www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 6 June 2015, Page No. 12417-12422

Graphical Modularity Analysis Using Force Directed Approach

Syed Hasan Saeed

SRM University, Computer Science & Engineering,NCR Campus,
Modinagar,Ghaziabad,Uttar Pradesh, India.
hasan.hasan.saeed8@gmail.com

Abstract: Nowadays graphical analysis has become the centrepiece of many of the studies & common day problems. Because of the ability of the graph to adept & contain data of various forms & then interpret information it is in demand everywhere, many of the big problems are being analysed through graphical networks. Graphical data mining a branch evolved out of data mining has today become such a vast subject of study, research, analysis & case-review. This in itself shows the broad scope on which graphs, graphical analysis & graphical mining is being carried out today. The basic reason for this popularity of graphs is their ease with which they can accommodate data of various types & then can give result across various domains. So in such an environment it becomes equally important to study various metrics of graph that help in such studies. Modularity is one of those metrics, whose importance has grown manifold in studying & researching social networks, behavioural pattern networks & geo-graphical networks. In this study a force directed approach has been used to analyse modularity in a graph since it facilitates the dynamic structure of graphs that changes with respect to force applied on it similarly like a real time graphical pattern changes. Modularity is one property which can be used in multiple networks to study & analyse various parameters on which efficiency & capacity of networks is evaluated. Also with the advent of big data, business intelligence the need for information which gives multi-dimensional & multi-value relationships has grown manifold & graphical analysis especially in terms of modularity or community detection can give us plenty of such information. In this paper a very novel & realistic approach has been taken that takes into account the ever changing & dynamic nature of real time networks, based upon that a force directed approach has been taken which involves calculating the force exerted on graph at any time then finding corresponding modularity for that force. Three case of force have been analysed all for same graph to differentiate the effect of force on same graph. Three different types of forces act in a graph – attractive forces, repulsive forces & displacement force.

Keywords: modularity, force directed, graphical networks, communities, maximum force, minimum force, Average force.

1. Introduction

Graphical analysis & information extraction from graph has become one of the most trending techniques for study of various graphs & graphical networks, a graph is analyzed on multiple metrics depending upon what kind of information we are looking for but modularity always forms one of the most important metric to be studied since it is modularity that tells us about the heterogeneity of data contained in graph, the amount of multi-dimensional information in a graph & extent of multi valued relationships in a graph. The community detection or modularity has become a very integral part of any graphical analysis almost all the graphical analysis or operations related to graphical data mining are incomplete without any reference to it, modularity in terms of graph is defined as the partition of the graph upon the basis of certain parameters which are actually communities of graph for example- a graph depicting population of a city can be divided into four directions-east, west, north & south then these four directions are four communities of the graph & the modularity of graph is four. Thus we can say that a community in a graph is set of nodes that have same attributes, shape, property & connected to each other in multiple ways [1].

There are also numerous practical utilities of finding modularity in a graph:

a) It helps in partitioning social networks on the basis of groups of users belonging to a particular type such that they can be held as one community based upon their likes, followings & friendships making social networks a more organized place.

b) Also in web search we can classify the type of information accessed depending upon whether it's related to technology, science, arts, history, sports etc thus enabling search engines & portals to enhance their searching time [1].

c) Frequently, the nodes in a densely knit community share a salient real-world property. For social networks, this could be a common interest or location; for web pages, a common topic or language; and for biological networks, a common function. Thus, by analyzing structural features of a network, one can infer semantic attributes.

d) By identifying communities, one can study the communities individually. Different communities often exhibit significantly different properties, making a global analysis of the network inappropriate. Instead, a more detailed analysis of individual communities leads to more meaningful insights, for instance into the roles of individuals [2].

2. Nature Of Work

The present paper deals with enhancing modularity of a graph by the use of Fruchterman & Reingold algorithm based upon the principle of force directed algorithm. Essentially this algorithm has its roots in physical system or to say in simple terms it has been derived from mechanical systems. When described in mechanical system it is like laying out a graph in physical plain where nodes are replaced by steel rings & edges are replaced by springs. The vertices are placed in some initial order such that they are not too far from each other & also to let the spring forces on the rings achieve a minimal energy state [3]-[4].

The two basic principles that have been kept in regards for this approach are :

a) Vertices that are connected through an edge should be drawn near to each other.

b) But two vertices should not be too close such that they repel each other[3].

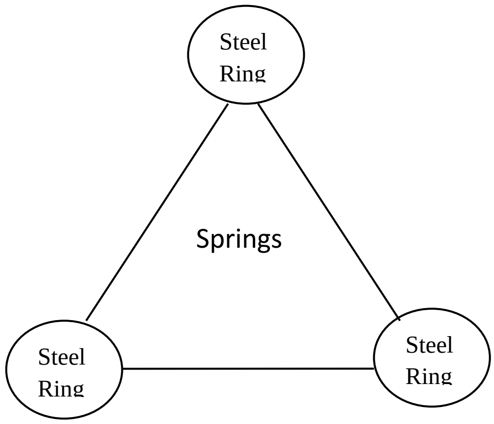

Fig 1.1 physical model of force directed approach.

The position of vertices placing will depend upon the nature & size of the graph if the graph is large then we are looking for higher number of communities so vertices will be nearer to each other whereas in case of smaller graph it will be vice-versa. Also how many communities we find in a graph depends upon the strength of forces acting in the graph if there is high amount of force then less chance of equilibrium being achieved so lesser communities whereas if force acting in the graph is less then more chances of equilibrium therefore more communities can be detected [3]-[4]-[5]-[6].

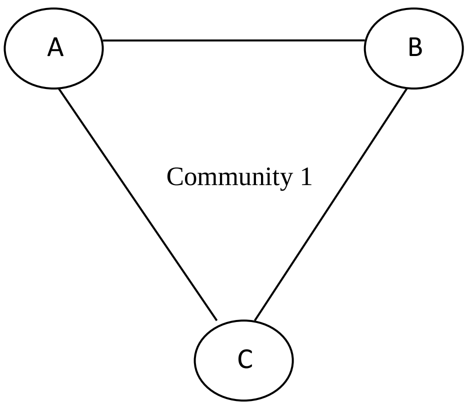

Fig 1.2 Sparse Graph with lesser communities.

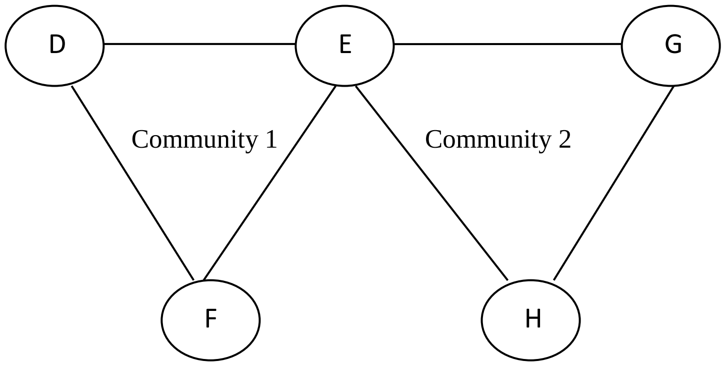

Fig 1.3 Dense Graph with more communities.

3. Previous Work

The previous work that has been done under this subject is of two categories:

i) Modularity maximization for community detection.

ii) Statistical & min-max model of finding modularity.

3.1. Modularity Maximization Method

In this method the approach used was the technique of solving and rounding fractional mathematical programs to the problem of community discovery, and propose two new algorithms for finding modularity-maximizing clustering. Two major algorithms have been used in this approach.

The first algorithm was based on a linear programming (LP) relaxation of an integer programming (IP) formulation. The LP relaxation put nodes "partially in the same cluster". A "rounding" procedure was used due to Charikar [7] for the problem of Correlation Clustering. The idea of the algorithm is to interpret "partial membership of the same cluster" as a distance metric, and group together nearby nodes [7]-[8].

The second algorithm is based on a vector programming (VP) relaxation of a quadratic program (QP). It recursively splits one partition into two smaller partitions while a better modularity can be obtained. It is similar in spirit to an approach recently proposed by Newman [9, 10], which repeatedly divides clusters based on the first eigenvector of the modularity matrix. Newman's approach can be thought of as embedding nodes in the interval [−1, 1], and then cutting the interval in the middle. The VP embeds nodes on the surface of a high-dimensional hyper sphere, which is then randomly cut into two halves containing the nodes. The approach is thus very similar to the algorithm for Maximum Cut due to Goemans and Williamson [11].

The method of modularity maximization is further classified into two types on which previously work has been done :

3.1.1 The LP Rounding Algorithm

Our LP rounding algorithm is essentially identical to one proposed by Charikar [7] for the Correlation Clustering problem. In correlation clustering, one is given an undirected graph

G = (V,E) with each edge labelled either '+' (modelling similarity between endpoints) or '−' (modelling dissimilarity).

The goal is to partition the graph into clusters such that few vertex pairs are classified incorrectly. Formally, in the Min-Disagree version of the problem, the goal is to minimize the number of '−' edges inside clusters plus the number of '+' edges between clusters. In the Max-Agree version, which is not as relevant to our approach, the goal is to maximize the number of '+' edges inside clusters plus the number of '−' edges between clusters. Using the same 0-1 variables xu,v as we did above, Charikar [7] formulate Min-Disagree as follows [9]-[10]-[12].

Minimize $P_{(u,v)\in E_+} x_{u,v} + P_{(u,v)\in E_-} (1 - x_{u,v})$ subject to $x_{u,w}$ _ $x_{u,v} + x_{v,w}$ for all u, v, w $x_{u,v}$ 2 {0, 1} for all u, v, where E_+ and E_- denote the sets of edges labelled '+' and '−', respectively. The objective can be rewritten as $|E_+| - P_{(u,v)\in E}\ \mu_{u,v}(1 - x_{u,v})$, where $\mu_{u,v}$ is 1 for '+' edges and -1 for '−' edges. The objective is minimized when $P_{(u,v)\in E}\ \mu_{u,v}(1 - x_{u,v})$ is maximized; thus, except for the shift by the constant $|E_+|$, Min-Degree takes on the same form as modularity maximization with $m_{u,v} = \mu_{u,v}$[10]-[13].

3.1.2 Vector Programming Based Algorithm

This algorithm used a 'heirarichal clustering', In the sense that the clustering is obtained by repeatedly finding a near-optimal division of a larger cluster. For two reasons, this clustering is not truly hierarchical: First, we do not seek to optimize a global function of the entire hierarchy, but rather optimize each split locally. Second,

we again apply a local search based post-processing step to improve the solution, thus rearranging the clusters. Despite multiple recently proposed hierarchical clustering algorithms, there is far from general agreement on what objective functions would capture a "good" hierarchical clustering. Indeed, different objective functions can lead to significantly different clustering. While our clustering is not truly hierarchical, the order and position of the splits that it produces still reveal much high level of information about the network and its clusters. This approach is aimed for the best division at each level individually, requiring a partition into two clusters at each level. Clusters are recursively subdivided as long as an improvement is possible. Thus, a solution hinges on being able to find a good partition of a given graph into two communities [14]-[15].

3.2 Max-Min Modularity Method

The idea of MM Modularity is based on the intuition that a good division of a network into communities is not merely one in which the number of edges between groups is smaller than expected, it is also one in which the number of unrelated pairs within group is smaller than expected. Only if both the numbers of between-group edges and within-group unrelated pairs are significantly lower than would be expected purely by chance, can we justify claim to have found significant community structure. Equivalently, we can examine the number of edges within communities and unrelated pairs between communities and look for divisions of the network in which this number is higher than expected. These two approaches are equivalent since the total number of edges/pairs is fixed and any edges/pairs that do not lie between communities must necessarily lie inside one of them [16].

Generally speaking, our evaluation attempts to maximize the number of edges within groups and minimize the number of unrelated pairs from the user defined unrelated pair set within groups at the same time, therefore we named it Max-Min Modularity. But maximizing the edge number within groups does not automatically minimize the unrelated pair number, e.g., if we have no network knowledge, thus have no related pairs, and unrelated pairs as disconnected node pairs, consider a node that only connects one member of a community with size n, maximizing the within-group edge number by including that node in this community would increase the unrelated pair number by n – 1. [17].

4. Present Work Undertaken

As we saw with the above both approaches that both of them requires a lot of preliminaries parameters for analysing modularity in any graph as well as lot of calculations have to be taken into account.

So the present approach is much more simplified version of finding modularity or community detection in any graph. The approach involves application of Fruchterman Reingold [3] algorithm on gephi tool to analyze modularity. The basic principle involved here is that the algorithm gives graph a circular shape so once the graph attains a final spherical shape the nodes that are in central are connected nodes with much higher degree connectivity & part of multiple communities. Whereas the nodes that are formed on outer rings are less connected nodes with lesser degree of connectivity. This alignment of nodes in a graph in spherical shape forms the basis of this research & study, the following figure explains how nodes are formed in this approach.

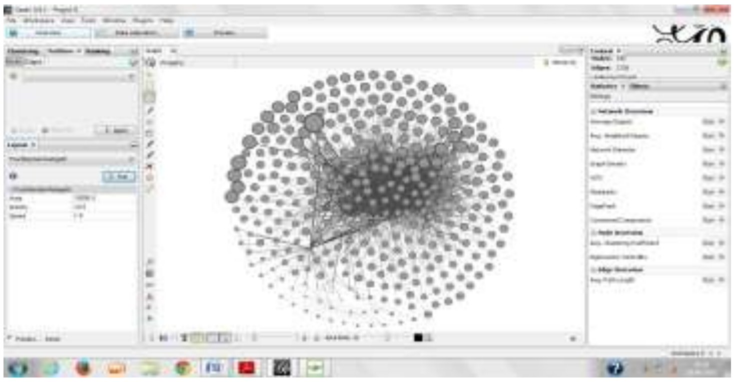

Fig 1.4 Alignment of nodes in the graph showing U.S airline flight plan.

In the above figure the nodes that are at centre are highly connected nodes having higher chance of being found out in community detection whereas the fringe nodes on outer lines are less connected nodes with lesser chance of being found out in a community.

Since the algorithm that has been used here is a force directed algorithm so the amount of force that exists in a graph or in other words that is exerted on graph plays a major role in deciding the number of communities that we can get for a graph. So the major emphasis that has been given here is to analyse how the force exertion in different conditions results in founding of different set of communities for same graph.

The three major kind of forces that exists in this algorithm is attraction, repulsion & displacement.

a)ATTRACTIVE FORCES : Force acting between nodes that are connected to each other.

b)REPULSIVE FORCES : Force acting between nodes that are disconnected to each other.

c)DISPLACEMENT FORCES : Force that finally settles the node after above two forces have acted upon it.

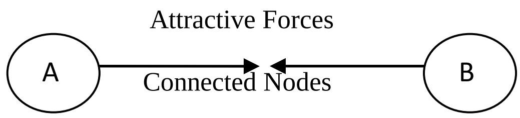

Fig. 1.5

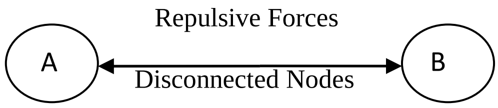

Fig. 1.6

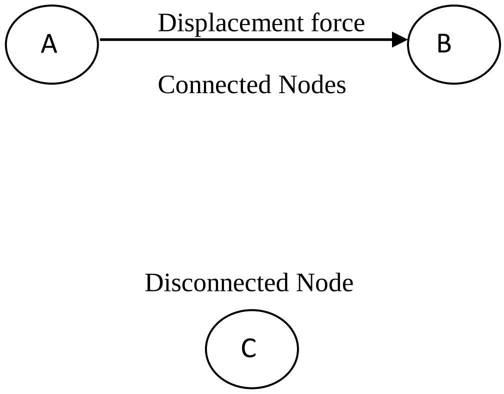

Fig. 1.7

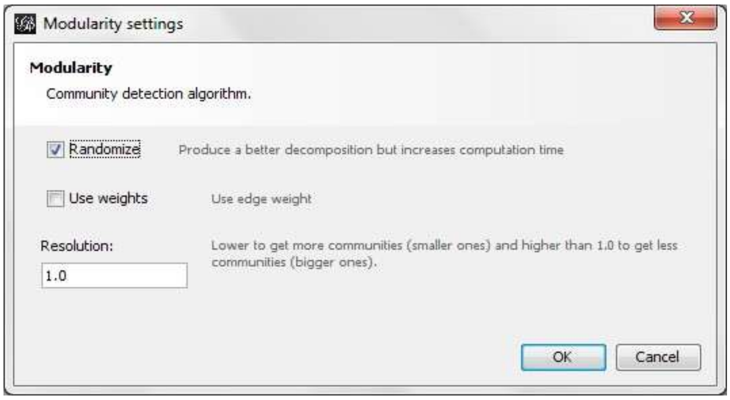

Fig. 1.8 Showing average force taken for calculation. Force = 1.0 [average]

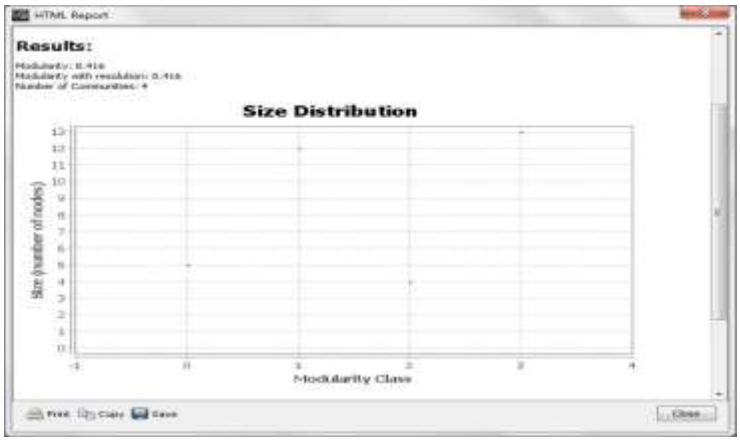

Fig. 1.9 Showing four communities obtained for average force calculation.

So basically three different cases where used for this study :
i)Average Force Calculation : In this case we try to find out the modularity or number of communities founded when all forces acting where of average nature.
ii)Maximum Force Calculation : In this case we try to find out the modularity or number of communities founded when all forces acting where of maximum nature.
iii)Minimum Force Calculation : In this case we try to find out the modularity or number of communities founded when all forces acting where of minimum nature.

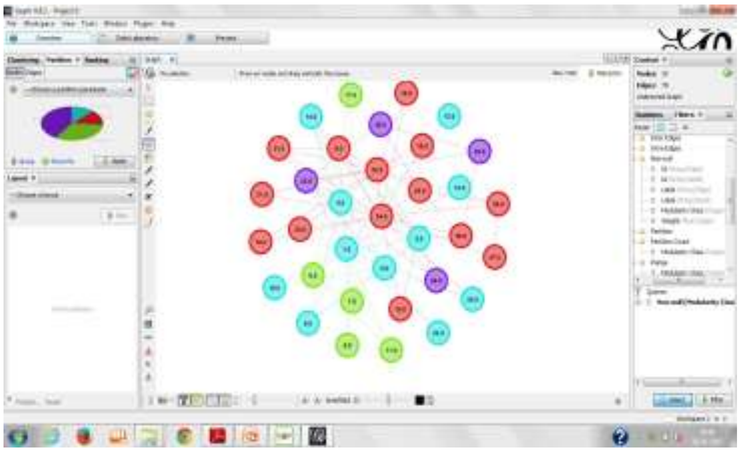

Fig. 1.10 Graphical representation of four communities obtained

5. Results

The karate club [18] interaction dataset has been taken for the purpose of analysing modularity for different cases. The communities obtained where according to interactions taking place in communities & force acting in the graph

5.1 Case 1 - Average force Modularity

5.1.1 Result Analysis

The result for average force shows that when average amount of force is acting within the graph then number of communities that will be detected is also in medium range.

Here we took force resolution= 1.0, so we got 4 communities which are represented in graphical manner through four different colours. In graphical format we can see the spherical alignment in which the algorithm adapts itself has connected communities like red & blue at centre whereas dispersed communities like violet & green are more settled on outer areas. Thus establishing the fact which formed the basis of this study.

5.2 Case 2 - Minimum force Modularity

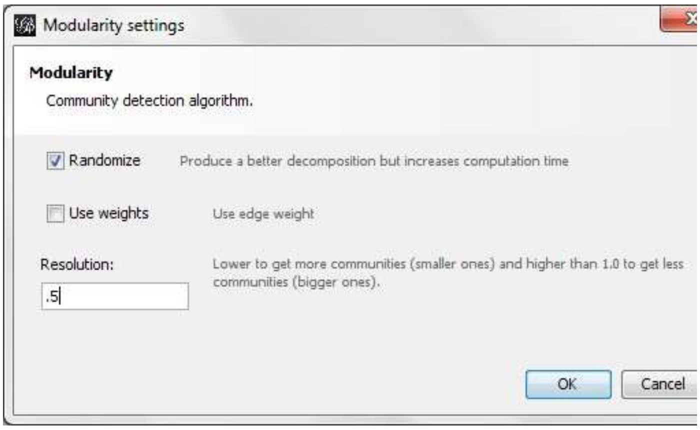

Fig. 1.11 Showing minimum force taken for calculation.
Force = 0.5 [minimum]

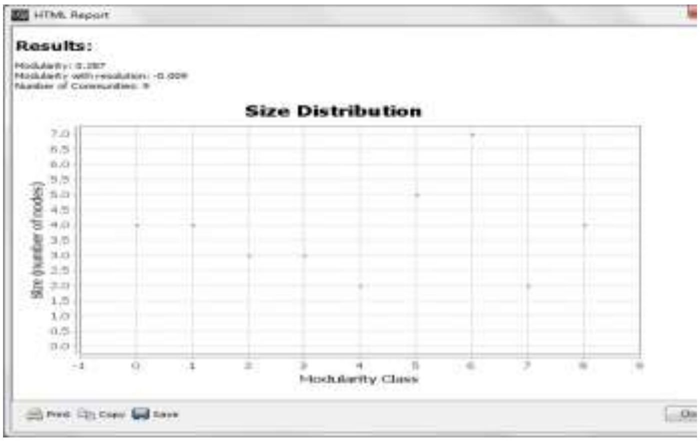

Fig. 1.12 Showing nine communities obtained for minimum force calculation.

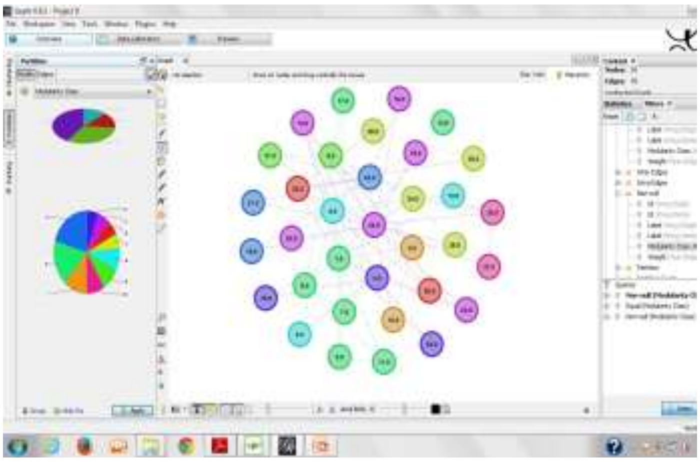

Fig. 1.13 Graphical representation of nine communities obtained.

5.2.1 Result Analysis

The result for minimum force shows that when minimum amount of force is acting within the graph then number of communities that will be detected is in maximum range. This inverse relationship between force applied & number of communities detected is because there is a direct relationship between nodes movement due to application of force, so when there is minimum force movement nodes do not change their position & they form maximum communities.

In this case the force resolution was taken 0.5 & nine communities were detected. The graphical representation depicts all of them but since the number of communities is much more this time so we do not get such a strong connected & disconnected node contrast like we got in average force case.

5.3 Case 3 - Maximum force Modularity

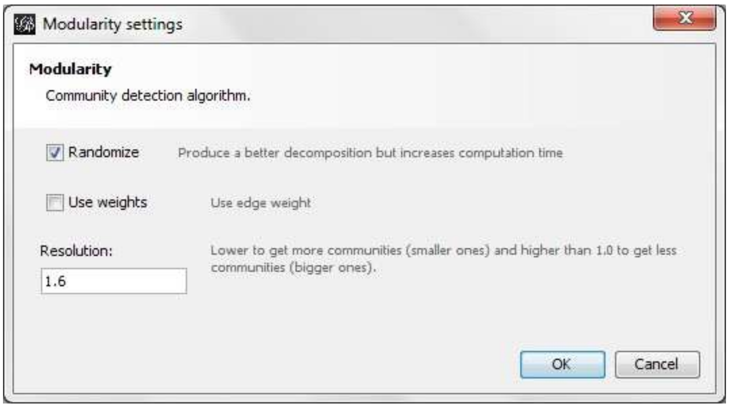

Fig. 1.14 Showing maximum force taken for calculation.
Force = 1.6 [maximum]

Fig. 1.15 Showing two communities obtained for maximum force calculation.

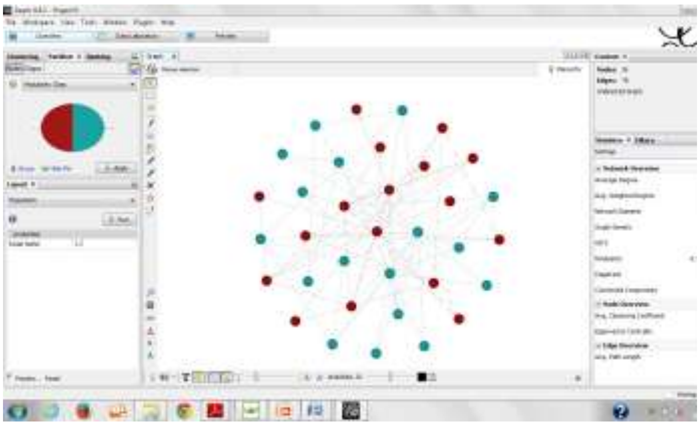

Fig. 1.16 Graphical representation of two communities obtained.

5.3.1 Result Analysis

The result for maximum force shows that when maximum amount of force is acting within the graph then number of communities that will be detected is in minimum range. The relationship is inverse & vice-versa to case of minimum forces. Reason is same only the order gets changed, because this time the force acting in graph leads to change in position or displacement of nodes thus they are not in their fixed position & very few communities are detected. Here we used force resolution = 1.6 but only two communities were obtained. It is also a case of submerging of communities leading to smaller communities getting dissolved & only larger communities remaining.

6. Conclusion

So we saw that we got three different results for three different kind of force intensity, but all belonging to same graph. For all the cases the .count of modularity or number of communities changes. In first case when force was "average" we got "four communities" since there was not much distortion in the graph & we got optimum number of communities. In second case when there was "minimum force" acting in the graph we got "nine communities" which were maximum this happened because the amount of distortion happening in the graph was minimum so all nodes retained their position & hence their communities, In third case it was vice-versa of second case because of maximum distortion nodes changed their position a lot & in last only "two communities" could be found which were minimum.

Thus the study showed that in force directed graph the amount of force exerting in the graph plays a major role in the forming of communities, the lesser the force, distortion in graph is also less & more communities are founded, whereas increasing force leads to inverse changes. In case of average forces the results also lie between maximum & minimum.

7. References

[1] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert G¨orke,"On Modularity Clustering".

[2] Mari´a C. V. Nascimento and Leonidas S.Pitsoulis,"Community Detection by ModularityMaximization using GRASP with Path Relinking".

[3] Thomas M. J. Fruchterman & Edward M. Reingold,"Graph Drawing by Force-directed Placement", *Department of Computer Science, University of Illinois at Urbana-Champaign, 1304 W.Springfield Avenue, Urbana, IL 61801-2987, U.S.A.*

[4] Boštjan Pajntar," OVERVIEW OF ALGORITHMS FOR GRAPH DRAWING*"* Department of Knowledge Technologies,Jozef Stafan Institute,Jamova 39, 1000 Ljubljana, Slovenia.

[5] P. Eades and R. Tamassia, "Algorithms for drawing graphs: an annotated bibliography", *Networks*,to appear.

[6] P. Eades, 'A heuristic for graph drawing', *Congressus Nutnerantiunt*, **42,** 149–160 (1984).7]

[7] David Kempe1, Gaurav Agarwal 2," Modularity-Maximizing Graph Communities via Mathematical Programming",1Computer Science Department, University of Southern California, Los Angeles, CA 90089, 2.Google Inc,Hyderabad,India.

[8] M. Charikar, V. Guruswami, and A. Wirth. "Clustering with qualitative information". Journal of Computer and System Sciences, pages 360–383, 2005.

[9] Santo Fortunato,"Complex Networks and Systems" Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino,I-ITALY.]

*[10]*M. Newman. "Finding community structure in networks using the eigenvectors of matrices". Physical Review E, 74,2006.

[11] M. Girvan andM. Newman. "Community structure in social and biological networks". Proc. Natl. Acad. Sci. USA, 99,2002.

[12] M. Newman. "Modularity and community structure in networks". Proc. Natl. Acad. Sci. USA, 103:8577–8582, 2006.

[13] M. Charikar, V. Guruswami, and " A. Wirth. Clustering with qualitative information". Journal of Computer and System Sciences, pages 360–383, 2005.

[14] N. Bansal, A. Blum, and S. Chawla. "Correlation clustering",Machine Learning, , 56:89–113, 2004.

[15] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. Physical Review E, 74, 2006.

[16] Jiyang Chen,Osmar R. Za • and Randy Goebel,"Detecting Communities in Social Networks using Max-Min Modularit".

[17] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. "Exploration of link structure and community-based node roles in network". In ICDM, 2007.

[18] Zachary's Karate Club data set, network data repository, university of California, networkdata.ics.uci.edu.

8. Author Profile

Syed Hasan Saeed received B.tech degree in Computer science & Engineering in 2013 from GBTU University. At present he is pursuing M.tech from SRM University, NCR Campus. He can be contacted at hasan.hasan.saeed8@gmail.com