

Visualization Through Knowledge Representation Model for Social Networks

D. M. Akbar Hussain, Muhammad Athar Javed and Zaki Ahmed *

Abstract—Knowledge management is a systematic and organizationally specified process and knowledge management system is all those technological components; software, hardware, people and processes supporting knowledge management initiative. These initiatives includes work flow maps, web sites, portals, document/team management system, data warehouses, data mining processes, databases, contact lists, virtual teams, collaboration tools, customer relationship management, applications and news (Davenport and Prusak 1998, Jashapara 2004) [1, 2]. Knowledge is not important per se [3] (Agostini et al 2003) instead the process of knowing, learning and creating knowledge is the relevant aspect [4] (Nonaka and Takeuchi 1995). In this paper knowledge representation is presented in 3D style for the understanding and visualization of dynamics of complex social networks by developing a TANetworkTool (Task Analysis Network Tool). The standard or normal representation of a typical social network is through a graph data structure in 2D. The dynamics of larger social networks is so complex some time it becomes difficult to understand the various levels of interactions and dependencies just by mere representation through a tree or graph. Although, many analytical methods provide relationship dependencies, role of different nodes and their importance in the network. In this paper we are presenting a visualization of networks by rotating the network through various dimensions to provide a more realistic view to understand the dynamics of complex social networks and complimenting the analytical results. This representation can also help authorities not necessarily having specific scientific background to understand and perhaps take preventive actions required in certain specific scenarios for example dealing with terrorist/covert networks.

Keywords: Social Networks Analysis (SNA), Terrorism, Models, Intellectual Simulation Model.

*Manuscript submitted January, 2011 Dr. M. Akbar Hussain is member of IEEE, IDA, IAENG and works at the Department of Electronic Systems Aalborg University, Niels Bohrs Vej 8, 6700 Esbjerg, Denmark. Email: akh@es.aau.dk. Dr. Zaki Ahmed is working as a chief Scientist at PILO (Pakistan Institute of Laser and Optics). Email: zaki786@ieee.org. Muhammad Athar Javed, Expert on South Asian Affairs, Copenhagen, Denmark. Email: ma_javed@hotmail.com.

1 Introduction

To have an insight visualization of any network, typical focus is on the characteristics of the network structure. Social Network Analysis is a mathematical method for 'connecting the dots'. SNA allows us to map and measure complex, and sometimes covert, human groups and organizations [5]. Given any network where the nodes/agents are individuals, groups, organizations etc., a number of network measures such as centrality or cut-points are used to locate critical/important nodes/agents. Many traditional social network measures and the information processing network measures can help in revealing importance and vulnerabilities of the nodes/agents in the network. Networks visualization is semantically presented in the form of a graph in which the nodes represent entities and the arcs represent relationship among nodes.

In this paper we are presenting a framework for visualizing the network. The reference point could be any of the well known centrality measures in the context of social network analysis, for example Degree, Betweenness or Closeness. Once the important nodes based on these measures are determined our developed novel technique, adjust and rotate the network in 3D for a better visualization view. The nodes distinctiveness classification is a challenging task. Typically, social network analysis identifies the following characteristics:

- Important individual, event, place or group.
- Dependency of individual nodes.
- Leader-Follower identification.
- Bonding between nodes.
- Vulnerabilities identification.
- Key players in the network.
- Potential threat from the network.
- Efficiency of overall network

Networks visualization is semantically presented in the form of a graph in which the nodes represent entities and the arcs represent relationship among nodes. Classification of nodes and its distinctiveness is a challenging task and one needs to discover the following characteristics [6].

- An individual or group that if given new information can propagate it rapidly.
- An individual or group that has relatively more power and can be a possible source of trouble, potential dissidents, or potential innovators.
- An individual or group where movement to a competing group or organization would ensure that the competing unit would learn all the core or critical information in the original group or organization (inevitable disclosure).
- An individual, group, or resource that provides redundancy in the network.

Many traditional social network measures and the information processing network measures can help in revealing importance and vulnerabilities of the nodes/agents in the network [7, 8, 9, 10]. Application of existing tools on these complex socio-technical networks/systems is very demanding to winkle out the required information. Most of the measures and tools work best when the data is complete; i.e., when the information is inclusive about the interaction among the nodes. However, the difficulty is that covert and terrorist networks are typically distributed across many boundaries for example from cities or countries and data about them is never complete-correct at a certain instant of time. Normally, a sampled snapshot data is available some of the links may be intentionally hidden. Also data is collected from multiple sources for example news (print/tv), open source Internet data, security agencies, etc., and at different time instants. In addition inclusive and correct information may be prohibitive because of secrecy. Obviously, there could be other difficulties but even these provide little guidance for what to expect when analyzing these complex socio-technical systems with the developed tools. The next section provides a survey of the standard centrality measures used in the social network analysis. These measures are important as they have been used as a reference point for visualization of the network. In this investigation we have considered many examples to demonstrate the impact of having this multi-dimensional visualization for knowledge representation to understand networks dynamics using our TANetworkTool.

2 Network Analysis Measures

Social networks provides mapping and the social network analysis measure relationships and movement between people, groups, events, organizations or other information/knowledge processing entities. People, organization and groups are represented as nodes in the network while the links show relationships or movement between the nodes. SNA provides both visual and mathematical analysis of human relationships. This methodology could also

be used by the management to perform Organizational Network Analysis [5].

2.1 Centrality (Degree)

To comprehend networks and their participants, we evaluate the location of participants in the network. Degree provides the relative importance and the location of a particular node in the network. Degree and similar measures indicate the various roles of the nodes in a network, for example leaders, gatekeepers, role models etc. A node is central if it is strategically located on the communication route joining pairs of other nodes [7, 8]. Being central it can influence other nodes in the network, in other words potentially it can control the flow of information. The potential of control makes the centrality conceptual model for these nodes.

The idea of centrality is not new it was first applied to human communication by Baveles in 1948 [7, 11]. In this study relationship between structural centrality and influence in group processes were hypothesized. Following Baveles it was concluded that centrality is related to group efficiency in problem-solving, perception of leadership and the personal satisfaction of participants [12, 13, 14]. In the fifties and sixties more research was conducted on these measures and it was concluded that centrality is relevant to the way groups get organized to solve problems. The following references provide a very deep and pioneering work on these measures [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26].

The centrality concept is not exclusive to deal with group problem tasks, it has been used in other discipline as well [27, 28]. A number of centrality measures have been proposed over the past years. Most of the centrality measures are based on one of two quite different conceptual ideas and can be divided into two large classes [29]. The measures in the first class are based on the idea that the centrality of an individual in a network is related to how it is near to the others. Second class of measures is based on the idea that central nodes stand between others on the path of communication [30, 31, 32]. A node being on the paths of others nodes communication highway has the potential to control what passes through it. The simplest and most straightforward way to quantify the individual centrality is the degree of the individual node, i.e., the number of its first neighbors. In figure 1 a graph with seven nodes and seven edges is shown, nodes 2 and 3 are adjacent to each other, the number of nodes to which a given node is adjacent is called the degree of that point. Therefore, node 5 for example has a degree of 3. In a graph if every node is reachable from any node in the graph it is called a connected graph, which is the case in figure 1. Each path in the graph is associated with a distance equal to the number of edges in the path and the shortest path to reach a given pair of nodes is geodesic distance. For example path from node 2 to node

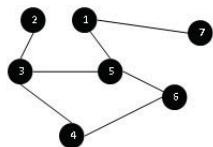


Figure 1: Network Example

1 through node 3 and 5 is a geodesic as the other path for the same pair is also reachable through node 3, 4, 6 and 5 but has longer distance.

Nieminen has provided a very systematic elaboration of the concept of degree [32]. Scott has extended the concept based on degree beyond immediate (first) neighbors by selecting the number of points an individual can reach at a distance two or three [9]. Similarly, Freeman produced a global measure based on the concept of closeness in terms of the distances among the various points [30]. The simplest notion of closeness is obtained from the sum of the geodesic distances from an individual to all the other points in the graph [33].

2.2 Betweenness

Betweenness measures to what extent a node can play the role of intermediary in the interaction between the other nodes. The most popular and simple betweenness measure based on geodesic path is proposed by Freeman and Anthonisse [30, 31]. In many real scenarios however, communication does not travel exclusively through geodesic paths. For such situations two more betweenness measures are developed first based on all possible paths between couple of nodes [34] and second based on random paths [10].

2.3 Closeness

Another more sophisticated centrality measure closeness based on geodesic distance can be defined, which is the mean geodesic (i.e., shortest path) distance between a node and all other nodes reachable from it. Closeness can be regarded as a measure of how long it will take information to spread from a given node to other nodes in the network.

3 Simulation Test

Although, we conducted tests for a number of networks ranging from 10 - 150 nodes however, we are only presenting results and observations from three networks with twenty, thirty and sixty nodes respectively. The connectivity density is also differently selected for these networks. Density means the numbers of edges a node can possess in other words the maximum degree of a node. It is selectable through a defined window range, set by the system designer (normally it has a default value of 6).

For the first two scenario networks density (degree) is in the range of 4 - 8 maximum and for the third network it is a bit higher. In the first two networks it is kept normal or low to show our information processing sequence for knowledge representation. Figure 2 shows a network with 20 nodes. The network data is generated by the data generated software engine DataGA, at present we have a small database with names and nationalities mostly known so far from open sources database like trackingthethreat.com etc. The DataGA takes the name and create links/edges (connectivity) with other nodes randomly, however the names are not shown on the transitional phases to avoid unnecessary text and lines mix up. The DataGA engine selects the positions of each node and the edges between them randomly from a normal distribution. After that the network visualization tool TANetworkTool determines the three centrality measures namely, degree, betweenness and closeness. These measures are necessary because the TANetworkTool needs one of these measures as a reference tool to place nodes at their appropriate locations to make visualization of the network more meaningful. For this network (figure 1) we are using degree as a reference measure. When degree is the reference measure the system tries to place nodes with higher degree at the periphery of the network and subsequently the remaining nodes inside of it. Obviously, during this process the network may go through rotation and skewing.

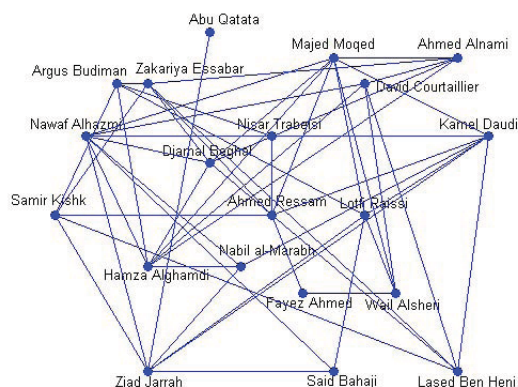


Figure 2: Network of 20 Nodes

The different selected phases for the network of figure 2 are shown in figure 3 (a - f). One can see that although this particular network was not a complex one to visualize even in its original form (figure 2) but still the final representation is visually more structured.

Figure 4 shows another network with 30 nodes. This time the reference measure is closeness. The TANetworkTool in this case tries to place immediately connected nodes side by side. However, this algorithm becomes computational very demanding in making decisions for those situations when many similar closeness values co-exist for

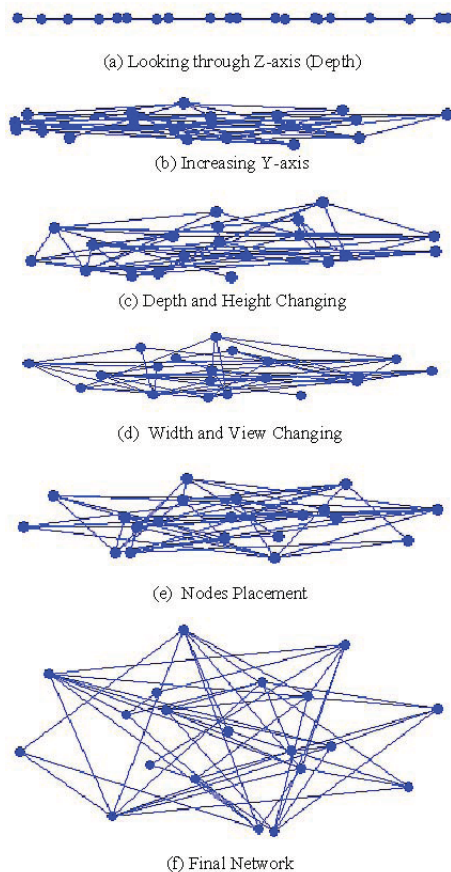


Figure 3: Selected Phases

multiple nodes. Therefore, to avoid system saturation in some cases it does not move the positions of those nodes.

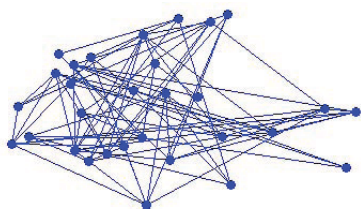


Figure 4: Network of 30 Nodes

Figure 5 shows the final phase, the intermediate phases are not shown here to avoid repetition as the idea is similar to figure 3. However, we can eventually see better visualization and also nodes closer to each other can be seen more clearly.

Finally we present a network of 60 nodes as shown in figure 6 with a relatively high density and one can see how complex the network looks. For this network we are using betweenness as a reference measure. The system tries to place nodes of higher betweenness in the inner periphery. As for this network, connectivity is a kind of worst case scenario, normally for a functional and viable

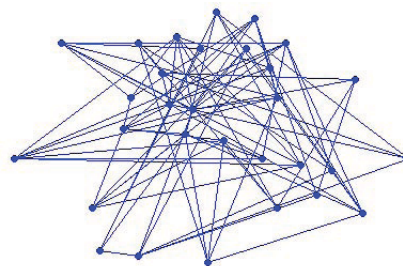


Figure 5: Final Phase of the Network

social network the average connectivity is 6 degree for a node. But the idea here is to show that our representation of the available information produce better results. The final phase after intermediate transitions is shown in figure 7. In addition to this processing the system can also blow up (Zoom) a particular area of the network to see the details more clearly. One such blow up of an area selected through a rectangle is shown in figure 8.

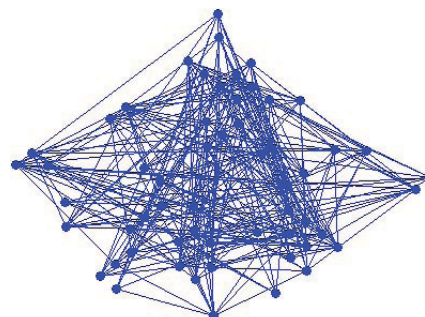


Figure 6: Network of 60 Nodes

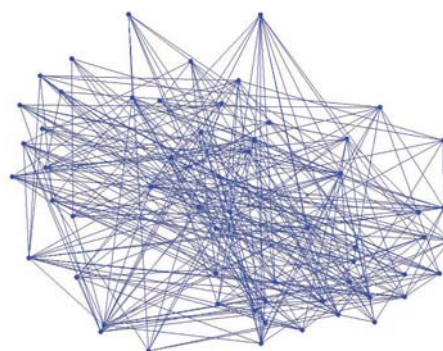


Figure 7: Final Phase of the Network

4 Conclusion

In this paper we have demonstrated by systematic representation of information to create knowledge representation for visualization of social networks. In doing so we have shown through various transitional phases that our method have significantly improved the visibility in order to understand the various dependencies/links. This kind

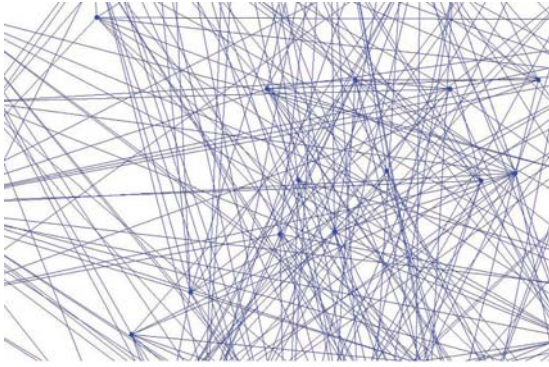


Figure 8: Blow-up of Selected Area is also Possible

of representation is very important and useful for those making critical decisions regarding destruction of terrorist/covert networks. Through simulations we have shown a better visualization of nodes connectivity. Although, our method is simple but it gives a more controlled approach to see the network from various dimensions and the system is quite flexible and interactive. However, we have found that when the connectivity density is very high, the visualization deteriorates and the only option left is to use the blow-up version of the algorithm.

References

- [1] Davenport T and Prusak L.: Working knowledge: Managing what your organization knows, harvard business school press, boston, ma, 1998.
- [2] Ashok Jashapara.: Knowledge management - an integrated approach, prentice hall, 2004, isbn-0273682989.
- [3] R. Boselli G. De Michelis F. De Paoli R. Dondi: Agostini, S. Albolino. Stimulating knowledge discovery and sharing, proceedings of the 2003 international acm siggroup conference on supporting group work, sanibel island, florida, usa, pages: 248 - 257, 2003 isbn: 1-58113-693-5.
- [4] Nonaka I. and Takeuchi H.: The knowledge creating company, oxford university press, oxford 1995.
- [5] Valdis Krebs.: Connecting the dots, tracking two identified terrorists, 2002.
- [6] Kathleen M. Carley Ju-Sung Lee David Krackhardt.: Destabilizing networks, dept. of social and decision sciences, carnegie mellon university, pittsburgh, pa 15143, november 2001.
- [7] Bavelas A.: A mathematical model for group structures". human organization 7: Pages 16 - 30, 1948.
- [8] Shaw M. E.: Group structure and the behaviour of individuals in small groups, journal of psychology, vol. 38, pages 139 - 149, 1954.
- [9] Scott J.: Social networks analysis. 2nd edition, sage publications, london, 2003.
- [10] Newman M. E. J.: A measure of betweenness centrality based on random walks, cond-mat/0309045, 2003.
- [11] Bavelas A.: Communication patterns in task oriented groups". journal of the acoustical society of america vol. 22, pages 271 - 282, 1950.
- [12] Leavitt Harold J.: Some effects of communication patterns on group performance. journal of abnormal and social psychology vol. 46, pages 38 - 50, 1951.
- [13] Smith Sidney L.: Communication pattern and the adaptability of task-oriented groups: an experimental study. cambridge, ma: Group networks laboratory, research laboratory of electronics, massachusetts institute of technology, 1950.
- [14] Bavelas A. and D. Barrett.: An experimental approach to organizational communication. personnel vol. 27, pages 366 - 371, 1951.
- [15] Mulder M.: Groepsstructuur en gedrag. nederlands tgdschrift voor de psychotogie i i: Pages 85 - 133, 1956.
- [16] Mulder M.: Groepsstnrectuur, motivatie en prestatie. den haag: C.o.p, 1958.
- [17] Glanzer M. and R. Glaser.: Techniques for the study of team structure and behaviour. part ii: Empirical studies of the effects of structure. technical report, pittsburgh, american institute, 1957.
- [18] Glanzer M. and R. Glaser.: Techniques for the study of group structure and behaviour. part ii: Empirical studies of the effects of structure in small groups. psychological bulletin 58, pages 1 - 27, 1961.
- [19] Cohen A. M.: Communication networks in research and training. personnel administration 27, pages 18 - 24, 1964.
- [20] Shaw M. E.: Communication networks", in I. Berkowitz (ed.), advances in experimental social psychology, vol. vi, pages 111 - 147, new york, academic press, 1964.
- [21] Stephenson K. A. and Zelen M.: Rethinking centrality: Methods and examples, social networks 11, pages 1 - 37, 1989.
- [22] Flament C.: Applications of graph theory to group structure. englewood cliffs, nj, prentice hall, 1963.
- [23] Burgess R. L.: Communication networks and behavioural consequences". human relations 22, pages 137 - 159, 1968.

- [24] Snadowski A.: Communication network research: an examination of controversies". human relations 25, pages 283 - 306, 1972.
- [25] Rogers D. L.: Socio-metric analysis of inter-organizational relations: application of theory and measurement. rural socioeconv 39, pages 487 - 503, 1974.
- [26] Rogers E. M. and R. Agarwala Rogers:.. Communication networks in organizations. communication in organizations, pages 108 - 148, 1976, new york, free press.
- [27] Cohn B. S. and M. Marriott:.. Networks and centres of integration in indian civilization. journal of social research i, pages 1 - 9, 1958.
- [28] Pitts F. R.: A graph theoretic approach to historical geography, the professional geographer 17, pages 15 - 20, 1965.
- [29] Latora V. and Marchiori M:.. A measure of centrality based on network efficiency, arxiv.org preprint cond-mat/0402050, 2004.
- [30] Freeman Linton C:.. Centrality in social networks: Conceptual clarification, social networks 1, page 215 - 239, 1979.
- [31] Anthonisse J. M:.. The rush in a graph, university of amsterdam mathematical centre, amsterdam, 1971.
- [32] Nieminen J:.. On centrality in a graph. scandinavian journal of psychology 15, pages 322 - 336, 1974.
- [33] Sabidussi G:.. The centrality index of a graph. psychometrika 31, pages 581 - 603, 1966.
- [34] Freeman Linton C. Stephen P. Borgatti and Douglas R. White:.. Centrality in valued graphs, a measure of betweenness based on network flow. social networks 13, pages 141 - 154, 1991.