**COMMUNITY STRUCTURE IN NETWORKS:**

**AN IMPROVED GIRVAN-NEWMAN ALGORITHM**

*A thesis in partial fulfilment for the degree of*

Master of Computer Applications

*Submitted By*

**ABHISHEK**

14/CA/638

*under the supervision of*

**Dr. Prasenjit Choudhury**

Associate Professor

NIT Durgapur



**Department of Computer Science and Engineering**

**National Institute Of Technology, Durgapur**

**May 2017**

**Declaration**

I certify that

1. The work contained in this thesis is original and has been done by me under the guidance of my supervisor.

2. The work has not been submitted to any other Institute for any degree or diploma.

3. I have followed the guidelines provided by the Institute in preparing the thesis.

4. I have confirmed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

5. Whenever I have used materials (data, theoretical analysis, figures and text) from other sources, I have given due credit to them by citing them it the text of the thesis and giving their details in the references.

ABHISHEK

****

NATIONAL INSTITUTE OF TECHNOLOGY

MAHATMA GANDHI AVENUE

*DURGAPUR, INDIA-713209*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CERTIFICATE OF RECOMMENDATION

This is certified that the work contained in this project entitled, has been carried out by ***ABHISHEK*** under my supervision and may be accepted in partial fulfilment of the undersign and this work has not been submitted elsewhere for any other degree.

Dr. Goutam Sanyal Dr. Prasenjit Choudhury

Professor and Head Assistant Professor

Department of Computer Sc. and Eng. Department of Computer Sc. and Eng.

NIT DURGAPUR NIT DURGAPUR

****

NATIONAL INSTITUTE OF TECHNOLOGY

MAHATMA GANDHI AVENUE

*DURGAPUR, INDIA-713209*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CERTIFICATE OF APPROVAL

The foregoing thesis is hereby approved as a creditable study of technological subject carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite with degree for which it has been submitted. It is to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

BOARD OF THESIS EXAMINERS

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 4.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

****

NATIONAL INSTITUTE OF TECHNOLOGY

MAHATMA GANDHI AVENUE

*DURGAPUR, INDIA-713209*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ACKNOWLEDGEMENT

I am highly grateful to Dr. Goutam Sanyal, Head of Department of Computer Applications, NIT Durgapur and Dr. Prasenjit Choudhury Assistant professor Department of Computer applications NIT Durgapur for providing this opportunity to carry out the major project.

I would like to express my gratitude to other faculty members of Department of Computer Applications NIT Durgapur, for providing academic inputs and encouragement throughout this period. The help rendered by PhD. Scholar Dhananjay Kumar Singh is greatly acknowledged. Without them it would not have been possible to carry out my project in this manner.  
  
 Finally I express my indebtedness to all who have directly or indirectly contributed to the successful completion of my major project.

ABHISHEK

NIT Durgapur

14/CA/638

INDEX

TOPICS Page number

1. Abstract 7

2. Introduction 8

3. Technologies Used 11

4. Literature Review 12

5. Algorithm and Work 14

6. Network Characteristics 22

7. Result 25

8. Conclusion and Future Work 28

9. Reference 31

# Abstract

In the study of complex networks, a network is said to have community structure if the nodes of the network can be easily grouped into (potentially overlapping) sets of nodes such that each set of nodes is densely connected internally. Communities, also called clusters or modules, are groups of vertices which probably share common properties and/or play similar roles within the graph.

In the particular case of non-overlapping community finding, this implies that the network divides naturally into groups of nodes with dense connections internally and sparser connections between groups. But overlapping communities are also allowed. The more general definition is based on the principle that pairs of nodes are more likely to be connected if they are both members of the same community(ies), and less likely to be connected if they do not share communities.

In this paper we investigate the community structure using edge-betweenness property and local maxima property defined in Newman-Girvan paper.

***Keyword - Community, edge-betweenness, Newman-Girvan, local maxima, overlapping-communities, cluster, module.***

## Why is it important?

Identifying clusters of customers with similar interests in the network of purchase relationships between customers and products of online retailers (like, e.g., www.flipkart.com, www.amazon.com) enables to set up efficient recommendation systems (Reddy et al., 2002), that better guide customers through the list of items of the retailer and enhance the business opportunities.

In a social network representing friendship between individuals, two people might be considered similar if they have similar professions, interests, or backgrounds. A good CDA helps in connecting people that share similar interest.

Identifying graph communities is a popular topic in computer science, too. In parallel computing, for instance, it is crucial to know what is the best way to allocate tasks to processors so as to minimize the communications between them and enable a rapid performance of the calculation. This can be accomplished by splitting the computer cluster into groups with roughly the same number of processors, such that the number of physical connections between processors of different groups is minimal.

# Introduction

## History

The modern science of networks has brought significant advances to our understanding of complex systems. One of the most relevant features of graphs representing real systems is community structure, or clustering, i. e. the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. Such clusters, or communities, can be considered as fairly independent compartments of a graph, playing a similar role like, e. g., the tissues or the organs in the human body. Detecting communities is of great importance in sociology, biology and computer science, disciplines where systems are often represented as graphs.

The origin of graph theory dates back to Euler's solution of the puzzle of Konigsberg's bridges in 1736 (Euler, 1736). Since then a lot has been learned about graphs and their mathematical properties (Bollobas, 1998). In the 20th century they have also become extremely useful as representation of a wide variety of systems in different areas. Biological, social, technological, and information networks can be studied as graphs, and graph analysis has become crucial to understand the features of these systems. For instance, social network analysis started in the 1930's and has become one of the most important topics in sociology (Scott, 2000; Wasserman and Faust, 1994). In recent times, the computer revolution has provided scholars with a huge amount of data and computational resources to process and analyze these data. The size of real networks one can potentially handle has also grown considerably, reaching millions or even billions of vertices. The need to deal with such a large number of units has produced a deep change in the way graphs are approached.

## Complex Networks

A complex network is a graph (network) with non-trivial topological features—features that do not occur in simple networks such as lattices or random graphs but often occur in graphs modelling of real system.

Most social, biological, and technological networks display substantial non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. Such features include a heavy tail in the degree distribution, a high clustering coefficient, assortativity or disassortativity among vertices, community structure, and hierarchical structure. In the case of directed networks these features also include reciprocity, triad significance profile and other features. In contrast, many of the mathematical models of networks that have been studied in the past, such as lattices and random graphs, do not show these features. The most complex structures can be realized by networks with a medium number of interactions.

## Community Structure

A network is said to have community structure if the nodes of the network can be easily grouped into (potentially overlapping) sets of nodes such that each set of nodes is densely connected internally. In the particular case of non-overlapping community finding, this implies that the network divides naturally into groups of nodes with dense connections internally and sparser connections between groups. But overlapping communities are also allowed.

Community structures are quite common in real networks. Social networks include community groups (the origin of the term, in fact) based on common location, interests, occupation, etc. Finding an underlying community structure in a network, if it exists, is important for a number of reasons. Communities allow us to create a large scale map of a network since individual communities act like meta-nodes in the network which makes its study easier. Individual communities also shed light on the function of the system represented by the network since communities often correspond to functional units of the system. In metabolic networks, such functional groups correspond to cycles or pathways whereas in the protein interaction network, communities correspond to proteins with similar functionality inside a biological cell. Similarly, citation networks form communities by research topic. Being able to identify these sub-structures within a network can provide insight into how network function and topology affect each other.

The community structure in graphs is very important and finding it has plagued the mathematicians and researchers alike, for a very long time.

## Significance of Community

The significance of community is basically why to go to all such lengths of finding a community in a graph. Why do we need to find communities in a graph?

To answer this question we can have many modern examples and needs where community detection can help in determining the right path:

* During election campaign finding out the voter’s support for a particular issue using community detection.
* Detecting the vulnerable nodes and links of a communication network and taking measures to protect it.
* Detecting external influence on the network because of which formation of community has happened.
* Also checking effects of external influences on the network’s which can cause new communities to form by breaking the previous communities.
* The communities formed can be interpreted to study about a networks topology and its functionalities.
* It may be so in biological networks that communities in it are basically representing the sub functionalities of the network i.e. functionalities of the network may vary completely if one of the communities breaks.
* In transport networks the community detection helps in identifying the key transport hubs as well as one can easily identify the demand for more transport link.
* The above are some real life examples where the community detection plays a significant role.

## Breaking of Community

The breaking of a community means that the community structure breaks down due to variety of reasons like change of external influences, removal of edges or nodes .etc. The effect of breaking a community can be both positive and negative or both. The number of communities in the network might change or their sizes will change or both.

The above can be explained by an example:-

A political party exists in a state. There are two major leaders in the party. Before the polls the party is in power and relationship (edge) between them is strong. After the polls when they are washed out, the relationship between the senior leaders end causing the party to split into two independent parties.

Advantages

* The advantage of breaking here is that since the party has now broken new leader will emerge in both parties as neither has sufficient candidates left for all seats.

Disadvantages

* The breaking of party will lead to the reduction in their vote bank as they’ll cut each other’s votes and thus overall performance of both parties will go down.

# Tools and Technology Used

## Python

Python is a widely used high-level programming language for general-purpose programming, created by Guido van Rossum and first released in 1991. An interpreted language, Python has a design philosophy which emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly braces or keywords), and a syntax which allows programmers to express concepts in fewer lines of code than possible in languages such as C++ or Java.

## PyCharm

PyCharm is an Integrated Development Environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains. PyCharm is cross-platform, with Windows, macOS and Linux versions.

## Gephi

Gephi is an open-source network analysis and visualization software package written in Java on the NetBeans platform. Gephi has been used in a number of research projects in academia, journalism and elsewhere, for instance in visualizing the global connectivity of New York Times content and examining Twitter network traffic during social unrest along with more traditional network analysis topics. Gephi can also import data to social networks also Facebook or Twitter and generate a graph and clusters.

# Literature Review

## Girvan-Newman

Girvan-Newman method is divisive method where edge weight is the number of shortest paths passing through the edge. That value is called edge betweenness and it is a generalization of central vertex betweenness which determines vertex influence on other vertices in network. Vertex betweennes is the number of shortest paths passing through the vertex, therefore, edge betweenness is the number of shortest paths passing through the endpoints of the edge.

The main problem of Girvan-Newman algorithm is the fact that it is not really an algorithm that has a graph as input and community structure as output. Namely, if the labelling of the vertices of the graph is rearranged, than the result of Girvan- Newman algorithm may change. Hence, there is no unique output for a given input (which should be a property of any algorithm).

The computational complexity of this part of algorithm is *O(mn),* where m is the number of edges and n is the number of vertices. After each edge betweenness calculation, the edge with highest edge betweenness is removed and the algorithm is repeated until there is no remaining edge. The algorithm complexity is therefore *O(m2n)*.

## ScaleFreeCDA

A novel parameter-free community detection algorithm based on the scale-free property of networks, named ScaleFreeCDA. The basic idea behind it is two mechanisms, i.e., node growth and preferential attachment.

The main ideas behind our algorithm are: (1) the hubs (nodes with a much higher degree) are the potential community centers for they may attract more nodes. (2) Two potential centers cannot be the real centers at the same time if they bear great resemblance, because in this case, they should be in the same community. (3) The communities are formed from hubs.

Power law distribution, i.e., some nodes have a huge number of connections, whereas most nodes have just a small number.

Preferential attachment means that a new node prefers to connect to nodes with more connections. These will lead to a few highly connected nodes and a majority lowly connected node, i.e., the degree distribution follows power law. This nonrandom feature called scale-free property.

ScaleFreeCDA runs in the worst-case time O(mlogn). Although ScaleFreeCDA is a little slower than and Louvain algorithm, it is simpler and has better performance in the quality of resulting communities in many test networks. Also, part of the efficiency limit is the way that we compute the pairwise similarities. Besides, we want to tailor our algorithm to analyze some type of biological networks.

## Louvian

Our method is a heuristic method that is based on modularity optimization. The problem of community detection requires the partition of a network into communities of densely connected nodes, with the nodes belonging to different communities being only sparsely connected.

The modularity of a partition is a scalar value between -1 and 1 that measures the density of links inside communities as compared to links between communities.

The algorithm that finds high modularity partitions of large networks in short time and that unfolds a complete hierarchical community structure for the network.

Our algorithm is divided in two phases that are repeated iteratively. Assume that we start with a weighted network of N nodes. First, we assign a different community to each node of the network. So, in this initial partition there are as many communities as there are nodes. Then, for each node i we consider the neighbours j of i and we evaluate the gain of modularity that would take place by removing i from its community and by placing it in the community of j. The node i is then placed in the community for which this gain is maximum (in case of a tie we use a breaking rule), but only if this gain is positive. If no positive gain is possible, i stays in its original community. This process is applied repeatedly and sequentially for all nodes until no further improvement can be achieved and the first phase is then complete.

This first phase stops when a local maxima of the modularity is attained, i.e., when no individual move can improve the modularity. The second phase of the algorithm consists in building a new network whose nodes are now the communities found during the first phase.

Time complexity of Louvian algorithm is O(nlogn).

# Algorithm and Work

## Detection of communities

Communities, also called clusters or modules, are groups of vertices which probably share common properties and/or play similar roles within the graph. We performed six CDA on Football Network, Zacchary’s Karate Network, Dolphin Network and Simple Networks.

#### SUMMARY OF NETWORK DATASETS AND THE NUMBER OF COMMUNITIES DETECTED BY **GIRVAN-NEWMAN**, **FAST GREEDY**, **LABEL PROPAGATION**, **LOUVAIN**, **INFOMAP, WALKTRAP AND IMPROVED GN ALGORITHM**

|  |  |  |
| --- | --- | --- |
| **Real Networks** | **Number of Nodes/Edges/Real Communities** | **Number of communities detected**  **by the seven algorithms** |
| **Football network** | 115/613/12 | 10/6/12/10/12/10/13 |
| **Zachary’s network** | 34/78/2 | 5/3/3/4/3/4/2 |
| **Dolphin’s network** | 62/159/2 | 5/4/5/5/6/5/4 |
| **Simple Network I** | 14/48/3 | 3/3/3/3/3/3/3 |
| **Simple Network II** | 13/22/2 | 4/4/3/2/2/4/3 |
| **Simple Network III** | 14/17/4 | 6/4/6/6/4/6/6 |

## Constant Communities

The vertex orderings produce very different sets of communities, some groups of vertices are always allocated to the same community for all different orderings. We define the group of vertices that remain invariant as constant communities.

We performed analysis on football network with six CDAs

GIRVAN-NEWMAN, FAST GREEDY, LABEL PROPAGATION, LOUVAIN, INFOMAP AND WALKTRAP

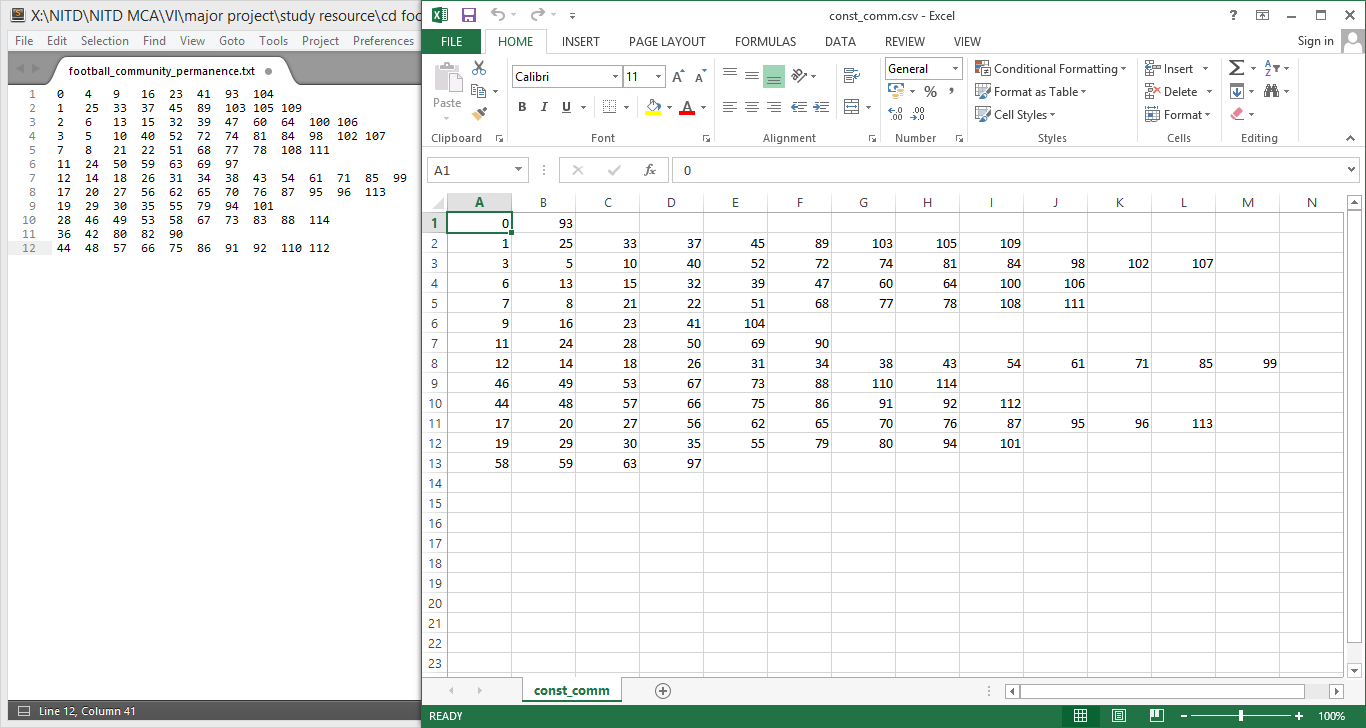


Figure : Left doc represents the constant community performed by third party and right doc represents the constant community returned by our python script.

We observed that the cluster size has increased from 12 to 13 giving rise to one extra cluster in the football network.

## Structural Property

### Degree Centrality

In the case of a directed network (where ties have direction), we usually define two separate measures of degree centrality, namely in-degree and out-degree. Accordingly, in-degree is a count of the number of ties directed to the node and out-degree is the number of ties that the node directs to others. The degree centrality *Cd* for node *vi* in an undirected graph is

*Cd(vi) = di*eq. 1

Simple normalization methods include normalizing by the maximum possible degree,

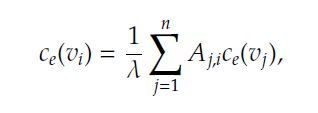
*Cnorm d(vi) = di / n-1* eq. 2

where *n* is the number of nodes.

### Eigen Vector Centrality

Eigenvector centrality (also called eigencentrality) is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

We want the centrality of *vi* to be a function of its neighbors’ centralities. We posit that it is proportional to the summation of their centralities

 eq. 3

where λ is some fixed constant. Assuming *Ce = (Ce(v1);Ce(v2); : : : ;Ce(vn))T*is the centrality vectors for all nodes, we can rewrite Equation 2 can be written as

Capture eq. 4

### Betweenness Centrality

#### Vertex Betweenness Centrality

Vertex betweenness is a centrality measure of a vertex within a graph. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

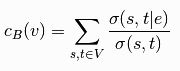
The betweenness of a vertex *v* in a graph *G:=(V,E)* with *V* vertices is computed as follows:

1. For each pair of vertices *(s,t)*, compute the shortest paths between them.
2. For each pair of vertices *(s,t)*, determine the fraction of shortest paths that pass through the vertex in question (here, vertex *v*).
3. Sum this fraction over all pairs of vertices *(s,t)*.

#### Edge Betweenness Centrality

Edge betweenness of an edge as the number of shortest paths between pairs of nodes that run along it. If there is more than one shortest path between a pair of nodes, each path is assigned equal weight such that the total weight of all of the paths is equal to unity. If a network contains communities or groups that are only loosely connected by a few inter-group edges, then all shortest paths between different communities must go along one of these few edges. Thus, the edges connecting communities will have high edge betweenness (at least one of them). By removing these edges, the groups are separated from one another and so the underlying community structure of the network is revealed.

Edge Betweenness centrality of an edge e is the sum of the fraction of all-pairs shortest paths that pass through e:

 eq. 4

where *V* is the set of nodes,`*sigma(s, t)`* is the number of shortest *(s, t)-*paths, and sigma *(s, t|e)* is the number of those paths passing through edge *e*.

### Closeness Centrality

Closeness is a measure of the degree to which an individual is near all other individuals in a network. It is the inverse of the sum of the shortest distances between each node and every other node in the network. Closeness is the reciprocal of farness.

C:\Users\Gagan Brar\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Capture.jpg eq. 5

where *d(y,x)* is the distance between vertices *x* and *y*.

### Modularity

Modularity is a measure that tells how good a partition is. It tells that a optimal decomposition of a network is reached.

Let us define *k×k* matrix *E* such that *eij* is the fraction of number of edges that connect vertices between community *i* to community *j* in the total edge number. Then on the diagonal of the matrix *E* is the fraction of edges that are located within the same community, so the trace of the matrix *tr(E)* is the fraction of edges that will not be removed in process of removing edges. A good partition is the partition with high value of the trace.

C:\Users\Gagan Brar\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Capture.jpg eq. 6

where || || is the sum of matrix elements, as the fraction of edges that connect vertices within communities decreased with the expected fraction of edges within partition in random graph. When the fraction of edges within the communities is higher than it is in random graph, the value is Q = 0. As Q is approaching the value 1, the community structure in network is better. In most cases, the value of Q is between 0.3 and 0.7

# Girvan-Newman improved

The algorithm's steps for community detection are summarized below

1. The betweenness of all existing edges in the network is calculated first.
2. The edge with the highest betweenness is removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until local maxima of modularity is achieved.

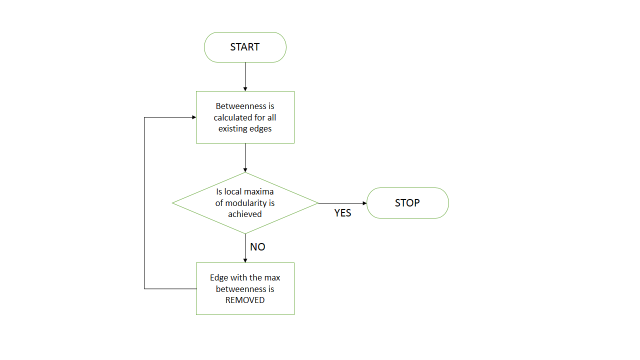
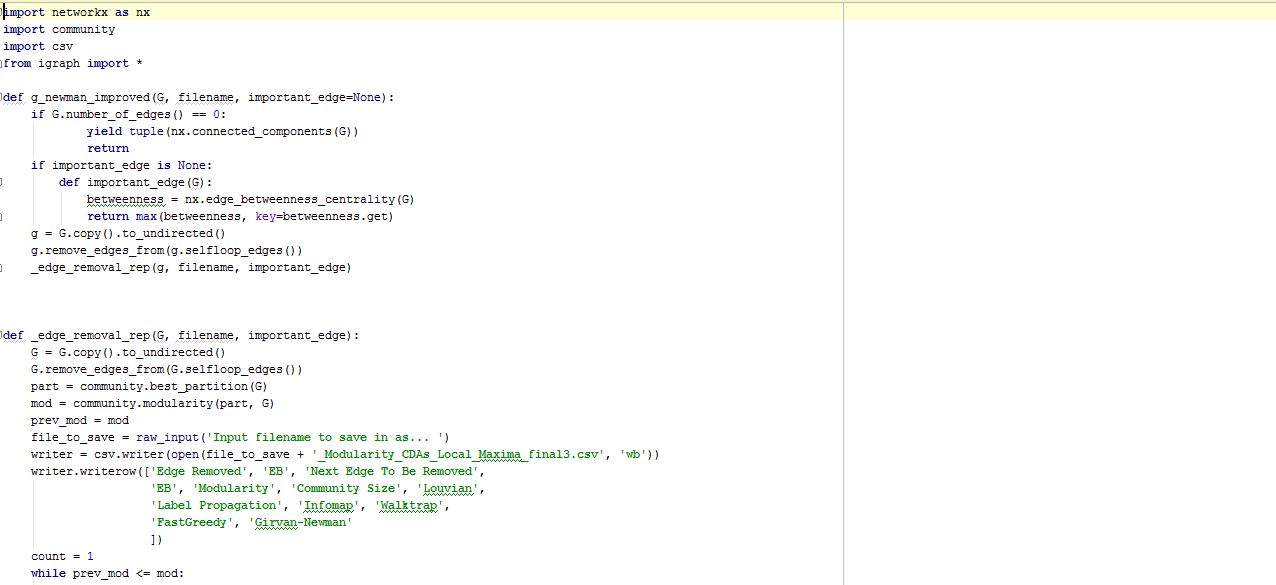
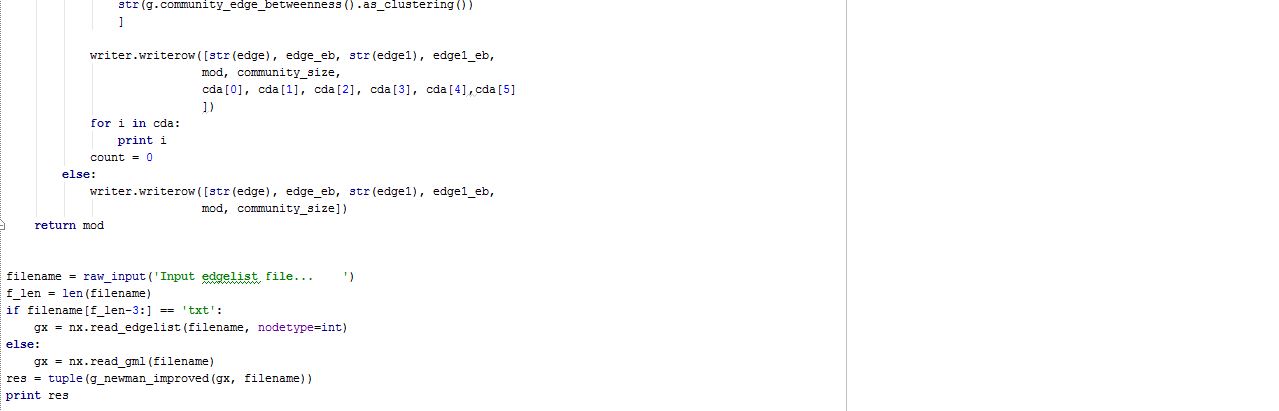


Figure : Flowchart of Girvan-Newman Improved

## Code



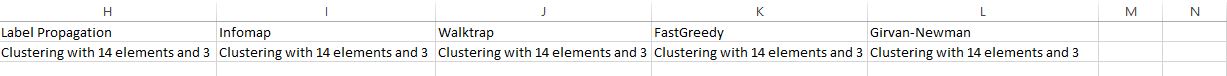




## Code Output

Showing result for Toy Network(Simple Network I). As mentioned our algorithm successfully identified 3 communities.

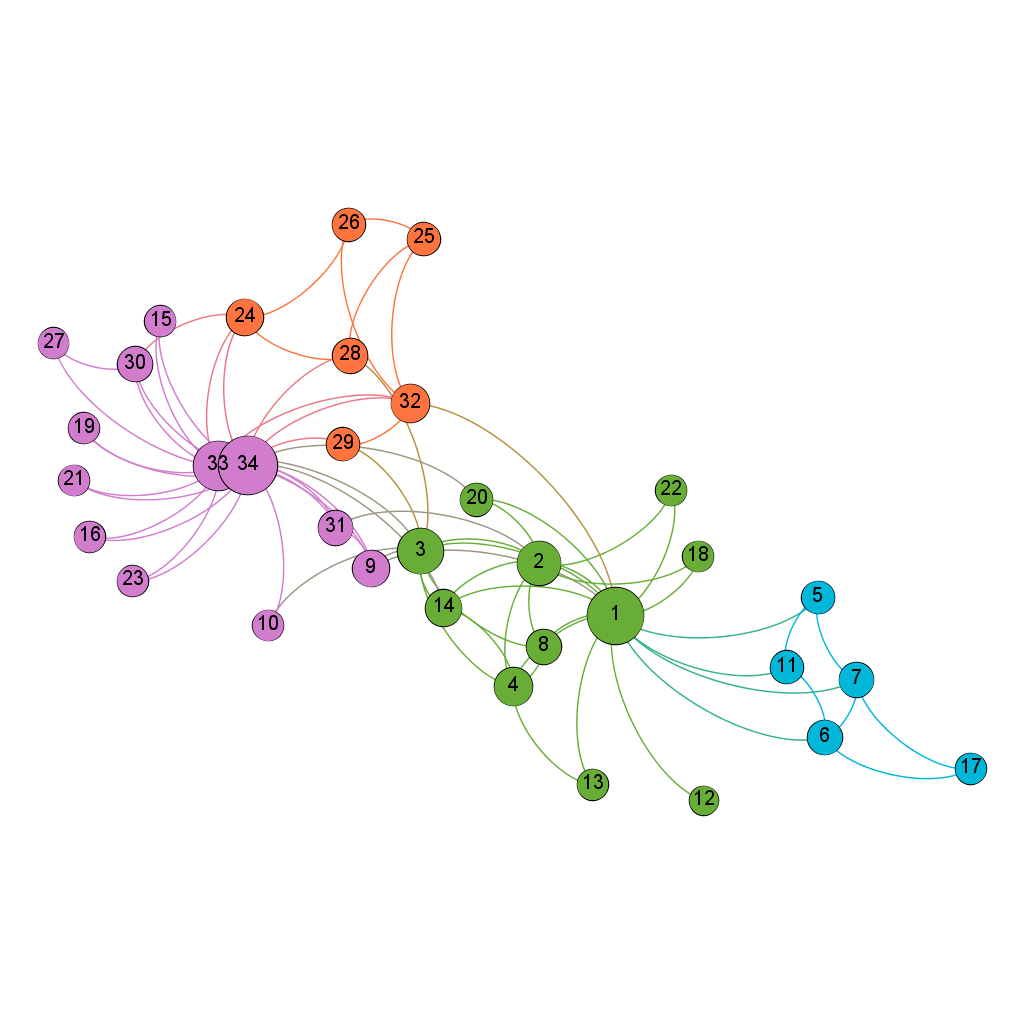




# Network Characteristics

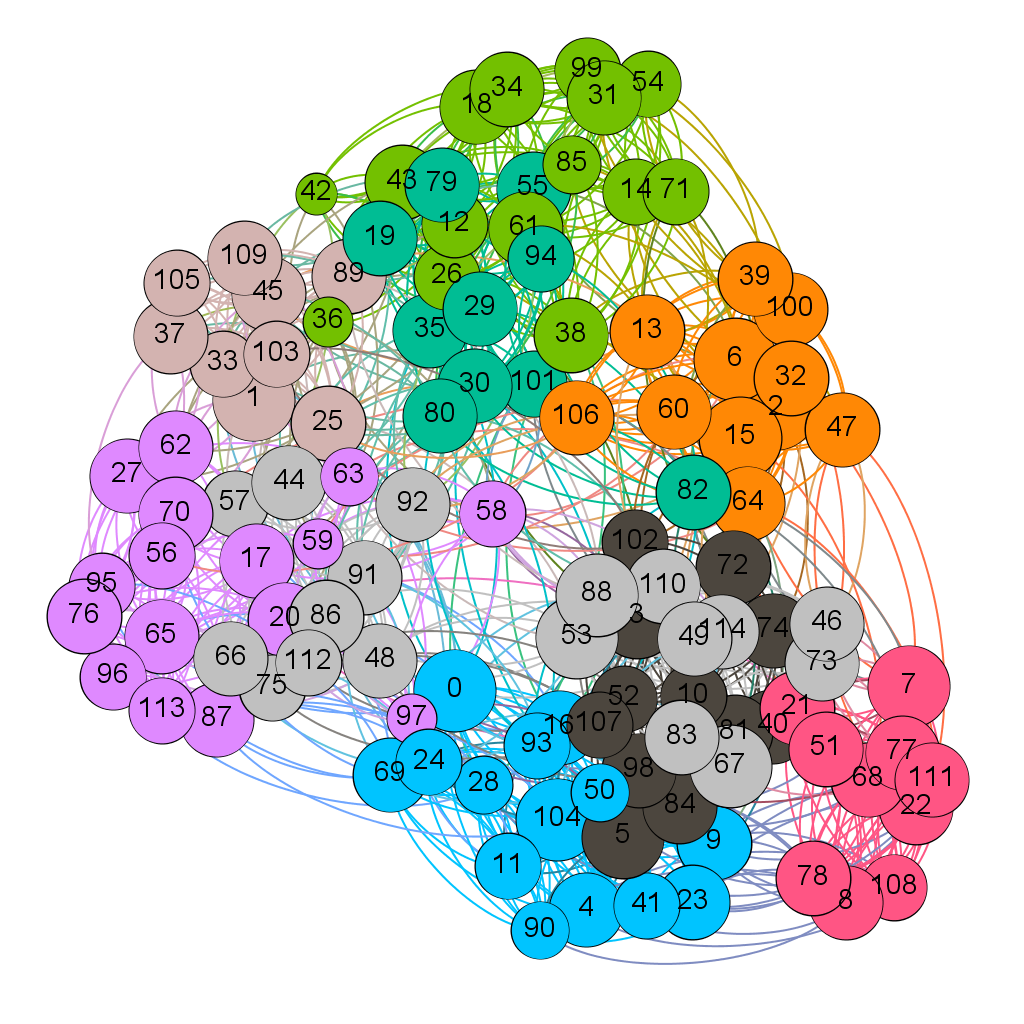
Karate club

The Zachary Karate Club is a well-known social network of a university karate club described in "An Information Flow Model for Conflict and Fission in Small Groups" paper by Wayne W. Zachary. The network captures 34 members of a karate club, documenting 78 pairwise links between members who interacted outside the club. During the study a conflict arose between the administrator "John A" and instructor "Mr. Hi" (pseudonyms), which led to the split of the club into two.



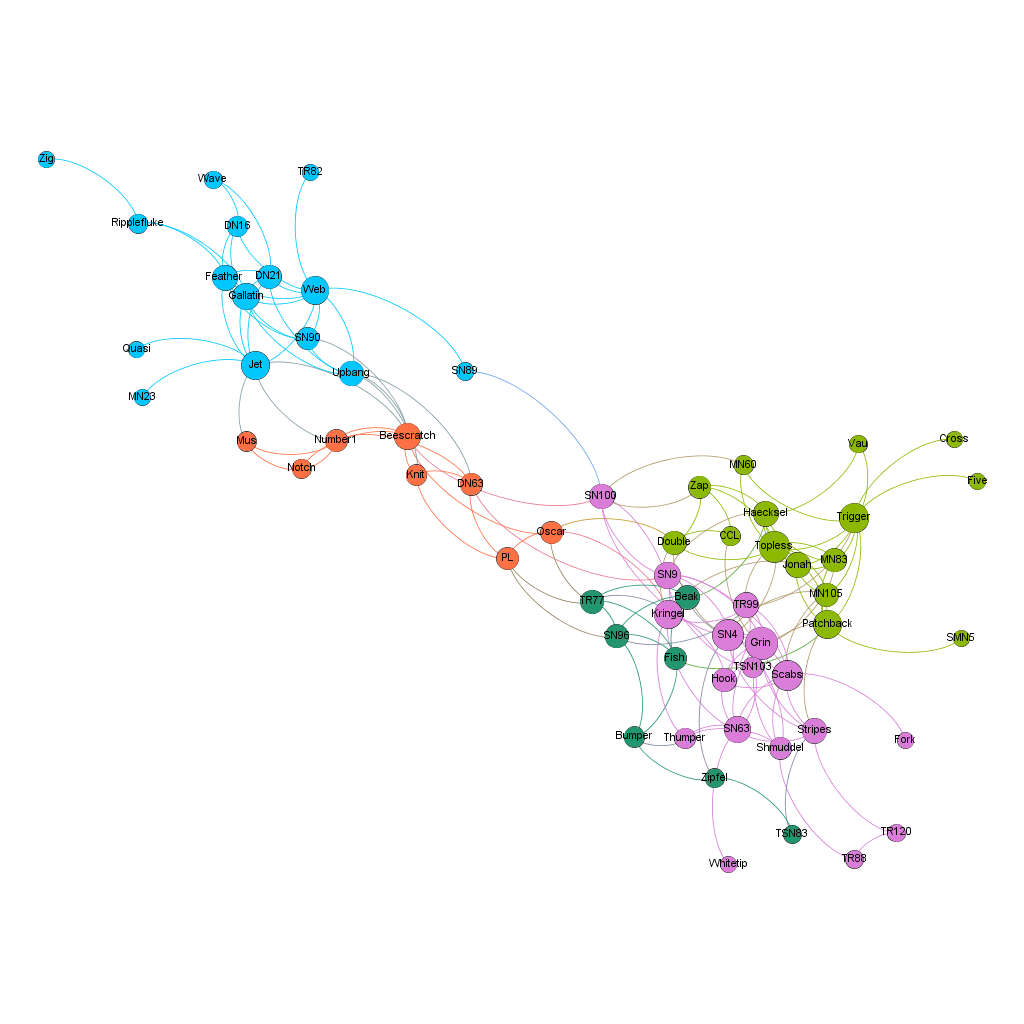
American Football College

Collected by Girvan and Newman, this dataset concerns football games that opposed American Football College in a given season. Nodes of the graph represent teams and edges mean that end-point teams have disputed a match. The Conference membership of each team is available, teams in the same conference played more games than with other ones.



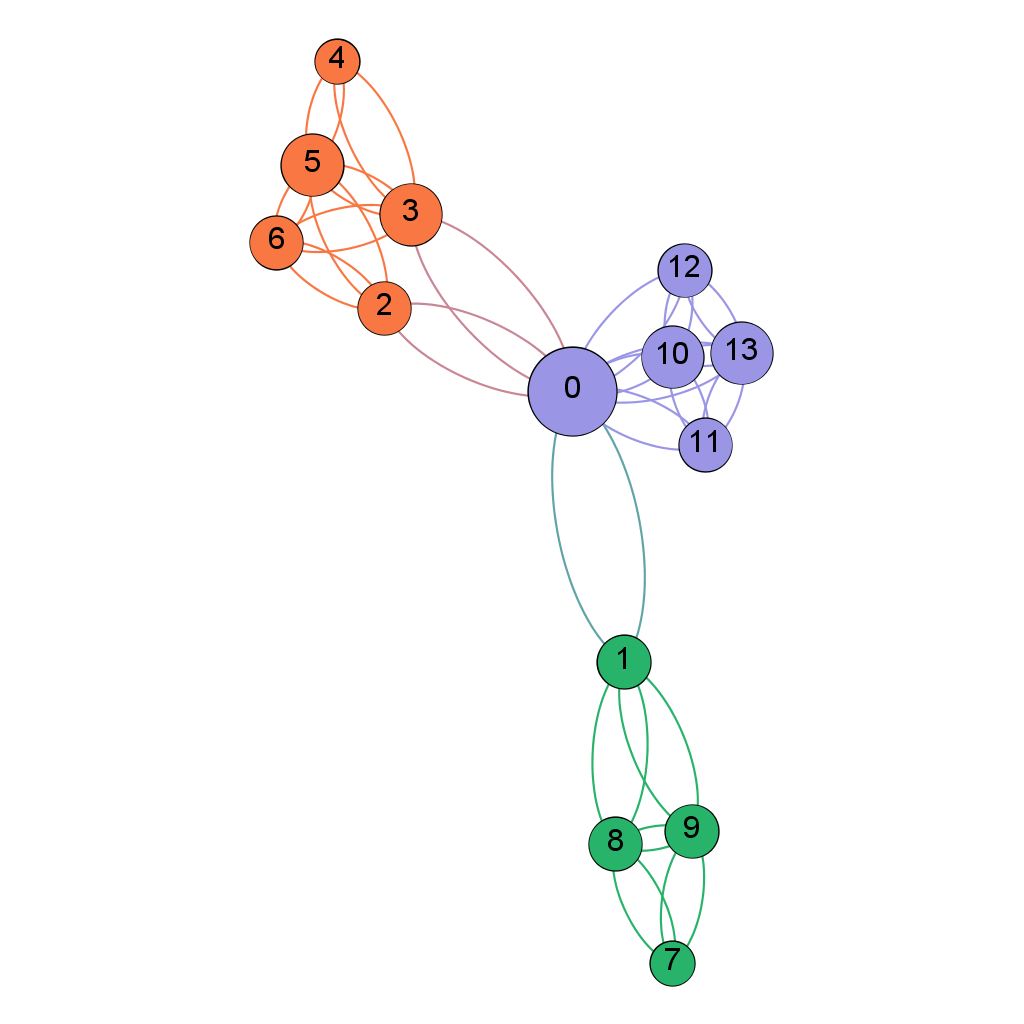
## Dolphins

It is a network of frequent associations between 62 dolphins living off Doubtful Sound in New Zealand compiled by David Lusseau. Edge between two dolphins represent their close association with one another.

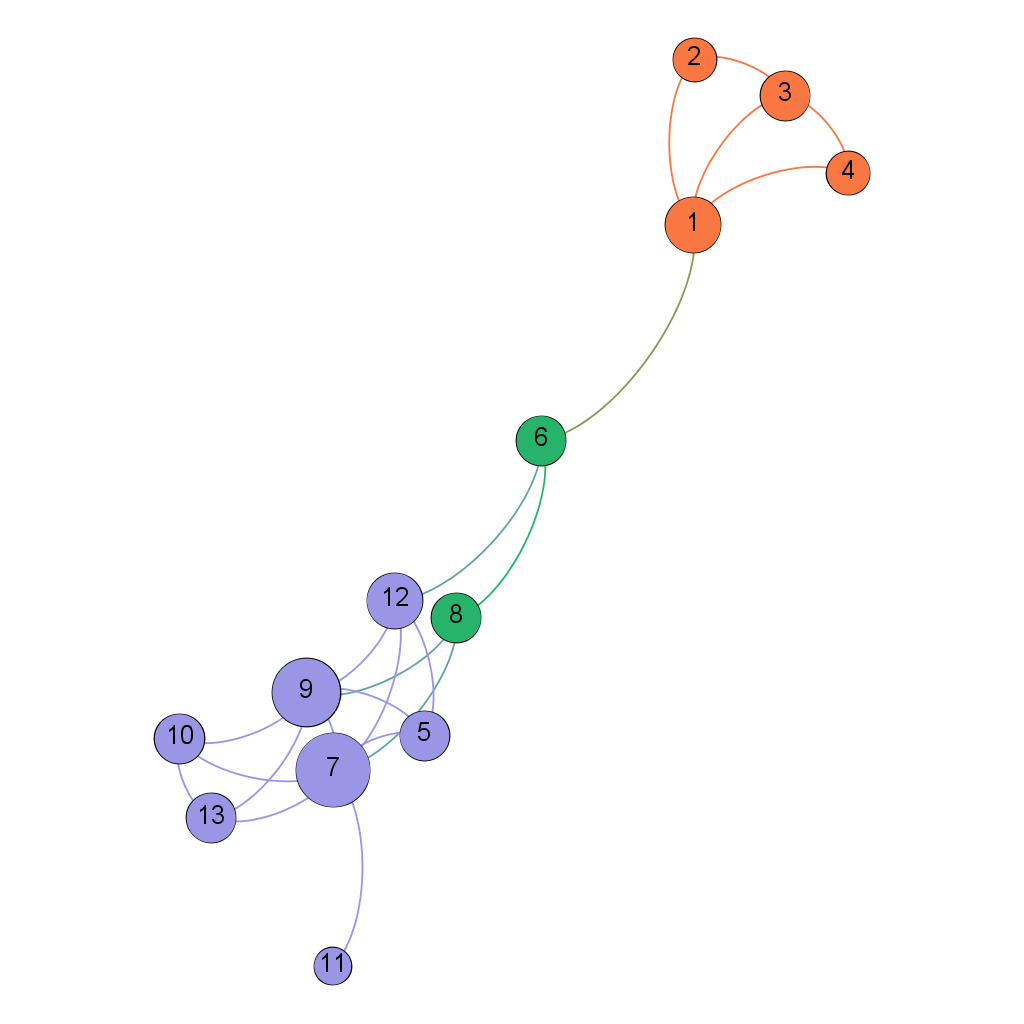


## Simple Networks

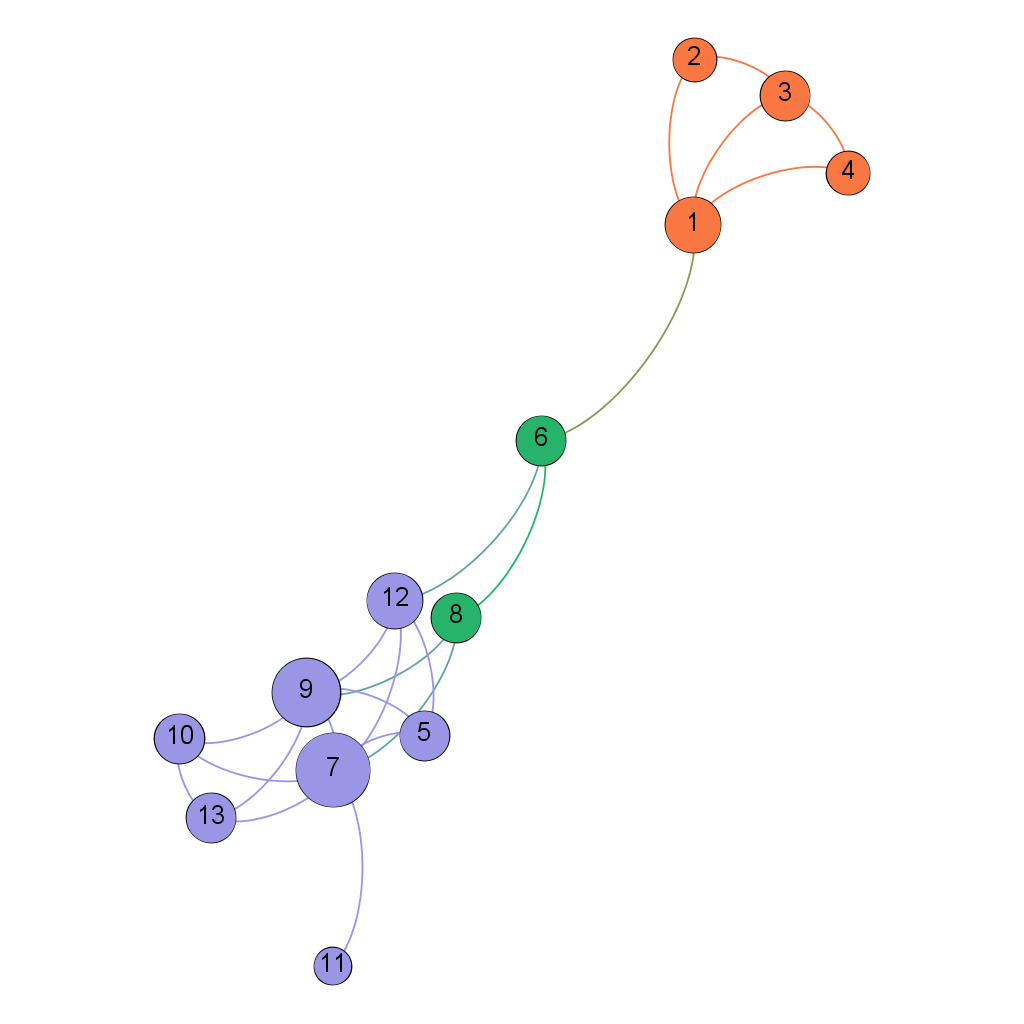
1. It has 14 nodes and 48 edges.



1. It has 13 nodes and 22 edges.



1. It has 14 nodes and 17 edges.



# Result

We tested the Girvan-Newman algorithm with the single edge removal and improved algorithm with the local maxima condition edge removal on three different complex networks. The first, Zachary’s karate club; the second, football network; and the third, simple network.

In Zachary’s karate club, the number of communities detected by improved Girvan-Newman Algorithm is 2, whereas GN gives 5 communities.

In American Football Network, the number of communities detected by improved Girvan-Newman Algorithm is 13, in comparison to GN’s 10 communities.

In Simple network(Toy Network), the number of communities detected by improved Girvan-Newman Algorithm is 3 and GN also outputs 3 communities.

Our improved algorithm successfully discovered community structure with reduced number of operations but retains the same computational complexity.

# Conclusion and Future Work

Girvan-Newman algorithm is one of the first algorithms that deals with detecting communities in networks and as such suffers from certain ”childhood diseases”. Number of operations is proportional to m2n, or n3 for sparse networks.

Our improved algorithm runs with reduced number of operation but retains the same time complexity.

To improve our algorithm, we want to do further work from two facets:

* We need to tailor our algorithm to analyze some type of biological networks.
* We need to test the reliability by using large dataset with 10000 or more number of nodes and edges.

# References

M. Girvan and M. E. Newman, "Community structure in social and biological networks," Proceedings of the National Academy of Sciences, vol. 99, pp. 7821-7826, 2002.

Xiyang Liu, Qiong Pan, Yingying An, Guimin Qin, “Community detection algorithm based on the scale free property of networks” Proceedings of the IEEE Computer Society, pp 2270-2275, 2015.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theor Exp 2008, 2008:P10008.

M. C. Nait-Hamoud, F. Didi, Y. Boualleg, “Core Community Detection Algorithm based on Edge Removal Learning” MedPRAI-2016

M. E. J. Newman. 2004. Fast algorithm for detecting community structure in networks, Phys. Rev. E, 69, 6 (2004), 066133. DOI: 10.1103/PhysRevE.69.066133.

Ljiljana Despalatovi´, Tanja Vojkovi´and Damir Vukiˇcevi´, ” Community structure in networks: Girvan-Newman algorithm improvement” Proceedings of MIPRO, pp. 996-1002

Tanmoy Chakraborty, Sriram Srinivasan, Niloy Ganguly, Sanjukta Bhowmick & Animesh Mukherjee, “Constant Communities in Complex Networks” proceedings of Scientific Reports 3, Article number: 1825 (2013)