Gephi : An Open Source Software for Exploring and Manipulating Networks

Mathieu Bastian and **Sebastien Heymann**
Gephi, Webatlas
Paris, France
{mathieu.bastian, sebastien.heymann}@gephi.org

Mathieu Jacomy
Webatlas founding member
R&D at TIC-Migrations program in
Fondation Maison des Sciences de l'Homme
mathieu.jacomy@gmail.com

Abstract

Gephi is an open source software for graph and network analysis. It uses a 3D render engine to display large networks in real-time and to speed up the exploration. A flexible and multi-task architecture brings new possibilities to work with complex data sets and produce valuable visual results. We present several key features of Gephi in the context of interactive exploration and interpretation of networks. It provides easy and broad access to network data and allows for spatializing, filtering, navigating, manipulating and clustering. Finally, by presenting dynamic features of Gephi, we highlight key aspects of dynamic network visualization.

Visualization and Exploration of Large Graphs

In the aim of understanding networks, the visualization of large graphs has been developed for many years in many successful projects (Batagelj 1998; Shannon 2003; Adar 2006). Visualizations are useful to leverage the perceptual abilities of humans to find features in network structure and data. However this process is inherently difficult and requires exploration strategy (Perer 2006). As well as being technically accurate and visually attractive, network exploration tools must head toward real-time visualizations and analysis to improve the user's exploratory process. Interactive techniques have successfully guided domain experts through the complex exploration of large networks.

We can identify some main requirements for a network exploration tool: high quality layout algorithms, data filtering, clustering, statistics and annotation. In practice these requirements must be included in a flexible, scalable and user-friendly software. Focusing on analysis clarity and on modern user interface, the Gephi project brings better network visualization to both experts and uninitiated audience. Inspired by WYSIWYG editors like Adobe Photoshop, we develop modules set around a center visualization window.

The Gephi Software

Gephi is an open source network exploration and manipulation software. Developed modules can import, visualize, spatialize, filter, manipulate and export all types of networks. The visualization module uses a special 3D render engine to render graphs in real-time. This technique uses the computer graphic card, as video games do, and leaves the CPU free for other computing. It can deal with large network (i.e. over 20,000 nodes) and, because it is built on a multi-task model, it takes advantage of multi-core processors. Node design can be personalized, instead of a classical shape it can be a texture, a panel or a photo. Highly configurable layout algorithms can be run in real-time on the graph window. For instance speed, gravity, repulsion, auto-stabilize, inertia or size-adjust are real-time settings of the *Force Atlas* algorithm, a special force-directed algorithm developed by our team. Several algorithms can be run in the same time, in separate workspaces without blocking the user interface. The text module can show labels on the visualization window from any data attribute associated to nodes. A special algorithm named *Label Adjust* can be run to avoid label overlapping (Figure 1).

Figure 1: Label Adjust algorithm avoid label overlapping

The user interface (Figure 2) is structured into Workspaces, where separate work can be done, and a powerful plugin system is currently developed. Great attention has been taken to the extensibility of the software. An algorithm, filter or tool can be easily added to the program, with little programming experience. Sets of nodes or edges can be obtained manually or by using the filter system. Filters can select nodes or edges with thresholds, range and other properties. In practice filter boxes are chained, each box take in input the output of the upper box. Thus, it is easy to divide a bi-partite network or to get the nodes that have an in-degree superior to 5 and the property "type" set to "1". Because the usefulness of a network analysis often comes from the data associated to nodes/edges, ordering and clustering can be processed according to these values. With

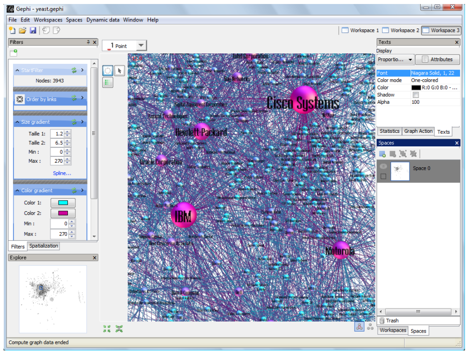

Figure 2: A screenshot of Gephi beta version 0.6

sets of nodes, graphical modules like *Size Gradient*, *Color Gradient* or *Color clusters* can then be applied to change the network design. Graphical modules take a set of nodes in input and modify the display parameters, like colors or size, to corroborate understanding of the network structure or content.

Though networks can be explored in an interactive way with the visualization module, it can also be exported as a SVG or PDF file. The vectorial files can then be shared or printed. A powerful SVG exporter named *Rich SVG Export* is included in Gephi. Many options are offered to users to set the design of nodes, edges and labels. Techniques are developed to increase networks clarity and readability. Special care is taken about fonts and labels. For instance, small labels can be drawn on edges to immediately see the neighbours of a node. The Figure 3 shows the brain network of the C. Elegans worm (Watts 1998) exported from Gephi.

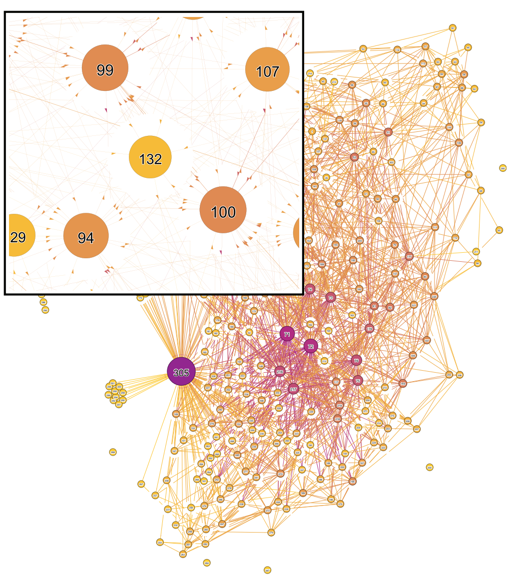

Figure 3: SVG File exported from Gephi

The current studies of network dynamics has brought some very interesting case study. Dynamic network visualization offer possibilities to understand structure transition or content propagation (Moody 2005). Exploring dynamic networks in an easy and intuitive way has been incorporated in Gephi from the beginning. The architecture supports graphs whose structure or content varies over time, and propose a timeline component where a slice of the network can be retrieved. From the time range of the timeline slice, the system queries all nodes and edges that match and update the visualization module. Hence a dynamic network can be played as movie sequences.

The dynamic module can get network data from either a compatible graph file or from external data sources. When running, a data source can send network data to the dynamic controller at any time and immediately see the results on the visualization module. For instance a web-crawler can be connected to Gephi in order to see the network construction over time. The architecture is interoperable and data source can be created easily to communicate with existing software, third parties databases or web-services.

Future work

Though the core of the software already exists, further work is required for the development of new features, especially filters, statitics and tolls. A special focus is made on clustering and hierarchical networks. Improvements will be integrated to the data structure to support grouping and navigation within a network hierarchy. As for spatialization algorithms, a framework will be able to host various classification algorithms.

As we continue to receive feedbacks, we are looking forward to better adapt the user interface to users' need. Gephi has been successfully used for Internet link and semantic network case studies. It is also frequently used for SNA. An effort has been made to speed up the analysis process, from data import to map export. Gephi is developed toward supporting the whole process with only user interface manipulation. The developement of dynamic features are also one of the top priorities.

Availability

Gephi is available at http://gephi.org

References

Adar, E. 2006. Guess: a language and interface for graph exploration. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, 791–800.

Batagelj, M. 1998. Pajek - program for large network analysis. *Connections* 21:47–57.

Moody, McFarland, B.-d. 2005. Dynamic network visualization. *American Journal of Sociology* 110(4):1206–1241.

Perer, S. 2006. Balancing systematic and flexible exploration of social networks. *Visualization and Computer Graphics, IEEE Transactions on* 12(5):693–700.

Shannon, Markiel, O. e. a. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.

Watts, S. 1998. Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442.