

# A New Interactive Platform for Visual Analytics of Social Networks

Quang Vinh Nguyen<sup>1</sup> and Mao Lin Huang<sup>2</sup>

<sup>1</sup>School of Computing & Mathematics, University of Western Sydney

<sup>2</sup>Faculty of Engineering & Information Technology, University of Technology, Sydney  
vinh@scm.uws.edu.au, maolin@it.uts.edu.au

**Abstract.** This paper presents a new interactive platform for visual analytics of large networks and graphs. The platform integrates multiple graph layouts, interactive navigations and clustering algorithms into an effective and flexible analytical visual environment for better understanding of the nature of variety of different networks. This could lead to the discovery and revealing of hidden structures and relationships among the network items as well as the attributes associated with particular focused elements. We provide a number of interactive navigation and exploration methods so that it can provide a flexible and controllable way to archive the preferable view for analytics. We are extending our visual analytics platform into a large and high-resolution display.

**Keywords:** Information Visualisation, Graph Visualisation, Interaction, Visual Analytics, Network, Visualisation Framework.

## 1 Introduction

Our knowledge-driven technology-mediated environments generate more than 5 exabytes ( $5 \times 10^{18}$  bytes) of new information each year [1]. Much of this information reflects the complex interconnected processes and systems that operate in these environments that span across ecological micro- and macro-systems to interconnected financial markets. A network, such as a social network, can be defined as the information associated with and the relationships between actors which are elements in the network, such as people, groups, corporate organisations, business partners, and others. Considering such systems in isolation is not longer feasible. For the last decade, there has been a rapidly increasing interest in modelling and understanding such systems and processes through network models [2]. Real world networks are often very large with thousands or even millions of elements and relationships. Examples of such networks range from biological and ecological networks through to socio-technological networks, such as telephone call graphs and the underlying interaction graphs on the Internet [3]. These heterogeneous networks present big challenges on how to discover reliable and significant patterns and irregularities within them? How do we holistically make sense of these discoveries and their network models? How do we harness these data at different levels of granularity in collective decision making processes? The challenge - and the fundamental focus of

this research - resides in creating a fast and effective platform in which analysts quickly analyse huge amount of data as well as share their insights and knowledge.

Network analysis brings a lot of benefits in a wide range of application domains, such as economics/finance analysis [4], homeland defence [5], and knowledge management, organisation development, social science, health informatics (see [6]). For example, analysing a terrorism network can help to understand the roles and relational structures among terrorists, their organisations, the associated events and other information that could reveal the potential vulnerabilities and identify any potential attack before it occurs. Another example is that the social network analysis for large cooperate organisations could identify the unofficial organisational roles, such as central connector, boundary spanner, information broker and peripheral specialists. The analysis of social network in organisations can provide business benefits including 1) identify and retention with vital knowledge and connections, 2) improvement of innovation, productivity and responsiveness, 3) lead to better decisions about organisational changes and establishment of key knowledge roles, and 4) insight into challenges of knowledge transfer and integration following restructuring, merges or acquisitions.

Many projects target the analysis of large information spaces, in which data sets are originated from many sources and they can contain millions of items, relations and attributes. Therefore, the scalability, efficiency and speed are the key issues in ensuring the effectiveness of analysis tools. Data is usually available from different sources, different formats and dimensions, the underlying system should provide the ability to access data seamlessly across data sources and formats and ingest the merged data into analysis tools. The seamless data access and data ingest processes should be seamless so that an analyst can perform his/her analysis tasks without concern about the underline data. The visual representation needs to be efficient enough for implementation in interactive large high resolution display systems. It is also a challenge to ensure the quality of large scale automated data analysis.

## 2 Large Network Visual Analytics

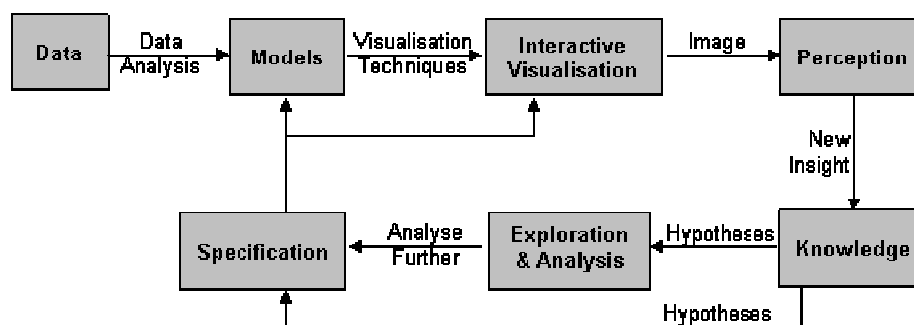


Figure 1. The process sequence of Visual Analytics based on sense-making loop (Adapted from [11])

Large network visualization has been considered as an advanced approach for data analysis that has attracted many research works and projects that attempted to use visualization to enhance the data analysis process. There are a number of good techniques that are quite capable for visual analytics of large graphs or networks of thousands to hundred thousands of nodes and edges, such as *Tulip* [7] and *ASK-GraphView* [8]. A comprehensive classification of analytical network visualisation tools or methods can also be found from the report written by A. Gort and J. Gort [9]. Although there are quite number of available tools for network and graph analysis, research for answering how to effectively integrating interactive visualisations with sense-making sequence (see Figure 1) is still one of the challenges in data analysis.

Visual analytics systems have been introduced to bridge the gap between human visual analysis and automated analysis. Visual analytics is formally defined as the science of analytical reasoning facilitated by the interaction with visual representation of abstract data [10]. Technically, it combines automated analysis techniques (such as data mining, statistical analysis, compression and filtering) with interactive visualisations, analytical reasoning and other related disciplines to achieve effective understanding, reasoning and decision making. Visual analytics can also be seen as an integral approach to decision making by combining interactive visualisation, human-factors and data analysis. Therefore, it is particular suitable for the analysis of large and complex networks and data sets in general. our goal is for creating a platform to enable people to 1) synthesise information and derive insight from massive, dynamic, ambiguous, and often conflicting data, 2) detect the expected and discover the unexpected, 3) provide timely, defensible, and understandable assessments, and 4) communicate assessment effectively for action [11]. As a result, visual analytics has been applied into a number of application domains, such as Homeland Security [12], Visual Analytics at the Boeing Company [13], Foreign Animal and Zoonotic Diseases [14], Finance [15], and many others.

Although some available frameworks, such as *Have Green* [16] and *Universal Visualization Platform* [17], can be used effectively for visual analytics with normal displays, these models or frameworks are insufficient to be extended for large high-resolutions displays. We are looking for an visual platform that ensures the effectiveness of visual analytics process with a large high-resolution display. This platform serves a central medium for data analysis, knowledge discovery and decision making through a combination of automated analysis techniques, interactive visualisation, and analytical reasoning that is based the human sense-making loop (see Figure 1). The future analysis will be a strong connection between central analysts with remote analysts in which the knowledge and insight are concurrently shared and updated among them for better knowledge discovery.

This paper presents a visual analytics platform that combines of different graph layouts, interactive navigations and automated analysis algorithms. This integration of these components makes it flexible for handling a variety of networks and hierarchical structures. The paper is organised as follow. Section 3 presents the architecture of visual analytics tool. Section 4 provides technical details of the interactive navigation and exploration methods. The following section discusses about our current and future work for large and high-resolution displays. The final section summarises and concludes our work.

### **3 The Architecture**

The architecture of our network analysis platform includes three main components: 1) data collection and conversion, 2) automated analysis, 3) visualization and interaction. These components are briefly described as below.

#### **3.1 Collection and conversion**

Raw data is collected from a variety of sources, such as email networks, protein networks, citation networks, and others. This process is responsible for converting the raw data of different formats into uniform XML format for automated data analysis and visualisation. This process is independent to other processes and is operated prior the automated analysis and visualization.

#### **3.2 Automated analysis**

This process analyses the structure of networks and finds out the clusters of similarity and connection based on the chosen algorithms. Network analysis includes concepts and data mining techniques based on degree distributions, centrality measures, clustering, network correlations, random graph models, and models of network evolution (see a summary at [18]). The automated analysis has a tight integration of interactive visualisation and related disciplines.

At this stage, graph clustering algorithms are used to discover the community structures embedded in a large graph and then divide them into densely connected sub-graphs. We use a fast graph clustering algorithm in which algorithm 1 can also achieve a consistent partitioning result in which a graph is divided into a set of clusters of the similar size [19], and algorithm 2 is used to discover communities with various size [20]. The clustering process works independently with the visualisation so that the findings can be updated as well as receiving feedback from visualisation.

#### **3.3 Visualisation and Interaction**

The interactive visualisation is more than just an interface. It provides users with an intelligent and appropriate display and users can also view, interact, and select different layout methods through the visualization. Interactive visualisation techniques usually perform well on particular application domains with similar data sets. Therefore, no particular technique is always superior to another. Intelligent interactive visualisation is one of the important approaches that takes the advantage of optimized algorithms by providing appropriate layouts and interactions based on the nature of applications, data sets and analysts' preferences.

A combination of several layout algorithms is employed to draw different clusters. We aim to optimize the utilization of geometrical space and so that a large network

with multiple clusters can be drawn in a normal screen size. This interactive interface is able to not only show the abstract view of the entire network with its clustering property and relational structures but also display the detail of any particular group, item and their associate connections and properties. We use rich graphic attributes to assist viewers to quickly identify the domain specific properties associated with data items.

The visualisation consists of a collection of multiple modified layout algorithms, such as *Encccon* [20], *Spring Embedder* [21], *Circular Drawing*, and the combination of multiple algorithms on one layout [22]. These algorithms are employed automatically or are selected manually by user to provide the optimal layout for the large network. Refining processes, including layout adjustment and node overlapping remove, are also applied onto the display to enhance the clarity, aesthetical niceness and cluster properties of drawings. At any stage, user can switch between different layouts and vice versa to obtain the optimal and/or preferable displays. Figure 2 shows the layout of the same network using different algorithms in our system. Figure 3 shows a multi-scale layout and display using *Encccon* and *Spring Embedder* algorithms.

In the future, we aim to provide an intelligent visualization that can produce the most suitable displays with corresponding layouts and interaction mechanism based on the nature of applications, data sets, analyst preferences and underline sense-making rules. The visualisation can be manually adjusted via user's feedbacks during the interaction.

The challenge lays on the creation an effective mechanism to integrate different interactive visualisation techniques together so that the visualization can provide a best-possible display to satisfy analysts' preference and observation styles. During the navigation, we allow users to interactively adjust views and layouts to reach an optimized representation of the network; from which users can obtain the best understanding of the data and structures they are observing. The visualization consists of real time human-computer interactions, fast graph layout and navigation algorithms for visual processing of hundred thousands of items within minutes or seconds using a personal computer with limited display space and computational power.

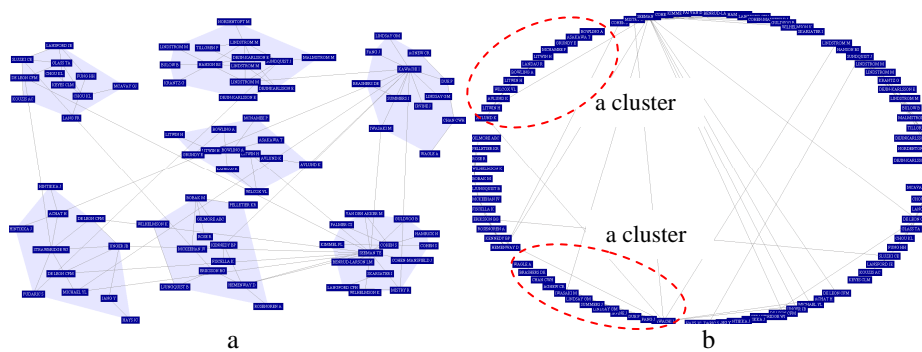
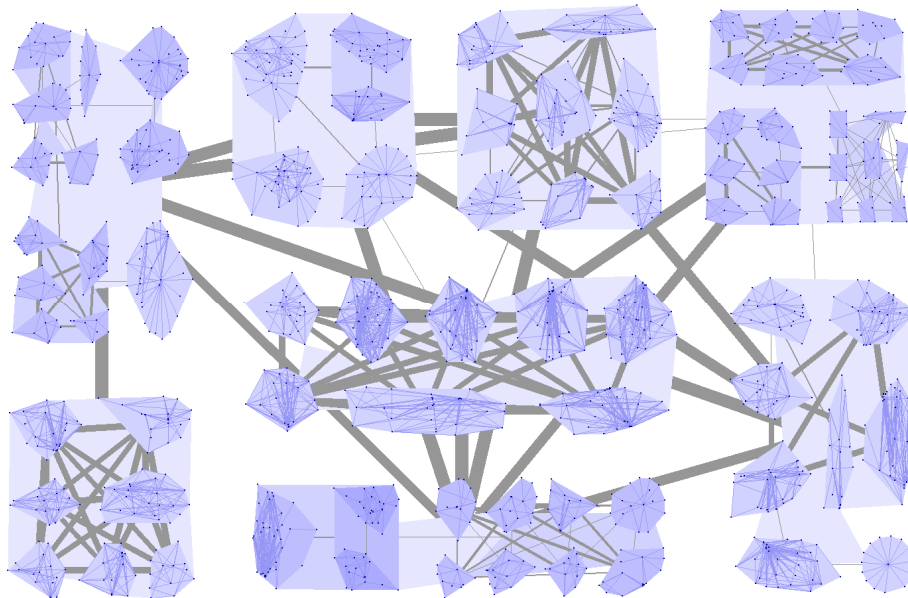
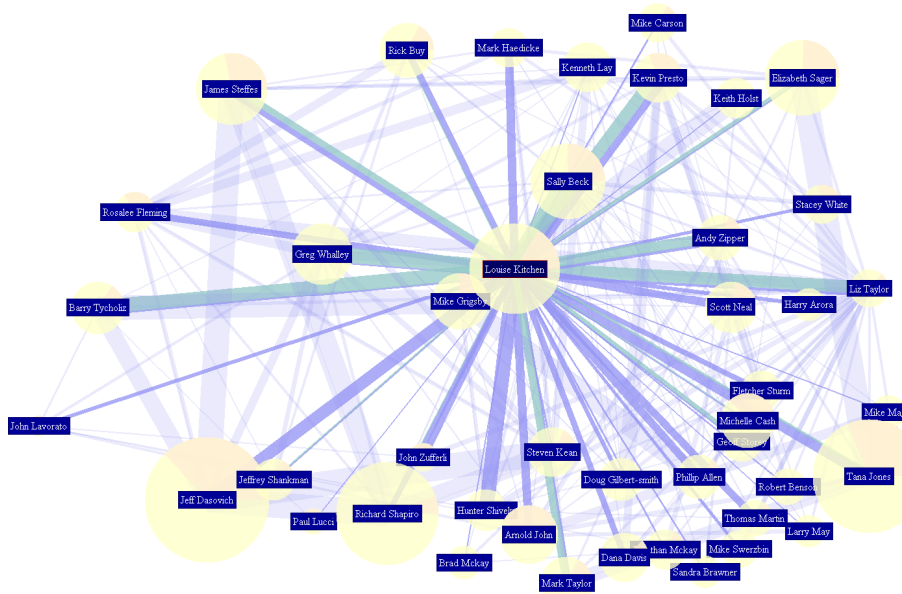


Figure 2. An example of two graph layouts of the same network in which a) using a Spring Embedder with Encccon partitioning, and b) using a circular layout algorithm.



a



b

Figure 3. An example of the visualizations of different networks where a) shows the abstract view of a hierarchical graph and b) shows the detail view associate with an item.

## 4 Interactive Navigation and Exploration

There is no pure layout technique that could assist data retrieval without providing users with an associate interactive navigation mechanism in graphic user interface design. In our visualisation, during the navigation user can interactively adjust the views to reach an optimised representation of the graph; from which user can obtain the best understanding of the data and its relational structures he/she are currently interesting in. The navigation is an integration of various viewing and interaction techniques so that it can provide a flexible and controllable solution to archive the preferable view. Details of these techniques are further described.

### 4.1 Chain-Context View

We applied an efficient interaction method, Chain-Context View [23] for the navigation of large networks. Instead of displaying a single or small number of context views as in traditional approaches, we provide users with a display of a progressive sequence of context views attempting to maximize the displayed contextual information. The display of rich context information produced in the exploration path could greatly increase the accuracy of user decisions and reduce the unsuccessful trips and unnecessary views during visual exploration of the large structures.

Conceptually, Chain-Context View allows users to trace each step of their interaction and makes it easy for them to jump or return to any level of the hierarchy they have already visited. This method maintains user's orientation during exploration by providing multiple navigational views. This also enables users to explore large graphs by arbitrary paths while moving towards the target view. In comparison with traditional approaches which only allow sequential navigation; our proposed navigation scheme is believed to be more efficient.

Technically, the entire display area is initially divided into two regions: 1) a small area in the left-hand-side for displaying all context-views and 2) a large area in the right-hand-side for the main-view (or focus view).

The main-view - displays a current focus sub-structure or the complete structure. This view occupies the main portion of the entire display. By showing the detail of the focusing structure, users can inspect particular data items, analyse the structure for taking further actions.

The context-views - display inside a small rectangular region with an efficient partitioning. The display of an intermediate chain of context views during the navigation also enables users to choose an arbitrary path for reaching the target quickly and efficiently. While the chain of context views is responsible to display contextual information, it should not occupy the major portion of display space. Some details of the substructures displayed in the context views are filtered through visual abstraction. Only the main structures remain for guiding navigation. The abstracted context views aim to maintain users' orientation of where they are, where they came from, and where they have been during visual exploration.

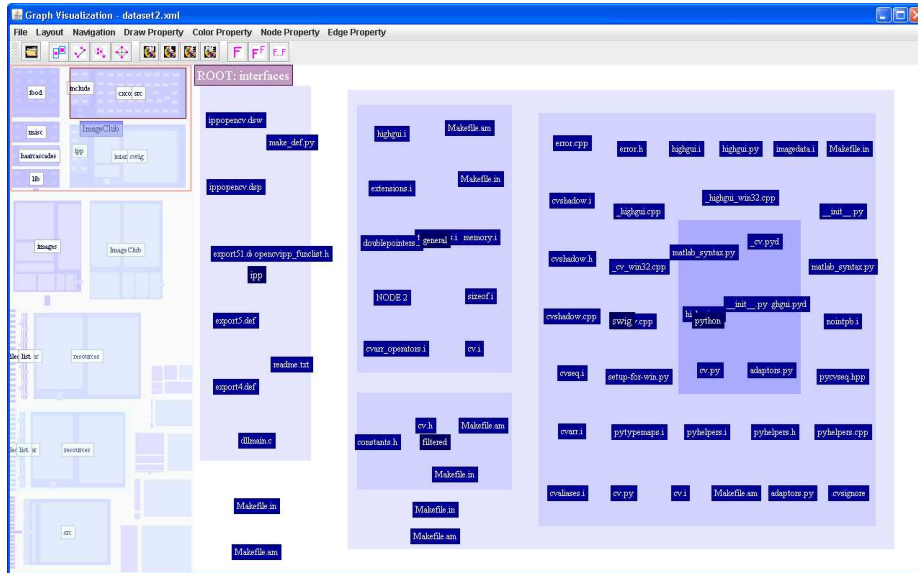


Figure 4. An example of the chain-context views when browsing a large file system. The current views are displayed on the small left-hand-side and the main-view is displayed a current navigation point on the large right-hand-side. The current mode is *highlight-and-select* method in which the first context view is highlight.

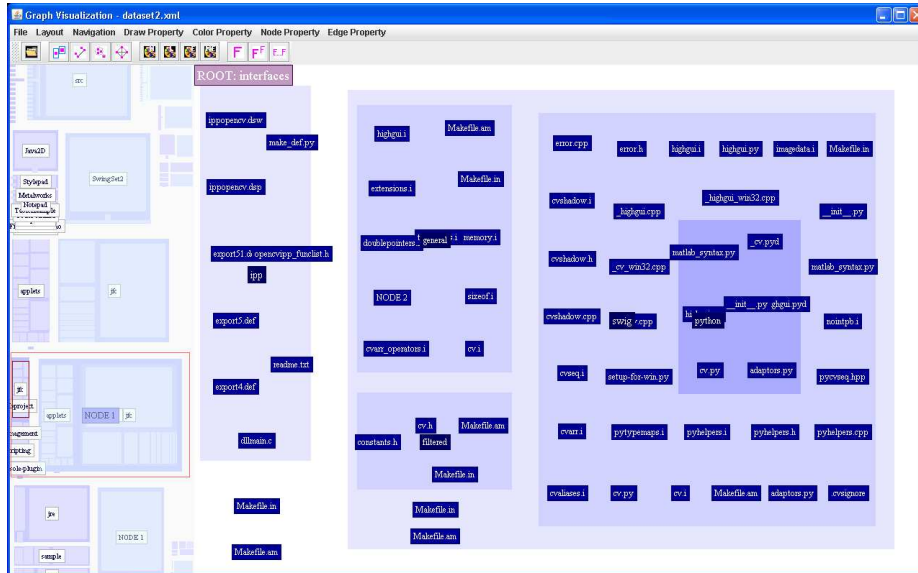


Figure 5. An example of *fisheye-browsing* of chain-context views using the same data set as Figure 4. This indicate the focus point is at the context view whose root is "NODE 1".



At any stage, users can interact with context views either using highlight-and-select or fisheye-browsing modes. Highlight-and-select includes two steps: 1) highlight and enlarge a context view in which more information can be displayed and 2) select the context to transform to the main-view. The highlight and enlarge step is aimed to provide a quick analysis at context views without showing these views at the focus panel (see Figure 4).

Fisheye-browsing enables the enlargement of context views dynamically using a modified fisheye distortion. In contrast to the highlight-and-select approach, in which the size of view boxes is decreased in order, the size of context views is now dependant on the position of the mouse pointer. The focus view is first defined at a current mouse pointer and then users can scroll down and up to change the focus point, and the position and size of all context views in the chain are then adjusted accordingly. We eliminate the distortion by making the uniform size at a particular context view (see Figure 5).

## 4.2 Group Selection

Group selection is the ability that a user can select a group of nodes for a quick analysis without retrieving all sub-networks that associating with these selected nodes. This technique is useful when the user want to inspect or analyse just a group of nodes. When a number of nodes in the *main-view* are selected, the program will recalculate the new layout of selected nodes and then display the main-view as a new structure (see Figure 6).

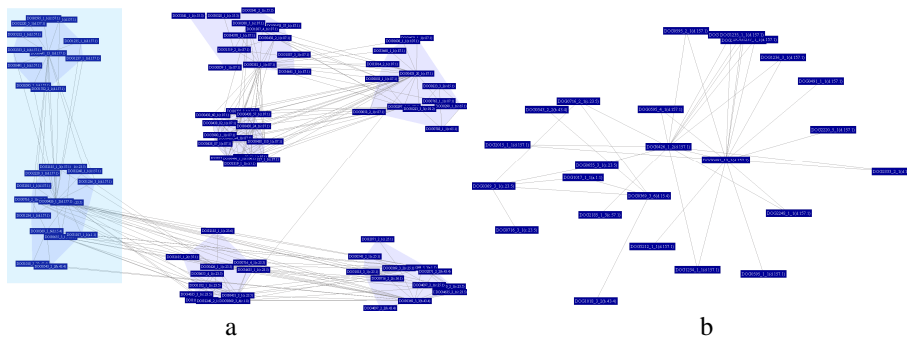


Figure 6. An example of group selection interaction. a) shows the network when the user is selecting a group of nodes at the left side and b) shows the corresponding visualisation for those selected nodes.

## 4.3 Node Centralised Selection

Node centralised selection is the ability that allows users to select one or more nodes for further analysis on sub-networks that are associated with these selected nodes.

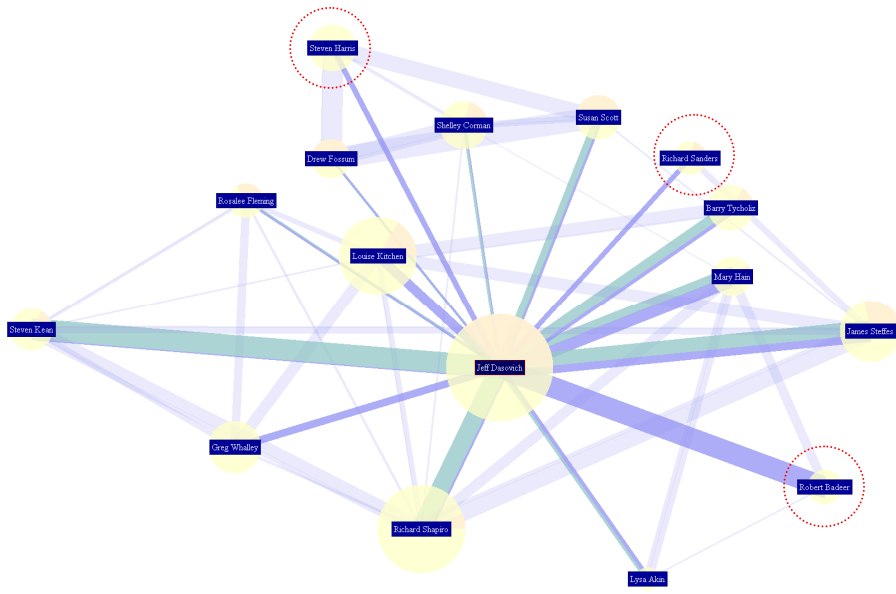


Figure 7. An example of node centralized selection for a particular node. The visualization of the sub-network associated with the focus node “Jeff Dasovich” is selected from the larger network for deep analysis.

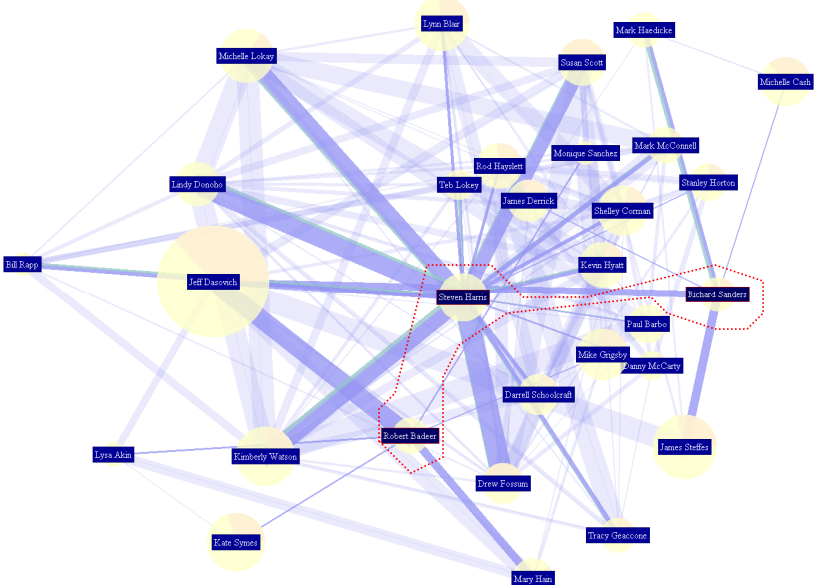


Figure 8. Another example of node centralized selection of a group of nodes. The visualization of the sub-network associated with three focused nodes: “Steven Harris”, “Richard Sanders” and “Robert Badeer”. These three nodes are red-circled that are selected from Figure 8.

This technique is particularly useful when the user wants to have a deep analysis on one or more particular nodes. Technically, when a number of nodes at the *main-view* are selected, the program will first find out the nodes from the entire network that link with the selected nodes. A layout algorithm is then applied to calculate the new layout of the sub-network. Rich graphics are also used to enhance the representation in which emphasised focus nodes and edges are highlighted while unfocused nodes and edges are less highlighted. For example, the selected nodes are drawn with brighter colours and the edges among un-selected nodes are painted with dimmer and transparent colours. Edges that are not associated with selected nodes are also painted in a de-emphasised manner. Figure 7 and Figure 8 show examples of node centralised selection of one or multiple persons respectively of an email network.

### 4.3 Interactive Visual Clustering

Interactive Visual Clustering is a process that enables users to create and delete clusters and reform the clustering trees in the whole network.

In the clustering process, the whole visualization is fully animated. Every change at one of the lower layers, whether triggered by the user or by another agent, is animated at the picture layer. The use of multiple animations may reduce the cognitive effort of the user in recognizing change of the clustering structures. This preserves the user's "mental map" [24]. There are six types of animation that are implemented in our system. These are detailed below.

- *Force animation* - The force model above is used in the animation of all changes to picture. The forces move the nodes toward a minimum energy state, and this movement gives the sequence  $D_1; D_2; \dots ; D_k$  of drawings to smoothly change the layout from a drawing to the next. Overall, the animation driven by the forces is the most important mechanism for the smooth transition between changes at the layers lower than the picture layer.
- *Animated addition and deletion of nodes and edges* - When a node is deleted or added to the drawing, we use animated *shrinking* or *growing* to help the user identify nodes that are disappearing or appearing. Disappearing nodes lose their connecting edges, then shrink until they disappear. Appearing nodes are displayed in a reduced size with their edges, then grow to be full-sized nodes. This avoids the sudden disappearing/appearing of node images on the screen, which can disturb the user's mental map.
- *Animated cluster regions* - Nodes move according to the force model, and the rectangular regions representing clusters follow the nodes. At all times, each cluster is the minimum enclosing rectangle of its children. These rectangles move smoothly as the positions of the children change. This is especially important for the gathering operation; the clusters separate smoothly.
- *Animated scaling* - Animated shrinking scales down a picture if it gets too large for the screen. Further, animated enlarging is used to increase the size of the picture; this enables the user to see details and makes direct

manipulation operations of the picture easier. In fact, animated scaling is implemented by simply changing the forces in the springs; for example, to shrink the whole picture, every spring is increased in strength.

- *Animated cluster closing/opening* - When closing a cluster, we firstly use animation to reduce the size of closed region of this cluster. As soon as the size reaches a certain threshold layer we smoothly replace the representation of the cluster from its opened form (a red line bounded rectangle) to its closed form (a small black line bounded rectangle). When opening a cluster, we firstly smoothly replace the representation of the cluster from its closed form to its opened form. The children and relevant links of the cluster are smoothly added into the layout. Then we smoothly enlarge the size of this cluster to its normal size.
- *Camera animation* - Camera animation moves the whole drawing. It is optional. It can be used, for example, to move specific nodes of interest to the center of the screen.

All these animations operate in parallel.

## 5 Future Work - Large and High-Resolution Displays

With technological advances, large high-resolution (or wall-size) displays are becoming more and more available in many fields. From multi-monitor configuration to tiled LCD panels to projection-based seamless displays, researchers have been constructing large displays with various hardware configurations. A large scientific system is able to produce a high resolution display of ten million pixels to even hundreds of million pixels, such as the near 287 million pixels display of *HIPerSpace* [25]. Two common features of such displays are increasing physical size and higher resolution in which a wall-size display can display easily large contextual and detail information. Therefore, it is particularly suitable for serving as a medium for presenting, capturing, exchanging ideas and interaction in supporting visual analytics of very large and complex information spaces (see an example at Figure 9).

We are currently extending our visual analytics system for large and high-resolution displays. This adoption of our future work aims to provide 1) a more effective visual analytics framework, 2) more intelligent visualisations and 3) tight integration of automated analysis and interactive visualisations.

The new framework required to ensure the effectiveness of the use of large-high resolution displays so that it can handle a vast amount of information in a fast speed. This platform serves a central medium for data analysis, knowledge discovery and decision making through a combination of automated analysis techniques, interactive visualisation, analytical reasoning based our human sense-making loop (see Figure 1). There should be a strong connection between central analysts with remote analysts in which the knowledge and insight are concurrently shared and updated among them for better knowledge discovery. We will also investigate a more effective integration of automated analysis and interactive visualisations. The integration is not only at intelligent visualisation but also at general context of automated analysis and interactive visualisations.

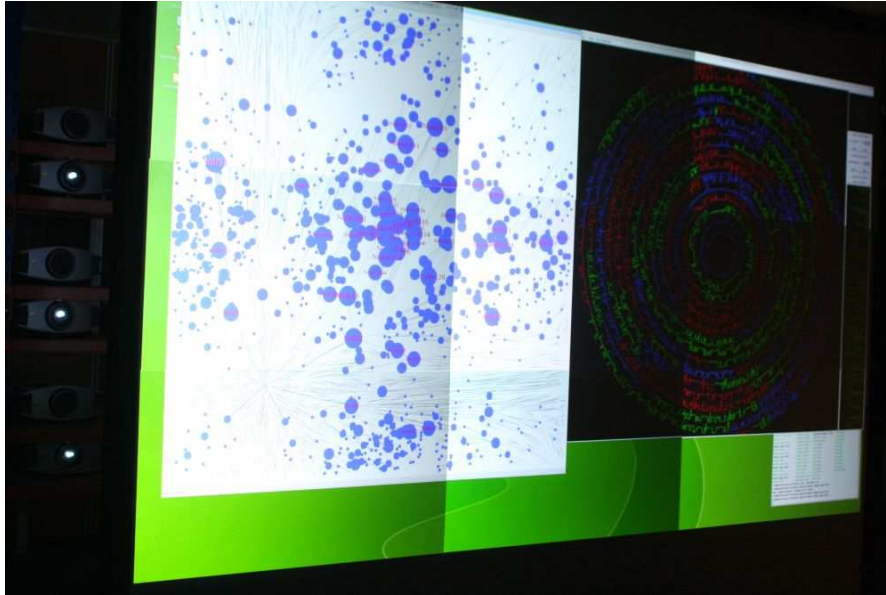


Figure 9. An example of multiple visualizations shown with large and high-resolution displays.

## 6 Conclusion

We have presented a new visual analytics tool for large networks and/or graphs analysis. Our system is a tight integration of different modules: data conversion, automated analysis and interactive visualisation. Each module is a combination of multiple algorithms so that an algorithm (or combination) can be selected to archive the optimal result. Although the work is still on progress, we believe that our work will be a valuable tool for the visual analytics of large networks and/or graphs. We will next extend our tool for large and high-resolution displays.

## References

1. Lyman, P., Varian H.R.: How Much Information? University of California at Berkeley, <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>
2. Newman, M. Watts, D., Barabási A.L.: The Structure and Dynamics of Networks, Princeton University Press (2006).
3. Almaas, E. and Barabási A.L.: The Architecture of Complexity: From the WWW to Cellular Metabolism. In: Dynamics of Complex Interconnected Systems: Networks and Bioprocesses, NATO Sciences Series (2006).
4. Malhotra, Y.: Enabling Knowledge Exchanges for E-Business Communities. Information Strategy: The Executive's Journal, pp. 26-31 (2005).

5. Ressler, S.: Social Network Analysis as an Approach to Combat Terrorism: Past, Present, and Future Research. *Homeland Security Affairs*, 2(2) (2006).
6. Patton, S.: Social Network Analysis Helps Maximize Collective Smarts. <http://www.cio.com/article/print/6956>.
7. Auber D.: Tulip: a Huge Graph Visualization Framework, *Graph Drawing Software. Mathematics and Visualization*, pp. 105-126 (2003).
8. Abello J., van Ham F., Krishnan N.: ASK-GraphView: A Large Scale Graph Visualization System. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006.
9. Gort, A., Gort, James.: Survey of Network Visualization Tools. Contract Report, DRDC Ottawa CR 2007-280, Defence R&D Canada, (2007).
10. Thomas, J.J., Cook, K.A.: *Illuminating the Path*. IEEE Computer Society Press, Los Alamitos (2005).
11. Keim D. et al: Visual Analytics: Definition, Process, and Challenges *Information Visualization*. LNCS 4950, pp. 154–175 (2008).
12. National Visualization and Analytics Center (NVAC), <http://nvac.pnl.gov>.
13. Kao, A.: Visual Analytics at the Boeing Company. VAC Consortium, (2008).
14. Clarke, N. and Nipp T.: Foreign Animal and Zoonotic Diseases: An Information Analysis Framework. VAC Consortium (2007).
15. Ribarsky, W.: Discovery Suspicious Financial Activity. VAC Consortium (2007).
16. Wong, P. C., Chin Jr, G., Foote, H., Mackey, P., Thomas J.: Have Green – A Visual Analytics Framework for Large Semantic Graphs. In: *IEEE Symposium on Visual Analytics Science and Technology (VAST2006)*, pp. 67-74 (2006).
17. Gee, A. G. et al. Universal visualization platform. In: *SPIE* (2005).
18. Newman, M.E.J.: The Structure and Function of Complex Networks. *SIAM Review*, 45, pp. 167-256 (2003).
18. Huang M.L. and Nguyen Q.V.: A Fast Algorithm for Balanced Graph Clustering. In: *11th Int'l. Conf. on Information Visualisation (IV'07)*, IEEE, pp. 46-52 (2007).
19. Newman M.E.J., Girvan M.: Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69:026113 (2004).
20. Nguyen, Q. V. and Huang, M. L.: EncCon: An Approach to Constructing Interactive Visualization of Large Hierarchical Data. *Information Visualization Journal*, Palgrave, 4(1), pp. 1-21 (2005).
21. Eades, P.: A Heuristic for Graph Drawing. *Congressus Numerantium*, 42, pp. 149-60 (1984).
22. Huang, M. L. and Nguyen, Q. V.: Large Graph Visualization by Hierarchical Clustering. *Journal of Software*, 19(8), pp. 1933-1946 (2008).
23. Huang, M. L., Nguyen, Q. V.: Space-Filling Interaction with Chain-Context View. In: *IEEE Symposium on Information Visualization 2008 (InfoVis 2008) - Poster Section*, IEEE, Columbus, Ohio (2008).
24. Eades, P., Lai, W., Misue, K. and Sugiyama, K.: Preserving the mental map of a diagram. In: *Compugraphics 91*, pp. 24-33 (1991).
25. Ramsey, D. Highest-Resolution Scientific Display System. <http://www.spafi.org>.