Intelligent Visualisation of Social Network Analysis Data

Phillip Higgins, Debbie Richards and Mike McGrath

Department of Computing
Division of Information and Communication Sciences
Macquarie University
Sydney, Australia
phiggins, richards, mmgrath@ics.mq.edu.au

Social Network Analysis is the analysis of network data gathered in a social context. It has been used to examine social phenomena in such diverse areas as local communities, international organizations and sporting clubs. However, while the data collected by Social Network Analysts is eminently computable it has taken until relatively recently for SNA to make it's way into the computing field. And while there is almost an abundance of empirical-analysis type programs, the methods for viewing them analytically are currently limited.

Existing SNA visualization software either shows too little detail or has poor functionality, especially when large numbers of nodes are concerned. Generally, producing a meaningful graph of such data can be time consuming. The main problems seem to be the lack of an intelligent method of displaying the graph. Most programs currently try to fit everything on to one screen. As a result, once the number of nodes has passed a critical point the graph becomes crowded and almost useless for extracting information. The confusion of nodes and edges may even imply structures that do not exist! In addition, the clustering algorithms seem to get confused when they have to deal with the edge of the screen resulting in clusters that are represented inappropriately.

To address these issues we plan to investigate two techniques. One technique is to use a more powerful visualization tool that will support scrolling, zooming, alternative layout algorithms and manipulation by users. We are in the process of acquiring the TomSawyer Software Corporation Graph Editor Toolkit for this purpose. We expect that the ability to display the graph over a number of screens and at various levels of abstraction will allow us to navigate around the graph and more clearly see the relationships that exist. The second approach includes the first technique but also involves adding an intelligent filter to the data before it is passed to the graph software. Our plan is to attack the problem from a different angle. If we have a way of removing the concepts that are not interesting or part of our current focus of attention, then the size of the graph and associated display problems will be reduced. The approach is to preprocess the SNA data to discover interesting relationships using machine learning. Initially we are using Quinlan's C4.5 algorithm to identify what features in the data are relevant to a particular classification. We will then display this subset of the data containing less examples with less features. Finding and displaying the most "significant" aspects of the graph while hiding the "unimportant" nodes should prove an effective method of delivering the important information to the user.

To test our approach we will use some undirected adjacency data collected for a separate study. We are currently writing programs to apply C4.5 to the data. Since we have already manually analyzed much of the data in a separate study we will be able to confirm the validity of the results from C4.5. It may be that some interesting and yet undetected pattern may be uncovered. Even if the method does not produce new insights into the data, we expect the smaller and more focused graph to dramatically reduce the time needed to gain a clear and detailed understanding of the data compared to manual inspection of the textual data.