ISSN No. 0976-5697

Volume 2, No. 5, Sept-Oct 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Survey on Applications nf Social Network Analysis

Dr.S.S.Dhenakaran
Department of Computer Science and Engineering
Alagappa University
Karaikudi. India.
ssdarvind@yahoo.com

Mrs. L.Rajeswari*
Computer Centre
Alagappa University
Karaikudi. India.
lr_eswari@yahoo.co.in

Abstract: Social network is a social structure made up of individuals (or organizations) called "nodes", which are tied (connected) by one or more specific types of interdependency. Social network analysis is the study of complex human systems through the mapping and characterizing relationships between people, groups or organizations. Social network analysis views social relationships in terms of network theory consisting of nodes and connections/ties. This paper discusses some of the applications of Social Network Analysis in the field of Communication Networks

Keyword: Social Network Analysis, Collaboration graph, Affiliation Network, Virtual Private Network, RePEc Author Service (RAS) Network.

I. INTRODUCTION

Recent research directions in computer science is closely related to two areas – machine learning and data mining. These areas are having developed methods for constructing statistical models of network data. Examples of such data include social networs, networks of web pages, complex relational databases, and data on interrelation people, places, things and events extracted from text documents. In its simplest form, a social network is a map of specified connections/ties, such as friendship, between the nodes being studied. The nodes to which an individual is thus connected are the social contacts of that individual. The network can also be used to measure social capital – the value that an individual gets from the social network.

Social Network Analysis (SNA) has emerged as key technique in modern sociology. It also gained significance in anthropology, biology, communication studies, economics, geography, information science, organizational studies, and social psychology and has become a popular topic of speculation and study. SNA is the study of complex human systems through mapping and characterizing relationships between people, groups, or organizations.

In section II, the Degree Centrality of Affiliation Network is presented. In section III, Delay, Throughput and Normalized Traffic Received of Virtual Private Network is discussed. In section IV, the Rank Correlation of centrality metrices of RAS Network is proved.

A. Structure of Social Network

The shape of a social network helps to determine a network's usefulness to its individuals. Smaller, tighter networks can be less useful to their members than networks with lots of loose connections [1] (weak ties) to individuals outside the main network.

More open networks, with many weak connections and social connections, are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant connections. In other words, a group of friends who only do things with each other already share the same knowledge and opportunities. A group of individuals with connections to other social worlds is likely to have access to a wider range of information. It is better for individual success to have connections to a variety of networks rather than many connections within a single network. Similarly, individuals can exercise influence or act as brokers within their social networks by bridging two networks that are not directly linked (called filling structural holes).

B. Collaboration Graph

A **collaboration graph** [2] is a graph modeling of some social network. Collaboration Graph can be used to illustrate good and bad relationships between humans, where the vertices represent participants of that network (usually individual people) and two distinct participants are joined by an edge whenever there is a collaborative relationship between them of a particular kind. A positive edge between two nodes denotes a positive relationship (friendship, alliance, dating) and a negative edge between two nodes denotes a negative relationship (hatred, anger). Signed social network graphs can be used to predict the future evolution of the graph. In graph theory, closed (simple) path, with no other repeated vertices or edges other than the starting and ending vertices is called cylcle.

Signed social network graphs can be used to predict future evolustion of the graph. In signed social networks, there is the concept of "balanced" and "unbalanced" cycles. A balanced cycle is defined as a cycle where the product of all the signs are positive. Balanced graphs represent a group of people who are unlikely to change their opinions of the other people in the group. Unbalanced graphs represent a group of people who are very likely to change their opinions of the other people in the group. For example, a group of 3 people (A, B, and C)

A and B have a positive relationship, B and C have a positive relationship, but C and A have a negative relationship is an unbalanced cycle. Figure 1 illustrate an unbalanced cycle.

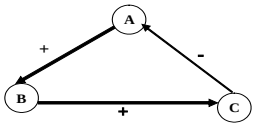

Figure 1. Unbalanced Cycle

B has a good relationship with A, and both A and B have a negative relationship with C. This group is a balanced cycle Figure 2 illustrate a balanced cycle.

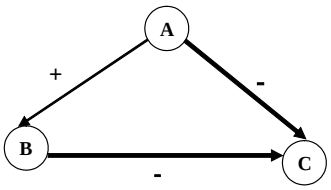

Figure 2. Balanced Cycle

By using the concept of balances and unbalanced cycles, the evolution of Signed social network graphs can be predicted.

C. Benefits of SNA

The Social Network Analysis might be used to:

- Identify the individuals, teams, and units who play central roles.
- Make out opportunities to accelerate knowledge flows across functional and organizational boundaries.
- Strengthen the efficiency and effectiveness of existing, formal communication channels.
- Raise awareness of and reflection on the importance of informal networks and ways to enhance their organizational performance.
- Improve innovation and learning.
- Refine strategies.

Development work, for one, is more often than not about social relationships. Hence, the social network representation of a development assistance project or program would enable attention to be quickly focused (to whatever level of complexity is required) on who is influencing whom (both directly and indirectly).

Understanding the structure and dynamics of social groups is a natural goal for network analysis, since such groups tend to be embedded within larger social network structures. That is, given a collection of individuals linked in an underlying social network, the groups and communities that they identify with can be thought of as corresponding to subgraphs of this network, growing and overlapping one another in a potentially complex fashion. A group that grows mainly through the agressive recruitment of friends by other friends would appear as a subgraph branching out rapidly over time along links in the network; a group in which the decision to join depends relatively little on the influence of friends might appear in stead as a collection of small disconnected components that grows in a "speckled" fashion that is such social networks are not themselves directly observable, on-line systems can provide rich data on large networks of interactions that are highly reflective of these underlying social networks.

II. AFFILIATION NETWORK

A social network analysis for the design and management of information network is described. One of the features is that the model of the social networks is the affiliation network, considering the group concept in the organization activity.

Degree is the number of ties to other person in the network. Degree centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has). Degree is often interpreted in terms of the immediate risk of node for catching whatever is flowing through the network (such as a virus, or some information). The degree of each node is plotted on the concentric circle in which the radius is the degree. The center of the circle indicates the highest degree in an organization.

A. Degree Centrality of Affiliation Network [3]

There are Clique analysis and Clan analysis which are useful for investigating the clustering structure in an organization[4]. A clique is defined as the maximal complete subgraph of three or more nodes, all of which are connected to each other. There are no other nodes in the network that are also connected to all of the members of the clique. When the largest geographic distance between any two nodes is no greater than N, the maximum complete sugbraph is called an N-Clique. An N-Clan is an N-Clique in which the geodesic distance between all nodes in the subgraph is no greater than N for links within the subgraph.

By using the group composition which was derived from the Clique & Clan analysis, a bipartite graph (as in figure 3) was formed to represent the affiliation network for an organization. The affiliation network consists of 113 nodes which mean 79 individual and 34 groups. The organizational activity, considering both individuals and groups simultaneously was then investigated using this bipartite graph. The Figure 3 below shows the degree centrality of the affiliation network.

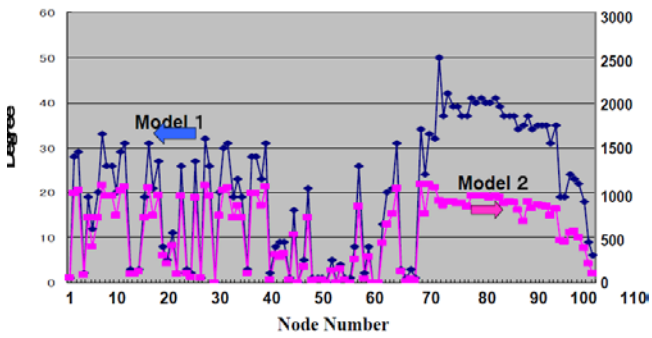

Figure 3. Degree Centrality of Affiliation Network

The centrality model derived from the entire picture of network including the individuals and group Model 1, and the centrality model in which groups and individuals are evaluated separately Model2. Two models are shown in this figure 3. Although the analytical result of Model 1 is almost equivalent to that of Model 2, there are some differing features. The degree centrality of Node 80 had the highest score (50) in the case of Model 1, whereas the same node had a value less than that of other nodes, such as Node 89 in the case of Model 2. The ranking of the degree sometimes changes among the members, such as Nodes 8 and 18. This is explained by the fact that some links evaluated in the original analysis were not included in the Clan distribution, because the bipartite graph was created on the basis of Clan analysis. It can also be said that links not found to be suitable for an organizational activity were considered to be noise. This indicates that it is possible to find the individuals who act throughout the entire network by analyzing the affiliation network. Those individuals who belong to a number of organizations can act as a bond between the different organizations, thereby enhancing their centrality in their organizations. By belonging to a higher-class organization, an individual can gain higher centrality.

III. VIRTUAL PRIVATE NETWORKS (VPN)

VPN[5] is a secure means of communicating over a public internet broadly employed by many companies for the purpose of confidentiality. VPNs are executed with a broad range of technologies, as such can be controlled by the service provider. This make the consumers to understand the cost advantage of a shared network while relishing its advantages of security, quality of service, reliability, efficiency and manageability. A VPN uses factual connections routed through the internet from the company's private network to the remote site. It is important to know that it's possible to transport VPN traffic over non-private networking substructure but has to be over standard protocols or over a service provider's private network with a predefined service level agreement. VPNs using the internet have the potential to solve many of today's business networking problems e.g. businesses today are finding that past solutions to wide area networking between the main corporate network and branch offices such as dedicated leased lines or frame relay circuits do not provide the flexibility required to quickly create new partner links or support project teams in the field. VPNs therefore allow many network managers to connect remote branch offices and project teams to the main corporate network economically and provide remote access to employees while also reducing the internal requirements for equipment and support. VPNs also offer direct cost savings over other communications methods such as leased lines and long distance calls. Other advantages of the VPN include: indirect cost savings as a result of reduced training requirements and equipment, increased flexibility and scalability.

A. Delay

Internet Protocols happened to give a larger delay when compared with ATM due to the absence of private virtual circuits (PVCs) [6]. PVCs will require that every packet be routed in each node of any IP network. This process will lead to increase in the propagation time that packet will be routed according to PVC definitions and not according to the network topology. This incident eventually cause the quality of real time traffic, especially voice data to be impaired with a high value of delay as depicted in Figure 4 below.

B. Throughput

Data throughput mostly depends on link speed and characteristic of the technology being used to transmit the data.

ATM and MPLS happened to give more data throughput compared to IP. This is due to congestion leading to heavy packet drop and also, to the connectionless feature that IP possess. However, there is a virtual path specified for each packet in ATM and MPLS cores networks that is responsible for the improvement of the overall network throughput performance as shown in Figure 5.

C. Normalized Traffic Received

The connectionless feature of IP turned to be a great deal for a normalized traffic received because there will be more packet drop at the time of congestion. However, IP device tries to avoid congestion by running congestion avoidance mechanism which most of the time works efficiently. In the case of heavy load there will be a progressive reduction in the received traffic due to capacity limit or poor performance at the physical layer as shown in Figure 6.

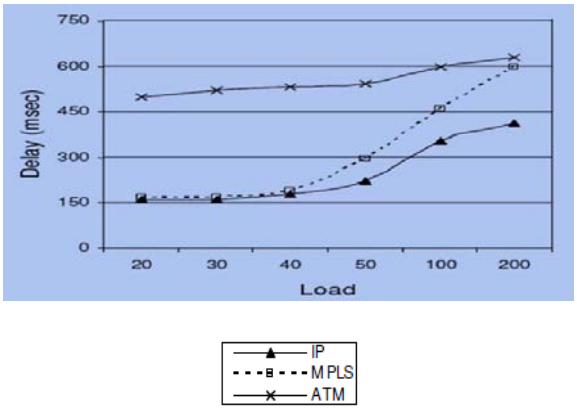

Figure 4. Average End-to-End Delay of the Network

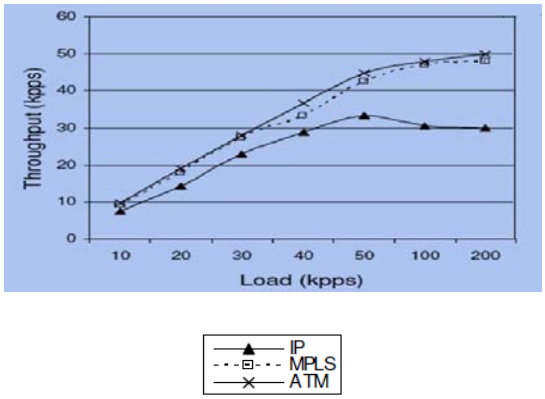

Figure 5. Throughput of the Network

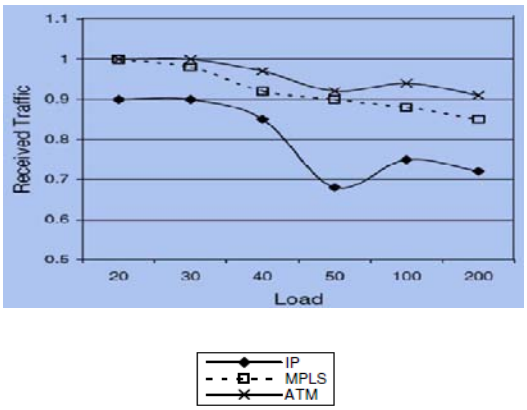

Figure 6. Normalized Received Traffic Comparison of IP, MPLS and ATM technologies

From the evaluations above, when IPv4 was considered along side with ATM and MPLS, there is a prior path establishment in ATM and MPLS which has a lot of preferences and this has made the two technologies to outstand typical IP technology. However, one of the reasons why ATM is rarely used for Internet cores is because of the difficulty to replace the existing IP infrastructure which is widely spread all over the Internet. MPLS allows service providers to converge their network into a single infrastructure while offering the service they currently support. MPLS enable new services and simplify services provisioning. It natively supports the rapid growth in IP applications and services allow the integration of the emulated service management into common network strategy and positioned to support integration of packet technologies and optical core.

IV. REPEC AUTHOR SERVICE (RAS) NETWORK

RePEc (Research Papers in Economics) offers the RePEc Author Service (RAS). They investigate the structure of research collaborations in economics, by applying social network analysis to the co-authorship network formed by RePEc Author Service (RAS).

RAS [7] network can be presented as a graph, which consists of points (or nodes) to represent actors and lines (or edges) to represent ties or relations. A component is a connected subset of a graph in which there are paths between the nodes. If two authors have written a joint paper, there is a path between them. If a third author has co-authored with any one of the first two authors, a chain of co-authorship path (henceforth "path") can be built connecting the first author with the third author and so on. If a path between two authors can be established, the two authors are said to belong to the same component of the network. One way to measure the centrality of an author is to look at how many immediate ties an author has. This is called the degree centrality of an author.

A. Rank correlation of centrality metrics

Spearman's correlation coefficient is used to measure the strength of the association between centrality metrics obtained by using different edge weighting schemes. Table 1 furnishes a list of the Spearman's correlations for all possible pairs.

First, it seems the association between any pair of betweenness scores (e.g. BIB vs. SYB) is stronger than the association between any pair of betweenness and closeness scores (e.g. BIB vs. SYC). This observation is similar for the closeness scores with the exception of random walk closeness [8] scores, which appears to be poorly correlated to all other scores.

Table 1: Rank correlation of centrality metrics

	NDO	DGE	BIB	BIC	SYB	SYC	RWB	RWC
NDO	1	0.7	0.68	0.55	0.71	0.6	0.7	0.19
DGE	0.7	1	0.84	0.67	0.85	0.57	0.87	0.3
BIB	0.68	0.84	1	0.6	0.9	0.52	0.89	0.3
BIC	0.55	0.67	0.6	1	0.54	0.81	0.61	0.57
SYB	0.71	0.85	0.9	0.54	1	0.54	0.91	0.23
SYC	0.6	0.57	0.52	0.81	0.54	1	0.56	0.42
RWB	0.7	0.87	0.89	0.61	0.91	0.56	1	0.41
RWC	0.19	0.3	0.3	0.57	0.23	0.42	0.41	1

Next look at the association between the numbers of papers an author has published (NDO) and the various centrality metrics. The moderately high correlation coefficient (i.e. 70%) between the NDO and the degree centrality rank (DGE) indicates productive authors have more collaborators than less productive authors. The association between the NDO rank and each of the betweenness centrality rank are moderately high whereas the association between the NDO rank and each of the closeness centrality ranks are surprisingly low. The lowest correlation coefficient (i.e. 19%) among all pairs is between NDO and random walk closeness rank (RWC). The authors with the highest numbers of betweenness centrality scores, labeled as RWB based on a random walk network model. In a list of top twenty authors with the highest numbers of betweenness centrality ranking, labeled as BIB, and closeness centrality ranking, labeled as BIC.

In general, betweenness scores are strongly correlated with one another, whereas closeness scores appear to be moderately correlated with one another.

V. CONCLUSION

This paper discussed some of the applications of Social Network Analysis in the field of Communication Networks such as Affiliation Network, Virtual Private Network and RAS Network. By analyzing the Affiliation Network, it is possible to find who has more affiliation towards the entire network. From the VPN applications, MPLS allows service provider to converge their network into a single infrastructure. From RAS applications, it is seen that Co-authorship centrality rankings seem to be a promising way to generate incentives for registered authors to promote RAS to their unregistered co-authors. It is concluded that Social Network Analysis is very efficient and it is very useful to improve the Quality of Service in Communication Networks.

VI.REFERENCES

[1] Granovetter, M.D. (2004). " The Impact of Social Structures on Economic Development Journal of Economic Perspectives (Vol 19 Number 1, pp. 33-50).

[2] Frank Harary. Topics in Graph Theory. New York Academy of Sciences, 1979. ISBN 0897660285.

[3] Noriaki. Yoshikai† Park Kyoung Hee† and Jun Kanemitsu‡ "Application of Social Network Analysis to Information Network Design" Nihon University , Advance Research Institute for the Science and Humanities, APNOMS2007

[4] Kanemitsu. "Explorations in Social Networks", Keiso Shobo (2003)

[5] Thomas Krichel , Long Island University College of Information and Computer cience ; Nisa Bakkalabasi , Yale University Kline Science Library " A Social network analysis of research collaboration in the economics community "

[6] Kang, Y., & Lee, J, *(*2005), The implementation of the premium services for MPLS IP VPNs. Proceedings of the 7th International Conference on Advanced Communication Technology, Vol 2, pp 1107 – 1110.

[7] A.L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, Evolution of the social network of scientific collaborations. arXiv:cond-mat/0104162v1, 2005.

[8] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel, Co-authorship networks in the digital library research community, Information Processing & Management, 41 (6): 1462-1480, 2005.