Categorization Of Social Networks Based On Multiplicity Constraints

Priya R[1], Chitra M T[2] and Elizabeth Sherly[3]

[1] Research Scholar, University of Kerala
Trivandrum, Kerala, India

[2] Research Scholar, University of Kerala
Trivandrum, Kerala, India

[3] Professor, Indian Institute of Information Technology and Management- Kerala
Trivandrum, Kerala, India

Abstract

In this paper, categories or groups are identified from a social network, which is defined as a graph with nodes and edges. Categorization is done on the basis of the attributes of the nodes and a set of priori defined constraints. Using fuzzy measures and multiplicity constraints, the node weight and edge weight for the graph is defined. An algorithm is proposed to identify the set of most optimal nodes from the social network which satisfies the priori defined input constraints.
Keywords: *alpha cutoff, multiplicity constraints, fuzzy edge weight, fuzzy weight, objective function.*

1. Introduction

Here, a social network model based on object constraints is proposed based on the inter-related patterns for a better categorization of groups in a social network. A social network is a network of individuals called nodes, which are connected by a set of relationships. Different clusters are formed on the basis of the attributes of the nodes, which represent the characteristics and relationships between the nodes, thus allowing to find a group of interacting users. Elements in a cluster are homogeneous but are distinct from other clusters.

A hybrid model is proposed to find the optimal solution of a highly reliable and strongly connected cluster of nodes. The method utilizes the concepts of object constraints technology to define the nodes and to connect these nodes to form the network based on attributes and multiplicity constraints. In order to identify the closeness of nodes in a cluster, fuzzy membership is calculated and also fuzzy edge weight is computed based on multiplicity constraints. In the proposed algorithm, an objective function is used to obtain an optimal solution based on the priori defined constraints, fuzzy weights of nodes and fuzzy edge weights.

2. Methodology

2.1 Concepts and Assumptions

The network is an undirected fully connected graph G(V,E) to form the cluster where V is the entity and E represent the relationship between entities. The network is defined with the assumption that the constraints are fixed, $C_1, \ldots, C_m$ and the characteristics of the nodes are also defined and fixed. A fully connected network is defined as

$$\sum X_{D_i,D_j} = 1 \,\forall\, D_i, D_j \in D \quad (1)$$

where the number of nodes or entities is fixed as $D_1, \ldots, D_n$. The objective function is defined on the basis that each node in the resultant set has satisfied all the conditions and constraints searched for. New clusters will be formed based on the changes in the searching constraints. Hence the objective function can be defined as

$$max(C_i)\,\forall\, C_i \in C$$

subject to

$$sizeof\left(\sum nC\left(D_i\right)\right) \geq C_k \forall entity D_i \in S$$

where C_k is the list of high priority constraints and $nC(D_i)$ is the number of constraints satisfied by D_i in the solution space, S.

2.2 Graph Pattern Clustering Based On Constraints

In this method, clusters are formed on the basis of the attributes of the nodes and the given set of constraints. Given a set $D=\{D_1,D_2,\ldots,D_n\}$ with n entities (nodes), a cluster of optimal entities is to be identified by a rule based clustering, for the set of constraints $C=\{C_1,C_2,\ldots,C_m\}$. The attributes of the entities are defined as $A_1,A_2,\ldots,A_s$ for each D_i for i=1,...,n. Based on the characteristics of the entities, and the constraints to be matched, an entity may belong to a cluster fully or partially, or may not belong to a cluster. Thus the membership of an entity in a cluster is fuzzy. Therefore, each cluster L, contained in G, can be defined as L(V,E,W,EW) where V={entity}, E={relationship between entities}, W={fuzzy weight associated with the entity} and EW={fuzzy edge weight}. The fuzzy weights are computed as follows.

$$W_i=\left(\sum C_i\right)/\left(\sum L\right)=\mu_L\left(D_i\right) \quad (2)$$

W_i is the weight associated with D_i, ΣC_i is the set of constraints satisfied by Di and ΣL is the total set of constraints to be satisfied to be in the cluster.

$$\mu_L\left(D_i\right)=\begin{cases} 1 & \sum C_i=\sum L \\ 0<x<1 & \sum C_i<\sum L \\ 0 & otherwise \end{cases} \quad (3)$$

Here the cluster formed is a fuzzy set, an entity is included in a cluster with a given degree of membership. The α_{cutoff} of the cluster is computed to find the set of entities that satisfy at least average of the total constraints to be satisfied. The α_{cutoff} of the cluster is computed as

$$\alpha_{cutoff}(L)=avg\left(W_i\right)=\frac{\sum W_i}{N_1} \quad (4)$$

The cluster thus formed is given by

$$L_\alpha=\left\{D_i \in S/\mu_L\left(D_i\right) \geq \alpha\right\} \quad (4a)$$

The entities whose fuzzy weights $W_i >= \alpha_{cutoff}$ will be taken to solution space of optimal entities. When a set of constraints or conditions are to be checked against attributes of entities, priority is set to the constraints. In such cases, only if the entity whose $W_i >= \alpha_{cutoff}$ and if the entity has satisfied all the high priority constraints, it will be taken to the optimal solution set.

Edge weights between the nodes are computed to identify the degree of closeness of the nodes, as

$$E_{i,j}\left(D_i,D_j\right)=nC_i \cap C_j \quad (5)$$

where $E_{i,j}$ is the edge weight between D_i and D_j, C_i is the fuzzy set of constraints of D_i . After computing the edge weight of all the edges in the cluster, the highest edge weight in the cluster is computed as follows.

$$HEW_L=max\left\{E_{i,j}\left(D_i,D_j\right)\right\} \forall D_i,D_j \in L \quad (6)$$

Those nodes connected by highest weighed edge are taken to check their fuzzy weights with the α_{cutoff} of the cluster and those entities whose fuzzy weights is greater than or equal to α_{cutoff} will be taken to solution space of optimal entities.

2.3 Proposed Algorithm

1: Input: Given G(V,E) and $C(c_1,c_2,\ldots,c_{m1})$

2: foreach $d \in V$ do

3: no_con = 0

4: foreach $C_i \in C$ do

5: if C_i exists in cf(d) then add one to no_con

[**cf is the function to check the attributes of object d]

6: endfor

7: Calculate f_wt(d) = no_con/m_1

8: Calculate α_{cutoff}

$$\alpha_{cutoff}=\frac{\sum f_w t(d)}{\sum(d)} \forall d \in V$$

9: foreach (d_i,d_j) do

10: calculate fuzzy_edge_wt of (d_i,d_j)

$$E_{ij}\left(D_i,D_j\right)=nC_i \cap C_j$$

11: endfor

12: if E_{ij} are distinct, find the largest E_{ij}

and the corresponding end vertices as d_i, and d_j

13: if f_wt(d_i) >= $\alpha_{cutoff,}$ then take d_i to the final cluster

14: if f_wt(d_j) >= $\alpha_{cutoff,}$ then take d_j to the final cluster

15: Repeat steps 2 to 14 for all input criteria.

16: Repeat steps 2 to 14 with the input as

G(V,E,W,EW) where G is the final cluster.

17: Output: Optimal list of nodes which satisfies all the given input criteria including high priority constraints.

3. Case Study with Results and Conclusion

A sample application of Medical System is chosen to experiment the method, which form the class diagram with its attributes, operations and multiplicity constraints. We consider the nodes as doctors and the characteristics of the nodes are identified and stored as attributes. The objective is to identify the most optimal doctors from the set of doctors, based on the constraints specified by the user.

The multiplicity constraints are used to depict the degree of closeness of the nodes within a network. Here, the multiplicity constraints are defined as the number of ways by which the two doctors in the social network are similar with each other.

The case taken for illustration is

Identify all the doctors who are Assistant Professors from Gynecology department.

The set of constraints to be satisfied are:

1. The doctors must be from the Gynecology department.
2. They should satisfy the priori constraints defined for Assistant Professor.

In the above example, the final set of constraints to be satisfied are:

1. The area of specialization of the doctor must be an allied area of the department.
2. The attributes of the doctor must match with the priori constraints defined for Assistant Professors viz., age, years of experience, and postgraduate qualification.

The edge weights are computed based on Eq.(5).

The graphical representation of the social network with the computed edge weights is shown in figure 1.

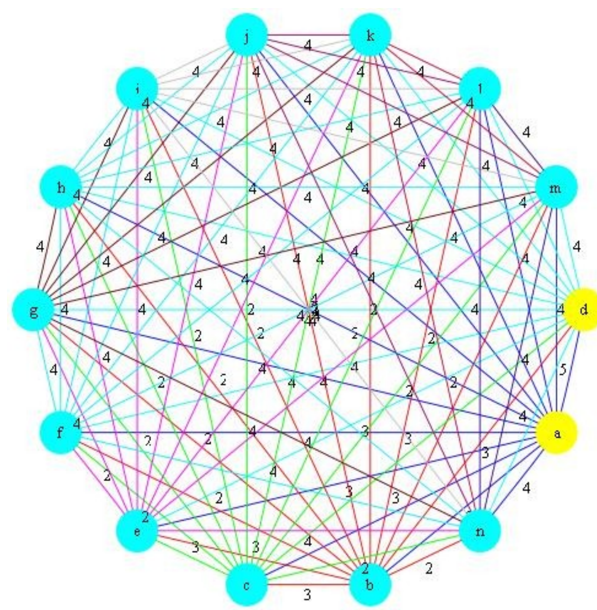

Fig. 1 Weighted graph of the Social network

The matrix representation is shown in Table 1.

Table 1: Computed Edge weights of the graph

	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a	-	3	3	5	3	4	4	4	4	4	4	4	4	4
b	3	-	2	3	3	3	2	2	2	2	2	2	2	2
c	3	2	-	3	3	2	2	2	2	2	2	2	2	2
d	5	3	3	-	3	4	4	4	4	4	4	4	4	4
e	3	3	3	3	-	2	2	2	2	2	2	2	2	2
f	4	3	2	4	2	-	1	4	4	4	4	4	4	4
g	4	2	2	4	2	4	-	4	4	4	4	4	4	4
h	4	2	2	4	2	4	4	-	4	4	4	4	4	4
i	4	2	2	4	2	4	4	4	-	4	4	4	4	4
j	4	2	2	4	2	4	4	4	4	-	4	4	4	4
k	4	2	2	4	2	4	4	4	4	4	-	4	4	4
l	4	2	2	4	2	4	4	4	4	4	4	-	4	4
m	4	2	2	4	2	4	4	4	4	4	4	4	-	4
n	4	2	2	4	2	4	4	4	4	4	4	4	4	-

In the above table, the nodes are labeled a to m, which represents the doctors in the final cluster. The figure shows the number of ways in which two doctors are similar with each other. The highest edge weight is computed and those doctors are taken to the list of optimal doctors. From Table 1, the highest edge weight value is 5. The doctors with fuzzy edge weight as 5 are {a, d}.

The fuzzy weights of all the nodes in the final list is shown in Table 2.

Table 2: Fuzzy weights of the nodes in the graph

Doctor_Id	Fuzzy_weight
a	**1**
b	0.25
c	0.25
d	**1**
e	0.25
f	0.75
g	0.75
h	0.75
i	0.75
j	0.75
k	0.75
l	0.75
m	0.75
n	0.75

The α_{cutoff} of the above table is 0.678571429. Hence the list becomes {a,d,f,g,h,i,j,k,l,m,n}. The attributes of these objects are compared against the final set of input constraints with high priority. As a result, the list becomes {a,d}. Since the fuzzy weight of a and d are 1 each, the final optimal list is {a,d}.

4. Conclusions

In this paper, a hybrid model based on constrained clustering using the attributes and relationship between the nodes in the cluster and fuzzy measures is proposed. In the algorithm, clusters of a group of nodes which have substantially similar characteristics are formed subject to the objective function. Finally a single cluster which satisfies all the constraints including high priority constraints specified by the user is formed. The algorithm of finding the optimal solution works efficiently for more than two input constraints.

In this model, only single layered network is considered to achieve the overall system reliability of the network. The proposed algorithm is tested by considering a set of constraints which are static in nature.

The future extension of this work includes the connection between the nodes in a multi layer network where the complexity will be much more than that in a single layered network model. More accuracy may be achieved by using dynamic multi-layered inter cluster interactions.

References

[1] Kellie Schneider, Chase Rainwater and Edward A Pohl, Assessing Multi-Layered Social Networks Using Reliability Models, Proceedings of Annual Reliability and Maintainability Symposium, IEEE RAMS 2011 : pp 103-107.

[2] Xiaoxiao Shi, Wei Fan, Philip S. Yu, Efficient Semi-supervised Spectral Co-clustering with Constraints. IEEE ICDM 2010: pp 1043-1048.

[3] Phanarut Srichetta, Sondsri Tangsripairoj, Using Social Network Analysis to Better Understand the Dependencies in UML Use Cases Collection. Proceedings of the Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/ Distribute Computing. 2008 pp 428-433.

[4] Jain, R.Chaudhari and N.S. Identification and Generation of Constraints in Social Network, Proceedings of International Conference on Emerging Trends in Networks and Computer Communications, ETNCC 2011 pp 11-14.

Priya R She received her Masters Degree in Computer Applications from Madurai Kamaraj University in 2001. Currently she is a Phd Research Scholar in University of Kerala, Trivandrum and working as Assistant Professor in Govt College Kariavattom, Trivandrum India. She has more than 6 years of experience in teaching graduate and postgraduate students of Computer Science. Her interest areas are Software Engineering, Object Oriented Analysis and Design, Modeling using Constrained Objects .

Chitra M T She received her Masters Degree in Computer Science from University of Kerala, Trivandrum in 2005. She is a Phd Research Scholar in University of Kerala, Trivandrum, India. She has 5 years of experience as Project Associate of various research projects in Indian Institute of Information Technology and Management – Kerala (IIITM-K), Technopark, Trivandrum, India. Her interest areas are Software Engineering and Object Oriented Analysis and Design.

Dr. Elizabeth Sherly She received Ph.D in Computer Science from University of Kerala, Trivandrum, 1995. She is currently working as Director and Professor in Indian Institute of Information Technology and Management – Kerala (IIITM-K), Technopark, Trivandrum, India. She has more than 14 years of teaching experience in Computer Science. Her research areas includes Data mining, Soft Computing - Artificial Neural Networks, Fuzzy Systems, Computational Linguistics, Object Oriented Software Engineering, Grid and Cloud Computing.