0.997	0.993	0.994	0.969
Eigenvector	0.999	0.995	0.992	0.997	0.998	0.997	0.998	0.972
Leach	0.804	0.779	0.709	0.820	0.883	0.941	0.939	0.903
Betweenness- Closeness	0.989	0.756	0.867	0.929	0.978	0.952	0.966	0.984
Betweenness- Eigenvector	0.962	0.611	0.724	0.859	0.947	0.889	0.918	0.952
Betweenness- Leach	0.846	0.738	0.941	0.951	0.894	0.955	0.949	0.927
Closeness- Eigenvector	0.992	0.978	0.965	0.986	0.992	0.984	0.986	0.990
Closeness- Leach	0.823	0.809	0.774	0.858	0.895	0.965	0.960	0.924
Eigenvector- Leach	0.789	0.746	0.623	0.780	0.871	0.923	0.923	0.902

5- Conclusion

Conventionally, social networks can be represented in the form of standard graphs, in which the nodes are equivalent to the actors of social networks, and the edges of the graph indicate the relationship between the actors. Research shows that the status of nodes in networks is not the same and there is a big difference between their importance. At the same time, nodes in different situations have different levels of influence on network survival. Therefore, evaluating the importance of nodes in complex networks is not only useful but also critical. In fact, the main challenge for social network analysis is to identify key people in social network. In most studies, these people are called "opinion leaders" and in other studies. In this paper, using the concept of centrality (degree, closeness, betweenness and eigenvector) and modifying the standard Leach algorithm, opinion leader of a social network has been determined. The main idea is applying the well-known Leach algorithm in the field of social networks. The main functionality of this algorithm in selecting the cluster head of wireless sensor network has been modified in order to find the opinion leaders in the social networks. The simulation results show that the proposed algorithm has a significant logical relationship with the standard centrality indicators and the correlation of their output values is more than 0.7.

6- References

1. Carrington PJ, Scott J, Wasserman S (2005) Models and methods in social network analysis, vol 28. Cambridge university press,
2. Benítez-Andrades JA, García-Rodríguez I, Benavides C, Alaiz-Moretón H, Rodríguez-González A (2020) Social network analysis for personalized characterization and risk assessment of alcohol use disorders in adolescents using semantic technologies. *Future Generation Computer Systems* 106:154-170
3. Pradhan T, Pal S (2020) A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity. *Future Generation Computer Systems* 110:1139-1166
4. Wei J, Meng F How opinion distortion appears in super-influencer dominated social network. *Future Generation Computer Systems* 115:542-552

5. Liu X, Zhu R, Anjum A, Wang J, Zhang H, Ma M (2020) Intelligent data fusion algorithm based on hybrid delay-aware adaptive clustering in wireless sensor networks. *Future Generation Computer Systems* 104:1-14
6. Fu C, Jiang Z, Wei W, Wei A (2013) An energy balanced algorithm of LEACH protocol in WSN. *International Journal of Computer Science Issues (IJCSI)* 10 (1):354
7. Wang T, Zeng J, Lai Y, Cai Y, Tian H, Chen Y, Wang B (2020) Data collection from WSNs to the cloud based on mobile Fog elements. *Future Generation Computer Systems* 105:864-872
8. Xu J, Jin N, Lou X, Peng T, Zhou Q, Chen Y Improvement of LEACH protocol for WSN. In: 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2012. IEEE, pp 2174-2177
9. Loscri V, Morabito G, Marano S A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH). In: IEEE vehicular technology conference, 2005. vol 3. IEEE; 1999, p 1809
10. Mahmood D, Javaid N, Mahmood S, Qureshi S, Memon AM, Zaman T MODLEACH: a variant of LEACH for WSNs. In: 2013 Eighth international conference on broadband and wireless computing, communication and applications, 2013. IEEE, pp 158-163
11. Rahmanian A, Omranpour H, Akbari M, Raahemifar K A novel genetic algorithm in LEACH-C routing protocol for sensor networks. In: 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE), 2011. IEEE, pp 001096-001100
12. Ahlawat A, Malik V An extended vice-cluster selection approach to improve v leach protocol in WSN. In: 2013 Third International Conference on Advanced Computing and Communication Technologies (ACCT), 2013. IEEE, pp 236-240
13. Okamoto K, Chen W, Li X-Y Ranking of closeness centrality for large-scale social networks. In: International workshop on frontiers in algorithmics, 2008. Springer, pp 186-195
14. Freeman LC, Roeder D, Mulholland RR (1979) Centrality in social networks: II. Experimental results. *Social networks* 2 (2):119-141
15. Elmi Sola Y, Pourjavad MA, Hasanpour H, Zojaji H, Analoui M Load balancing effects in DCUR QoS routing algorithm. In: 2009 2nd IEEE International Conference on Computer Science and Information Technology, 2009. IEEE, pp 210-213
16. Fumanal-Idocin J, Alonso-Betanzos A, Cordón O, Bustince H, Minárová M (2020) Community detection and social network analysis based on the Italian wars of the 15th century. *Future Generation Computer Systems* 113:25-40
17. Jensen D, Neville J (2003) Data mining in social networks. na,
18. Nettleton DF (2013) Data mining of social networks represented as graphs. *Computer Science Review* 7:1-34
19. Abdulsalam HM, Kamel LK W-LEACH: Weighted Low Energy Adaptive Clustering Hierarchy aggregation algorithm for data streams in wireless sensor networks. In: 2010 IEEE international conference on data mining workshops, 2010. IEEE, pp 1-8
20. Sasirekha S, Swamynathan S (2017) Cluster-chain mobile agent routing algorithm for efficient data aggregation in wireless sensor network. *Journal of Communications and Networks* 19 (4):392-401
21. Kirsan AS, Al Rasyid UH, Syarif I, Purnamasari DN (2020) Energy Efficiency Optimization for Intermediate Node Selection Using MhSA-LEACH: Multi-hop Simulated Annealing in Wireless Sensor Network. *EMITTER International Journal of Engineering Technology* 8 (1):1-18
22. Baghoury M, Chakkor S, Hajraoui A (2014) Ameliorate Threshold Distributed Energy Efficient Clustering Algorithm for Heterogeneous Wireless Sensor Networks. *arXiv preprint arXiv:14084112*