ISSN: 0975-766X
CODEN: IJPTFI
Available Online through ***Research Article***
www.ijptonline.com

A STUDY ON SNA: MEASURE AVERAGE LENGTH OF KNOWLEDGE DIFFUSION IN GEPHI

Ayyappan.G[1], Dr.C.Nalini[2]
Research Scholar, Department of CSE, Bharath University, Chennai[1]
Professor, Department of Computer Science and Engineering, Bharath University, Chennai[2]
Email: ayyappan.it@bharathuniv.ac.in

Received on: 15.10.2016 *Accepted on: 22.11.2016*

Abstract

Social Media is a term that covers the platforms Facebook, twitter, blogs and other things typically thought of as social networking. ONS's are primarily Internet based tools for sharing and discussing information among human beings. In this paper focuses mainly to study social network parameters commonly used to explain social structures.

In this paper, we extract academic researcher data from an academic social network, analyze and evaluate network parameters on some widely recognized graph topology using GEPHI social network tool. Using these parameters we find out the strength of research spectrum, authors contribution in academic knowledge diffusion.

In this research work we have collected this data set from ArnetMiner (http://www.arnetminer.org). In this massive real time dataset we have taken co-author training set randomly it contains 100 records from the 4258615 records.

This research community we identified collaborations of authors and co authors published in research article using gephi tool we find out the maximum and minimum research community groups in the academic knowledge diffusion. We measure betweenness centrality distribution, Eccentricity distribution, Harmonic closeness centrality distribution of research community.

Keywords: Graphs, edges, lines, arcs, betweenness centrality distribution, Eccentricity distribution, Harmonic closeness centrality distribution.

I. Introduction

A graph is *vertices*, *nodes*, or *points* which are connected by *edges*, *arcs*, or *lines*. Simple graph - does not have loops (self-edges) and does not have multiple identical edges. Simply we define graphs are connections of vertices and edges.

We don't consider graph shape, we consider only nodes are connected or not. An edges connected two vertices.

Two types of edges:

1. Symmetrical,
2. Arcs (Directed).

Two types of graph frameworks:

1. Graph,
2. Digraph.

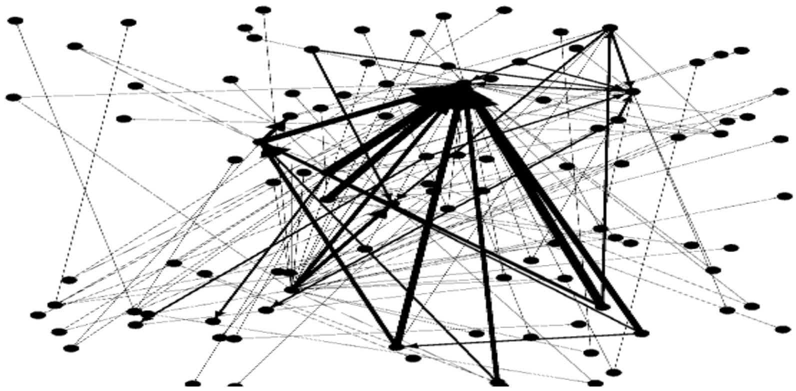


Fig1: Cluster of Research community without Label.

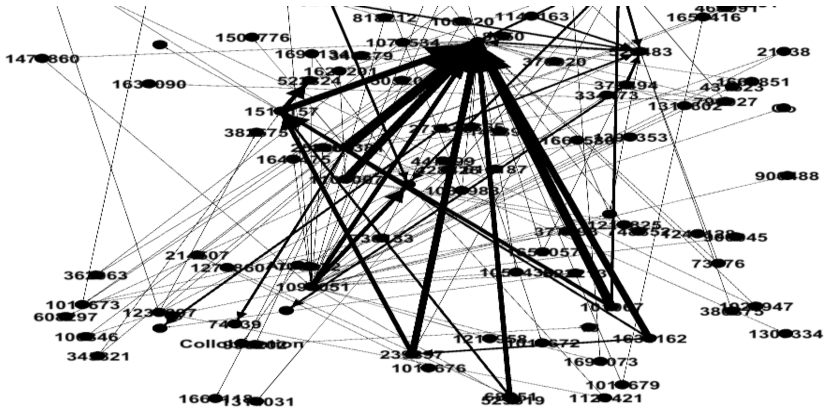


Fig2: Cluster of Research community With Label Representation.

II. Related Work

Dynamic network analysis can provide an aid to longitudinal SNA research. However, as a relatively young field, many aspects have not been explored and there are few standards that have been established.

One path taken was to treat network edges as probabilistic, and use multi agent systems to study network evolution (Carley 2003). Carley redefined the traditional sociogram by adding probabilistic parameters on the edges, providing a quantification of the likelihood they will form. Individual nodes were also given more emphasis; they are treated as agents, and can potentially impact how a network will develop. Terroist networks are a good example where prediction is vital (Krebs 2002), and one in which link prediction has important applications(Carley, Dombroski, Tsvetovat, Reminga and Kamneve 2003; Liben-Nowell and Kleinberg 2006). Another visualization method is to explicitly use the time and order of social interactions to build the network (Berger-Wolf and Saia 2006). Researchers have also applied DNA methods to study other forms of longitudinal networks. Kossinetts and Watts(2006)examined the stability of bridges, defined as connections outside one's circle of acquaintances, and measured how social ties were created and dissolved over time. Barabasi and colleagues (2002) applied DNA methods to find unique properties in an evolving citation network that differed from classic models. Indeed, various forms of citation and co-authorship networks offer a wealth of reliable data with which to study the evolution of networks. New layouts and metric computations are constantly being developed from these dynamic network studies (Brandes and Pich 2012).

The study of dynamic networks greatly benefits from visualizations that can illustrate ideas and concepts not immediately visible in a static sociogram. In fact, "The ability to see data clearly creates a capacity for building intuition that is unsurpassed by summary statistics" (Moody, McFarland and Bender-deMoll 2005). Moody and others research emphasizes how the ability to see data can be superior to summary statistics, and illustrates the need to visualize how networks develop and change over time. Additionally, they lay the foundation of how dynamic network visualizations should be presented (e.g. differentiating between discrete and continuous time), and recommend visualization and analysis be interactive. These theoretical ideas were developed in parallel with SoNIA, a software package forvisualizing dynamic network data (Bender-deMoll and McFarland 2006).Other researchers have continued to focus on studying different properties of dynamic networks, e.g. the evolution of subgroups(Falkowski, Bartelheimer and Spiliopoulou 2006), effects of network topology and organizational structure over time (Kossinets and Watts 2006), detecting and predicting statistically significant changes in a network over time (McCulloh and Carley 2011), and new visualization methods using shortest-path computations (Brandes and Pich 2012).

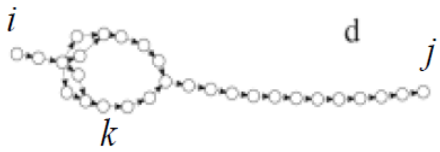
III. Materials and Methods

Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchial graphs. Gephi is a tool for people that have to explore and understand graphs. Like Photoshop but for data, the user interacts with the representation; manipulate the structures, shapes and colors to reveal hidden properties. The goal is to help data analysts to make hypothesis, intuitively discover patterns, isolate structure singularities or faults during data sourcing. It is a complementary tool to traditional statistics. This is a software for Exploratory Data Analysis. Gephi tool provides an fastest graph visualization engine to speed-up understanding and pattern discovery in large graphs. Gephi is powered by ad-hoc OpenGL engine, it is pushing the envelope on how interactive and efficient network exploration can be.

- Networks up to 50,000 nodes and 1,000,000 edges
- Iterate through visualization using dynamic filtering
- Rich tools for meaningful graph manipulation Gephi is a modular software and can be extended with plugins. Plugins can add new features like layout, filters, metrics, data sources, etc. or modify existing features. Gephi is written in Java so anything that can be used in Java can be packaged as a Gephi plug-in.

In this research work we find out average length of research community like as betweenness centrality distribution, Eccentricity distribution, Harmonic closeness centrality distribution.

a) Betweenness centrality distribution



• Find all the shortest paths between nodes i and j - C(i,j)

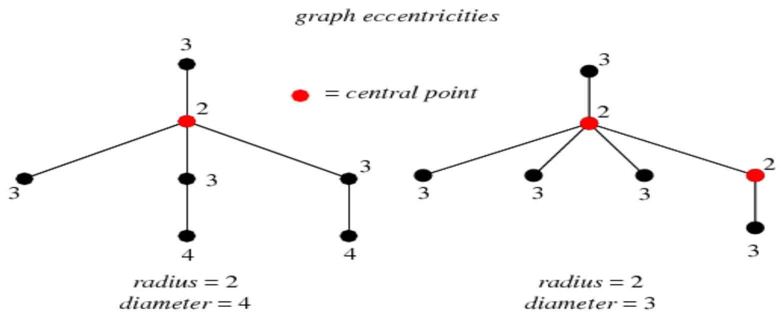
• Determine how many of these pass through node k - Ck(i,j)

• The betweenness centrality of node k is

$$g_k = \sum_{i \neq j} \frac{C_k(i,j)}{C(i,j)}$$

b) Eccentricity Distribution

The eccentricity $\epsilon(v)$of a graph vertex $v$in a connected graph $G$is the maximum graph distance between $v$and any other vertex $u$of G. For a disconnected graph, all vertices are defined to have infinite eccentricity.



c) Harmonic closeness centrality distribution

The maximum value of the harmonic centralization is defined through the study of the index on a star graph.

$$C_H = \frac{2(n-1)\sum_i \left(c_H^* - c_H(x_i)\right)}{n}$$

with $c_H^* = \max_j c_H(x_j)$ and $n = |V|$.

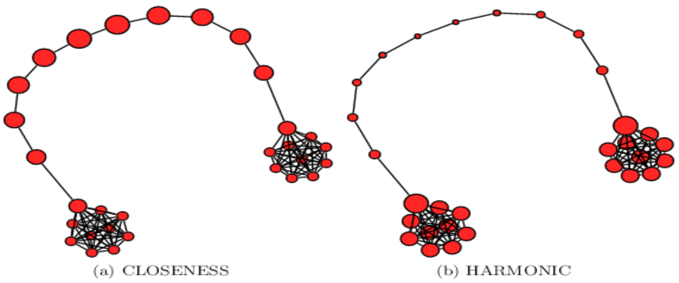


Fig:1 Closeness and Harmonic.

IV. Implementation and Result

In this Harmonic closeness centrality it has performed in this directed graph and the diameter is. Radius is zero. The Average Path length is 1.8819875776397517.

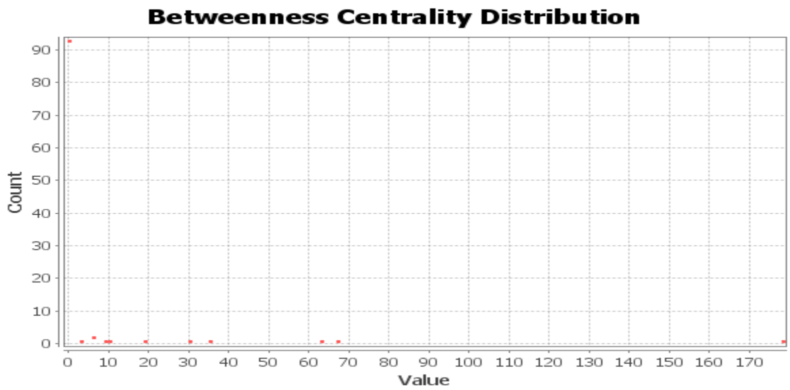


Fig:2. Betweenness Centrality Distribution.

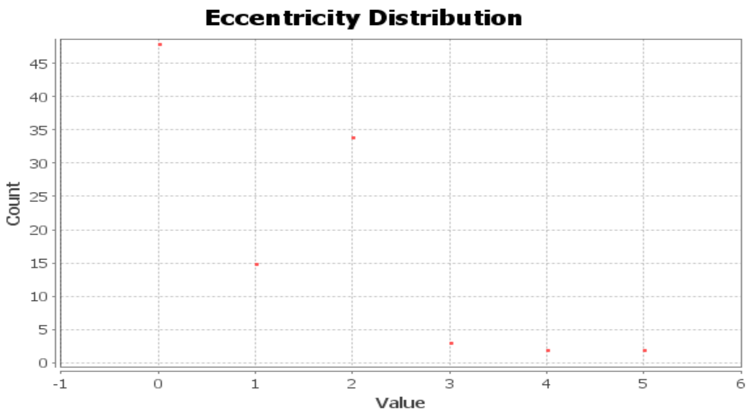


Fig:3. Eccentricity Distribution.

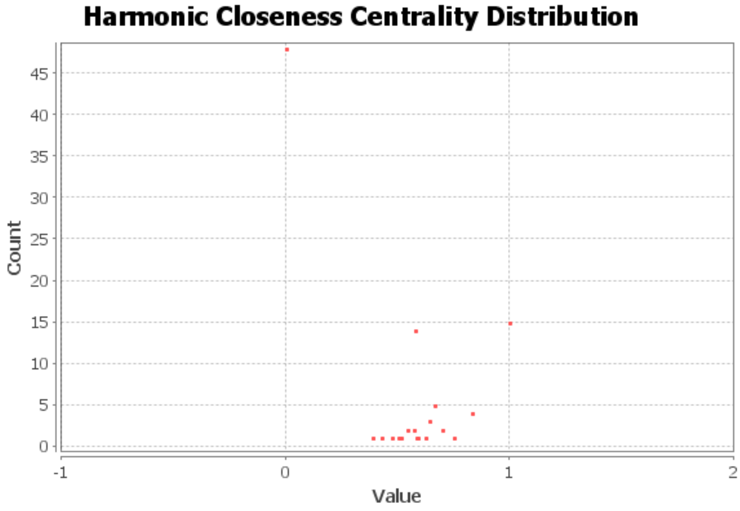


Fig:4 Harmonic Closeness Centrality Distribution.

V. Conclusion

This result an alternative method of computing centrality has been defined and studied. The High correlation of harmonic centrality with closeness centrality makes it a good alternative, in particular because it can be computed and interpreted on unconnected graphs in this research spectrum. The Limitations and unexpected behavior of the index find their origin from what we have seen only from degenerate and highly improbable cases and then comparisons should be done on other types of graphs, especially those reproducing properties of social networks. The definition of the harmonic centrality index uses inversions, but not in a similar way as the closeness centrality index.

VI. References

1. http://matthieu-totet.fr/Koumin/2013/12/16/understand-degree-weighted-degree-betweeness-centrality/
2. http://users.phys.psu.edu/~ralbert/phys597_09/c02_graph_conc.pdf
3. Jie Tang, Jing Zhang, Ruoming Jin, Zi Yang, Keke Cai, Li Zhang, and Zhong Su. Topic Level Expertise Search over Heterogeneous Networks. Machine Learning Journal, Volume 82, Issue 2 (2011), Pages 211-237.
4. Kar-Hai Chu,Heather Wipfli,Thomas W. Valente,Using Visualizations to Explore Network Dynamics , Journal of Social Structure,Volume 14,
5. https://sites.google.com/site/ucinetsoftware/home
6. http://www.r-project.org/
7. Wipfli, H., K. Fujimoto and T. W. Valente (2010). "Global Tobacco Control Diffusion: The Case of the Framework Convention on Tobacco Control." American Journal of Public Health100(7): 1260.

8. Suh, B., L. Hong, P. Pirolli and E. Chi (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. 2010 IEEE Second International Conference on Social Computing (SocialCom), IEEE.
9. Correa, C. D. and K.L. Ma (2011). Visualizing Social Networks. Social Network Data Analytics, Springer: 307-326.
10. Kairam, S., D. J. Wang and J. Leskovec (2012). The Life and Death of Online Groups: Predicting Group Growth and Longevity. In Proceedings of the fifth ACM International Conference on Web Search and Data Mining, ACM.
11. McCulloh, I. and K. Carley (2011). "Detecting Change in Longitudinal Social Networks." Journal of Social Structure12(3).
12. Suh, B., L. Hong, P. Pirolli and E. Chi (2010).Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. 2010 IEEE Second International Conference on Social Computing (SocialCom), IEEE.
13. http://www.academia.edu/16776131/GEPHI_Introduction_to_Network_Analysis_and_Visualization
14. http://www.slideshare.net/digitalmethods/gephi-rieder-23834788
15. http://www.ijera.com/papers/Vol2_issue6/DB26703707.pdf
16. Ulrik Brandes, *A Faster Algorithm for Betweenness Centrality*, in Journal of Mathematical Sociology 25(2):163-177, (2001)
17. Harary, F. *Graph Theory*. Reading, MA: Addison-Wesley, p. 35, 1994.
18. Skiena, S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, p. 107, 1990.
19. West, D. B. *Introduction to Graph Theory, 2nd ed.* Englewood Cliffs, NJ: Prentice-Hall, 2000.
20. https://infoscience.epfl.ch/record/200525/files/[EN]ASNA09.pdf