A STUDY ON SNA: MEASURE THE MODULARITY OF KNOWLEDGE DIFFUSION IN GEPHI

1Ayyappan.G, 2Dr.C.Nalini, 3Dr.A.Kumaravel
Research Scholar, Department of CSE, Bharath University, Chennai-600073
Professor, Department of Computer Science and Engineering, Bharath University, Chennai-600073
Professor, Department of Information Technology, Bharath University, Chennai-600073
1Email.ID:ayyappan.it@bharathuniv.ac.in,
2Email.ID:drnalinicidambaram@gmail.com
3Email.ID:drkumaravel@gmail.com

Abstract
Social media is a term that covers the platforms of Face book, 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 Arnet Miner (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. We measure modularity of research community. 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.

Keywords: graphs, edges, lines, arcs, modularity.

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 edge connected two vertices[1-6].
Two types of edges:
1. Symmetrical
2. Arcs (directed)
Two types of graph frameworks:
1. Graph
2. Digraph.
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[7-10].

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 Socio Gram 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.

Terrorist 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 [11-16].

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 for visualizing 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).

**Materials and methods**

Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical 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, and isolate structure singularities or faults during data sourcing. It is a complementary tool to traditional statistics.

This is 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 modular software and can be extended with plug-in. Plug-in can add new features like layout, filters, metrics, data sources, etc. Gephi is written in Java so anything that can be used in Java can be packaged as a Gephi plug-in.

**Modularity:**

To measure the strength of division of a network into modules (also called groups, clusters or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Modularity algorithm implemented in Gephi looks for the nodes that are more densely connected together than to the rest of the network (it's well explained in the paper they published on the website by the guy who created the algorithm - Google scholar it - Blondel, V. D., Guillaume, J., & Lefebvre, E. (n.d.).

When you implement this measure the colors indicate different communities determined by this algorithm and basically it'll show, in your case, which routers are more densely connected between each other than to the rest of the network.

**Implementation and result**

**Parameters:**
- Randomize: On
- Use edge weights: On
- Resolution: 1.0

**Results:**
- Modularity: 0.436
- Modularity with resolution: 0.436
- Number of Communities: 14

---

Authors: Ayyappan.G, Dr.C.Nalini, Dr.A.Kumaravel

Period from (Jan - Jun: 2017)
community detection is modularity maximization. Modularity is a benefit function that measures the quality of a particular division of a network into communities. The modularity maximization method detects communities by searching over possible divisions of a network for one or more that have particularly high modularity.

References
[5] https://sites.google.com/site/ucinetsoftware/home