

MultiNode-Explorer: A Visual Analytics Framework for Generating Web-based Multimodal Graph Visualizations

S. Ghani, N. Elmqvist, and D. S. Ebert

Purdue University, USA

Abstract

We propose MultiNode-Explorer, a visual analytics framework that is capable of transforming multidimensional datasets into an entity-relationship (E-R) model and visualizing the data as node-link diagrams. The framework accepts an E-R schema, a set of relational data tables, and an interface specification file, and generates a multimodal and multivariate graph and a corresponding interactive applet for viewing the graph in a web browser. We show examples for a large research organization, a visual search engine for a Wiki, and for NSF funding data.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

1. Introduction

Multidimensional data can easily be represented using the relational model, and many practical visualization applications tend to use this model for internal representation [Fek04]. In the relational model, tables describe sets of relations between attributes: for example, a table describing cars may involve attributes for manufacturer name, model name, gas mileage, and weight. Actual relations could be a car made by "Volvo" with the "XC 90" model name, 18 MPG mileage, and 3000 lbs in weight. Other relations describe connections between cases in other tables. In practice, rows link values in the same table for a particular case (such as a car), and keys link tables together (such as the manufacturer name in a car table linking to a separate sales table).

While multidimensional data are typically stored in this relational fashion, visualizations of this data rarely use an intrinsically relational representation (some techniques, such as parallel coordinates [Ins85], do show links between values, but this is not the focus of those techniques). Node-link diagrams, on the other hand, are common visual representations for relational data, but they are rarely, if ever, used for multidimensional visualization. There are many reasons for this, such as the fact that general multivariate visualizations typically emphasize values and clusters of values instead of relations between them, and that effective multivariate graph visualization is an open research challenge [BCD*10]. However, the primary reason may very well be the conceptual

shift needed to start thinking about multidimensional data as graphs consisting of entities of different types (modes) and with different attributes. We call these graphs **multimodal** and **multivariate**, and only a select few works in visualization have begun to tackle such graphs [LNS11, DRL*12].

In this work, we propose an approach for modeling multidimensional data stored in a relational form as multimodal and multivariate graphs, and show how to effectively visualize them by adapting basic node-link diagrams for this purpose. We implemented our technique in a generic framework we call the MultiNode-Explorer that is capable of taking a database schema and a set of data tables, and generates a multimodal/multivariate graph of this data that can be interactively analyzed. While our approach can be applied to virtually any multidimensional dataset, we show three examples here: (1) COE-Explorer, a large research network consisting of centers, universities, projects, and investigators; (2) Assimi, a visual search engine for a large-scale student Wiki; and (3) the Grant-Explorer, a visual database of research funding allotted to various projects and PIs by NSF.

This paper is structured as follows: We first survey the related work in the area. We then present the technical description of converting multidimensional datasets into multimodal/multivariate graphs using an E-R model. We describe the design and specifications of the interface and its implementation. We present three practical examples of our idea. We close with a summary and our plans for future work.

2. Related Work

Graphs are used to represent relationships in a dataset. However, many graph datasets now exist where different attributes are associated with each node. Such types of datasets are represented as *multivariate graphs*. Another class of data are *multimodal graphs*, where the nodes and edges have different *modes*, or types. Both types of graphs have become increasingly important in the last few years, but visualizing them is a difficult challenge [BCD*10].

2.1. Multivariate Graphs

Several approaches exist for visualizing general multidimensional data that could conceivably be applied to multivariate graphs. One popular such approach is Online Analytical Processing (OLAP) [CCS93], which uses the concept of a “data cube” as a set of items with values in different dimensions. Wattenberg [Wat06] uses and extends the OLAP concept with his tool PivotGraph for multivariate graphs. PivotGraph employs a grid-based approach for showing relationships between node attributes and links, allowing users to generate summary graphs by selecting a subset of attributes.

Semantic substrates [SA06, AS07] is another framework for exploring multiple attributes of a network and it allows the user to view several panes in the same window so that different node attributes can be analyzed in parallel. Dynamic query sliders are provided for specifying neighborhood around user-selected nodes such that the relationship between node attributes and edges can be studied.

Pretorius and van Wijk [PvW08, PW06] describe a hierarchical clustering technique for visualizing multivariate graphs. Here, nodes and edges are grouped based on their data and users can interactively query and inspect the graph.

GraphDice [BCD*10] is another recent tool for visualizing multivariate graphs. Similar to PivotGraph, this tool is based on the OLAP model that visualizes attributes by using an attribute plot for information display and an overview plot matrix for navigation and exploration.

2.2. Tabular Data Visualization

A number of systems for analyzing tabular data could potentially be adopted for multivariate graphs [HP11, LNS11]. For example, Table Lens [RC94], Polaris [STH02], FOCUS [SBB96], and Tableau [Tab] visualize tabular data in the form of plots such as bar charts, box plots, scatter plots, etc. They are interactive systems with features like searching, filtering, and pan and zoom to better enable users to understand the data.

ManyNets [FPSG10] builds on the Table Lens metaphor by visualizing network data as the rows and columns in a spreadsheet, showing summary graphics for cells using a

combination of histograms, bar charts, and numbers. Similarly, NodeXL [HSS10] is an Open Source tool that integrates with Microsoft Excel to enable users to visualize and analyze network data. It allows the import of network data in formats such as edge lists, matrices, GraphML, CSV, and other workbooks. The tool generates network statistics and creates visualizations of the graph. Filtering and display attributes can be used to highlight important graph structures.

NetLens [KPLB07] draws on the entity-relationship (E-R) model [Che76] to transform multidimensional data into a network, and applies this idea to scientific authorship data. However, the system does not actually employ graph visualization to render the resulting networks. The more recent GraphTrail [DRL*12] system takes this idea further by visualizing a multimodal and multivariate graph. GraphTrail incorporates familiar statistical charts, smooth interaction, and a history mechanism into this visualization.

Most relevant to our work is Ploceus [LNS11], which visualizes tabular data in network form and supports multi-level network analysis. Ploceus allows the users to not only build new networks but also transform existing ones to explore the data at different levels of abstraction. However, whereas the main focus of Ploceus is a model for converting tabular data into network form, our focus in this paper is on designing a content pipeline that also includes a customizable interface for visualizing the resulting graph.

3. The MultiNode-Explorer

The MultiNode-Explorer is a visual analytics framework designed for creating concrete instantiations of a graph visualization applet that has been customized to a particular type of multimodal and multivariate data. In other words, the purpose of MultiNode-Explorer is to define a general framework that can be tailored to any knowledge domain, such as a scientific discipline, research funding, or document collections. The framework consists of components that accept tabular input data and generate a web-based visualization tool for overviewing, interacting with, and analyzing the data. Figure 1 shows an overview of this framework. In the next subsections, we discuss each step in detail.

3.1. Framework Components

The MultiNode-Explorer framework is implemented in Java and consists of the following components: (1) a Content Generation Pipeline, which builds a multivariate and multimodal graph from a database and an E-R schema; (2) the Applet Generator, which instantiates a specific Explorer applet given an interface specification; and (3) the Explorer Applet Prototype, which uses prefuse [HCL05] to interactively visualize a multimodal and multivariate graph in a web browser.

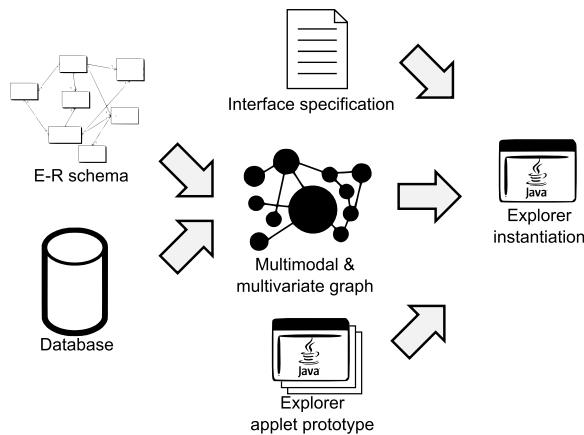


Figure 1: The MultiNode-Explorer framework. Tabular data encoded in a database is transformed into a multimodal and multivariate graph using an entity-relationship model. This graph is then combined with an interface specification to create a graph explorer instantiation as a Java applet.

3.2. Data Model

One of the core contributions of the NetLens [KPLB07] project was the observation that an entity-relationship (E-R) model [Che76], consisting of *domain concepts* and their *relations*, can be used to view multidimensional data as network data. E-R models are at the core of database design, and also suggest a way for how to treat tabular data in a database in a truly relational fashion: as graphs consisting of vertices for the entities, and edges for the relations. For example, in the car domain, relevant concepts include physical objects such as chassis, wheels, passengers, driver, engine, etc, and relations define how these concepts are connected.

Our MultiNode-Explorer framework builds on these ideas by providing a Content Generation Pipeline for transforming tabular data, e.g., from a relational database, into a multimodal and multivariate graph using a *E-R schema* as input. Unlike the recent Ploceus system [LNS11], where the schema generation is seamlessly integrated within an interactive dialogue with the system, the focus of our pipeline is less on generating the schema and more on the resulting graph and visualization. For this reason, the E-R schema is defined in an XML file that specifies which fields in different tables should be mapped to **entities** (i.e., node types in the resulting graph), which fields are **keys** that relate entities in different tables, and which fields should become **attributes** of a particular entity. Our framework also allows computing **derived values** from the tabular data, such as calculating tf-idf terms for a free text field across all entries, the degree for a particular relation, or the aggregated sum of a related field.

Executing the Content Generation Pipeline mechanism on a database (currently represented as separate .CSV files, one for each table) and an E-R schema will extract all of the enti-

ties and attributes, calculate the derived values, and save the resulting graph as a GraphML file for later visualization.

3.3. Creating the Explorer Instantiations

Visualizing the resulting multivariate and multimodal graph is done using an Explorer that has been instantiated from the core Explorer Applet Prototype. This prototype supports all of the basic graph visualization and interaction functionality, but can be tailored to a specific domain using an interface specification. The interface specification is an XML file that describes the views to include, their layout, and interface elements such as search fields, filters, and navigation controls.

The framework currently supports the following basic views and controls (easily extensible):

- **Multimodal node-link visualization:** A node-link diagram where nodes and edges are color-coded based on their mode (type). The view supports several layouts—including force-directed, radial, and attribute-based—and interactions for panning, zooming, and dragging nodes.
- **Wordle:** Our Wordle [VWF09] view (using a deterministic Wordle layout algorithm [KKEE11]) summarizes the words in a specified graph attribute for the currently visible nodes. Clicking on a word in the view will automatically filter out the nodes that do not contain the word.
- **List and details:** A simple list view that shows a textual list of the visible nodes. Clicking on a node will bring up the node details (its attributes) in a specific detail view.
- **Modal filters and layout:** A mode interface panel that allows for toggling the visibility, layout, size, and color mapping for different modes in the multimodal graph.
- **Search query:** A text field for directly querying the multimodal graph. Nodes that do not match the query will be filtered out so that they are no longer visible.

Actually creating an Explorer instantiation is done by running the Applet Generator with the interface specification and the URL of the GraphML data source. This will create an Explorer applet ready for distribution on a website.

4. Application Examples

4.1. Research Network Organizational Structure

The motivation for the MultiNode-Explorer came in 2009 from a project to visualize the U.S. Department of Homeland Security's network of Centers of Excellence. The network had grown to twelve centers at that time, and the DHS program managers needed a way to overview this large research organization. The purpose would be to both see details of individual centers, projects, and investigators in the network, as well as analyze the general themes in by the network.

Figure 2 shows a screenshot of the COE-Explorer, the applet we created in response to this need. The tool has been very positively received by the DHS program managers and has been used for analysis as well as presenting

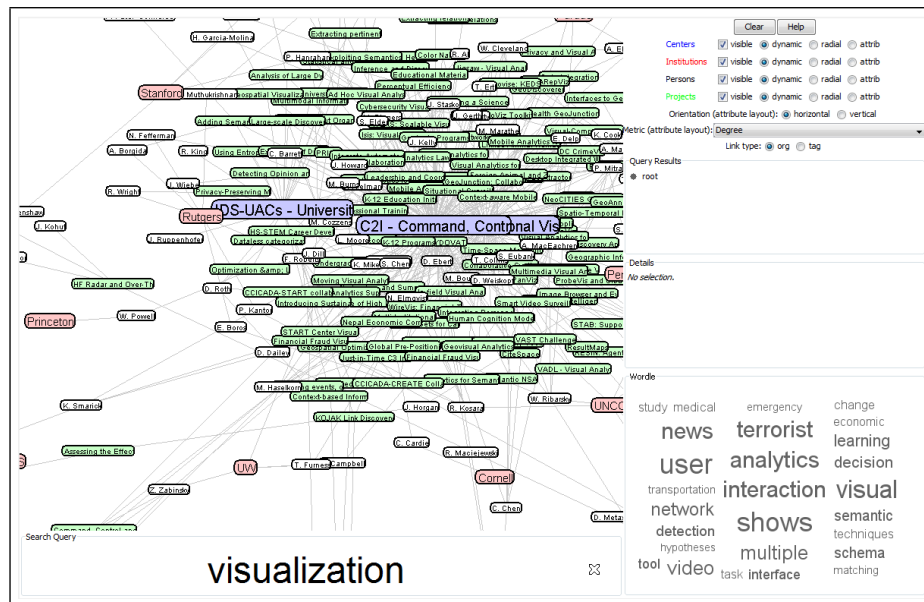


Figure 2: The COE-Explorer applet visualizing the U.S. DHS University Programs research organization. Blue nodes are centers, red nodes are institutions, white nodes are investigators, and green nodes are projects.

research within the network to external stakeholders and policy makers. We have since generalized the applet into the MultiNode-Explorer framework, but the COE-Explore remains the flagship implementation. Versions exist for tabletops as well as for local (non-Internet) use.

4.2. Wiki Search Engine

Another use of the MultiNode-Explorer framework has been to visualize the content and structure of search results in a Wiki designed for student learning [HB10]. This tool, called Assimi (<http://www.assimi.com/>), shows not only direct search results are shown (thick borders), but also pages that link to those results. Assimi has been deployed since 2010 and is used daily by Purdue students.

4.3. Research Funding

Our third example is the NSF-Explorer, shown in Figure 3, where we have used the framework to generate a tool to visualize the research funding from the NSF for Purdue University. Here the purpose is to give both faculty as well as administrative staff (such as the university's office of research) an overview of the research being conducted at the university. However, we have not yet deployed the tool with any such end users, so we lack any feedback on its usefulness.

5. Conclusion and Future Work

In this work, we describe a pipeline for converting multidimensional data into the E-R model and visualizing the result-

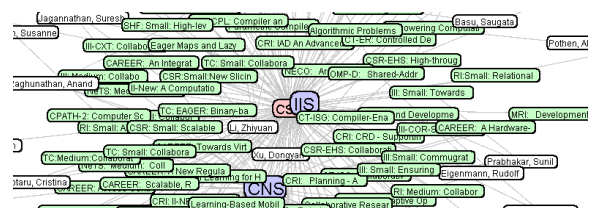


Figure 3: The NSF-Explorer visualizing a subset of the active NSF funding for Purdue University. Red nodes are directors, blue nodes are divisions, white nodes are investigators, and green nodes are funded proposals.

ing graph in terms of node-link diagrams. We implemented an interactive web-based framework that takes multidimensional data as input and generates visualization of this data in the form of multivariate and multimodal graphs. We also give three practical examples for different types of data.

In the future we want to design aggregation and summarization strategies to combat very large datasets. In particular, we want to implement strategies for on-the-fly hierarchical aggregation [EF10] of these multimodal/multivariate graphs to support interactive browsing.

Acknowledgments

This work was partly supported by the U.S. Department of Homeland Security's VACCINE Center under award no. 2009-ST-061-CI0001.

References

- [AS07] ARIS A., SHNEIDERMAN B.: Designing semantic substrates for visual network exploration. *Information Visualization* 6, 4 (2007), 281–300. 2
- [BCD*10] BEZERIANOS A., CHEVALIER F., DRAGICEVIC P., ELMQVIST N., FEKETE J.-D.: GraphDice: A system for exploring multivariate social networks. *Computer Graphics Forum* 29, 3 (2010), 863–872. 1, 2
- [CCS93] CODD E. F., CODD S. B., SALLEY C. T.: *Providing OLAP to User-Analysts: An IT Mandate*. Technical report, E. F. Codd & Associates, 1993. 2
- [Che76] CHEN P. P.-S.: The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems* 1, 1 (1976), 9–36. 2, 3
- [DRL*12] DUNNE C., RICHE N. H., LEE B., METOYER R., ROBERTSON G.: GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the ACM Conference on Human Factors in Computer Systems* (2012). to appear. 1, 2
- [EF10] ELMQVIST N., FEKETE J.-D.: Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 439–454. 4
- [Fek04] FEKETE J.-D.: The InfoVis Toolkit. In *Proceedings of the IEEE Symposium on Information Visualization* (2004), pp. 167–174. 1
- [FPSG10] FREIRE M., PLAISANT C., SHNEIDERMAN B., GOLBECK J.: ManyNets: an interface for multiple network analysis and visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2010), pp. 213–222. 2
- [HB10] HADDAD A. W., BOUTIN M.: Rhea: a student-driven tool for enhancing the educational experience. *Journal of Computing Sciences in Colleges* 26, 1 (2010). 4
- [HCL05] HEER J., CARD S. K., LANDAY J. A.: prefuse: a toolkit for interactive information visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2005), pp. 421–430. 2
- [HP11] HEER J., PERER A.: Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology* (2011), pp. 51–60. 2
- [HSS10] HANSEN D., SHNEIDERMAN B., SMITH M. A.: *Analyzing social media networks with NodeXL*. Morgan Kaufmann, 2010. 2
- [Ins85] INSELBERG A.: The plane with parallel coordinates. *The Visual Computer* 1, 2 (1985), 69–91. 1
- [KKEE11] KIM K., KO S., ELMQVIST N., EBERT D. S.: Word-Bridge: using composite tag clouds in node-link diagrams for visualizing content and relations in text corpora. In *Proceedings of the Hawaii International Conference on System Sciences* (2011), pp. 1–8. 3
- [KPLB07] KANG H., PLAISANT C., LEE B., BEDERSON B. B.: NetLens: iterative exploration of content-actor network data. *Information Visualization* 6, 1 (2007), 18–31. 2, 3
- [LNS11] LIU Z., NAVATHE S. B., STASKO J. T.: Network-based visual analysis of tabular data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology* (2011), pp. 39–48. 1, 2, 3
- [PvW08] PRETORIUS A. J., VAN WIJK J. J.: Visual inspection of multivariate graphs. *Computer Graphics Forum* 27, 3 (2008), 967–974. 2
- [PW06] PRETORIUS A. J., WIJK J.: Visual analysis of multivariate state transition graphs. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 685–692. 2
- [RC94] RAO R., CARD S. K.: The Table Lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (1994), pp. 318–322. 2
- [SA06] SHNEIDERMAN B., ARIS A.: Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 733–740. 2
- [SBB96] SPENKE M., BEILKEN C., BERLAGE T.: FOCUS: the interactive table for product comparison and selection. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (1996), pp. 41–50. 2
- [STH02] STOLTE C., TANG D., HANRAHAN P.: Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 52–65. 2
- [Tab] Tableau software. <http://www.tableausoftware.com/>. Aug. 2010. 2
- [VWF09] VIÉGAS F. B., WATTENBERG M., FEINBERG J.: Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 1137–1144. 3
- [Wat06] WATTENBERG M.: Visual exploration of multivariate graphs. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2006), pp. 811–819. 2