VISUALIZATION OF COMPLEX NETWORK COMMUNITIES BASED ON NETWORK MEASUREMENT

*I. Suriana, I. Roslan,
[1]Universiti Kuala Lumpur, Malaysian Institute of Information Technology, 1016 Jalan Sultan Ismail, 50250 Kuala Lumpur, Malaysia
*For Correspondence; suriana@unikl.edu.my

ABSTRACT: *Community detection is a common problem in graph mining that consists of finding similar member in a group. Most majority falls within its own group than other groups. Various community detection approaches have been designed and developed to identify hidden community lies within a complex network. Details information on the community found can be derived through visualization of the network based on the network measurements. In this paper, various measurements are studied and visualization comparison of the network measurements between them is presented.*

Keywords: Complex Network, Visualization, Network Measurement, Community Detection, Modularity, Gephi.

1.   INTRODUCTION

Recent research focused much on networks because of their suitability to represent many real world complex networks. These networks invite a huge and wide range of knowledge discovery research, particularly to know the hidden information of the networks. Many workers have focused on discovering communities in large networks, mostly, dedicated to acquiring the true meaning of a community [1].

Some of the real research interest in community detection in particular evolved from a wide range of real –world network. A community can be defined as a subgraph of a network having higher number of similar nodes tightly connected with each other than with the nodes outside the subgraph. Hence, a good community has a high interaction of nodes within the same community while a low interaction of nodes between communities denotes a less quality of communities found [2, 3, 4]. Its important contributions in determining the underlying information of a community can be further explored to understand the true meaning and valuable information of a particular network.

Many types of systems are evolving from the smallest to enormous network such as social network [5], technological network [6] or biological network [4] are currently being studied researcher. These networks invite a huge and wide range of knowledge discovery of research, particularly to know and understand the hidden information of the networks. These systems are organized as a complex network, which can be further grouped into modules or clusters.

2.   MATERIAL AND METHODS

Network Measurement

Network measurements are computation of different interaction type between edges and nodes in a network. These measurements indicate the strengthness or tightness of a newly formed community or group. The most important centrality measures are degree centrality and betweenness centrality.

Degree of centrality denotes the number of direct connections a node has, measures a node's local connectivity, and identifies connectors or hubs in a network. It can be defined as the number of links incident upon a node (i.e., the number of ties that a node has). Degree centrality can represent level of activity or popularity depending on the number of link or ties it has per node. Lots of ties coming in and lots of ties coming out of a node would increase degree centrality.

Betweenness centrality is the number of times a node connects pairs of other nodes, who otherwise would not be able to reach one another. The centrality value denotes the number of shortest paths from all nodes to all others that pass through a node. It measures load placed on a node in which a node with high betweenness has great influence over what flows in the network.

Complex Network

Complex network can be considered as a network consists of highly interconnected units. For example, there are online community, technological, biological, and social networks among networks that often have been understudied by researchers [7].

The goal of community detection is to identify sets of nodes with common function based only the connectivity structure of the network. The community found by an algorithm can be further examine and analyze in term of its strengthness interaction between its own node. These are made possible by the network measurement. In particularly for biological networks, by applying the network measurement and applying further computation of modularity function, it could provide additional and useful information for important species discovered, thus contributed to a new knowledge for phylogenetic studies [8].

Gephi Tool

Gephi is an interactive visualization and exploration which can handle well the visualization of very large and dynamic networks. Gephi provides state-of-the-art layout algorithms which cover both efficiency and quality [9]. Its goal is to assist data analyst to produce hypothesis, finds pattern which can handles up to 50,000 nodes and 1,000,000 edges. The statistics and metrics framework offer the most common metrics for social network and scale-free networks.

Gephi facilitate network analysis by calculating complex network parameters like average clustering coefficients, shortest paths, and node degrees, as well as centrality measures like stress centrality, betweenness centrality, and closeness centrality. The calculated parameters can be visualized as histograms and mapped as visualization attributes on the network in the form of node size and color.

Gephi tends to focus on network visualization. It is based on the principle What You See Is What You Get (WYSIWYG) which able to generate powerful visualizations. Since it is an open source tools, make it desirable as an alternative network exploration, manipulation, and analysis by providing filtering, layout, coloring, and clustering capabilities.

3.   DATA NETWORK SETUP

The objectives of this experiment are to perform network measure analysis on the network as mentioned in the previous section. To illustrate the visualization capabilities of Gephi, we collected known dataset from Ucinnet and Snap which normally being used being used a knowledge base dedicated to large network datasets (Ucinnet 2016; Snap 2016). Dataset from various type of network as used in our study as depicted in Table 1 below.

Table 1: Network Details

Network	Number of Node	Number of Edge
Yeast	2361	7182
C.elegans	297	2359
Karate	34	78
Les Miserables	77	254

4.   VISUALIZATION

We applied several force-directed layout algorithms to those networks. We chose force-directed layouts as they rely on the structure of the network and thus do not require domain-specific information. These layout types furthermore are visually more appealing as they reduce edge crossing and reveal symmetries within the graph. The idea of a force directed layout algorithm is to consider a force between any two nodes. In this algorithm, the nodes are represented by steel rings and the edges are springs between them. The basic idea is to minimize the energy of the system by moving the nodes and changing the forces between them [10].

Force Atlas 2 uses a different set of options that provide more control over the output by enabling you to set parameters for hubs and gravity, as well as the aforementioned repulsion [11]. Gravity draws nodes to the center, while dissuade hubs push nodes out toward the borders of the graph. Following are outcome of the networks on different layout algorithm as depicted in Figure (1a-4c) below.
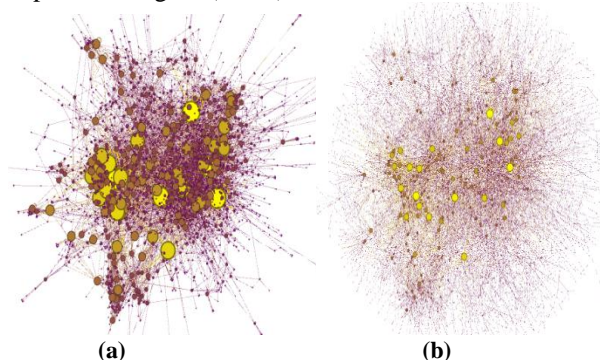

(a)   (b)

Figure 1a Yeast Network with Force Atlas2
Figure 1b Yeast Network with Fructherman Reingold
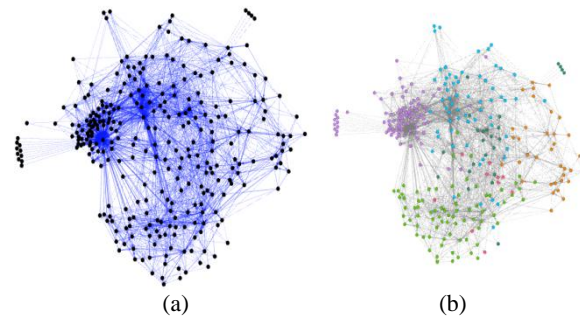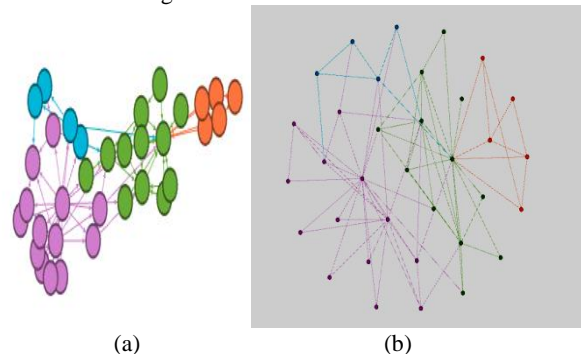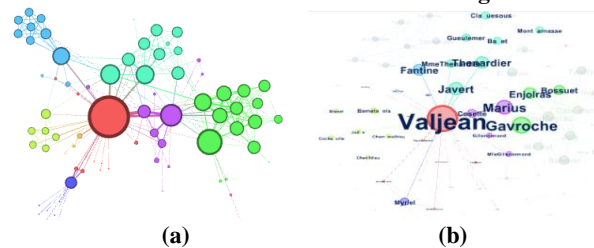

(a)   (b)

Figure 2a C.elegans Network with Force Atlas2, 2b C.elegans with Fruchterman Reingold

(a)   (b)

Figures 3a Karate Network with Force Atlas2, 3b Karate Network with Fruchterman Reingold

(a)   (b)

Figures 4a Les Miserables with Force Atlas2, 4b Fruchterman Reingold layout
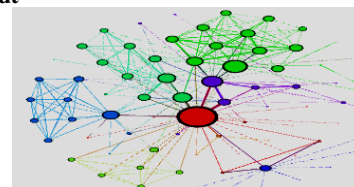

Figure 4c Applied Modularity

5.   RESULTS AND DISCUSSION

Based on the experimental study on the available dataset as in Table 1, following are results gathered by utilizing two different layouts of Force Atlas2 and Fruchterman Reingold. Number of major communities found as depicted in Table 2 show a strong interaction between communities as it falls between 0.3-0.7 ideal modularity values.

Table 2: Modularity value by various network

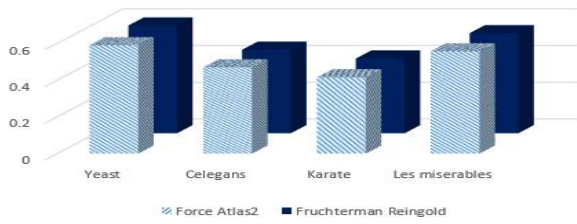| Network	Modularity value	Community Found
Yeast	0.59	122
C.elegans	0.47	5
Karate	0.416	4
Les Miserables	0.557	5

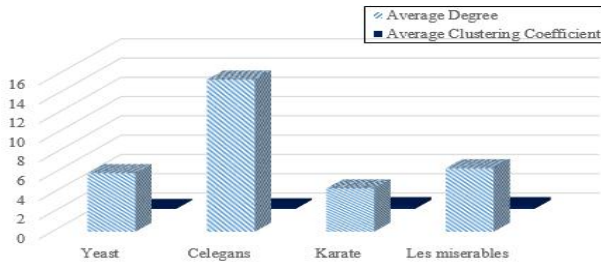**Fig 5: Modularity value in comparison with different layout**

Fig 6: Community found in various networks

6.   CONCLUSION

We found the Force Atlas2 layout as the most appealing, as it provides some form of structure for very large networks with a slightly higher modularity value. Both chosen layout options are separating the small network in several, more tightly connected clusters, with the Force Atlas2 layout performing best with respect to visual separation of sub clusters.

We were able to analyze and compare results from different layout on network size, number of communities, average degree distribution, and average clustering coefficient in which serves as an input to our future work.

7.   ACKNOWLEDGEMENTS

8.   REFERENCES

[1]   Fortunato, S., Community detection in graphs. *Physics Reports*, 486(3-5): 75-174, 2010.

[2]   Girvan M and Newman M E J. Community structure in social and biological networks. *Proc. Natl Acad. Sci*. USA99:7821, 2002.

[3]   Newman M E J and Girvan M. Finding and evaluating community structure in networks. *Phys. Rev*., E 69: 026113, 2004.

[4]   Yang J and Leskovec J. Defining and Evaluating Network communities based on Ground-truth. In *ICDM*, 2012.

[5]   Clauset A, Moore C, and Newman M E J. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453: 98, 2008.

[6]   Newman M E J. Detecting community structure in networks. *Eur. Phys. J. B*. 38: 321–330, 2004.

[7]   Blondel V D, Guillaume J L, Lambiotte R, and Lefebvre E. Fast unfolding of communities in large networks. *J. Stat. Mech*. P10008, 2008.

[8]   Hao X, Jiang R, Chen T., 2011. Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. *Bioinformatics*. 27: 611–618.

[9]   Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Web and Social Media**,** North America, 2009.

[10]   Fruchterman, T. M. J., & Reingold, E. M. Graph Drawing by Force-Directed Placement Software: Practice and Experience, 21(11) , 2011.

[11]   Mathieu J., Tommaso V., Sebastien H., Mathieu B. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software, 2014.