

Exploring Semantic Social Networks Using Virtual Reality

Harry Halpin¹, David J. Zielinski², Rachael Brady², and Glenda Kelly²

¹ School of Informatics
University of Edinburgh
2 Buccleuch Place
EH8 9LW Edinburgh
Scotland, UK

H.Halpin@ed.ac.uk

² Visualization Technology Group
Duke University
130 Hudson Hall Box 90291
Durham, NC 27708, USA

djzielin@duke.edu, rbrady@duke.edu, glenda@ee.duke.edu

Abstract. We present Redgraph, the first generic virtual reality visualization program for Semantic Web data. Redgraph is capable of handling large data-sets, as we demonstrate on social network data from the U.S. Patent Trade Office. We develop a Semantic Web vocabulary of virtual reality terms compatible with GraphXML to map graph visualization into the Semantic Web itself. Our approach to visualizing Semantic Web data takes advantage of user-interaction in an immersive environment to bypass a number of difficult issues in 3-dimensional graph visualization layout by relying on users themselves to interactively extrude the nodes and links of a 2-dimensional graph into the third dimension. When users touch nodes in the virtual reality environment, they retrieve data formatted according to the data's schema or ontology. We applied Redgraph to social network data constructed from patents, inventors, and institutions from the United States Patent and Trademark Office in order to explore networks of innovation in computing. Using this data-set, results of a user study comparing extrusion (3-D) vs. no-extrusion (2-D) are presented. The study showed the use of a 3-D interface by subjects led to significant improvement on answering of fine-grained questions about the data-set, but no significant difference was found for broad questions about the overall structure of the data. Furthermore, inference can be used to improve the visualization, as demonstrated with a data-set of biotechnology patents and researchers.

Keywords: Visualization, Virtual Reality, RDF, Semantic Web, 3-D Interaction, User Interface, Network Analysis, CAVE.

1 Introduction

While researchers have become interested in the large amounts of network-structured data available on the Web, intuitive understanding of the structure

of networks remains more of a black art than science. Mathematical frameworks developed to analyze networks can be difficult to interpret, so visualizing networks has become a common tool for users to gain intuitive understanding of the data. While research into network visualization algorithms is extensive on the Semantic Web, much of this research has focused on two dimensional visualization which often produces dense and confusing “spaghetti-like” structures that elude visual analysis and comprehension.

Redgraph, our virtual reality-based network visualization program for Semantic Web data, relies on user-directed 3-D extrusion to transform the visualization from 2 into 3-dimensions. While this application builds on previous virtual reality work that allows users to reposition nodes [6], Redgraph differs from previous applications by allowing users to “extrude” nodes from 2-D into 3-D in a fully immersive and interactive CAVETM-like environment, as shown in Figure 1 [7]. Redgraph is not customized for any particular data-set or application, but is a generic toolset is capable of visualizing any network that can be described using the RDF (Resource Description Framework) data model, which naturally maps its subjects and objects to nodes in a network and properties to links [13].

2 Related Work

Visualization of Semantic Web data is nearly synonymous with graph visualization, which has a long history prior to the advent of the Semantic Web [15]. Visualization is necessary since “implicit information embedded in semantic web graphs, such as topography, clusters, and disconnected subgraphs, is difficult to extract from text files” [18]. Almost all graph visualizers for Semantic Web data, as exemplified by IsaViz, produce 2-D graphs [25]. These tools have in turn been built on top of tools developed for generic graph visualization such as GraphViz [10], so most researchers have simply applied pre-existing 2-D visualization algorithms from applications such as GraphViz to Semantic Web visualization [23]. Researchers who study user interfaces in the Semantic Web have begun criticizing graph visualization as the primary method of visualizing Semantic Web data. In particular, Karger and Schraefel have noted that the “Big Fat Graph” approach is popular because visualization researchers are “allowing the computer’s internal representation of data to influence the way their tools present information to users, when instead they should be developing interfaces that are based on the users’ needs, independent of the computer’s particular information representation” [18]. In other words, just because RDF has a graph structure does not mean a graph structure should be used to interface with data. To the extent that Karger and Schraefel believe that more human-centric user interfaces are needed for Semantic Web, we agree wholeheartedly. However, work on alternative methodologies for visualizing Semantic Web data suffers from its fair share of limitations as well. If graph visualization can be characterized as the “Big Fat Graph” approach, the alternatives cited by Karger and Schraefel such as mSpace [18] and the Tabulator browser [3] can be characterized as the “Lots of Confusing Little Menus” approach. It is the careful choice of menus and other

interface components that creates a functional user-interface. Therefore prior to designing any application-centric user interface, an overall intuitive understanding of the particular data-set and what users want out of this data is needed. For data sets that researchers are just beginning to explore, visualization can be vital for gaining an intuitive understanding of the data set so that these questions can even be asked. Graph visualization is then an important default mode of visualization for Semantic Web data, as it makes minimal assumptions about the data.

A pragmatic approach to visualization of Semantic Web data recognizes that there is no one perfect visualization technique, much less a fully generic user interface for the entire Semantic Web. Despite its critics, there are definitely cases where graph visualization is the most appropriate technique. First, graph visualization is appropriate when what is being graphed is an actual network, such as a social networks. Visualization is important in network analysis, as most networks are less amendable to analytic analysis than other types of data, primarily due to their violation of the Gaussian distribution. For identifying clusters, there are wide variety of statistics, each with their own drawbacks and advantages, which can lead to confusion for the user [21]. Typical networks in “the wild” (as opposed to Erdoš-Renyi random graph models [2]) obey a power-law distribution of connectivity between nodes and links [1] so that typical descriptive statistics such as means cannot be computed since these rely on properties that do not hold in power-law distributions [24]. Due to the above-mentioned factors, visual inspection can be crucial when understanding social networks in particular. While Karger and Schraefel complain that one problem with graph visualization is that it puts “next to each other the things that are connected by links,” if the link represents something important, such as a relationship in a social network, it is crucial for user understanding that the node be placed in close proximity with the other node. While more traditional user interfaces may be developed once a data-set is understood thoroughly, for the initial exploration of the data-set it is actually useful for the visualization to follow the data quite closely, as to not give any a priori bias to the presentation of subsets of the data when creating interfaces using tools such as Exhibit [16]. Due to this request from researchers in the social networks of patents, we decided to use graph visualization as our method of exploring the data before developing a more traditional user interface.

In order to address some of the traditional problems with graph visualization, in particular the limitations of two-dimensions in visualizing complex networks, many researchers have studied graph visualization in virtual reality [15]. Studies in virtual reality have found that viewing a graph using three-dimensions allows “three times as much information can be perceived in the head coupled stereo view as in the 2-D view.” [29]. Moving beyond mere 3-D visualization to fully immersive virtual reality, with head-tracking and stereoscopic vision, allows users to gain an intuitive understanding of the network by literally letting them walk “around” and “into” the data, facilitating kinesthetic comprehension [5]. This is not to underestimate the effect of the quality of the visual display for

understanding complex data in the presence of information clutter. The high fidelity graphics provided by virtual reality systems like the CAVE have been shown to help performance, as immersion “provides many depth cues that other technologies do not; in particular, stereo images and head tracking let users exercise their built-in capacity for understanding stereopsis¹ and motion parallax. Thus a higher level of immersion can lead to greater spatial understanding, which can result in greater effectiveness for many applications such as scientific visualization” [5].

Despite these advantages, no previous graph visualization research in the Semantic Web has used immersive virtual reality technology in order to enable the employment of an interactive third dimension. This stands in contrast to research in Topic Maps, where immersive virtual reality technology has been demonstrated [19]. Researchers have developed *OntoSphere3D*, a Protege plugin that in 2-D visualizes an ontology projected on a 3-D sphere that allows “panning, rotation, and zoom” but does not allow interaction with the graph’s 3-D layout itself, much less immersion and stereoscopic 3-D [4]. In particular, the usage of virtual reality technologies can confront one of the primary objections of Karger and Schraefel to using graphs as a way of visualizing Semantic Web data: the fact that graphs “are flat” [18].

To explore the advantages and disadvantages of immersive and interactive visualization, a particular data-set featuring a social network of people and institutions involved in filing patents was selected. It is social since researchers are interested in the co-authoring of patents and movements of researchers through institutions, and a network since these relationships can be represented as links between people and institutions, as well as patents. The literature on social networks, including their interaction with the Semantic Web, is immense, and those interested in exploring the interface of social networks with the Semantic Web are encouraged to reference the work of Mika and others [21]. Note that this data-set is a semantic social network since the connections between the various actors in the network have been given a semantic basis by being formulated using Semantic Web standards [13]. For our purposes, it is enough that this network is a social network where the data naturally takes a graph format, and so graph visualization is a sensible visualization for this data.

3 Virtual Reality Vocabulary

There is currently no standardized Semantic Web vocabulary for graph visualization, including virtual reality. Proposals for virtual reality markup-languages on the Web have proliferated since the beginning of the Web itself. The current ISO standard is the Web3D’s consortium’s X3D, the successor of VRML (Virtual Reality Markup Language) [30]. X3D is far too complex for our application, as it is meant to cover all possible cases of virtual reality modeling from humanoid movement to landscape scenes, and furthermore, it is defined only via XML and

¹ In other words, “depth perception.”

XML Schema, although there are proposals to introduce the use of Semantic Web standards to X3D [26].

One alternative is to use a Semantic Web version of GraphXML, an easy-to-use XML language meant to describe and annotate graphs in both 3-D and 2-D [14]. GraphXML is used in several open source projects and has parsers available in a number of languages. While merely creating a Semantic Web version of an existing XML vocabulary is not a contribution in of itself, creating a Semantic Web vocabulary for 3-D visualization allows already existing Semantic Web data to be easily annotated with the properties needed for 3-D visualization using the same format. Users can then store and query both visualization meta-data and the data itself using the same tools, and seamlessly merge formerly separate graphs of data into a single visualization, offering new and useful capabilities not provided by GraphXML itself. We developed an XSLT transformation capable of transforming GraphXML to a RDF(S) vocabulary we call Vis3D. This mapping is given in Table 1, illustrating the equivalence between GraphXML elements and attributes with Vis3D classes and properties, where a property’s domain is the class it is listed with and its range is given in parenthesis with XML Schema data-typing assumed. If there is no class listed, the property gives both its domain and range as *property(domain,range)*. If multiple properties are listed with the same domain and range, only the final property explicitly lists these. The advantage of Vis3D over GraphXML is that while GraphXML is a “closed-world” for sharing graph visualizations, arbitrary data-sets can be visualized by just giving Vis3D visualization properties or Vis3D classes if the data is or can be converted to the Semantic Web, a byproduct of the “meta” nature of having the visualization vocabulary and data share the same underlying model. A `visualize` property turns visualization of that element on or off, allowing a user to visualize employee relationships and not co-author relationships. In this manner, users can customize exactly what subset of their data is to be visualized, and configure the precise details.

Table 1. Mapping of GraphXML to Vis3D

GraphXML Element: attribute(s)	Vis3D Class: Property(Range)
Graph: vendor, version	Graph: vendor(string), version(positiveInt)
node: isMetaNode	Node, isMeta(boolean)
edge: source, target	link(Node,Node)
position: x, y, z	x, y, z(Node,int)
size: width, height, depth	width, height, depth(Node,positiveInt)

4 Redgraph Capabilities

Redgraph is the first generic cross-platform Semantic Web virtual reality tool for visualization. It uses a custom parser to load the data file into a virtualized data structure, and has optimizations for speed over large data-sets. Vis3D annotations are used to determine which nodes (subjects and objects) and edges

(properties) are rendered, and which nodes and edges are considered meta-data that is shown only as “details on demand” when a user “touches” the node. When the user touches the node, data associated with the node is presented to the user. Whatever schema is available can be retrieved, and the XML Schema data type provides the information needed to format the data in a human-readable form and provides captions for the data as well. The network visualization and extrusion technique is implemented in the Syzygy virtual reality library [27]. GraphViz [15] or Boost (only available on Linux) can both be used for the initial 2-D network layout. Vis3D is used to save and load visualizations. Pictures showing Redgraph in action are shown rightmost in Figure 1 and a movie may be viewed on the Web at <http://www.redgraph.org>.



Fig. 1. DiVE: Left - Person in DiVE, Right - Exploring Social Networks in the DiVE

5 Network Layout

From the point of view of the user, the initial network layout is of primary importance. A large number of algorithms have been developed for both 2-D and 3-D network visualizations. Most 2-dimensional layout algorithms consider links to be simulated “springs” and nodes to have some repulsive force, and these simulated springs are used to model forces between each pair of nodes. “Springer embedding” algorithms like the Kamada-Kawai algorithm minimize a function which is the sum of the forces on all nodes in the network [17]. This algorithm has been shown to produce diagrams that accurately model the structure of networks like social networks and also are aesthetically pleasing [23]. However, Redgraph separates the initial 2-D layout of the network, which can use in theory any algorithm to present data to the user. This is because the algorithm used for network layout often takes much longer to layout. Even using a relatively fast algorithm such as Fruchterman-Reingold, which has been optimized for large data-sets, the network layout takes much more time than the actual loading of the data and virtual reality materialization [11]. For example, for a network of 15734 nodes, loading and visualizing the data took only 2.6

seconds, while layout took 54.7 seconds. By separating the two steps, a user can lay out large data-sets only once, and then easily and quickly load the data, manipulate node locations, and save any modifications made, and not have to repeat the network layout step. For the current user study, the Kamada-Kawai algorithm was utilized [17]. For the data-set used in our foray into using inference in Section 11 the Fruchterman-Reingold algorithm was used [11].

6 Retrieving Data through Extrusion

One common challenge in graph visualization is the complexity in understanding unfamiliar data-sets in order to conceptualize underlying patterns and points of interest. This problem is especially pronounced in data sets in which either the higher-dimensional structure does not map well onto 2 or even 3 dimensions. This becomes apparent when the number of connections is so high that visualizing the network as a graph leads to confusing “spaghetti-like” clusters that are so dense as to not be amendable to interpretation [6]. Furthermore, while algorithms like Kamada-Kawai can be extended to three dimensions, the problem of determining where precisely a user should be placed in the 3-D visualization remains an open area of research [8]. Placing the user’s view point “in the middle” of a three-dimensional visualization may hinder rather than help the visualization process, since the viewpoint may hide parts of the network. One effective way to circumvent both these problems in one fell swoop is the use of extrusion. Extrusion allows the user to select nodes of interest from a 2-dimensional network and then pull these out into a 3-dimensional space, which literally extrudes the 2-dimensional data into 3 dimensions.



Fig. 2. Redgraph Snapshots: Left - Node Metadata, Center - 2-D Network Layout, Right - Interactively extruding a Node into 3-D

In detail, the user will first see the data in 2-D on a plane of the virtual environment, as shown in the center picture in Figure 2 above. The subject can then use the virtual wand, which they control through a handset, and their point of view that is monitored by head-tracking, to rapidly get an overview of the network and “zoom in” onto relevant details. Upon intersection between the wand and a node, the visualization program will display all information contained in the Semantic Web data model as shown in the leftmost picture of Figure 2. When the subject discovers a node that they are interested in, they can

continue holding down a button on the wand, and pull or push the node into the third dimension, as shown in the rightmost picture in Figure 2 above. As a result of this movement of the node, the subject can stretch the links between the nodes into three dimensions. Furthermore, the user can also use a key combination on the virtual wand to do “group pullout,” where a node and its children nodes are pulled out into 3-D together.

This method has a number of advantages for data interpretation. First, many users can quickly identify clusters and other interesting phenomenon in 2-dimensional data-structures. By beginning the visualization in 2-D, this capitalizes on users ability to conceptualize the data quickly in 2-D and then use 3-D only as needed. Extrusion avoids the problem of the 3-D layout of data by using well-understood and optimized 2-D visualization algorithms for the initial layout. Most importantly, extrusion allows users to interact with the data using both their visual and proprioceptive-motor abilities, thus giving them potential to optimally position the data display according to their preferences. This interaction with network data via extrusion allows users to dynamically re-cluster data using parameters difficult or impossible for computers alone to detect, and thus offers the third dimension as a “sketch pad” for the placement of nodes according to their particular preferences and task at hand.

7 Exploring Social Networks in Patents

The data-set used was available at no cost from the United States Patent and Trademark Office (USPTO) and allows exploration of the social networks of innovation in computer science. The goal was to map the data available in the USPTO to the Semantic Web, and then use a visualization of this data-set to expose the social networks of patent inventors through their affiliation with various institutions and their co-authoring of patents. Researchers with backgrounds in history of science had been using non-interactive 2-D visualization but found it insufficient when confronted with these dense networks, and so thought that visualizing them in immersive 3-D would help. Although they did not have a background in the Semantic Web, they correctly thought the use of Semantic Web technologies would allow them to “mash-up” data from several sources in order to help their visualization. Using Semantic Web technologies, the U.S. Patent Data was integrated with pictures, video interviews, employee records, and other material of interest [12], and stored in the OWL ontology given briefly in Table 2. This table uses the same conventions as used in Table 1, with dates always given as month, day, and year. If the exact date is not known, the date will default to the 1st of January.

Since the USPTO patent database totaled over seven million patent records in 2006, a subset of patents were selected that relate to computing history linking work in personal computing at Xerox PARC by gathering a list of employees from Xerox PARC and then searching the entire patent database for any patents filed under their names. This resulted in a data-set of patents with 7667 RDF triples.

Table 2. Social Network Ontology for Patents

Class	Properties(range)
Person	imageURI, interviewURI, movieURI(URI) born, died, started, ending, worked(date) name, university, location, bio(string)
Institution	foundedYear, endedYear(date) location, name(string) employees, creditScore(positiveInt) NAIC, sales(float)
Patent	number(positiveInt) name, classification, title, abstract, fulltext(string) dateFiled, dateIssued(date) cited, citedBy(Patent) inventor(Person) assignee(Institution)

After the visualization was developed and explored using Redgraph, the researchers used their discoveries in 3-D to architect a 2-D exhibit using Exhibit [16].

8 User Study

A user study was conducted to assess the efficacy of the 3-D data extrusion technique when compared with the 2-D method of data presentation as applied to the social network data-set described in Section 7. In other words, did pulling out and interacting with the data in an immersive 3-D modality help any tasks? Comparison of the time taken by subjects to correctly answer 8 quantitatively-measured questions was made between these two methods of data presentation. While the 3-D method allowed full interaction, including extrusion, the 2-D method, while also taking place in the DiVE, but did not allow extrusion, allowing only re-positioning the nodes on a 2-D plane using the virtual wand. It was hypothesized that subjects would give correct answers faster when allowed to use the 3-D technique as compared to the 2-D technique. Subjects were also asked 4 additional qualitative questions to provide formative evaluative feedback to better understand the ways that subjects interacted with the interface. A total of 21 subjects completed the study, 15 male and 6 female, with a mean age of 19.6 years and ranging from 18 to 28 years. Only 3 subjects had used immersive interaction before, and of these, their average number of times in an immersive virtual reality environment was 1.6. The study was conducted in the DiVE (Duke Immersive Virtual Environment), a 6-sided CAVE-like system, as shown in Figure 1 [7]. The structure is 3m x 3m x 3m with a rigid ceiling and floor, flexible walls, and a sliding door. The DiVE uses 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 to provide the visual interface, while head-tracking is supported by the Intersense IS-900 system.

For this network visualization, patents, inventors, and institutions (colored green, blue, and red respectively) were all represented as cubes and the links connecting the nodes represented relationships between these elements. The effect of using different colors and shapes was not investigated, but standard results from visualization were assumed to hold, and hence basic shapes and primary colors with high contrast were used. Auxiliary information for each element extracted from the Semantic Web encoding was displayed when the user touched that node with the virtual wand. For example, when subjects touched a given patent node with their virtual wands, the patent abstract, patent filing date, any images associated with the abstract, and other associated information were displayed on the wall of the virtual chamber, as in the rightmost picture of Figure 2. Other nodes also displayed relevant information when touched. Before the experiment was run, subjects were given a tutorial to help familiarize them with the data display. Subjects were shown how the virtual 3-D extrusion, head-tracking, and virtual wand controls worked and allowed 3-5 minutes to experiment with the system. They were asked to find and name a company, an inventor and a patent to see if they understood the node representation system, and then received two simple basic warm-up questions. After their training, they were given instructions to answer each question aloud as soon as they could using the visualization. The experimenter recorded the length of time that was taken by the subject to correctly answer the question. After each question, the DiVE was re-set and the visualization re-loaded in 2-D without any 3-D extrusion left over from previous question-answering.

8.1 2-D and 3-D Experimental Conditions

For the quantitatively-measured questions, subjects were assigned to alternating 2-D versus 3-D conditions per question. Assignment was made so that there were equal numbers of subjects assigned to both the 2-D and 3-D conditions for each question. Many of these quantitatively-measured questions are essentially queries of the data, while the qualitative questions help measure the utility of the visualization for more free-form exploration and understanding of the network.

1. What patent had 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 answering the qualitative questions, subjects were given the task of exploring the data network for a few minutes using whatever strategies they

chose (including 2-D and 3-D) via these instructions: “See how you can use this network to get a better understanding of the history and flow of ideas in Computer Science.” After exploration, subjects gave written responses to the following questions:

- What discoveries did you make using this virtual data display today?
- What did you discover about what was or is important in the area of computer science using this data display system today?
- What were the most helpful features of this display system for helping you learn and/or make discoveries today?
- What suggestions do you have for making this system more helpful as a tool?

9 Data Analysis

For each of the 8 quantitatively-measured questions, separate t-tests were computed comparing subjects’ time to answer correctly the question posed by the experimenter when allowed to use 2-D versus 3-D pull-out strategies. F-tests for equality of variances were computed and found to be significant at $p < .05$ for questions 3, 4, and 5 so for these questions t-tests were computed using the unequal variance model. Subjects using the 3-D condition were found to correctly answer Questions 3 and 5 were significantly faster ($p < .05$) than subjects using only the 2-D condition, although for Question 4 2-D was faster. Examination of the mean answer time for these 8 questions in Table 3 below indicates that for all questions except Question 4, subjects were faster in giving correct answers using 3-D versus 2-D strategies, though these differences were not statistically significant for three-quarters of the questions due to high individual differences between users in their ability to exploit the third dimension. However, the mean time of subjects using the 3-D condition were quicker for all except one question, suggesting that this effect should be assessed further utilizing a larger sample size.

The answers to the qualitative questions listed in Section 8.1 about the discoveries subjects made using this display indicated that when allowed to experiment with both 2 and 3-D displays to explore the data, all subjects preferred the 3-D display. Particular aspects of the 3-D extrusion that subjects found most helpful were “being able to bring everything into three dimensions allowed the

Table 3. Mean Time to Solve Tasks in 2-D and 3-D

Question	2-D Mean	2-D S.D.	3-D Mean	3-D SD
1	190.6	230.54	95.18	55.31
2	25.18	22.17	37.13	36.85
3	85.40	39.77	45.90	20.21
4	38.90	14.00	54.20	23.89
5	44.20	24.68	22.00	10.35
6	31.55	17.38	29.70	13.74
7	42.90	37.97	33.55	17.21
8	100.45	91.26	70.70	64.83

connections between company, inventor and patent to be seen very clearly” and “crucial in sorting overlapping lines that connected patents and their creators.” Furthermore, the use of 3-D extrusion “allows the user to separate and identify how nodes are related to each other-it was like one of those cognitive thought maps” and “since the entire 2-D workspace was filled with data it was difficult to do any sort of mass organizing with lack of open space - the 3-D did just that. The added dimension gave me a lot of free space for me to use in my organization.”

10 Discussion

These findings suggests that in general questions users were faster using 3-D extrusion to answer directed timed questions than 2-D inspection, although more further studies with a greater number of subjects are needed to assess this effect. In particular, the effect is more pronounced (i.e. significant) on questions that focus on finding fine-grained structure, such as Question 5, which requires “tracing” a route in a dense cluster, and that require searching through and “picking-apart” a dense cluster either for particular information (Question 5) or for purposes of counting (Question 3). Question 1 and 2 were questions whose answer could be deduced by just looking at the overall structure of the network, and for these there was no significant difference. The one exception where subjects achieved a correct answer faster using 2-D than 3-D was on Question 4, which is a simple question in a non-dense section of the visualization that requires no fine-grained tracing of nodes or any inspection of multiple nodes, so in this case extrusion served significantly as a distraction. In conclusion, for tasks involving navigating dense networks for fine-grained results that involve tracing connections between nodes and information search, extrusion into the third-dimension is useful, while it may not be useful and can even be a distraction in making broad observations or finding information that is not hidden in dense clusters. The qualitative feedback from subjects suggests that the added value of the 3-D data extrusion technique lies in the area of being able to explore the organization of the data and the relationships in the underlying structure of the data.

11 Inference and Large Data-Sets

Once the utility of exploring social networks in patent data in 3-D were demonstrated, researchers wanted to use the same technique on larger data-sets. A substantially larger data-set, consisting of 47202 triples in comparison with the data-set used in Section 8 that consisted of 7667 triples, was created by querying the United States Patent Office with the names of all biotechnology companies either based or having research facilities in North Carolina, the “Silicon Valley” of biotechnology. This data-set was aggregated, using GRDDL [12], with large amounts of data about employees, net income of the institution per year, and other information stored in various traditional spreadsheets. By converting

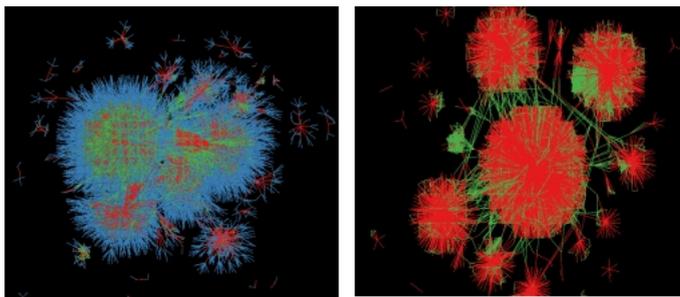


Fig. 3. DiVE: Left - Before inference, Right - After inference and parameter change

these diverse data-sets to RDF and combining them with visualization annotations given by Vis3D, the data was visualized, as shown in Figure 3. The users of the Redgraph were happy that it “quickly showed them the hubs and clusters” but felt that the data-set, even in an immersive environment, was “just overwhelming.” In response to this, inference was used to add a new property that let the patent creator be directly linked to the company they filed a patent for. After the inference, the Vis3D annotations were changed to only visualize this inferred property as opposed to the earlier `assignee` and `inventor` properties. In this manner, the data-set was “filtered” for easier browsing and manipulation, as shown in the contrast between the leftmost and rightmost pictures in Figure 3. As one researcher noted “this makes everything easier.” Ongoing work with these researchers is using Redgraph to understand value chains and platforms in biotechnology that without visualization would be difficult to extract from masses of diverse textual records [20]. Further work aims to elucidate the general principles of abstraction that can be used to help visualize inference.

12 Conclusion and Future Work

There are a number of improvements that can be made to Redgraph. The next visual component to be implemented is a hyperbolic browsing mode that would make nodes and links diminish more rapidly in size the farther away they are from where the user’s point of view as determined by the head-tracker, thus bringing closely related items into higher resolution. This may increase the ability of users to utilize the space of virtual reality systems more effectively, since “the volume of hyperbolic 3-space increases exponentially, as opposed to the familiar geometric increase of Euclidean 3-space” [22]. Some work has begun using hyperbolic browsers for the Semantic Web such as Ontorama, but it allows only very limited types of data to be displayed (i.e. only class hierarchies in OWL ontologies, not properties or instances), and it is unsuitable for many tasks, including the social network visualization done in our study where a large amount of interaction is wanted by the users and when the network is very large [9]. Development of better filtering techniques via dynamic SPARQL querying

and inference animation are also being investigated. Modifications in further user studies will incorporate the findings of the current study to apply Redgraph to understanding connections and underlying structures in other types of data as well as assessing user efficiency at discovering types of information.

Although Redgraph is available as open source,² the full advantages of using fully immersive techniques virtual reality programs like Redgraph are only available to those institutions with a CAVE-like virtual reality environment, although some advantages may be gained by using the program in “2-1/2 dimension” high-resolution environments as well [7]. Yet in order for the advantages of interactive and immersive environments to be more widely available, we are planning to develop a version of Redgraph for conventional desktop use, and ideally a version that would allow it to be incorporated in popular environments like Second Life³, which while lacking immersion, provide popular and three-dimensional forums for the social and collaborative creation and manipulation of visualizations. The lack of open standards in these virtual environments makes development more difficult, but the recent progress of open source 3-D environments like OpenCroquet is encouraging [28]. Standards-based virtual reality integrated with the Semantic Web is the long-term goal for our project. Despite the long path ahead, the future of bringing the Semantic Web into 3-D and immersive environments is bright. The users of the patent innovation data-set, all of whom were ordinary people well-acquainted neither with the Semantic Web nor virtual reality, found using Redgraph to be an enjoyable and exciting way to discover relationships in Semantic Web data.

References

1. Barabasi, A., Albert, R.: Emergence of scaling in random networks. *Science* 286, 509–512 (1999)
2. Bollobas, B.: *Random Graphs*. Academic Press, London (1985)
3. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanara, R., Hollenbach, J., Lerer, A., Sheets, D.: *Tabulator: Exploring and Analyzing Linked Data on the Web*. In: *Proceedings of the Third International Semantic Web User Interaction Workshop*, Athens, Georgia, USA (2006)
4. Bosca, A., Bonino, D.: *OntoSphere3D: a multidimensional visualization tool for ontologies*. In: *The Proceedings of 5th International Workshop on Web Semantics*, Krakow, Poland (2006)
5. Bowman, D., McMahan, R.: *Virtual Reality: How Much Immersion Is Enough?* *Computer* 40(7), 36–43 (2007)
6. Crutcher, L.A., Lazar, A.A., Feiner, S.K., Zhou, M.X.: *Managing networks through a virtual world*. *IEEE Parallel Distributed Technology* 3(2), 4–13 (1995)
7. Cruz-Neira, C., Sandin, D., Defanti, T.: *Surround-screen Projection-Based Virtual Reality: the design and implementation of the CAVE*. In: *The Proceedings of SIGGRAPH 2000*, Anaheim, CA (2000)

² See <http://www.redgraph.org>

³ <http://www.secondlife.com>

8. Eades, P., Houlse, M., Webber, R.: Finding the best viewpoints for three-dimensional graph drawings. In: *The Proceedings of Symposium on Graph Drawing*, Rome, Italy (1997)
9. Eklund, P., Roberts, N., Green, S.: *OntoRama: Browsing RDF Ontologies using a Hyperbolic-style Browser*. In: *The Proceedings of First International Symposium on Cyber Worlds, Theory and Practices*, Toyko, Japan (2002)
10. Ganser, E., North, S.: An open graph visualization system and its applications to software engineering. *Software Practice and Experience* 30(11) (1993)
11. Fruchterman, T., Reingold, E.: Graph drawing by force-directed placement. *Software Practice and Experience* 21(11), 1129–1164 (1991)
12. Halpin, H.: *Social Semantic Mashups with GRDDL and Microformats*. In: *The Proceedings of XML 2006*, Boston, Massachusetts, USA (2006)
13. Hayes, P.: *RDF Semantics, W3C Recommendation* (2004), <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/>
14. Herman, I., Marshall, M.: *GraphXML - An XML-Based Graph Description Format*. In: *The Proceedings of the 8th International Symposium on Graph Drawing*, Williamsburg, Virginia, USA (2000)
15. Herman, I., Melancon, G., Marshall, M.: Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics* 6(1), 24–43 (2000)
16. Huynh, D., Miller, R., Karger, D.: *Exhibit: Lightweight Structured Data Publishing*. In: *The Proceedings of the World Wide Web Conference (WWW 2007)*, Banff, Canada (2007)
17. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. *Information Processing Letters* 31(1), 7–15 (1989)
18. Karger, D., Schraefel, M.: *Position Paper: The Pathetic Fallacy of RDF*. In: *The Proceedings of the International Semantic Web User Interaction Workshop*, Athens, Georgia, USA (2006)
19. Le Grand, B., Soto, M.: *Information management: Topic maps visualization*. In: *The Proceedings of XML Europe 2000*, Paris, France (2000)
20. Lenoir, T., Giannella, E.: *Networks of Innovation: The Emergence and Diffusion of DNA Microarray Technology*. *Journal of Biomedical Discovery* 11(1) (2006)
21. Mika, P.: *Social Networks and the Semantic Web*. Springer, London, United Kingdom (2007)
22. Munzer, T.: *H3: Laying out Large Directed Graphs in 3D Hyperbolic Space*. In: *The Proceedings of the IEEE Symposium on Information Visualization*, Phoenix, Arizona, USA (1997)
23. Mutton, P., Golbeck, J.: *Visualization of semantic metadata and ontologies*. In: *The Proceedings of Seventh International Conference on Information Visualization*, Zurich, Switzerland (2003)
24. Newman, M.: Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46, 323–351 (2005)
25. Pietriga, E.: *IsaViz: a Visual Environment for Browsing and Authoring RDF Models*. In: *The Developer's Track of the World Wide Web Conference (WWW 2002)*, Honolulu, Hawaii, USA (2002)
26. Pittarello, F., Faveri, D.: *A Semantic description of 3D environments: a proposal based on web standards*. In: *The Proceedings of the Eleventh international Conference on 3D Web Technology*, Columbia, Maryland (2006)
27. Schaeffer, B., Goudeseene, C.: *Syzygy: native PC cluster VR*. In: *The Proceedings of Virtual Reality (VR 2003)*, Los Angeles, CA, USA (2003)

28. Smith, D., Raab, A., Reed, D., Kay, A.: Croquet: A Menagerie of New User Interfaces. In: The Proceedings of Conference on Creating, Connecting and Collaborating through Computing (C5), Kyoto, Japan (2004)
29. Ware, C., Franck, G.: Viewing a graph in a virtual reality display is three times as good as a 2D diagram. In: The Proceedings of IEEE Conference on Visual Languages and Human-centric Computing, Rome, Italy (1994)
30. Web3D-Consortium. X3D Specification (2005),
<http://www.web3d.org/x3d/specifications>