

Maitreyi College,

Chanakyapuri-110021

University of Delhi

**PROJECT REPORT**

**ON**

**Community Detection in Signed Social Networks**

# **Under the guidance of: Submitted By:**

Dr. Vinti Agarwal Manisha Keim, 2984

Assistant Professor Oseen Gupta, 2214

Dept. of Computer Science Pallavi Pannu, 1666

Maitreyi College B.Tech Computer Science

University of Delhi 4th year

**ABSTRACT**

Complex networks are widely applied in every aspect of human society, and community detection is a research hotspot in complex networks. Many algorithms use modularity as the objective function, which can simplify the algorithm. In this paper, a community detection method based on modularity and genetic algorithm is put forward.

While studying online social networks, relationships can be either positive (friendship) or negative (enmity). In contrast, to the research been carried on social signed networks has focused entirely on positive elucidations of link between people. We study how to develop a better understanding of the role that network structure plays when some links between people are positive while others are negative (Patidar, *et al.*, Leskovec, *et al.*, 2010).

The GA is a famous adaptive heuristic search algorithm inspired by the evolutionary ideas of natural selection and genetics and pioneered by Holland at the University of Michigan. The GA is designed to simulate processes in natural system necessary for evolution and to follow the principles first laid down by Charles Darwin of survival of the fittest.

**TABLE OF CONTENTS**

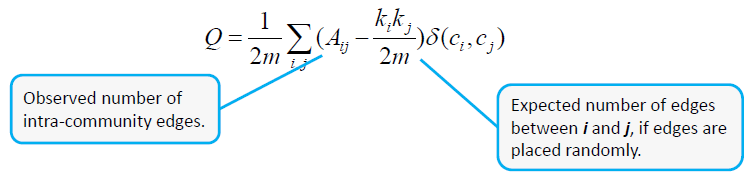
1. Community Detection in Signed Networks
2. Factors for determining communities
   1. Modularity, Q
   2. Consensus
   3. Balance Index
3. Dataset
4. Theories in Signed Networks
   1. Social Balance Theory
   2. Social Status Theory
5. Consensus Evaluation
6. Preference Relations
7. Fuzzy Binary Adjacency Relations
8. Fuzzy m-ary Relations
9. Variable Length Genetic Algorithm
   1. Chromosomes
   2. GA Operators
      1. Selection
      2. Crossover
      3. Mutation
   3. Objective Function
   4. Stopping Criteria

**Introduction**

Community is formed by individuals such that those within a group interact with each other more frequently than with those outside the group. The problem that community detection attempts to solve is the identification of groups of vertices that are **more densely connected** to each other than to the rest of the network.

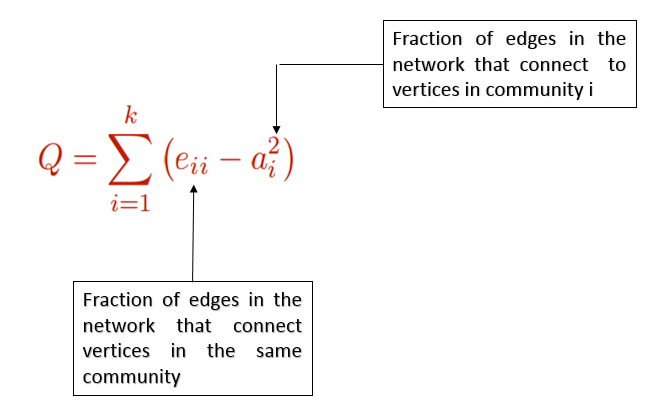
To provide a measure of the quality of a community structure, we make use of **modularity**.

Modularity quantifies the extent to which a given graph partition into communities presents a systematic tendency to have more intra-community links than the same community structure would present if the links would be rewired under ER (Erdos-Renyi) graph model.

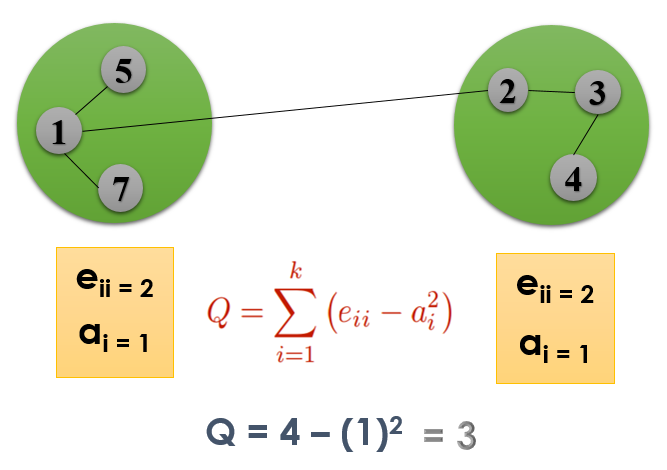


* **Αij:** adjacency matrix
* **ki:** degree of node **i**
* **ci**: community of node **i**
* **δ(ci,cj) = 1** if **i**, **j** belong to the same community
* **m**: number of edges on the graph

Generalized Formula:



Example for understanding Modularity:



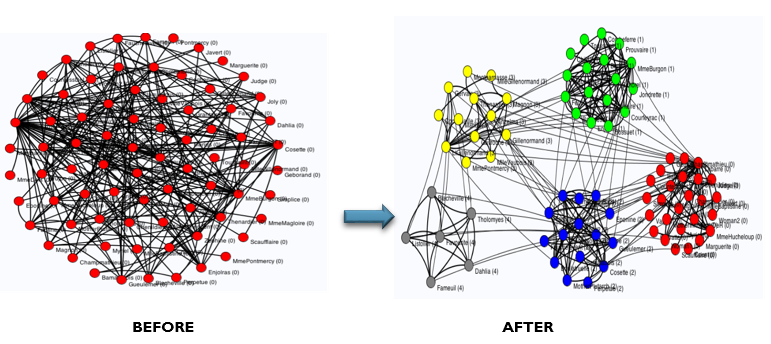


Fig. 1 Community Detection

To investigate positive and negative links between people, we need two fundamental components:

1. Datasets from social applications where the sign of each link whether it is positive or negative can be consistently determined, and
2. Theories of signed networks that help us in determining how different patterns of positive and negative links provide different kinds of relationships.

Datasets:

* No. of users = 9
* No. of communities = 3
* Each community matrix is of size 3\*3.

Theories in Signed Networks:

* **Structural Balance Theory:** It originated in social psychology in the mid-twentieth century. As formulated by Heider in the 1940s (Heider 1946), and subsequently cast in graph-theoretic language by *Cartwright and Harary* in 1956. It states that structural balance considers the possible ways in which triangles on three individuals can be signed and posits that triads with odd number of “+” are balanced (***T0, T2***), and triads with even number of “+” edges (***T1, T3***) are unbalanced (Fig. 2). Balanced triangles with three positive edges exemplify the principle that “the friend of my friend is my friend,” whereas those with one positive and two negative edges capture the notions that “the friend of my enemy is my enemy,” “the enemy of my friend is my enemy,” and “the enemy of my enemy is my friend.”

(T0)

**+**

**+**

**+**

(T2)

**-**

-

**+**

(T3)

**-**

**-**

**-**

(T1)

+

+

**-**

Fig. 2 Configurations of Triads

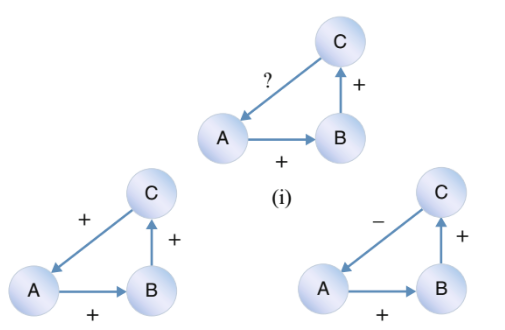
* **Social Status Theory:** In status theory (Patidar, A., Agarwal, V., et al), which provides a different organizing principle for directed networks of signed links. In this theory of status, we consider a positive directed link to indicate that the creator of the link views the recipient as having higher status; and a negative directed link indicates that the recipient is viewed as having lower status. Status theory often leads to different predictions than balance theory.

B has lower status than A

A has higher status than B

Fig.3 Social Status Theory

In Fig 3, where user A positively linked to user B, and B in turn links positively to user C, then what kind of link from C to A should we expect to have? Balance theory states that since C is a friend of B and B is a friend of A, therefore C will also be a friend of A, whereas according to status theory, A regards B as having higher status and B regards C as having higher status, so C should regard A as having lower status and hence the link directed from C to A would be negative.



(ii) Balance (iii) Status

Fig 4 Predicting relationship from C to A according to

(ii) Balance Theory and (iii) Status Theory

**BALANCE INDEX**

**Balance Index:**

This concept discovers a measure of balance, called balance index, which gives the proportion of triads that are balanced. So, the global Balance Index (*BI*) can be written as;

where denotes the number of balanced triads, denotes the total number of triads in whole interpersonal network.

Total number of triads ( ) = 3

Balanced triads ( ) = 1

***BI*** = 0.3333

Fig. 5 Calculating Balance Index

**CONSENSUS**

* **Consensus** is a group discussion where everyone's opinions are heard and understood, and a solution is created that respects those opinions.
* **Consensus** is not what everyone agrees to, nor is it the preference of the majority.
* **Consensus** results in the best solution that the group can achieve at the time.

To evaluate consensus among group of users, we are using Fuzzy m-ary preference relations.

* **Preference Relation:** A preference relation on a set is characterized by a function → [0, 1], such that and +.

Let us assume there are 3 items I1, I2, I3 and users have given preferences on them. To calculate preference relations on these items:

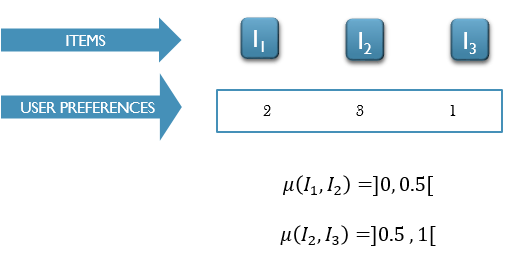


Fig. 6 Calculating Preference Relations

* A fuzzy binary relation representing the degree of relationship exist between and , is defined using the member function→ [0, 1] as:
* A fuzzy -ary relation on the set is defined by the membership function → [0, 1],

**GENETIC ALGORITHM**

***“A method for moving from one population of "chromosomes" to a new population by using a kind of "natural selection" together with the genetics−inspired operators of crossover, mutation, and inversion”.***

* John Holland, 1975

***“Genetic Algorithms are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and natural genetics”.***

- David E. Goldberg, 1989

When a GA is used to solve a problem, first, a population of chromosomes or individuals should be maintained where each chromosome represents a potential solution to the problem. Second, each chromosome is evaluated to give some measure of its fitness. And then some chromosomes undergo stochastic transformations by means of genetic operators (crossover and mutation) to form new chromosomes called offspring, where crossover creates new chromosomes by combining parts from two chromosomes and mutation creates new chromosomes by making changes in a single chromosome. Then, a new population is formed by selecting the more fit chromosomes from the parent population and the offspring population. Further iterations are carried out until the stopping criterion is satisfied.

**Genetic Algorithm Operators:**

Genetic operators are used to alter the genetic composition of chromosomes during representation. Crossover and mutation are two common genetic operators. Following are the details of the genetic operators:

**SELECTION**

Selection replicates the most successful solution found in a population.

* **Roulette−wheel Sampling**, which is conceptually equivalent to giving each individual a slice of a circular roulette wheel equal in area to the individual's fitness. The roulette wheel is spun, the ball comes to rest on one wedge−shaped slice, and the corresponding individual is selected.

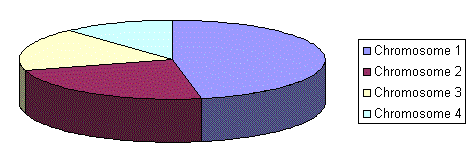


Fig. 7 Roulette-wheel Sampling

**CROSSOVER**

Crossover operates on two chromosomes at one time and creates new chromosomes by combining parts from two chromosomes. So the offspring generated by the crossover operator combine the features of both of the chromosomes [20]. Traditionally, the process of the crossover operator generally includes three steps: (1) it selects two chromosomes, (2) takes them together, and (3) outputs two new chromosomes. A crossover point is selected in a chromosome, and then the two chromosomes’ elements after that selection point are exchanged. The traditional GA based algorithms include different crossover operators such as the one-way crossover, two-way crossover, flat crossover, blend crossover, and so on.

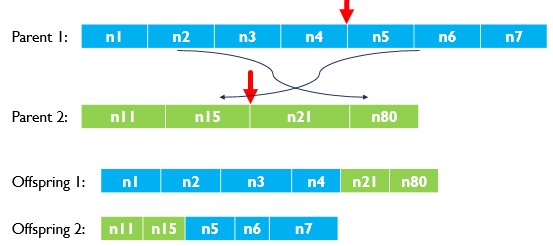


Fig. 8 Crossover operation

**MUTATION**

The mutation operator is the main operator in the GA; it produces spontaneously random changes in various chromosomes. So, the mutation operator can increase the diversity of the population and speed up the convergence. The mutation operator is to make the genic values in some certain genic positions of chromosomes inverse with a certain probability. After a crossover is performed, mutation take place. This is to prevent falling all solutions in population into a local optimum of solved problem. Mutation changes randomly the new offspring.

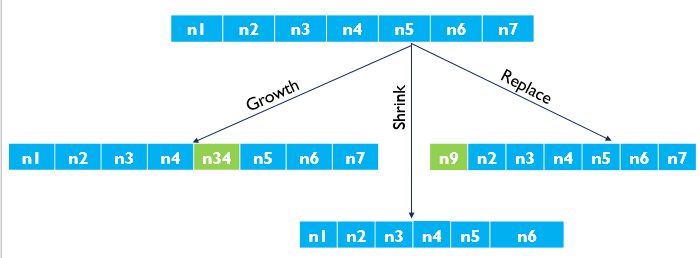


Fig. 9 Mutation Operator

**FITNESS FUNCTION**

A fitness function simply defined is a function which takes the solution as input and produces the suitability of the solution as the output. In some cases, the fitness function and the objective function may be the same, while in others it might be different based on the problem.

Objective function =

We want high value of Consensus and Balance Index to get a high value of objective function.

**STOPPING CRITERIA**

* When a maximum number of generation elapsed or
* a desired level of fitness value is achieved.

**RELATED WORK**

In past, many algorithms have been proposed to detect network communities only in positive networks. They may be classified into three groups:

* + Graph Theoretic Methods

Some example of Graph Theoretic Methods are Random walk methods, physics-based methods, and Spectral methods.

* + Divisive Algorithms

Algorithms of Girvan and Newman Tyler algorithm which they divide the whole network into smaller groups depending on the betweenness.

* + Algorithms based on Modularity which measures when the division is a good one. But these algorithms are just good for positive networks but cannot apply signed networks. Now, some algorithms have been proposed for detecting communities in signed networks. Some people improved the above algorithms so that they could apply these networks, such as GN-H algorithm which contains two phases. The first phase, we divide the positive sub network; and on the second phase, it governs the final community structure based on the information of negative edges.
  + Complex networks considering both positive and negative links have gained considerable attention during the past several years. Community detection is one of the main challenges for complex network analysis. Most of the existing algorithms for community detection in a signed network aim at providing a hard-partition of the network where any node should belong to a community or not. However, they cannot detect overlapping communities where a node is allowed to belong to multiple communities. The overlapping communities widely exist in many real-world networks.

**PROPOSED FRAMEWORK**

There has been a surge of interest in community detection in homogenous networks which contains only one type of nodes and edges until now work has been done on determining only positive relationships among users. However, many real-world systems are heterogenous which contain multiple type of nodes and edges. In proposed work, we propose an approach for detecting community in signed networks by modelling both positive and negative relationships (Patidar, *et al.*, Leskovec, *et al.*, 2010) among users. Our method works by computing modularity Q, which is fast and simple.

Further to increase the accuracy of our detection, we incorporate consensus building among a group of networked decision makers using Fuzzy *m-ary* relationships (Brunelli, *et al*., 2011). Also, our technique uses Variable Length Genetic Algorithm (“Genetic Algorithms are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and natural genetics”) assimilating Fuzzy m-ary consensus and balance index to find highly balanced and consensual community.

**CONCLUSION**

In the presented work, we are proposing an approach for detecting community in signed networks by modelling both positive and negative relationships among diverse people by computing Modularity Q with genetic algorithm determining the future relationship among people.

**RESULTS**

Fig. 10 Modularity Applied on Population

Fig. 11 Modularity after applying Genetic Algorithm Operators

**REFERENCES**

* J.Leskovec, D.Huttenlocher, J.Kleinberg, Predicting Positive and Negative Links in Online Social Networks, ACM CHI (2010 April).
* P.Esmailian and M.Jalili, Community Detection in Signed Networks: The role of negative ties in different scales.
* M.Brunelli, M.Fedrizzi and Mi.Fedrizzi, Fuzzy m-ary adjacency relations in social network analysis: Optimization and consensus evaluation.
* Patidar, A., Agarwal, V., Bharadwaj, K.K. , Predicting Friends and Foes in Signed Networks using Inductive Inference and Social Balance Theory.
* Agarwal, V., Bharadwaj, K.K. (2015) Predicting the Dynamics of Social Circles in Ego Networks using Pattern Analysis and GA K-means Clustering. WILEY Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(3), pp. 113–141.
* Agarwal V., Bharadwaj K.K (2014) Friends Recommendations in Dynamic Social Networks, Encyclopedia of Social Network Analysis and Mining, Springer Science
* Predicting Friends and Foes in Signed Networks using Inductive Inference and Social Balance Theory
* Agarwal, V., Bharadwaj, K.K. (2015) Recommending Diverse Friends in Signed Social Networks based on Adaptive Soft Consensus Paradigm Utilizing Fuzzy m-ary Adjacency Relations.
* Lee, C.Y., Antonsson, E.K.: Variable Length Genomes for Evolutionary Algorithms. In L. Darrell Whitley; David E. Goldberg; Erick Cantú-Paz; Lee Spector; Ian C. Parmee& Hans-Georg Beyer, ed., 'GECCO' , Morgan Kaufmann, , pp. 806 (2009)
* Kim, I.Y., de Weck, O.L.: Variable chromosome length genetic algorithm for progressive refinement in topology optimization. Struct. Multidisc. Optim. 29, 445–456 (2005).
* An Introduction to Genetic Algorithms, Melanie Mitchell, MIT Press, 1999
* Zeng Y., Liu J., (2015) Community Detection from Signed Social Networks Using a Multi-objective Evolutionary Algorithm, Springer International Publishing
* Xue R. R., Ma. Y., Li W., Community Detecting in Signed Networks based on Modularity
* [Chen](https://arxiv.org/find/cs/1/au:+Chen_Y/0/1/0/all/0/1) Y., [Yuan](https://arxiv.org/find/cs/1/au:+Yuan_B/0/1/0/all/0/1) B., [Tang](https://arxiv.org/find/cs/1/au:+Tang_B/0/1/0/all/0/1) B., (2014) Overlapping community detection in signed networks