Volume 8, No. 4, May 2017 (Special Issue)
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info

Different Aspects of Visualizing Social Network Data

Mandeep Kaur
M.Tech Scholar, CSE Dept
SBBSIET, Padhiana
Jalandhar, India

Harpreet Kaur
Assistant Professor, CSE Dept
SBBSIET, Padhiana
Jalandhar, India

Abstract: A social network is a set of people (or organizations or other social entities) connected by a set of social relationships, such as friendship, co-working or information exchange. Social network analysis focuses on the analysis of patterns of relationships among people, organizations, states and such social entities. Social network analysis provides both a visual and a mathematical analysis of human relationships. Networks are critical to modern society, and a thorough understanding of how they behave is crucial to their efficient operation. Fortunately, data on networks is plentiful by visualizing this data, it is possible to greatly improve our understanding. In this paper a state of the art survey of the works done on social network analysis ranging from pure mathematical analyses in graphs to analyzing the social networks in Semantic Web is given. The main goal is to provide a road map for researchers working on different aspects of Social Network Analysis.

Keywords: data visualization, data analysis, Network visualization, social media Analysis, Network data analysis.

I. INTRODUCTION

Social media enables users to generate content by sharing their knowledge, opinions and experiences on a variety of issues. Social media has changed the way customers engage with organizations, brands, products and services. It influences customer attitudes, perceptions and buying decisions. Social media provides organizations with many opportunities. It provides a new and powerful low cost marketing channel that can be harnessed to increase customer awareness of organizations and associated brands, products and services. Also, it enables organizations to improve their customer relationships through better engagement on a real time basis. [1]

We are currently in the midst of a networking revolution. Data communications networks such as the Internet now connect millions of computers; cellular phones have become commonplace, and personal communications networks are in the developmental stages. In parallel with the ever increasing network sizes has been a concomitant increase in the collection of network measurement data. Understanding this data is of crucial importance as we move to a modern, information-rich society.

Unfortunately, tools for analyzing network data have not kept pace with the data volumes. More network measurement data is available today than ever before, yet it is useless until it is understood. [2] Traditional network analysis software and graphs cannot cope with the size of today's networks and their data collection capabilities. In 2016, it was estimated that there will be around 2.13 billion social network users around the globe; up from 1.4 billion in 2012.Social network penetration worldwide is ever-increasing. In 2012, 63.1 percent of internet users were also social network users and these figures were expected to grow.

Twitter gained 42 million users this year. The site welcomes 32 percent of Internet users age 18 to 24, and about 24 percent of that demographic has the app downloaded on a mobile device. About 86 percent of users access the site through mobile devices. Users spend an average of 17 minutes per day on the site, and 37 percent of users say they will buy products from a brand they follow on Twitter.

Following is the Network Analysis Background:

A. Network Analysis Background:

SNA (Social Network Analysis) has its origins in both social science and in the broader fields of network analysis and graph theory Network analysis concerns itself with the formulation and solution of problems that have a network structure; such structure is usually captured in a graph. Graph theory provides a set of abstract concepts and methods for the analysis of graphs. These, in combination with other analytical tools and with methods developed specifically for the visualization and analysis of social (and other) networks, form the basis of what we call SNA methods.

But SNA is not just a methodology; it is a unique perspective on how society functions. Instead of focusing on individuals and their attributes, or on macroscopic social structures, it centers on relations between individuals, groups, or social institutions.

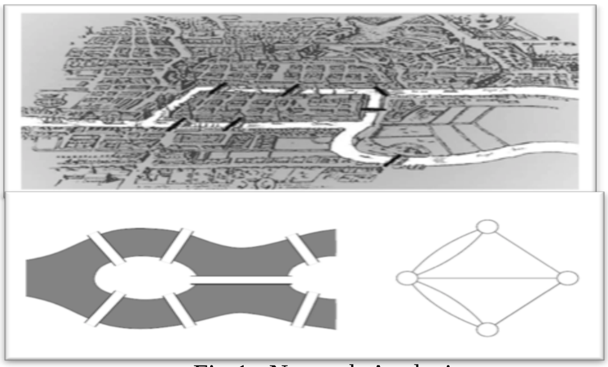

Fig.1. Network Analysis

Above is a very early example of network analysis comes from the city of Königsberg (now Kaliningrad). Famous mathematician Leonard Euler used a graph to prove that there is no path that crosses each of the city's bridges only once (Newman et al, 2006). Studying society from a network perspective is to study individuals as embedded in a network of relations and seek explanations for social behavior in the structure of these networks rather than in the individuals alone. This 'network perspective' becomes increasingly relevant in a society that Manuel Castells has dubbed the network society.

SNA has a long history in social science, although much of the work in advancing its methods has also come from mathematicians, physicists, biologists and computer scientists (because they too study networks of different types)

The idea that networks of relations are important in social science is not new, but widespread availability of data and advances in computing and methodology have made it much easier now to apply SNA to a range of problems. Network Analysis has found applications in many domains beyond social science, although the greatest advances have generally been in relation to the study of structures generated by humans Computer scientists for example have used (and even developed new) network analysis methods to study WebPages, Internet traffic, information dissemination, etc. One example in life sciences is the use of network analysis to study food chains in different ecosystems Mathematicians and (theoretical) physicists usually focus on producing new and complex methods for the analysis of networks that can be used by anyone, in any domain where networks are relevant.[4]

II.SOCIAL NETWORK MODELS

There are various social network models which are as follows:

A. Using formal methods to show Social Networks:

One reason for using mathematical and graphical techniques in social network analysis is to represent the descriptions of networks compactly and systematically. A related reason for using (particularly mathematical) formal methods for representing social networks is that mathematical representations allow us to apply computers to the analysis of network data. The third, and final reason for using "formal" methods (mathematics and graphs) for representing social network data is that the techniques for graph processing and the rules of mathematics themselves suggest things that we might look for in our data. In the analysis of complete networks, a distinction can be made between network data is that the techniques for graph processing and the rules of mathematics themselves suggest things that we might look for in our data. In the analysis of complete networks, a distinction can be made between

- Descriptive methods, also through graphical representations
- Analysis procedures, often based on a decomposition of the adjacency matrix
- Statistical models based on probability distributions

B. Using Graphs to Represent Social Relations:

Network analysis uses (primarily) one kind of graphic display that consists of points (or nodes) to represent actors and lines (or edges) to represent ties or relations. When sociologists borrowed this way of graphing things from the mathematicians, they renamed their graphs as"sociograms". There are a number of variations on the theme of sociograms, but they all share the common feature of using a labeled circle for each actor in the population we are describing, and line segments between pairs of actors to represent the observation that a tie exists between the two.

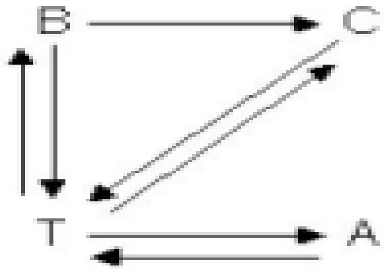

Fig.2. Using Graphs to Represent Social Relations

Visualization by displaying a sociogram as well as a summary of graph theoretical concepts provides a first description of social network data. For a small graph this may suffice, but usually the data and/or research questions are too complex for this relatively simple approach.

C. Using Matrices to Represent Social Relations:

The most common form of matrix in social network analysis is a very simple one composed of as many rows and columns as there are actors in our data set, and where the elements represent the ties between the actors. The simplest and most common matrix is binary. That is, if a tie is present, a one is entered in a cell; if there is no tie, a zero is entered. This kind of a matrix is the starting point for almost all network analysis, and is called an "adjacency matrix" because it represents who is next to, or adjacent to whom in the "social space" mapped by the relations that we have measured. By convention, in a directed graph, the sender of a tie is the row and the target of the tie is the column. Let's look at a simple example. The directed graph of friendship choices among Bob, Carol, Ted, and Alice looks like figure 1. Since the ties are measured at the nominal level (that is, the data are binary choice data), we can represent the same information in a matrix that looks like Table 1:

TABLE I. Using Matrices to Represent Social Relations

	Bob	Carol	Ted	Alice
Bob	---	1	0	0
Carol	1	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

D. Statistical Models for Social Network Analysis:

Statistical analysis of social networks spans over 60 years. Since the 1970s, one of the major directions in the field was to model probabilities of relational ties between interacting units (social actors), though in the beginning only very small groups of actors were considered. Extensive introduction to earlier methods is provided by Wasserman and Faust. Two of the most prominent current directions are Markov Random Fields (MRFs) introduced by Frank and Strauss and Exponential Random Graphical Models (ERGMs), also known as p. There are several useful properties of the stochastic models. Some of them are:

• The ability to explain important properties between entities that often occur in real life such as reciprocity, if i is related to j then j is more likely to be somehow related to i; and transitivity, if i knows j and j knows k, it is likely that i knows k.

• Inference methods for handling systematic errors in the measurement of link.
• General approaches for parameter estimation and model comparison using Markov Chain Monte Carlo methods.
• Taking into account individual variability and properties (covariates) of actors.
• Ability to handle groups of nodes with equivalent statistical properties.

There are several problems with existing models such as degeneracy analyzed by and scalability mentioned by several sources. The new specifications for the Exponential Random Graph Models proposed in attempt to find a solution for the unstable likelihood by proposing slightly different parameterization of the models than was used before.[3]

III. SOCIAL NETWORK PROPERTIES

There are some properties of social networks that are very important such as size, density, degree, reachability, distance, diameter, geodesic distance. Here we describe some more complicated properties which may be used in social network analysis.

A. Maximum flow: One notion of how totally connected two actors are, asks how many different actors in the neighborhood of a source lead to pathways to a target. If I need to get a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak - even if the person I send it to may have many ways of reaching you. If, on the other hand, there are four people to whom I can send my message, each of whom has one or more ways of retransmitting my message to you, then my connection is stronger. This" flow" approach suggests that the strength of my tie to you is no stronger than the weakest link in the chain of connections, where weakness means a lack of alternatives.[4]

B. Hubbell and Katz cohesion**:** The maximum flow approach focuses on the vulnerability or redundancy of connection between pairs of actors - kind of a" strength of the weakest link" argument. As an alternative approach, we might want to consider the strength of all links as defining the connection. If we are interested in how much two actors may influence on one another, or share a sense of common position, the full range of their connections should probably be considered. Even if we want to include all connections between two actors, it may not make a great deal of sense (in most cases) to consider a path of length 10 as important as a path of length 1. The Hubbell and Katz approaches count the total connections between actors (ties for undirected data, both sending and receiving ties for directed data). Each connection, however, is given a weight, according to its length. The greater the length, the weaker the connection.

C. Taylor's Influence: The Hubbell and Katz approach may make most sense when applied to symmetric data; because they pay no attention to the directions of connections (i.e. A's ties directed to B are just as important as B's ties to A in defining the distance or solidarity – closeness– between them). If we are more specifically interested in the influence of A on B in a directed graph, the Taylor influence approach provides an interesting alternative. The Taylor measure, like the others, uses all connections, and applies an attenuation factor. Rather than standardizing on the whole resulting matrix, however, a different approach is adopted. The column marginals for each actor are subtracted from the row marginals, and the result is then normed. Translated into English, we look at the balance between each actor's sending connections (row marginals) and their receiving connections (column marginals). Positive values then reflect a preponderance of sending over receiving to the other actor of the pair -or a balance of influence between the two-. D. Centrality and Power All sociologists would agree that power is a fundamental property of social structures. There is much less agreement about what power is, and how we can describe and analyze its causes and consequences. Table I summarizes some of the main approaches that social network analysis has developed to study power, and the closely related concept of centrality. [5]

TABLE II. Comparing three aspects of power in sociograms (degree, closeness, and betweenness)

Power Aspect Name	Definition	Influences
Degree	Number of ties for an actor	Having more oppurtunities and alternatives
Closeness	Length of paths to other actors	Direct bargaining and ex-change with other actors
Betweenness	Lying between each other pairs of actors	Brokering contacts among actors to isolate them or pre-vent connections

IV TYPES OF NETWORKS

Networks are categorized by the nature of the sets of actors and the properties of the ties among them.

A. One-mode network: a single set of actors + one or more types of relations between pairs of the actors + actor attributes.

B.two-mode network:

- Dyadic two-mode network: two sets of actors + one or more types of relations between actors in the two sets

- Affiliation network: one set of actors and one set of events + attendance/membership + attributes of the actors and the events

C.Ego-centered and special dyadic networks: couples; mothers-children; ego-centered network.[7]

V.CHARACTERISTICS OF SOCIAL NETWORK DATA

Human social activities generate a lot of social network data, such as Sina publishes tens of thousands of micro-blog every day. How to analyze the data of these social networks has become a hot research field. Visualization of social network data using mathematics diagram, graph vertices represent people, links show the relationship between people's activities. Because social networks are a kind of complex networks, so social network data is in line with the general characteristics of complex networks, mainly in the following four points:

A. social network data is sparse as a whole, dense local: In the Social networks and cooperation networks between scientists, the activities between people who mutual understand or cooperate are very frequently. The data is used to represent these activities is very large, contact between these data will be very close, and these social network data presents locally dense features

B. Social network data is sparse as a whole, dense local:In the Social networks and cooperation networks between scientists, the activities between people who mutual understand or cooperate are very frequently. The data is used to represent these activities is very large, contact between these data will be very close, and these social network data presents locally dense features. From the whole region or country to observe, social network and collaborative International Symposium on Social Science (ISSS 2015) © 2015. The authors - Published by Atlantis Press 275 network of scientists is less contact between the various regions, social network data as a whole experience the feature of sparse.

C. Social network data have time characteristic:
Social network activity is often occurred with a time, this time records the whole process of activities, studying of social networks often start from this time. For example, when analyzing the micro-blog public opinion, when, who issued, the number of people forwarding, each time of these actions. Then analyzing the information in chronological order, we can have a certain understanding of the propagation of public opinion.[7]

D. Social network data have hierarchical attribute**:**
The social network of social structure and family relationships has grade level. A hierarchy contains the following sub-level, as well as the second son of the following sub-level hierarchy. For example, a company's organizational structure, generally divided into chairman, middle manager, the general staff of the three levels, these social networks have a property of hierarchical structure.

E. *Social network data have network attribute:*
The main participants of the Internet network, social networks, collaboration networks, disease transmission networks is human, exchanging activities between people staggered overlap. The links of visual image between nodes mutually cross, intricate, such social network have network properties.[8]

VI. CONCLUSION AND FUTURE SCOPE

Visualizing social networks is of immense help for social network researchers in understanding new ways to present and manage data and to effectively convert the data into meaningful information.

Social Network Analysis (SNA) is becoming an important tool for investigators, but all the necessary information is often distributed over a number of Web servers. Currently there are developing information system that helps managers and team leaders to monitor the status of a social network. This paper presented an overview of the basic concepts of social networks in data analysis including social network analysis metrics and performances. Different problems in social networks are discussed such as uncertainty, missing data and finding the shortest path in a social network. Community structure, detection and visualization in social network analysis were also discussed. The current implementation includes analyzing one social network connection map. As a future enhancement additional modules can be included to derive similar pattern from a group of heterogeneous social network.

VII. ACKNOWLEDGEMENT

The satisfaction that accompanies that the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. The paper entitled “Different Aspects of Visualizing Social Network Data “is done under the guidance of my teacher “Er. Harpreet Kaur”. I am very thankful to my teacher for the inspiration and constructive suggestions that helpful us in the preparation of this paper.

REFERENCES

[1] Implementation of Enhanced Graph Layout Algorithm for Visualizing Social Network Data using NetworkX Library” in “International Journal of Advance research in Computer Science” Volume 8 Issue II (March-April 2017).

[2] Marcin Mincera and Ewa Niewiadomska-Szynkiewicza, Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland Research and Academic Computer Network (NASK), Warsaw, Poland Application of Social Network Analysis to the Investigation of Interpersonal Connections, journal of telecommunications and information technology 2012.

[3] Caroline Haythornthwaite Graduate School of Library and Information Science University of Illinois, Urbana-Champaign Social Network Analysis: An Approach and Technique for the Study of Information Exchange LISR 18, 323-342 (1998)

[4] IEEE Transactions on Visualization and Computer Graphics, Vol. 1, No. 1, pages 16-21, March 1995. Visualizing Network Data Richard A. Becker Stephen G. Eick Allan R. Wilks.

[5] Mohsen Jamali and Hassan Abolhassani Web Intelligence Research Laboratory Computer Engineering Department Sharif University of Technology, Network Analysis Tehran, Iran.

[6] Application of Social Network Analysis to the Investigation of Interpersonal Connections Marcin Mincera and Ewa Niewiadomska-Szynkiewicza,b a Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland b Research and Academic Computer Network (NASK), Warsaw, Poland in Interpersonal Connections, journal of telecommunications and information technology 2012.

[7] International Symposium on Social Science (ISSS 2015) A Survey on Social Network Visualization Jiang Du1 & Yafei Xian1, Jiayu Yang2 1College of Computer Science and Technology, Chongqing University of Posts and Telecommunications 2Chongqing iSoft Network Security Information Technology Co., Ltd.

[8] Social Network Analysis of Online Marketplaces Pushpa Kumar Kang Zhang Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083 {pkumar, kzhang}@utdallas.edu.

[9] http://scalar.usc.edu/works/querying-social-media-with-nodexl/using-graph-layout-algorithms-in-nodexl.