www.ijecs.in
***International Journal Of Engineering And Computer Science ISSN:2319-7242***
***Volume 6 Issue 12 December 2017, Page No. 22302-22307***
***Index Copernicus Value (2015): 58.10, 76.25 (2016) DOI: 10.18535/ijecs/v6i12.09***

# A Study of Visualizing Social Network Data.

**Bugga Bharath kumar*, Perika Kalanwesh , D.Hima Bindu**

Department of Computer Science and Engineering, Osmania University.
Hyderabad-07.T S.INDIA.

**Abstract**

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 (SNA) is becoming an important tool for investigators, but all the necessary information is often available in a distributed environment. Currently there is no information system that helps managers and team leaders monitor the status of a social network. This chapter presents 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 is also illustrated. This chapter bridges the gap among the users by combining social network analysis methods and information visualization technology to help a user visually identify the occurrence of a possible relationship amongst the members in a social network. 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 The main goal is to provide a road map for researchers working on different aspects of Social Network Analysis.The resulting social networks have high value to business intelligence, sociological studies, organizational studies, epidemical studies, etc.

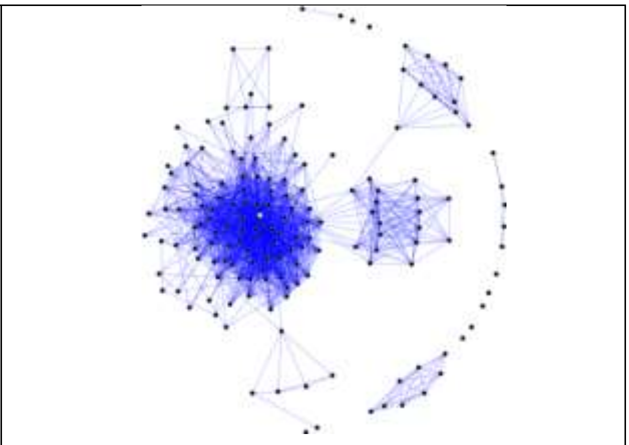
**Keywords:** Social networks, , network analytics. Data visualization, data analysis, Network visualization, Network data analysis.

## Introduction

Social network analysis (SNA) is a mathematical or computer science theory which consist to visualize and modeled the individuals of a network and the relationship between these individuals keep on through algorithms and statistics to glimpse a number information according to the user need. 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.

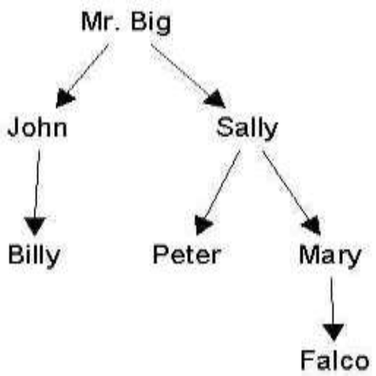
**Social network as a graph:**

- Ties (Contain One or More Relationships)
- Friendship (with possibly many relationships)
- Affiliations (Person – Organization)
- Works for IBM; ACM Member; Football Team
- One-Mode, Two-Mode Networks



**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.
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]

**Social Network Models**
**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
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. 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
A binary relation R is a set of ordered pairs (x,y). Each element in an ordered pair is drawn from a (potentially different) set. The ordered pairs relate the two sets: together, they comprise a mapping, which is another name for a relation. Functions, like $y = 2x^2$ are relations. The equation notation is just short hand for enumerating all the possible pairs in the relation (e.g., (1,2), (3,18), (-2,8), etc.). Social relations, which are at the heart of the network enterprise, correspond well to mathematical relations. To represent a social relation, such as "is the boss of", as a mathematical relation, we simply list all the ordered pairs:

| | |
|---|---|
| (Mr. Big, John),<br>(Mr. Big, Sally),<br>(John, Billy),<br>(Sally, Peter),<br>(Sally, Mary),<br>(Mary, Falco) |  |
| Figure 1a. The "is the boss of" relation, represented as a binary relation. | Figure 1b. Same relation represented as a directed graph. |

**Graphs**

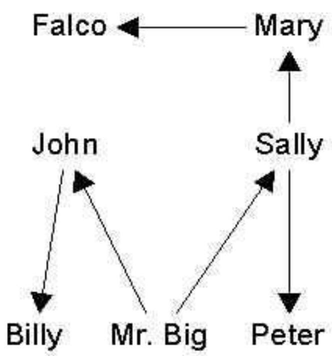
A graph G(V,E) is a set of vertices (V) together with a set of edges (E). Some synonyms:

| | **Vertices** | **Edges** |
|---|---|---|
| **Mathematics** | node, point | line, arc, link |
| **Sociology** | actor, agent | tie |

There are several cross-cutting dimensions that distinguish among graphs.

| | |
|---|---|
| Directed vs Undirected: | • Directed graphs (also called digraphs) consist of ordered pairs. The links in a directed graph are called arcs. Can use these to represent non-symmetric relations like "is the boss of" or "is attracted to"<br>• Undirected graphs (also known simply as "graphs") consist of unordered pairs. They are used for the relations which are necessarily symmetric, such as "is the sibling of" or "lives with" |
| Valued vs Non-Valued | • In non-valued graphs, nodes are either connected or not. Either Sally and Bill are siblings, or they're not.<br>• In valued graphs, the lines have values attached to represent characteristics of the relationships, such as strength, duration, capacity, flow, etc. |
| Reflexive vs Non-Reflexive | • Reflexive graphs allow self-loops. That is, a node can have a tie to itself. This is mostly useful when the nodes are collectivities. For example, if the nodes are cities and the ties represent phonecalls between people living in those cities, it is possible (a virtual certainty) that there will be ties from a city to itself. |
| Multi-graphs | • If more than one edge connects two vertices, this is a multigraph. In general, we do not use multigraphs, preferring to use either valued graphs (to represent the number of interactions between A and B) or wholly separate graphs (to represent substantively different relations, such as "does business with" and "is married to" |

We represent graphs iconically as points and lines, such shown in Figure 1b. It is important to note that the location of points in space is arbitrary unless stated otherwise. This is because the only information in a graph is who is connected to whom. Hence, another, equally valid way to represent the graph in Figure 1 would be this:




All representations in which the right people are connected to the right others are equally valid.

The ***degree*** of a node is the number of nodes it is adjacent to; or, equivalently, it is the number of edges that are incident upon it. A node with no degree (degree 0) is an ***isolate***. A node with degree 1 is called a ***pendant***. The graph in Figure 1 contains two pendants and no isolates.

In a digraph, the ***indegree*** of a node is the number of arcs coming in to a node from others, while the ***outdegree*** is the number of arcs from the node to all others. More technically, the indegree of a node *u* is |{(x,u): (x,u) E}|, where the vertical bar notation |S| gives the number of elements in set S. Similarly, the outdegree of *u* is given by |{(u,x): (u,x) E}|. A node with positive outdegree but no indegree is called a ***source***. A node with positive indegree but no outdegree is called a ***sink.*** A ***subgraph*** H of a graph G is a graph whose points and lines are also in G, so that V(H) V(G) and E(H) E(G). If we select a set of nodes S from a graph G, and then select all the lines that connect members of S, the resulting subgraph H is called an ***induced subgraph*** of G based on S. Two points are adjacent in H if and only if they are adjacent in G.A ***walk*** is a sequence of adjacent points, together with the lines that connect them. In other words, a walk is an alternating sequence of points and lines, beginning and ending with a point, in which each line is incident on the points immediately preceding and following it. The points and lines in a walk need not be distinct. For example, in Figure 1, the sequence *a-b-c-b-c-d* is a walk. A ***path*** is a walk in which no node is visited more than once. The sequence *a-b-c-d* is a path. A ***cycle*** is like a path except that it begins and ends with the same node (*e.g. c-d-e-c*).

A set of walks that share no points is called ***vertex-disjoint***. A set of walks that share no edges is called ***edge-disjoint***. Obviously, vertex-disjoint paths are also edge-disjoint. If a pair of nodes are connected by three vertex-disjoint paths, this means that there are three, completely independent ways of getting from one point to the other. In Figure 1, there is only one edge-disjoint path from *a* to *f*, but two edge disjoint-paths from *c* to *d*.

The ***length*** of a walk is given by the number of lines it contains. A ***geodesic*** path is a shortest path between two points. There can be more than one geodesic path joining a given pair of points. The ***graph-theoretic distance*** or ***geodesic distance*** between two points is the length of a shortest path between them. In a diffusion process, one expects faster diffusion among nodes that are close together than among nodes that are far apart.

A graph is ***connected*** if there exists a path from every node to every other. A maximal connected subgraph is called a ***component***. A *maximal subgraph* is a subgraph that satisfies some specified property (such as being connected) and to which no node can be added without violating the property. For example, in Figure 1, the subgraph induced by {a,b,c} is connected, but is not a component because it is not maximal: we could add node *d* and the subgraph induced would still be connected. The graph in Figure 1 has just one component, which is the whole graph.

A digraph that satisfies the connectedness definition given above (*i.e.* there exists a path from every node to every other) is called ***strongly connected***. That is, for any pair of nodes *a* and *b*, there exists both a path from *a* to *b* and from *b* to *a*. A digraph is ***unilaterally connected*** if between every unordered pair of nodes there is at least one path that connects them. That is, for any pair of nodes *a* and *b* (and *a b*), there exists a path from either *a* to *b* or from *b* to *a*. A digraph whose underlying graph is connected is called ***weakly connected***.

A maximal strongly connected subgraph is a ***strong component***. A maximal weakly connected subgraph is a ***weak component***. A maximal unilaterally connected subgraph is a ***unilateral component***.

A ***cutpoint*** is a node whose removal would disconnect the graph. Alternatively, we could define a cutpoint as a node whose removal would increase the number of components of the graph. In the figure, nodes *b, c,* and *e* are cutpoints, while *a, d,* and *f* are not. A network that contains a cutpoint will break apart if the person who occupies the cutpoint leaves. A ***block*** is a subgraph which contains no cutpoints. A ***cutset*** is a set of points whose removal would disconnect a graph. A ***minimal cutset*** is a cutset that contains the minimum possible number of nodes that disconnect the graph. The size of a minimal cutset is called the ***vertex connectivity*** of the graph, and is denoted . The smaller the value of , the greater the vulnerability of the network to disconnection. We can also define a pairwise version of vertex connectivity, (*u,v*), which gives the minimum number of nodes that must be removed in order to disconnect vertex *u* from vertex *v*.

A ***bridge*** is a line whose removal disconnects the graph. An ***edge cutset*** is a set of lines whose removal disconnects the graph. A ***minimum weight cutset*** is an edge cutset which, for non-valued graphs, contains the fewest possible number of edges or, for valued-graphs, contains edges whose values add to the smallest possible value. The size of a minimum weight cutset is called the ***edge connectivity*** of the graph, and is denoted . As with point connectivity, we can also define a pairwise version of vertex connectivity, (*u,v*), which gives the minimum number of edges that must be removed in order to disconnect vertex *u* from vertex *v*.

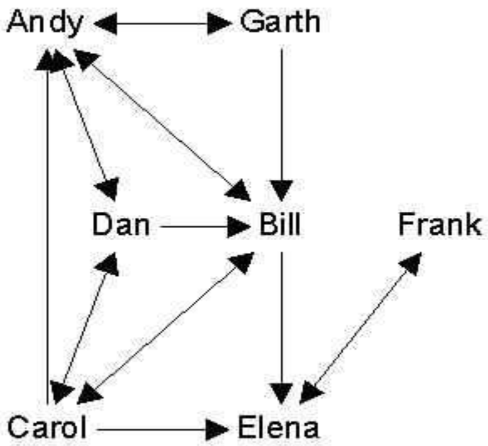
***Menger's theorem*** states that the minimum number of nodes that must be removed to disconnect node *u* from node *v*, (*i.e.* (*u,v*)) is equal to the maximum number of vertex-disjoint paths that join *u* and *v*. This also works for lines: the edge-connectivity (*u,v*) is equal to the maximum number of edge-disjoint paths that join *u* and *v*.

A good reference on graph theory is Frank Harary's 1969 book, *Graph Theory*, from Addison-Wesley.

**Matrices**

We can record who is connected to whom on a given social relation via an adjacency matrix. The adjacency matrix is a square, 1-mode actor-by-actor matrix like this:

|  | And | Bil | Car | Dan | Ele | Fra | Gar |
|---|---|---|---|---|---|---|---|
| Andy |  | 1 | 0 | 1 | 0 | 0 | 1 |
| Bill | 1 |  | 1 | 0 | 1 | 0 | 0 |
| Carol | 1 | 1 |  | 1 | 1 | 0 | 0 |
| Dan | 1 | 1 | 1 |  | 0 | 0 | 0 |
| Elena | 0 | 0 | 0 | 0 |  | 1 | 0 |
| Frank | 0 | 0 | 0 | 0 | 1 |  | 0 |
| Garth | 1 | 1 | 0 | 0 | 0 | 0 |  |




If the matrix as a whole is called X, then the contents of any given cell are denoted $x_{ij}$. For example, in the matrix above, $x_{ij} = 1$, because Andy likes Bill. Note that this matrix is not quite symmetric ($x_{ij}$ not always equal to $x_{ji}$).
Anything we can represent as a graph, we can also represent as a matrix. For example, if it is a valued graph, then the matrix contains the values instead of 0s and 1s.
By convention, we normally record the data so that the row person "does it to" the column person. For example, if the relation is "gives advice to", then xij = 1 means that person i gives advice to person j, rather than the other way around. However, if the data not entered that way and we wish it to be so, we can simply transpose the matrix. The transpose of a matrix X is denoted X'. The transpose simply interchanges the rows with the columns.

**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]

**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.

**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. Linton Freeman. Visualizing social networks. Journal of Social Structure, 1(1), 2000. [14] Linton C Freeman. Centrality in social networks: Conceptual clari¿cation. Social Networks, Vol. 1, No. 3(3):215 –239, 1979.
6. Abadie, A., Gardeazabal, J.: Terrorism and the world economy. Eur. Econ. Rev. 52(1), 1–27 (2008)
7. Freeman, L.: Visualizing social network. J. Soc. Struct. 1(1) (2000)
8. Boyd, D., Ellison, N.: Social network sites: definition, history and scholarship. J. Comput. Mediat. Commun. **13**(1), 210–230 (2007)
9. L. Kahney, "Making friendsters in high places," July 2003, the Wired News.
10. S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in Proceedings of the 7th World-Wide Web Conference (WWW7), BrisBane, Australia, April 1998.
11. J. Kleinberg, "Authoritative sources in a hyperlinked environment," in Proceedings of Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, California, January 1998.
12. J. Dean and M. R. Henzinger, "Finding related pages in the world wide web," in In Proceedings of the 8th World Wide Web Conference, Toronto, Canada, May 1999.
13. K. Bharat, A. Broder, M. Henzinger, P. Kumar, and S. Venkatasubramanian, "The connectivity server: Fast access to linkage information on the web," Brisbane, Australia, April 1998.