Community Analysis on Coauthorship network

Arvind Murugan

VSB – TUO

Contents

[1 Introduction 2](#_Toc41843582)

[1.1 Dataset description 2](#_Toc41843583)

[1.1.1 Nodes: 2](#_Toc41843584)

[1.1.2 Edges: 2](#_Toc41843585)

[2 Analysis 3](#_Toc41843586)

[2.1 Connected Component Analysis 3](#_Toc41843587)

[2.1.1 Description 3](#_Toc41843588)

[2.1.2 Network 3](#_Toc41843589)

[2.1.3 Size Distribution 4](#_Toc41843590)

[2.2 Community Analysis 5](#_Toc41843591)

[2.2.1 Description 5](#_Toc41843592)

[2.2.2 Network 5](#_Toc41843593)

[2.2.3 Size Distribution 6](#_Toc41843594)

[3 Conclustion 7](#_Toc41843595)

# Introduction

A cluster analysis on Coauthorship network of scientists working on network theory and experiment, as compiled by M. Newman in May 2006 using Gephi software.

A figure depicting the largest component of this network can be found here. M. E. J. Newman, Phys. Rev. E 74, 036104 (2006).

## Dataset description

The graph dataset was officially provided by Gephi on their own website (<https://github.com/gephi/gephi/wiki/Datasets>) as a GML file.

### Nodes:

* Consists of Id and Label
* Total Nodes: 1589

### Edges:

* Undirected edges
* Consists of Source, Target, Id and Weight
* Total Edges: 2742

# Analysis

## Connected Component Analysis

### Description

The connected components statistics in Gephi calculates the number of weakly connected components in a network using Depth-First Search and Linear Graph Algorithms proposed by Mr. Robert Tarjan.

### Network

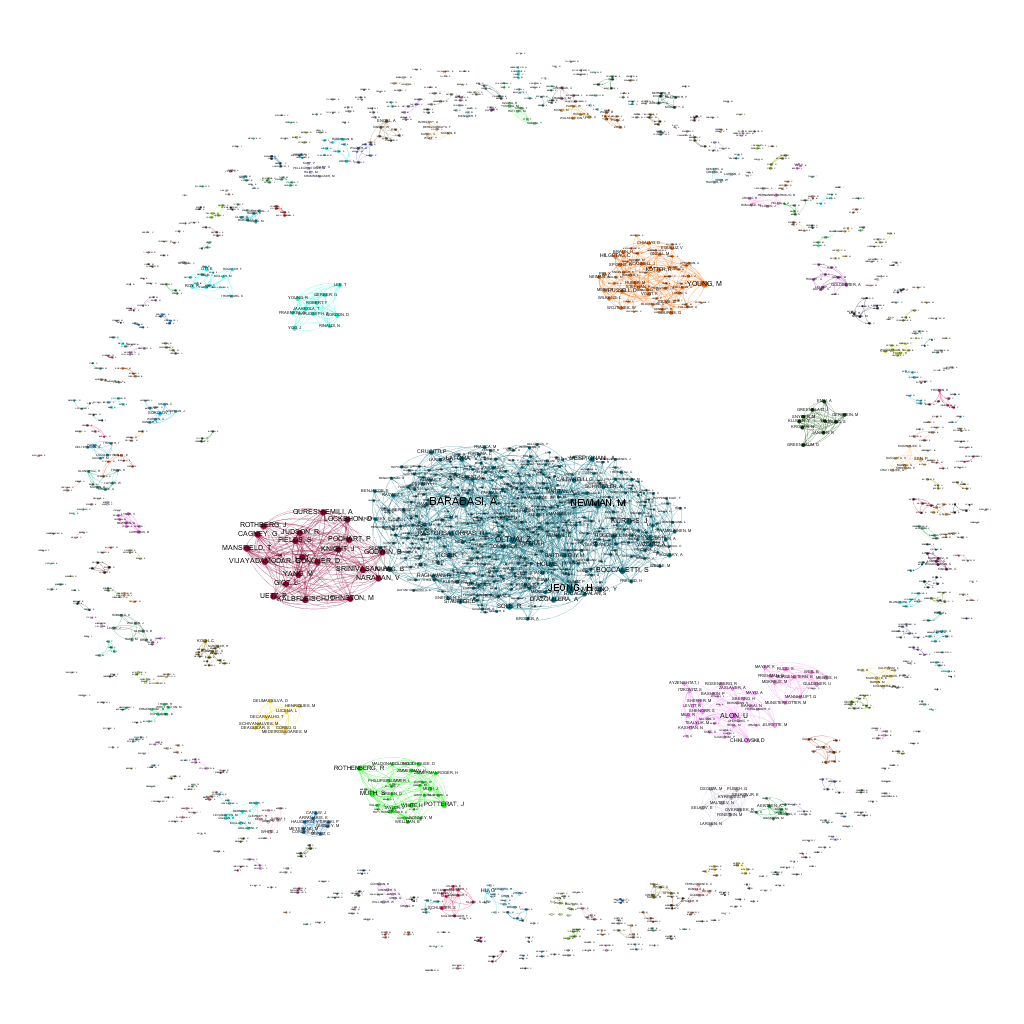
* Node size: The size of the node represents the degree of the node.
* Node color: The color of the node represents the component to which the node is connected.

Figure 2‑1 Connected Components

### Size Distribution

* Number of Weakly Connected Components: 396

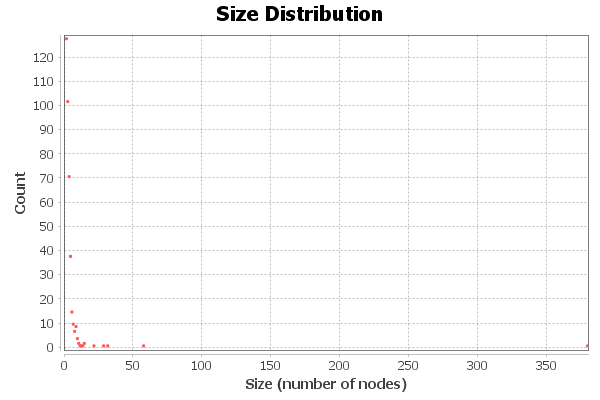


Figure ‑ Size Distribution CC

## Community Analysis

### Description

The cluster of communities can be detected using the Modularity Statistics in Gephi. The algorithm used was Fast unfolding of communities in large networks by Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre.

### Network

* Node size: The size of the node represents the degree of the node.
* Node color: The color of the node represents the component to which the node is connected.
* Label color: The color of the label represents the community in which the node belongs to.

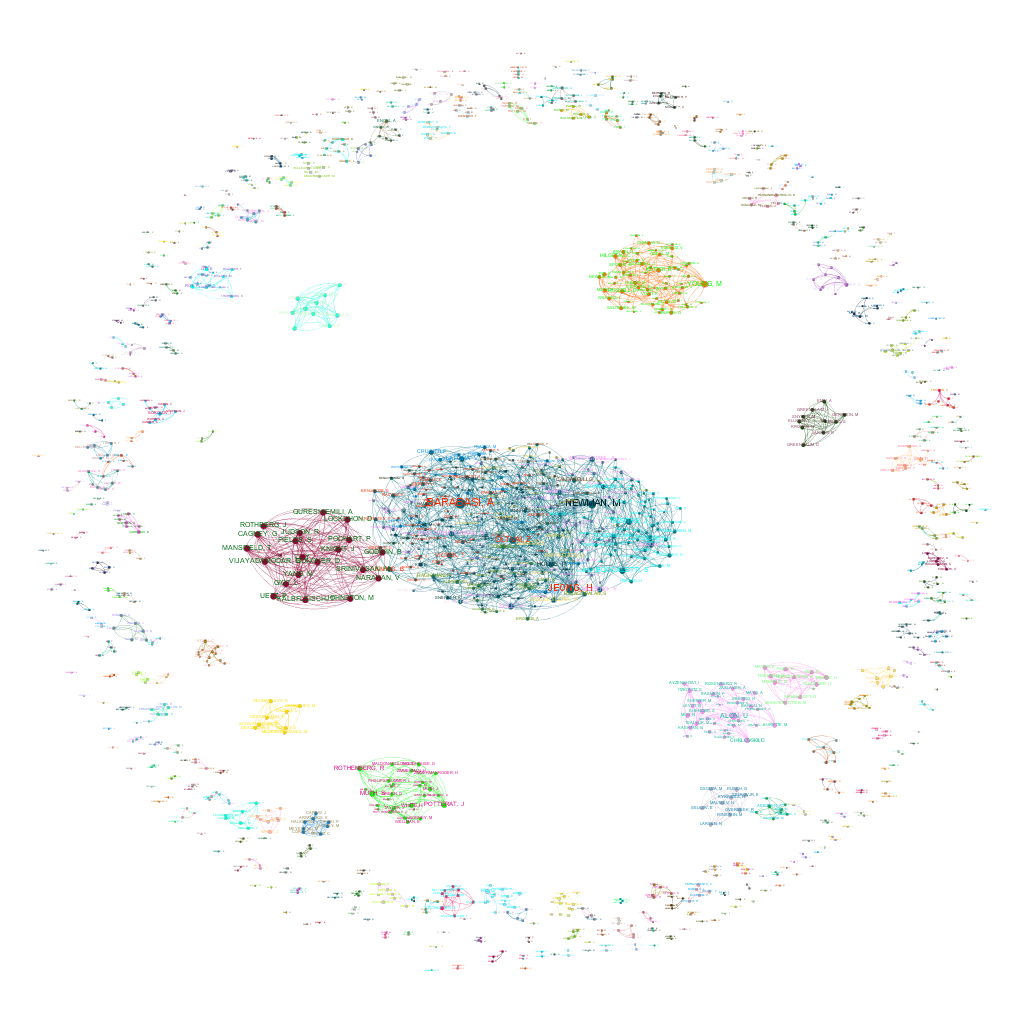


Figure ‑ Community Network

### Size Distribution

* Modularity: 0.955
* Modularity with resolution: 0.955
* Number of Communities: 406

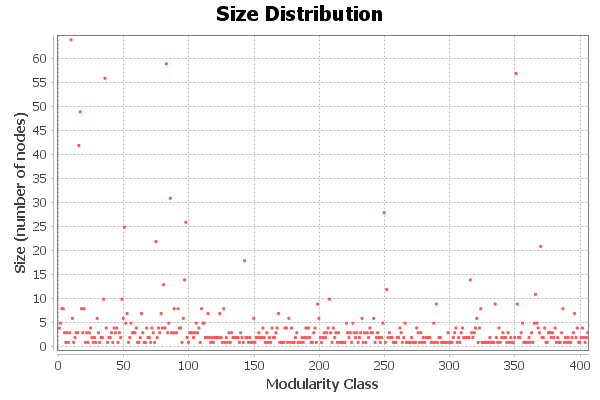


Figure ‑ Size Distribution Communities

# Conclustion

From the size distribution of connected components it is clear that there are only few components that contains high number of nodes. Most of the components have very less number of nodes.

The size distribution of communities shows the sparse distribution of communities with higher number of nodes and more densely populated with communities with fewer nodes.

The networks of communities also reveal there are many overlapping communities exists in this graph. We can see that the label color and the node color in some communities differ. This represents there are few authors who combine with authors from different fields of networking.

Barabasi, Newman, Oltvai, Jeong, Holme, Vicsek are some of the important authors in this network who are connected to different communities and different components.

For an example, Jeong is connected to Barabasi and Holme, who are from different communities and they both don’t have a connection. Altogether they form a giant component of the network.