Redgraph: Navigating Semantic Web Networks using Virtual Reality

Harry Halpin
Institute for Collaborative and Communicating Systems
University of Edinburgh

David J. Zielinski, Rachael Brady and Glenda Kelly
Visualization Technology Group
Duke University

ABSTRACT

We present *Redgraph*, a generic virtual reality (VR) visualization program for network data, based on Resource Description Framework (RDF), the primary standard of data underlying the Semantic Web. *Redgraph* bypasses a number of problems in 3D graph visualization by relying on users to interactively "extrude" a 2D network into the third dimension. This pilot study applies *Redgraph* to data from the U.S. Patent and Trademark Office to explore innovations in the history of computer science. Comparison of subjects' response times utilizing 3D pull-out vs. 2D strategies on tasks involving fine-grained connectivity or broad network observation found that subjects were faster correctly answering questions involving fine-grained connectivity using 3D strategies, particularly when data was densely clustered. Subjects' qualitative feedback suggests that the most valuable application of this 3D technique lies in untimed exploration to discover relationships in the underlying data structure.

CR Categories and Subject Descriptors: H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities—applications; H.5.2 [Information Interfaces and Presentation, (e.g. HCI]: User Interfaces - Interaction Styles. I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods–Semantic networks
Keywords: Visualization, Network Data Analysis, RDF, Semantic Web, 3-D Interaction, User Study

1 INTRODUCTION

A challenge in network visualization is the complexity in understanding new, unfamiliar datasets in order to conceptualise underlying data structures. This problem is pronounced in data sets in which either higher-dimensional structure does not map well onto 2 dimensions or when the number of connections is so dense that visualizing networks leads to confusing "spaghetti-like" structure. One effective way to circumvent this problem is the use of "extrusion" which allows the user to be in charge and select nodes of interest from a 2D structure, and then "pull" these into 3D space, which literally extrudes the 2D data into 3 dimensions. *Redgraph*, our VR-based network visualization program, relies on user-directed 3D extrusion. While this application builds on previous 3-dimensional work that allowed users to reposition nodes [1], *Redgraph* differs from this type work by allowing users to extrude nodes from 2D into 3D in fully immersive and interactive CAVE(TM)-like environments [2]. This generic toolset is not customized for any particular dataset and is capable of visualizing any network that can be described using the standardized RDF (Resource Description Framework) data model (http://www.w3.org/RDF/), a World Wide Web Consortium (W3C).

[1] email: H.Halpin@ed.ac.uk   [2] email: dzielin@duke.edu
[3] email: rbrady@duke.edu   [4] email: glenda@ee.duke.edu

2 DATA DISPLAY SETUP- EXTRUSION TECHNIQUE

As shown in Figure 1 (left below), the user first sees data on a 2D plane of the virtual environment and can then use the virtual wand to rapidly browse the nodes. Upon intersection between the wand and a node, the visualization program will display all information contained in the RDF data model (e.g. patent names
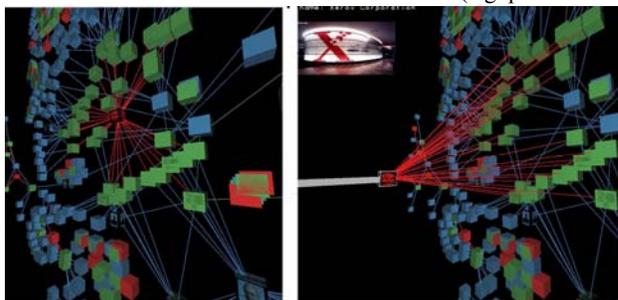
Figure 1: *Redgraph* photos
Left: 2D data   Right: 3D Pullout

and abstracts, company names, etc.). When users discover a node of interest, they can continue holding down a button on the wand, and pull or push the node into 3D, as shown in the rightmost picture (Figure 1).

Many users can quickly identify clusters and other interesting phenomenon in 2-dimensional data-structures. Extrusion (3D pull-out) avoids many problems of 3D layout by using well-understood and optimized 2D visualization algorithms for the initial layout. Most importantly, 3D pullout allows users to interact with the data using both visual and proprioceptive-motor abilities, thus giving them potential to optimally manipulate data displays according to individual preferences.

Furthermore, interaction with network data via 3D virtual extrusion allows users to dynamically re-cluster data using parameters often difficult or impossible for computers to detect, and thus offers the third dimension as a "sketchpad" for placement of nodes. This addresses the problem of *a priori* computer node placement that might block user viewpoint in a 3D data set.

5 STUDY DESIGN AND PROCEDURE

To assess efficacy of the 3D pull-out when compared to 2D presentation (e.g. did pulling out and interacting with data in an immersive 3D mode help any tasks?), a pilot study comparing subjects' time to correctly answer 8 quantitative questions between these two methods of data presentation was conducted. It was hypothesized that subjects would be faster using 3D than 2D. To better understand subjects' interactions with the data display, an additional 4 qualitative questions were asked.

There were 21 subjects, 15 male and 6 female with a mean

---

2 CAVE is a registered trademark of the Board of Trustees of the University of Illinois.

[3] *Redgraph* uses generic C-based Redland API (*http://librdf.org*) to load the data file into the virtualized data structure. Network visualization and extrusion technique is implemented using the Syzygy [3] virtual reality library. *GraphViz* [4] is used for the initial 2D network layout using the Kamada-Kawai algorithm.

age of 19.6 years. Ages ranged from 18- 28. Two-thirds had used VR headgear before and three had experienced an immersive VR room before. Two-thirds were engineering majors and other majors included biology, statistics, economics, psychology and administration.

The study was conducted in the Duke Immersive Virtual Environment (DiVE), a 6-sided CAVE-like system. The DiVE is 3m x 3m x 3m with rigid ceiling and floor, flexible walls, a sliding door and using a 7 PC workstation Cluster running XP with NVidia Quadro FX 3000G graphics cards as the graphics-rendering engine. Christie Digital Mirage 2000 projectors connect with Stererographics active stereo glasses. Tracking is supported by the Intersense IS-900 system.

The data set used from the U.S. Patent and Trademark Office (USPTO) allows exploration of social networks of innovation in computer science. Since this database totalled over seven million patent records (2006), a subset of patents selected resulted in a data set of patents with 7667 nodes and links.

Patents (shown as green), inventors (blue), and institutions (red) were all represented as nodes (cubes), and lines (links) connecting nodes represented relationships between patents and inventor, and patent and linked institutions. When subjects touched a patent node with the virtual wand, the patent abstract/associated images appeared. Inventor nodes displayed inventor's name, location(s), and, for a subset of inventors, the inventor's photo was displayed. Institution nodes displayed institutional names and ,when available, building images.

Prior to the experiment, subjects were shown how the 3D pullout and virtual navigation wand worked and allowed to find and name a company, an inventor and a patent. Next, subjects were asked 2 warm-up questions:
1. Name the company with the largest number of patents.
2. Name a patent filed by Xerox.

For quantitative questions, equal numbers of subjects were assigned to alternating 2D versus 3D conditions per each question below:
1. Which patent has the most inventors?
2. What company has the 2nd largest number of patents?
3. How many patents did Charles Thacker have?
4. Name a patent by Robert Kahn?
5. Name the patent by Robert Metcalfe that has "collision detection" in the title.
6. Name the company that filed a patent by Vinton Cerf.
7. Name the company that Ivan Sutherland works for.
8. Find the name of the inventor who filed patents for both BBN and Xerox.

Prior to qualitative questions, subjects, given 3-5 minutes, were instructed: **"**See how you can use this network to get a better understanding of the history and flow of ideas in Computer Science". Subjects then gave written responses to:
1.What discoveries did you make using this virtual data display?
2. What did you discover about what was or is important in the area of computer science using this data display system today?
3. What were the most helpful features of this display system for helping you learn and/or make discoveries today?
4. What suggestions do you have for making this system more helpful as a tool for discovery of connections and understanding this patent data?

6 DATA ANALYSIS AND RESULTS

Separate t-tests were computed for subjects' time to correct answer comparing the 2 and 3D strategies. Since F-tests (for equality of variances) attained significance (p< .05) for

questions 1, 3, 5, 7, t-tests for these questions were computed using an unequal variance model.

Table 1 below indicates that for all except questions 2 and 4, subjects attained correct answers faster using 3D versus 2D strategies. For questions 3 and 5, these differences were statistically significant (p<.05). For question 4 only (a simple question based on data in a non-dense area requiring no tracing), subjects using 2D were significantly faster (p<.05) than those using 3D pullout.

Table 1: Means (secs.), standard deviations & t-tests for 2D & 3D

Ques. p<.05*	2D Mean	2D SD	3D Mean	3D SD	T-test	P-value (1-tail)
1	**190.6**	230.5	**95.2**	55.3	1.28	0.115
2	**25.2**	22.2	**37.1**	36.9	-0.92	0.184
3*	**85.4**	39.8	**45.9**	20.2	2.83	**0.007**
4*	**38.9**	14.1	**54.2**	23.9	-1.81	**0.043**
5*	**44.2**	24.7	**22.0**	10.4	2.64	**0.011**
6	**31.6**	17.4	**29.7**	13.7	0.27	0.396
7	**42.9**	38.0	**33.6**	17.2	0.72	0.244
8	**100.5**	91.3	**70.7**	64.8	0.85	0.202

When untimed and allowed to experiment with both 2 and 3D, 100% of subjects preferred 3D to 2D. Particular 3D aspects found most helpful included: **"allowed connections** between company, inventor and patent **to be seen** very clearly; "*crucial in sorting* overlapping lines that connected patents and creators"; "allowed user to *separate and identify* how nodes are *related to each* other…like one of those *cognitive thought maps*"; "*movement of the nodes really helpful-2D was moving ideas and 3D is using those ideas*; "since entire 2D space was filled w/data… difficult to do mass *organizing* with lack of open space-3D did just that-*added dimension gave me free space to use in my organization"*.

7 DISCUSSION AND CONCLUSIONS

These findings suggest that for answering specific fine grained questions, particularly in densely clustered data which require tracing a route and either "picking-apart" a cluster for particular information or counting, users are faster using 3D data pullout than 2D in this VR environment. High individual differences between users in their ability to use 3D as well as the small sample size in this pilot could have contributed to only 3 of the 8 questions attaining statistical significance for this initial study.

Efficiency in attaining correct answers was not the only value of this 3D display. Rather, subjects' qualitative feedback suggests that added value of 3D extrusion in VR also lies in untimed exploration to discover relationships in the underlying data structure. We will incorporate these findings to continue the study with a larger group of subjects and with application to discovery in other types of datasets.

REFERENCES

[1] L.A Crutcher, A.A. Lazar, S.K. Feiner, and M.X. Zhou, Managing Networks Through a Virtual World. *IEEE Parallel Distribution Technology,* 3(2): 4–13, 1995.

[2] C. Cruz-Neira, D. Sandin, and T. Defanti, Surround-screen Projection-Based Virtual Reality: The Design and and Implementation of the CAVE. In *Proceedings of SIGGRAPH* ACM Press / ACM SIGGRAPH, Anaheim, CA, ACM, 2000.

[3] B. Schaeffer, and C. Goudeseene, Syzygy: Native PC Cluster VR. In *Proceedings of Virtual Reality 2003*, IEEE, 15–22, 2003.

[4] E. Ganser, and S. North, S. An Open Graph Visualization System and Its Applications to Software Engineering. *Software - Practice and Experience* 30(11), 1993.