# Interactive Visual Graph Analytics on the Web

**Nesreen K. Ahmed**
Purdue University
nkahmed@purdue.edu

**Ryan A. Rossi**
Purdue University
rrossi@purdue.edu

## Abstract

We present a *web-based network visual analytics* platform called GRAPHVIS that combines interactive visualizations with analytic techniques to reveal important patterns and insights for sense making, reasoning, and decision-making. The platform is designed with simplicity in mind and allows users to visualize and explore networks in seconds with a simple drag-and-drop of a graph file into the web browser. GRAPHVIS is fast and flexible, web-based, requires no installation, while supporting a wide range of graph formats as well as state-of-the-art visualization and analytic techniques. In particular, the multi-level network analysis engine of GRAPHVIS gives rise to a variety of new possibilities for exploring, analyzing, and understanding complex networks interactively in real-time. Finally, we also highlight other key aspects including filtering, querying, ranking, manipulating, exporting, partitioning (community/role discovery), as well as tools for dynamic network analysis and visualization, interactive graph generators (including two new block model approaches), and a variety of multi-level network analysis and statistical techniques.

## Introduction

Network analysis and graph mining play a prominent role in providing insights and studying phenomena across various domains, including social, behavioral, biological, transportation, entertainment, and financial domains. This paper presents a *web-based network visual analytics* platform called GRAPHVIS (http://graphvis.com) that integrates powerful statistical analysis, graph mining, and machine learning techniques with interactive visualization to aid in the discovery of important patterns and insights for sense making, reasoning, and decision-making.

GRAPHVIS is a visual analytics tool for network data, designed for rapid interactive visual exploration and graph mining (Figure 1). Starting from a simple drag-and-drop of a graph file into the browser, users can move from data to insights within seconds. Unlike other network visualization software (such as Gephi (Bastian et al. 2009) and Tulip (Auber 2004)) that requires installation and updates,

GRAPHVIS is web-based working directly from the browser. Furthermore, another key difference is that GRAPHVIS is designed to be consistent with the way humans learn via *immediate-feedback* upon every user interaction (e.g., change of a slider for filtering) (Ahlberg, Williamson, and Shneiderman 1992; Thomas and Cook 2005; Rossi and Ahmed 2015a). Thus, users have rapid, incremental, and reversible control over all graph queries with *immediate* and *continuous visual feedback*.

While GRAPHVIS serves as a flexible *web-based* platform for *visual interactive graph mining*, it also has a number of features that differentiate it from other offline visualization tools. We summarize a few of the important and differentiating features below[1].

- Drag-and-drop graph file(s) to quickly visualize and interactively explore networks in seconds.

- Support for a wide variety of graph formats such as edge-lists (txt, csv, tsv, mtx, etc), XML-based formats (gexf, graphml), and a variety of others (gml, json, net/pajek).

- Online profile and system to help manage your data, settings, visualizations, etc.

---

[1]Note that while some of the features below seemingly overlap with those found in other visualization platforms, the key difference is the simplicity and degree of interactivity provided, as the features in GRAPHVIS are designed for rapid interactive visual feedback after each human interaction/event.
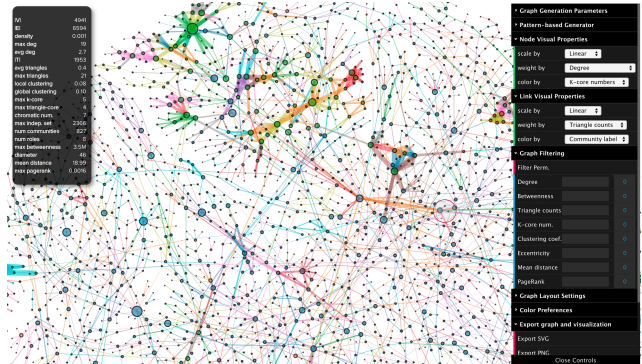


Figure 1: Screenshot of GRAPHVIS

- Export visualizations as high-quality images (SVG, PNG) as well as (transformed/filtered) graph data, attributes, etc (Figure 2).

- All macro and microscopic graph properties and statistics (triangles, graphlets, kcore, etc.) are automatically updated in an efficient manner after each graph manipulation is performed such as inserting and deleting nodes and links.

- Mouseover nodes to analyze their microscopic properties (e.g., betweenness, PageRank, number of triangles).

- Real-time visual graph filtering and querying capabilities including chaining multiple filters as well as exporting the resulting data for other tasks.

- To aid the analytic process, both nodes and edges may be *colored* and *sized* according to a variety of network properties (k-core number, eccentricity, etc) or customized by the user.

- Subgraphs may be selected by brushing over interesting regions of the network visually (i.e., click-and-drag while holding shift). Multiple selections from different regions of the graph are also supported.

- Brushed nodes, edges, and subgraphs may be deleted (by pressing 'D'), or manually adjusted by clicking on a selected node and dragging to the desired location.

- Interactive graph generation including three newly proposed block model approaches that capture community structure.

- Nodes, edges, and subgraph patterns (cliques, stars, chains, cycles, as well as probabilistic patterns via block models, etc) are also easily added with a simple click to allow for easy exploration, hypothesis testing (e.g., what if scenarios, as well as simulations to understand the impact of insertions/deletion on the local and global network), or simply to quickly generate a benchmark network for testing correctness of an algorithm.

- Tools for dynamic network analysis and visualization.

- Powerful graph mining and learning techniques are developed including *interactive visual role discovery* as well as advanced network analysis methods such as *triangle-core decomposition*.

- Node information may also be updated easily via double-clicking the node.

- Nearly all visualizations are interactive and support brushing, linking, zooming, panning, tooltips, etc.

- Multiple visual representations of the graph data are supported, including the multi-level graph properties (e.g., interactive scatter plot matrix, and other statistical plots).

- Network may also be searched via textual query (e.g., node name).

- There are many other features including full customization of the visualization (color, size, opacity, background, fonts, etc), text annotation, graph layouts, collision detection, fish eye, and many others.
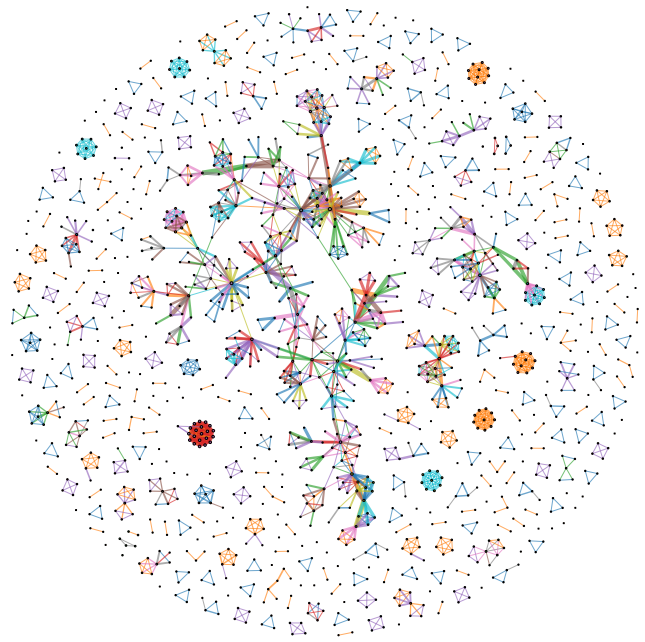


Figure 2: Exported SVG image from GRAPHVIS

## Interactive Multi-level Visual Graph Analytics

Visual analytic tools need to allow for interacting and reasoning across multiple simultaneous scales of data representations (Ebert et al. 2014). Thus, we developed GRAPHVIS with a multi-scale visual analytics engine to support (visual) interactive network exploratory analysis at both the global macro-level as well as the local microscopic level. Visual graph mining and machine learning techniques lie at the heart of GRAPHVIS and provide the analysts with a set of powerful tools to discover key insights and reveal important structural patterns interactively in real-time. Such an approach is vital for interactively exploring big data in real-time by summarizing its patterns, statistics (binning, distributions, etc), as well as spotting anomalies. Statistical techniques are used to find interesting nodes, allowing the user to sort through the top-k most interesting nodes for further investigation.

Every update, insertion, or deletion of a node, edge, or subgraph is immediately reflected in the visualization window. Furthermore, the visualization and analytics are also updated immediately upon any parameter change via sliders or other interface controls. This allows to quickly test a hypothesis as well as investigate the impact of certain actions on the network structure and its properties/statistics. For instance, suppose we use betweenness to filter the graph, as we adjust the slider, the analyst receives visual feedback immediately at each change in the slider (in contrast to adjusting the slider to the desired value, then receiving feedback on the selection[2].

---

[2]This type of visual network analysis is also extremely useful for learning and education, since students can quickly grasp the behavior of the statistical techniques through interactive exploration.
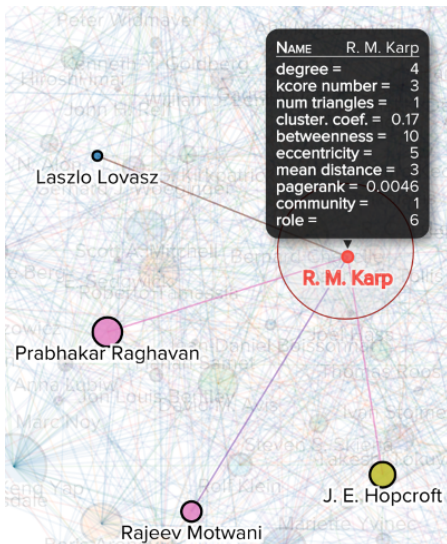
Figure 3: Node profile tools highlight important properties including microscopic and neighborhood information

## Macro-level Interactive Graph Analysis

At the macroscopic level, we use a variety of key network properties. A few of these include max/mean degree, total triangles, global clustering, max k-core number, diameter, mean distance, approx. chromatic number, number of communities/roles, and max triangle-core number. To help guide the interactive exploration, we display many of the important macro properties that help characterize the global structure of the network in the visualization window.

Moreover, statistical aggregates (mean, max, mode, sum, var) are used to summarize the global structure and behavior of the network.

## Micro-level Interactive Node & Edge Analysis

To facilitate the discovery process, GRAPHVIS provides interactive exploration at the microscopic level, e.g., using edge and node degree, eccentricity, k-core number, and triangle-core number. In addition, many important social network analysis measures are used in the interactive visual analytics including betweenness, number of triangles, clustering coefficient, path lengths, PageRank, and many others.

These node and edge properties are displayed in visual form and can be explored/manipulated directly by the user in a free-flowing manner (e.g., using brushing, linking, zooming, mouseover, filtering, etc). For instance, the neighborhood of a node can be highlighted as well as its micro-level statistics and properties (Figure 3).

Multiple visual representations of the graph data are also provided. For example, GRAPHVIS leverages an interactive scatter plot matrix for analysis of the correlation between pairs of node statistics (Figure 4). We also support brushing to allow users to highlight interesting nodes (and links) across the various measures. Furthermore, semantic zooming can be used to drill-down in order to understand the differences between individual nodes and links.

Links may also be analyzed more closely using a similar approach. A few important link measures are available including triangle-core numbers, number of triangles incident to an edge, as well as a variety of other measures.

## Graph Partitioning Methods

GRAPHVIS provides a diverse collection of visual interactive graph partitioning methods. For example, community detection, role discovery (Rossi and Ahmed 2015b), and graph coloring. All graph partitioning methods are designed to be efficient taking at most linear time in the number of edges to compute.

## Distributions of Measures

Node and link summarization techniques (e.g., binning/histograms, statistical distributions) are used to obtain fast, meaningful and useful data representations. For example, we leverage binning methods to interactively compute and maintain the frequency distribution of some graph properties (e.g., degree distribution) upon any update, insertion, or deletion of a node, edge, or subgraph. Furthermore, we also interactively plot the cumulative distribution function (CDF) and the complementary CDF, which are easily computed from the frequency distribution. These are known to be important for networks, capturing interesting structural properties (e.g., heavy-tailed distributions). Furthermore, we also utilize sampling (Ahmed et al. 2014) as well as fast ranking algorithms for displaying top-k nodes, links, and subgraphs to the user for further exploration. In addition to distributions, the macro-level measures are also useful for big graph data and vital to the multi-level strategy offered by GRAPHVIS.

## Interactive Graph Generation

Graph generators are useful for simulations, testing algorithms, assumptions, benchmarks, etc (Barabási and Crandall 2003). The interactive graph generators developed in
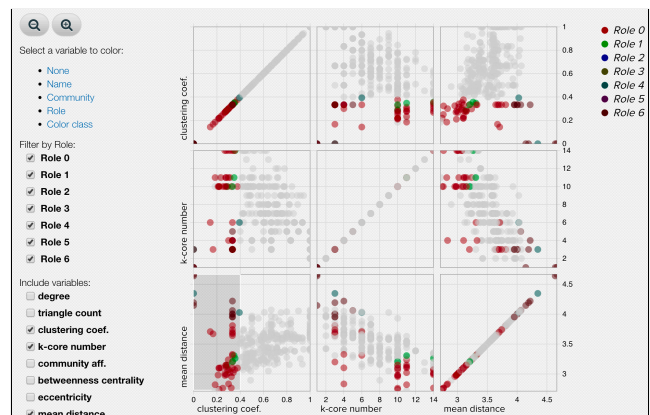


Figure 4: Microscopic network properties may be interactively analyzed using scatter plot matrices. GRAPHVIS supports interactive techniques such as brushing, linking, highlighting, as well as semantic zooming.

this work are broadly categorized into

1) Model-based synthetic graph generation and visualization, using standard models such as Erdős-Rényi, Chung-Lu (CL) (Aiello, Chung, and Lu 2001), and preferential attachment (PA) (Barabási and Crandall 2003).

2) Pattern-based synthetic graph generation and visualization, using subgraph patterns such as nodes, edges, cliques, stars, cycles, and chains.

3) Hybrid synthetic graph generation and visualization that allows users to generate graphs using a standard model (such as Erdős-Rényi) in addition to adding certain patterns to the generated graph (e.g., cliques, and stars).
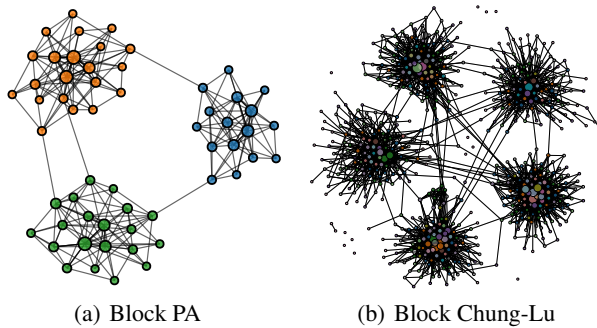


(a) Block PA      (b) Block Chung-Lu

Figure 5: Networks generated from the proposed Block-PA (BPA) and Block-CL (BCL) models.

For capturing community-structure, we proposed three additional block model approaches that combine multiple one-stage probabilistic models such as Chung-Lu, Erdős-Rényi, or preferential attachment by probabilistically creating inter-community edges. See Figure 5 for an illustration. Notably, these block-model approaches may also be used as patterns allowing the user to interactively experiment with adding various patterns to already existing networks in the visualization window.

## Dynamic Network Exploration

Dynamic networks arise in many settings (e.g, email communications). To understand the evolution and dynamic patterns in networks, GRAPHVIS provides tools to interactively analyze the evolution of the graph over time (Figure 6). In particular, users can filter temporal networks by date and time. Additionally, the time scale can be selected (via brushing) and adapted based on the application or data properties. This controls the range of time being visualized. Using these tools, analysts can begin to understand the dynamics and trends present in the network (e.g., seasonality, spike, trends).

## Conclusion

This paper introduced a *web-based interactive visual analytics platform* for graph and network data. Our work is based on a multi-level visual analytics engine, designed for rapid
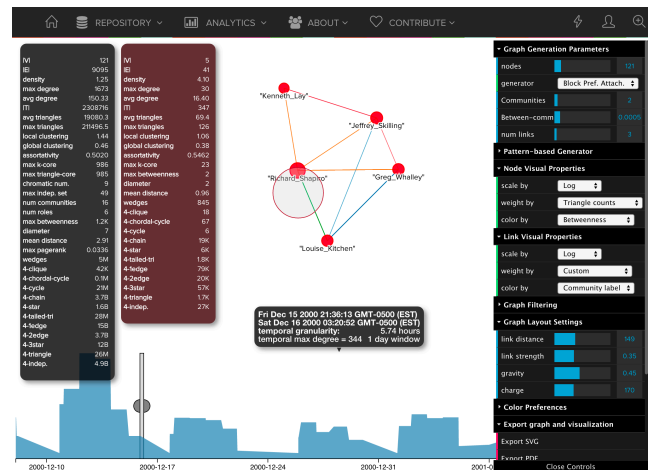


Figure 6: Interactive visual exploration of the Enron email communication dynamics. Using a window of approx. six hours, we immediately see many of the major players involved in the Enron scandal.

interactive visual exploration and graph mining. Furthermore, it integrates macro and microscopic statistical techniques as well as graph mining and machine learning with interactive visualization to aid in the discovery of important patterns and insights for sense making, reasoning, and decision-making.

## References

Ahlberg, C.; Williamson, C.; and Shneiderman, B. 1992. Dynamic queries for information exploration: An implementation and evaluation. In *Proc. of SIGCHI*, 619–626.

Ahmed, N. K.; Duffield, N.; Neville, J.; and Kompella, R. 2014. Graph sample and hold: A framework for big-graph analytics. In *SIGKDD*, 1–10.

Aiello, W.; Chung, F.; and Lu, L. 2001. A random graph model for power law graphs. *Exp. Math.* 10(1):53–66.

Auber, D. 2004. Tulip – a huge graph visualization framework. In *Graph Drawing Software*. Springer. 105–126.

Barabási, A. L., and Crandall, R. E. 2003. Linked: The new science of networks. *Amer. J. of Physics* 71(4):409–410.

Bastian, M.; Heymann, S.; Jacomy, M.; et al. 2009. Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8:361–362.

Ebert, D.; Gaither, K.; Jang, Y.; and Lasher-Trapp, S. 2014. Cross-scale, multi-scale, and multi-source data visualization and analysis issues and opportunities. In *Scientific Visualization*. Spr. 353–360.

Rossi, R. A., and Ahmed, N. K. 2015a. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Rossi, R. A., and Ahmed, N. K. 2015b. Role discovery in networks. *TKDE* 27(4):1112–1131.

Thomas, J. J., and Cook, K. A. 2005. *Illuminating the Path: the research and development agenda for visual analytics*. IEEE Computer Society.