Overview of Social Network Analysis and Tools

[1]Jyoti Shokeen, [2]Partibha Yadav
[1,2]Ph.D Scholar
[1]Department of Computer Science & Applications, M.D. University, Rohtak, India
[2]I.G.U., Meerpur, Rewari, Haryana, India

Abstract- A social network is a set of people or other social entities which are connected by set of social relationships such as friendship, interests, behavior or information exchange. Social network analysis focuses on the analysis of the pattern of relationships among people, organizations, states and such social entities. This paper presents the overview of social network analysis starting from the introduction and the methods to represent social network by different means like matrix, formal methods and graphs and then followed by social network metrics. To analyze and mine the information from such network, we have various tools used for the analysis and visualization of social network.

Keywords: social network analysis, social network models, centrality, social network tools.

I. INTRODUCTION

A social network is a social structure made up of a set of actors called nodes (such as individuals or organizations or things within the network) and the ties or edges (relationships and interactions) that connect them. Social network analysis is the mapping and measuring of relationships and flows between people, groups and organizations. The nodes in the network are the people and groups, while the links show relationships or flows between the nodes. [1, 2] Social network analysis is the process of investigating social structures through the use of network and graph theories. Different tools are required to analyze the information, extract the meaningful patterns and visualize the network data.

In this paper, Section II provides the various ways of representing social networks. Section III deals with the metrics used in the analysis of social networks. Section IV deals with the tools used in social network analysis.

II. SOCIAL NETWORK MODELS

1. Formal Methods to represent Social Relations

A social network consists of few or large number of nodes or actors and different kinds of relationships exist among those actors. Analyzing the patterns of these relationships can be too complicated due to the great amount of information. Formal methods are used for the analysis of social networks systematically and compactly [1, 3]. The reason behind using formal methods for representing social networks is that mathematical representations make it easy to allow computers for analyzing the network data accurately [1]. Matrices are useful for the manipulation of network data and graphs are significant for visualizing patterns.

2. Matrix to represent Social Relations

Matrices, a rectangular arrangement of elements, are used for the analysis of social relations. Square matrices are the frequently used matrices in social networks in which there are as many rows and columns as there are actors in the data set. A binary matrix contains 0's and 1's indicating presence and absence of relationship between actors respectively [1, 4]. This type of matrix is the starting of network analysis and is known as adjacency matrix, denoted by g. g_{ij} indicates relationship status for nodes i and j, where one is entered if a tie is present, i.e. if there is a social relation between them, otherwise zero is entered indicating no tie. Each node has a degree that shows the number of edges it is sharing with other nodes.

3. Graphs to represent Social Relations

Social network analysis uses graphs which consist of nodes to represent actors and lines or edges to represent relations or ties. The use of graphs in social networks is termed as "sociograms". A graph is a set of nodes or vertices *V* and a set of edges or lines. If an edge exists *{a,b}* then we can say that nodes *a* and *b* are related to each other. The edges themselves can be unordered pairs of nodes or in a directed graph (digraph), ordered pairs of nodes where each edge has a direction. In this case *{a,b}* is an arc from a *to* b [5]. A graph can represent a single type of relationship such as a friendship graph or may consist of different kinds of relationships [1] such as friendship and spouse graph. There are different levels of measurement in graphs such as binary, signed and valued graphs. Presence of an edge indicates the relationship between actors and the absence of edge shows no relationship in binary graphs. Graphs with signed data uses + sign on edge to denote a positive choice, zero for don't care and – sign to indicate a negative choice. Further, weights can be assigned to the edges in graphs to indicate the strength of the relationship.

III. METRICS IN SOCIAL NETWORK ANALYSIS

Centrality:

Centrality measure determines the importance of a node in the network based on how well it connects the network. The centrality measure identifies most prominent nodes or actors in the network. Betweenness centrality, degree centrality and closeness centrality are some of the important centrality measures. Other centrality measures include power centrality, stress centrality,

Betweenness Centrality:

This type of centrality measures the number of times a node occur on the geodesic path (shortest path) [8]. The node connects a pair of nodes which are not able to reach each other. The betweeness measure takes into account the location of a node in network and how well it acts as a "hub". The node which is more connected or centralized is the more powerful, efficient and popular [9]. Betweenness centrality of a node can be computed as:

$$C_B = \sum_{i=1, i\neq x}^{N} \sum_{j=1, j<i, j\neq x}^{N} \frac{Gij(x)}{Gij}$$

Where G_{ij} is the number of shortest paths from node i to j and $G_{ij}(x)$ is the number of these paths which passes through the node *x*.

Degree Centrality:

This measure counts the number of direct connections a node has in the network. A node having high degree has high influence in the network and considered as the popular node. The incoming and outgoing ties of a node increase the degree centrality [9]. Degree centrality of a node can be computed as:

$$C_D = \frac{\deg(x)}{N-1}$$

Where N is the total number of nodes in the network and *deg(x)* is the number of links it has.

Closeness Centrality:

This centrality measure is based on the idea of distance that nodes having shortest distance (geodesic distance) with other nodes have higher closeness and spread information more productively [9]. It measures the node's independence or efficiency [6]. Closeness centrality of a node can be computed as:

$$C_C = \frac{N-1}{\sum_{y \neq x} d(x,y)}$$

Where d(x,y) is the shortest-path distance between nodes x and y.

Eigenvector centrality:

This type of centrality measure determines the influence of a node in the network. Relative scores are assigned to all the nodes in network based on the notion that ties with high-scoring nodes contribute more to the score of the node in comparison to the equal ties to low-scoring nodes. One of the variant of the eigenvector centrality measure is PageRank [7].

Clique:

A clique is the maximum number of nodes or actors consisting of all possible ties among them. These groups of nodes share views, interests, and patterns of behavior or purposes. Broadly speaking, actors are more intensely and closely related to one another as compared to other actors in the network [1]. Dyad is the smallest clique that can be present in the network which can be extended to become more inclusive.

Density:

The density of a network is the number of edges existing over the total possible number of edges. A high dense network is more connected compared to a less dense network. This measure is used to show if the underlying network or graph is dense or sparse.

HITS algorithm:

Hyperlink-Induced Topic Search (HITS) proposed by Kleinberg [10] is a link analysis algorithm used to rate Web pages. The algorithm calculates two scores: hub score and authority score. These score are assigned to each page where authority value estimates the value of the content of a page and hub value estimates the value of its links to other pages [10]. A node is considered a hub if it points to many other nodes. A node is considered an authority if it has many pages linked with it. A node having more outgoing links has high hub score and a node having high incoming links has high authority score. Every node is considered as hub and authority scores are fixed to a constant at the beginning. Then the scores are updated and they converge after little iteration [8]. If u is one of the n vertex connected to v, the scores auth(v) and hub(v) for the new iteration as computed as follows:

$$auth(v) = \sum_{i=1}^{n} hub(u)$$
$$hub(v) = \sum_{i=1}^{n} auth(u)$$

IV. SOCIAL NETWORK ANALYSIS TOOLS

Social network analysis tools are used to identify, analyze, visualize or simulate nodes (organizations, or knowledge) and edges (relationship or interaction) from various types of input data including mathematical models of social networks. There are several tools available for analysis of social networks. The International Network for Social Network Analysis (INSNA) [12] maintains a large list of software packages and libraries. Network analysis tools are either GUI based or packages/libraries which can be used in a programming language. The GUI packages are easier to learn, while scripting tools are more powerful and extensible. Pajek and Gephi are some widely used and well-documented GUI-based network tools [13]. IGraph and Networkx are package based network tools [11]. Visualization tools facilitate qualitative interpretation of network data. We have discussed five tools in brief which are as follows:

Pajek [14] is widely used Software for drawing networks. It is a program for Windows for analysis of large networks. It is available free for noncommercial use. The main motive behind the development of Pajek was the existence of various types of large networks having multiple sources. Pajek also has analytical capabilities, and can be used to compute most centrality measures, identify structural holes, blockmodel, etc.[15] .NET is the commonly used internal format supported by Pajek. Pajek also supports other formats like UCINET DL, GML, GDF, CSV, MOL (MDL MOL file), Spreadsheet, GraphViz DOT, etc.

Gephi [16] is an interactive visualization and exploration software for all kinds of networks and complex systems, dynamic and hierarchical graphs. It is a tool for graph exploration and manipulation. The user interacts with the representation; manipulate the shapes, colors and structures to reveal the hidden properties. Some of its most attractive features are its support of many different native graph formats, real time interactive features, and easy to use interface. Most importantly, it has many supporting features built for dynamic network analysis that incorporate functions such as live filtering, a combination of static and dynamic metrics, a multitude of layouts, and a timeline component that can generate various longitudinal reports [17]. The flexible and multi-task architecture allows working on complex data sets and produce valuable visual results.

Netdraw [18] is a program for visualizing social networks. The tool allows reading multiple relations, valued relations, node attributes, 1-mode and 2- mode network data [8]. The program can read its VNA format, Pajek files, UCINET files and UCINET DL files. Further, the tool provides facilities such as coloring nodes, sizing nodes, changing label text and positioning, printing, appearance options and layout. The software allows visualization of tie strength; filtering of data by the strength and attributes; saving maps in the form of pictures (jpeg, bitmap, metafile). The user can create a text file using any editor. The software has some analytical capabilities that partially overlap with UCINET.

NetworkX [19] is a Python language software package for the creation, manipulation and the study of structure and functions of the complex networks. It is suitable for operation on large real-world graphics. With this tool you can load and store networks in standard and nonstandard data formats, can generate many types of random and classic networks, analyze network structure, build network models, draw networks, and much more. Networkx has many features like language data structures for graphs, dIGraphs, and multIGraphs [20]. Nodes can be "anything" (e.g. text, images), Edges can hold arbitrary data (e.g. weights, time-series), Standard graph algorithms, Network structure and analysis measures etc.

IGraph [21, 22] is a free software package for creating and manipulating graphs. It includes implementations for classic graph theory problems like minimum spanning trees and network flow, and also implements algorithms like community structure search. The efficient implementation of IGraph allows it to handle graphs with millions of nodes and edges. IGraph can be installed as libraries for C, R, Python and Ruby.

Table 1. Classification of Sna Tools

Tool Name	Authors	Language	Purpose
Pajek	Vladimir Batagelj, Andrej Mrvar	.NET	Large network analysis and visualization
Gephi	Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy	JAVA	Network analysis and visualization
NetDraw	Steve Borgatti	DOS	Network visualization
NetworkX	Lanl and others	Python	Exploration and analysis of complex networks and graphs
IGraph	Gabor Csardi and Tamas Nepusz	C	Graph analysis

V. CONCLUSION

Social networks are abundant and useful sources of information. In this paper we have described different models to represent social networks in various forms. We also reviewed a few social network metrics used for the analysis by various tools. We have also described some tools used in exploring, analyzing and visualizing social networks. The tools facilitate qualitative and quantitative analysis of social networks, by describing features of a network either through numerical or visual representation. We have compared the tools on the basis of language or platform and their purposes.

REFERENCES

[1] A. Hanneman and M. Riddle "Introduction to Social Network Methods", online at http://faculty.ucr.edu/hanneman/nettext, 2005.
[2] The Development of Social Network Analysis", Vacouvver: Empirical Press, 2005.
[3] M.Jamali and H.Abolhassani, "Different Aspects of Social Network Analysis", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'O6), 2006.
[4] S.Wasserman and K.Faust, "Social network analysis: methods and analysis", Cambridge University Press, November 1994.
[5] Rupam Some, "A Survey on Social Network Analysis and its Future Trends", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 6, June 2013.
[6] Agnieszka Rusinowska, Rudolf Berghammer, Harrie De Swart and Michel Grabisch, "Social networks: Prestige, centrality, and inuence", RAMICS, Springer, pg. 22-39, 2011.
[7] David Combe, Christine Largeron, Elod Egyed-Zsigmond and Mathias Gery, "A comparative study of social network analysis tools", International Workshop on Web Intelligence and Virtual Enterprises 2, 2010
[8] M. E. J. Newman, "A measure of betweenness centrality based on random walks," arXiv:cond-mat/0309045v1, 2003.
[9] Raju Enugala, Lakshmi Rajamani, Kadampur Ali and Srevanthi Kurapati, "Community Detection in Dynamic Social Networks: A Survey", International Journal of Research and Applications, pg. 278-285, 2015.
[10] J.M. Kleinberg, "Authoritative sources in a hyperlinked environment", Journal of the ACM, 46:668-677, 1999. Agnieszka Rusinowska, Rudolf Berghammer, Harrie De Swart, Michel Grabisch, "Social networks: Prestige, centrality, and influence", RAMICS, Springer, pp.22-39, 2011.
[11] Ioana-Alexandra, "An overview of Software Application for Social Network Analysis", International Review of Social Research, Vol. 3, 2013.
[12] International network of Social Network Analysis INSNA, www.insna.org.
[13] N.Akhtar, "Social Network Analysis Tools", IEEE, Computer Society, 978-1-4799-3070-8/14 IEEE 2014.
[14] De Nooy, Wouter, Andrej Mrvar and Vladimir Batagelj, "Exploratory social network analysis with Pajek", Cambride University Press, 2011.
[15] V. Batagelj and A. Mrvar (2004). "Pajek: Program for Large Network Analysis", http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
[16] Gephi https://gephi.org
[17] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks", in International AAAI Conference on Weblogs and Social Media, pg. 361–362, 2009.
[18] S.P. Borgatti, "NetDraw Software for Network Visualization: Documentation & FAQs",(2002). https://sites.google.com/site/netdrawsoftware/documentation-faqs Retrieved: October 20, 2013.
[19] Networkx, http://Networkx.lanl.gov
[20] A.A. Hagberg, D.A. Schult, and P.J. Swart, " Exploring network structure, dynamics, and function using NetworkX" In Proc. 7th Python in Science Conference, Varoquaux G, Vaught T, and Millman J (Eds), pg. 11-16, 2008.
[21] IGraph, http://igraph.sourceforge.net
[22] G. Csárdi and T. Nepusz, "The igraph software package for complex network research", InterJournal Complex Systems, 2006.