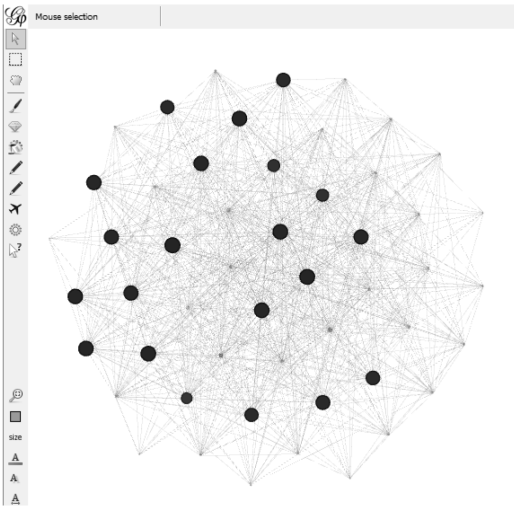
APPLYING VISUAL ANALYTICS IN SOCIAL NETWORK ANALYSIS AND LARGE SCALE DATA

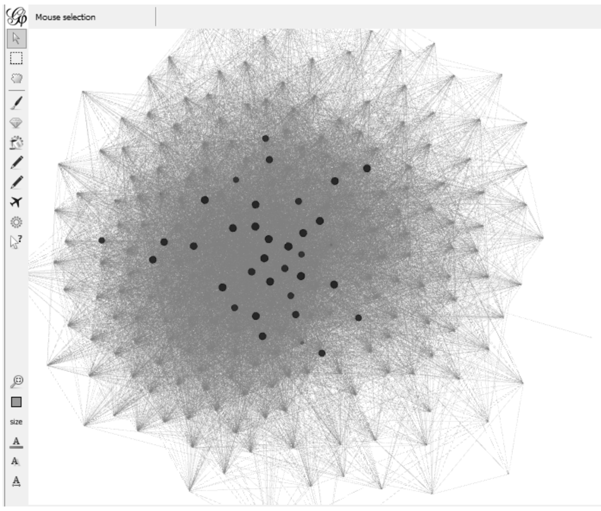
Akalanka Galappaththi
Board of Study in Statistics and Computer Science

We are living in the world where information found in structured or unstructured way and the information or data is in large scale. Generally we call them "Big Data". Analysing big data is a tedious task which consumes time. Time is a big hurdle in data analysis where analysts required to produce results within an acceptable time period. Compare to text based analysis visual techniques speed up the process since the human cognition can identify the patterns in visualized data much easily.

In the context of social network analysis, the visualizing tools are categorized into different generations according to their characteristics. First generation tools only provide a visual representation of entities in the network where network is drawn manually. Anacapa map is an example for the first generation tool. The paper published by Valid E. Krebs in 2001 at Connections 24-3 contains information of hijackers of 9/11 attack (Krebs, 2001). It is clearly observable that visualizing information is very much helpful in detecting relationships among patterns compare to textual presentation. Second generation tools have the ability to create the network automatically from data stored in a spread sheet file or a text file. Analyst's Notebook, Netmap and XANALYS link explorer are some of the tools in this category (Xu and Chen, 2005). However, still they do not provide analytical capabilities. Third generation tools provide the capability of analysing social network data to identify relational and positional information of entities. Relational analysis can find entities with higher degree, node which lies between geodesies of nodes and node which has the shortest path to many other nodes. Such information can reveal the importance of a particular person in a network for its logistic viability. Positional analysis finds entities which has similarities according to their features and associations between other nodes in the network. Hence possible substitute characters in the network can be identified.



A user interaction of a group in facebook with 51 nodes and 802 edges using Gephi tool



A user interaction of a group in facebook with 238 nodes and 12620 edges using Gephi tool

Visualizing social network data can provide deep insight of people in the network and rolesthey play in the network compared to simple statistical reasoning. A graphical representation of entities in the network and their interactions as a graph with nodes and edges is very convenient to the human. It provides the capability of analysing level of interaction by visualizing weights on edges, relative position of nodes to represent central members of network with many links to other nodes, nodes which lay between shortest path of other nodes and node which starts shortest paths to other nodes. First generation tools are practical in drawing relatively small networks. When drawing large scale complex networkssuch asfacebook, twitter and linkedin it becomes impractical. Further network structure is dynamic in social networks due to its growing or shrinking nature. Each small change requires tracking to understand the network structure since they change the cluster size and centrality measurements.

One of the challenges in graph creation in network analysis is the layout mapping. An efficient placement of nodes provides meaningful visualization. Force-directed approach, fast multipole multilevel mode and graphic drawing with intelligent placement are some algorithms used to layout static graphs (Ma and Muelder, 2013). Space filling curve method can represent clusters in a static graph. Clusters contain nodes with similar features.

Visualizing dynamic network suffers from the trade-off between stability and layout quality. One common method used is the animating the changes of dynamic graph which results appearance or disappearance of nodes and changes in placement of nodes. Another method is placing vertices on parallel vertical lines with directed edges from left to right. The changes to graph displayed in time step graph where two consecutive vertical lines show the time step. Incremental clustering based layout method and global clustering based layout method provide quality layout and stable network in each time step. As for the limitations, incremental clustering doesn't guarantee that the layout is the ideal compared to initial cluster while global clustering has a large computational burden (Ma and Muelder, 2013).

The whole purpose of applying visual analytics in social network analysis is to provide a means of analysis. Hence semantics of the network data need to be considered. It is much more convenient to show only relevant entities and their related association with other nodes rather than displaying the full network. For example, association weight can be used as negatively related to the distance of the edge. This will result some nodes in tightly associated while others are associated loosely. Further applying colours accordingly to represent links between group members and mix of colours to represent links between different group members will provide more semantics to the user.

With the overwhelming large amount of data with large scales in social networks, use of visual analytics has more advantagesin analysing them. Depends on the dynamic nature of the networks it is a very difficult task to find efficient method for creating network layout preserving the stability of the network over time. The future of this field is to combine different methods for dynamic network visualization to eliminate the disadvantages of each method. Clustering methods are heavily used in the domain to identify structural and feature level similarities of entities to create layout to represent semantics of the network. Centrality measure plays a major role finding importance of entities in the network and their impact to other entities and to the network overall. Visualizing these features provides the means of analysis to the analyst of the network.

References:

Krebs, V. E. (2001). Mapping Networks of Terrorist Cells, *Connections*. 24(3), 43-52.

Ma, K. and Muelder, C. W. (2013). Large-Scale Graph Visualization and Analytics, *Computer*, 46(7), 39-45. (Mar 14, 2014).

Xu, J. J. and Chen, H. (2005). CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery, *ACM Transaction on Information Systems*, 23(2), 201-226.