IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Analysis of Social Network Based on Graph Theory and Fuzzy Logic

Dr.A.Kavitha[2], Dr. G.S.G.N.Anjaneyulu[*1], Shobhika Panda[3], AnneMary Joy[4]
[*1] School of Advanced Sciences, VIT University, Vellore-632014, India
[2,3,4] School of Computing Science and Engineering, VIT University, Vellore-632014, India

Abstract

A social network consists of a set of individuals connected by a binary relationship, which can be represented as friendship between them, together with the communities that they join and the information that they exchange through forums at a global level. Presently, there are more than 200 social networking websites. This paper will help researchers to implement search optimization. In this paper, we analyze the behavior of people in joining social communities and incentive of friendship between set of users, the weak and strong ties between user and community and the relationship between communities. Social network analysis provides visual analysis of human relationships by graph coloring and mathematical analysis by fuzzy C-means.

Keywords: Fuzzy logic, Social Network, Planar graph, K-Color theorem.

Introduction

In the knowledge society, social networking and technical skills go hand in hand. It has become important for people to branch out and expand their social connections for inter and intra personal activities.

It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familiar bonds. Search optimization techniques can be modeled based on the analysis. Although resources and opportunities may be available, one may not necessarily be aware of their existence, or even have direct access to them. In those cases, knowing people from different backgrounds, grades of expertise, and social levels turns out to be essential. For instance, who among us has never required a letter of recommendation in order to get a job, or had to rely on somebody else's knowledge in order to find a specific doctor? Isn't it through word of mouth that many kids get to know about the Clubhouse? International students get to know how to contact the students in other universities through forums available on social networks. Throughout our lives, we tend to form different kinds of relationships with different people. With people like us, our greatest buddies and closer relatives, we establish strong ties that reinforce our beliefs and provide us with the support required to endure life's challenges. With other people, with whom we usually do not have such in-depth connections, we develop weaker ties that help us address a larger variety of perhaps more specific needs. Even though we usually don't pay too much attention to our weak ties, some authors say that it is through them that we become aware of new opportunities and broaden our horizons. Some studies even demonstrate that, when searching for a job, in some cases having a wide network of weak ties is equally or more important than having strong personal skills.

This paper shows visual representation of social network using k-color theorem. The coloring of communities is essentially a topological problem, in the sense that it depends only on the connectivity's between the communities not on their membership strength. We can just as well represent each individual and community by a single point (vertex), and the adjacency between same individual belonging to different communities can be represented by a line (edge) connecting those two points. It's understood that connecting lines cannot cross each other. A drawing of this kind is called a planar graph. The graph is shown below.(fig.1).
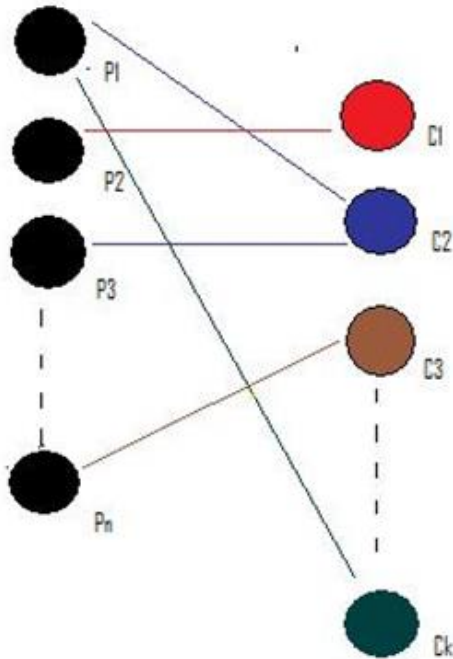
Fig1.

This paper makes use of fuzzy c-means to determine the membership of users belonging to different communities. The most probable question arises as to why we should use fuzzy logic. Fuzziness describes the uncertainty in the occurrence of an event for example a user can either belong to one or more communities or cannot belong to a community at all. So the membership function decides as to how strong or weak the user-community relationship is.

Methods
1) K-Color Theorem
How many different colors are sufficient to color the communities in such a way that no two communities have the same color?

A graph is said to be k-colorable if it's possible to assign one of k colors to each vertex in such a way that no two connected vertices have the same color.

We have taken K-colorable theorem from graph theory for visual representation of our social network scenario

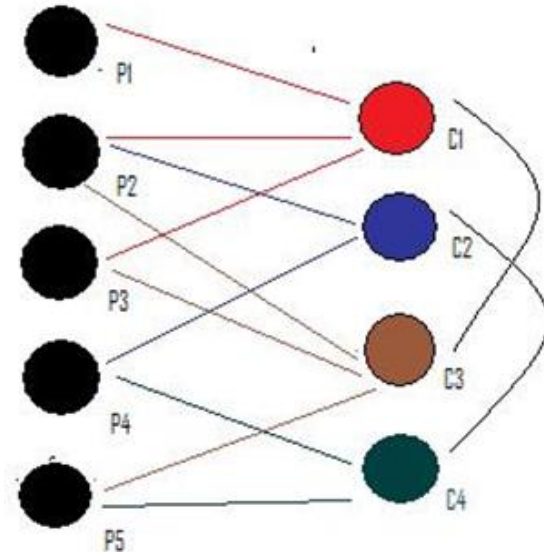The figure below shows a typical arrangement of colored communities.

fig 2.

In the above diagram we have shown C1, C2, C3, C4 that represents communities in different colors. No two adjacent edges from Pi to Ck is of same color. (i and k are the numbers) .If there are k communities then as in fig 1, we would have k different colors to represent them. We have distinguished communities based on different color. The relationship between the user and community is shown by the edges from Pi to Ck. Each edge is assigned a weight on basis of which the membership is calculated of user in a community. If there exist a edge between two communities then they belong to same category like ethnic, social, education, technology etc.

2) Fuzzy c means clustering
The Algorithm
Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is developed by Dunn in 1973 and improved by Bezdek in 1981. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \quad , \quad 1 \le m < \infty$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i(no. of users)in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \quad , \qquad c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This iteration will stop when

$$\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon,$$

where ε, is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

The algorithm is composed of the following steps:

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

3. Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

4. If $\| U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP; otherwise return to step 2.

Lets take 'n' users and two communities(A and B), We have categorized the communities into religious, ethnic ,technical and non-technical etc. The communities are divided into clusters based on fuzzy c means clustering.

$$U_{MxC} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ .. & .. \\ 0 & 1 \end{bmatrix} \qquad U_{MxC} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ .. & .. \\ 0.9 & 0.1 \end{bmatrix}$$

(a)   (b)

Fig(a) represents a crisp set. It contains membership values of each users for both communities. In crisp set each user can belong to only one community also referred as cluster.

Fig(b) is determined from fig(a) by above algorithm as in practical a user can belong to more than one community but has different membership in different communities.

The above graph represents fig(a), where users belong to either A or B.

The above graph represents fig(b), where users can belong to both A and B.

The red dot of user represent that he/she is more active in group A as the membership function of B is 0.2.

Evaluation Of Above Methods
1) Weak and strong ties between user & community:
This paper uses Fuzzy –C means to determine the relationship between user and community. It takes membership values of each user belonging to different community to determine weak and strong ties. If user has membership values < 0.5 then user has a weak tie with the corresponding community else it has a strong tie. In case a user has no membership values grater than 0.5 then the community to which user belong having highest membership value has a strong tie with user.
2) Incentive of friendship between users:
If there exists two or more common communities between users represented by edges then there are more chances for the users to becomes friends.

Consider two users and four communities, if one user P1 belongs to C1,C2,C3 (means has edges between P1 and respective communities) and another user P2 belongs to C1,C2,C3,C4 (means has edges from P2 to respective communities), both have C1,C2,C3 common. So most frequent support count is 3 whereas we assume the minimum support to be

2. So there is more likely that the two users may become friends or are friends.

3) Relationship between Communities:

If there exists a edge between two communities then those two communities belong to same categories. We can classify communities into various categories like ethnic, social, technical, media, brands, education etc.

4) Search Optimization:

For search optimization the order followed is given below:

a)The priority is given to people who are friends i.e. having common edges between them. b) Then, the priority is given to the list of users belonging to same strong tie communities. c) Then, the list of users belonging to same weak tie community is displayed in search.

Conclusion

In this paper, we have closely analyzed the relationship between the user and the communities in a social network and how strongly the user is involved in the community. With the help of membership values calculated on the basis of fuzzy c means algorithm we can provide search optimization and this can increase the reliability of the social network system thereby alleviating the security issues. In future we want to improve search optimization technique by including Hamiltonian path to find shortest path to cover all the vertices (representing friends or communities).

References

[1] Coffman, Thayne; Greenblatt, Seth; Marcus, Sherry: Graph based technologies for intelligence analysis. In: Commun. ACM, Vol.47(3): pp. 45-47,2004, ISSN 0001-0782.

[2] Carley,K.M., & Reminga,J., & Kamneva,N. Destabilizing treeorist networks. NAACSOS Conference proceedings. Pittsburgh, P.A. 2003.

[3] S. Buchegger, D.Schioberg, L.H.Vu and A.Datta, "Peerson:P2P Social Networking," in Social Network Systems, 2009.

[4] T. Takagi and M. Sugeno, "Fuzzy identification of a system and its application to modeling and control", IEEE Trans. Syst. Man and Cybern., vol.15, pp.116-132, Jan./Feb. 1985.

[5] Lotfi Asker Zadeh, "Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers", World Scientific, 01-Jan-1996 - Computers - 826 pages.

[6] George J. Klir, Bo Yuan, Fuzzy Sets and Fuzzy Logic, Prentice Hall, New Jersey, 1995.

[7] J.Mendel. Fuzzy Logic systems for engineers: a tutorial. In Proceedings of the IEEE, volume 83, pages 345-377, March 1995.

[8] http://mathpages.com/home/kmath266/kmath266.htm

[9] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html

[10] http://eng.uok.ac.ir/esmaili/teaching/spring2012/sna/slides/Lecture2.pdf

[11] https://www.cs.sfu.ca/~oschulte/teaching/socialnetwork/papers/SNA-intro-mohsen.pdf

[12] http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch03.pdf