



Web-Based Visualization of Large Hierarchical Graphs Using Invisible Links in a Hyperbolic Space

Ming C. Hao, Meichun Hsu, Umesh Dayal, Adrian Krug*

Software Technology Laboratory

HP Laboratories Palo Alto

HPL-2000-2

January, 2000

E-mail:(mhao, mhsu, dayal)@hpl.hp.com

akrug@hpugrca.grc.hp.com

large hierarchical space, invisible links, navigation, hyperbolic space.

Many real-world data warehouse applications involve navigation of large highly connected hierarchies e.g. web pages, product catalogs, and document hot topics hierarchies. Quite often users are confused and overwhelmed with many complex displays. This paper discusses a new invisible link technique for linking a large highly connected graph in a hyperbolic space without cluttering the display. Only the primary links are shown to the user. All other cross-links are hidden in the property of each node and invisible to the user. These invisible links only appear when the user focuses on the node. "Invisible link" allows a user to freely focus on the hierarchy of interest.

Internal Accession Date Only

* Hewlett-Packard GmbH, Pempelfurt Strasse 1, 40882 Ratingen, Germany

© Copyright Hewlett-Packard Company 2000

Web-Based Visualization of Large Hierarchical Graphs Using Invisible Links in a Hyperbolic Space

Ming C. Hao, Meichun Hsu, Umesh Dayal, Adrian Krug
(*mhao, mhsu, dayal*) @*hpl.hp.com*; *akrug@hpugrca.grc.hp.com*
Hewlett Packard Laboratories

Abstract Many real-world data warehouse applications involve navigation of large highly connected hierarchies e.g. web pages, product catalogs, and document hot topics hierarchies. Quite often users are confused and overwhelmed with many complex displays. This paper discusses a new invisible link technique for linking a large highly connected graph in a hyperbolic space without cluttering the display. Only the primary links are shown to the user. All other cross-links are hidden in the property of each node and invisible to the user. These invisible links only appear when the user focuses on the node. “Invisible link” allows a user to freely focus on the hierarchy of interest.

Keywords: Large Hierarchical Space, Invisible Links, Navigation, Hyperbolic Space.

1. INTRODUCTION

Today many web applications involve navigation of large highly connected hierarchies e.g. web pages, product catalogs, and document topic hierarchies [4,5,6]. For example, in data mining, there is an immediate need for users to visualize the content and usage of the web. A difficult problem to solve is how to navigate through millions of interconnected documents to access information on one display.

Hyperbolic space provides an elegant solution to display large hierarchies on a user screen. Hyperbolic space is different from the conventional approaches of laying trees on a Euclidean space. In Euclidean space, the area of the circle to contain nodes grows linearly. In a hyperbolic space, the area of a circle grows exponentially with respect to its radius. As a result, the approach described in [2,4] can handle a graph of over 20,000 documents on the web by using a focus and context scheme. Hyperbolic space allows a user

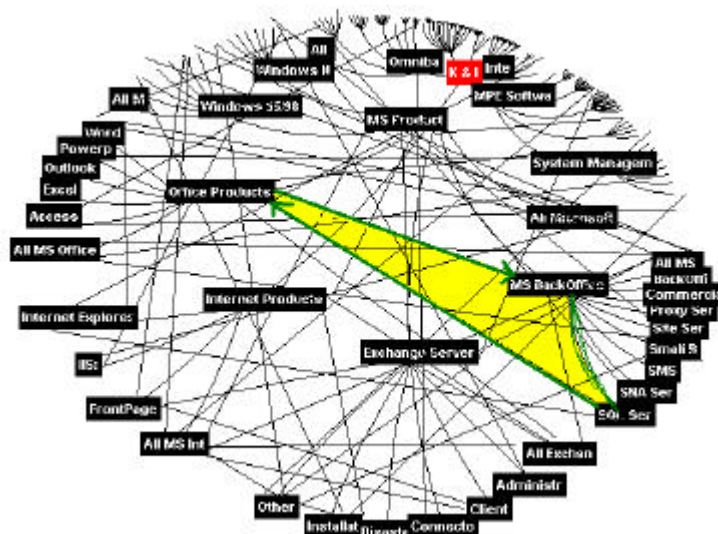
to navigate through the nodes and to see the relationship of the visible portion of the space to the entire structure on one single display. It is not like MindMan [3], which requires multiple displays to represent large amounts of data. The user has to click through display after display to find the information he/she needs.

Based on our practical usage and design experience, we have discovered that hierarchical parent-child tree structures are too restrictive. Often, there are relationships that need to be linked to different branches of a tree [1]. They may even form cycles. For example, in a customer support web application, hyperbolic trees are used to organize large numbers of questions and answers in a hierarchical structure. Questions are parent nodes; Answers are arcs (links) to child nodes. It is sometimes necessary for an answer to link to a set of question/answer in another group that does not directly belong to the hierarchical hyperbolic tree. As a result, the generalization to a hyperbolic space is necessary. However, there will be multiple links with additional lines among all connected nodes in a hyperbolic space. Thus this introduces thousands of lines and intersections. Also, for far away nodes that are off the screen, it would introduce “broken” lines. The hyperbolic space becomes very cluttered and difficult to visualize.

The graph in Figure 1 (A) illustrates the difficulties of visualizing the multiple paths with existing methods. This graph is a hyperbolic space with many highly connected paths. It contains n nodes, $n-1$ primary edges, cycles, and many extra cross edges and intersections. This hyperbolic graph becomes very cluttered.

To date, many practical applications have shown the usefulness of hyperbolic trees on navigating large hierarchies with millions of nodes. This paper generalizes a hyperbolic tree into a hyperbolic space and provides techniques to unclutter the display.

(A) A Cluttered Hyperbolic Tree



* The links among nodes: "Office Products", "MS Back Office" and "SQL Server" form a cycle.

(B) New Method: A Cyclic Hyperbolic Space with Invisible Links

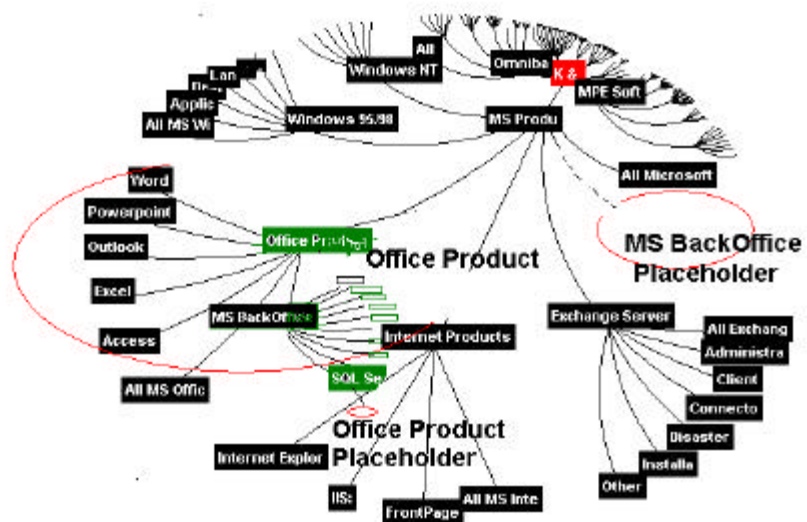


Figure 1: Hyperbolic Space with & without Multiple Paths

2. A NEW TECHNIQUE

In this paper we discuss a new “invisible link ” technique with a placeholder. This technique is used for visualizing a large highly connected graph in a hyperbolic space without cluttering the display. During the visualization, only the primary links are shown. All other types of links (non-tree/secondary links) are made invisible and hidden in the property of a node. The placeholder is designed to track the links including cyclic conditions. These invisible links only appear when the user navigates/clicks on a node. This technique allows a user to freely focus on the hierarchy of interest. The graph in Figure 1 (B) illustrates a new hyperbolic space with multiple invisible links and placeholders. It contains the same number of n nodes, $n-1$ primary edges, and cycles as Figure 1 (A).

2.1 Definitions

In a directed hierarchical space with or without cycles, there is a primary graph, which links all the nodes in a tree form. These links are primary tree links. The others are non-tree/secondary links in a highly connected graph. A node can have one incoming primary link and many incoming non-tree/secondary links, which are called invisible links. A hyperbolic space is defined as a directed hierarchical hyperbolic graph with cross-links and cycles. Figure 2 illustrates the definition of a hyperbolic space with cycles.

Primary Path: (tree-link) “YX”: A directed non-cyclic graphic link in a hierarchical hyperbolic space. Every node has exactly one primary parent (except for the root); the link from a node’s primary parent to the node is a primary path. The primary parent is the first parent of a node. It is defined at the time the hyperbolic graph is created. The other parents linked to that same node are called secondary parents.

Secondary Path (non-tree/cross link) “DX”: A node may have additional secondary parents; the link from a node’s secondary parent to the node is a secondary path.

Invisible Link Node: Node D that serves as an invisible link node contains both the primary path and the secondary path.

Primary Sub-Space Nodes: Defines sub-space parent node and its child nodes linked by a primary path. For example, node D and child nodes E & F, linked by a primary paths DE, DF, are the primary sub-space nodes.

Secondary Sub-Space Nodes: Defines sub-space parent node and its child nodes linked by a secondary path. For example, node X and its child nodes A and B linked by a secondary path DX are the secondary sub-space nodes.

Placeholder: Contains the secondary path between the node and secondary sub-space nodes. For example, as illustrated in Figure 3, node D has a placeholder X, that represents an invisible link to node X. Node B has a placeholder D that contains an invisible secondary link to node D.

As a result, a node (except for the root) has exactly one primary parent in a directed cyclic hierarchical hyperbolic space. If a node has more than one parent, then exactly one is designated as the primary parent, and the others are called secondary parents. For example, in an employee database, the path from the regular/first manager is called the primary path. The path from the employee to the temporary /second manager is called the secondary path.

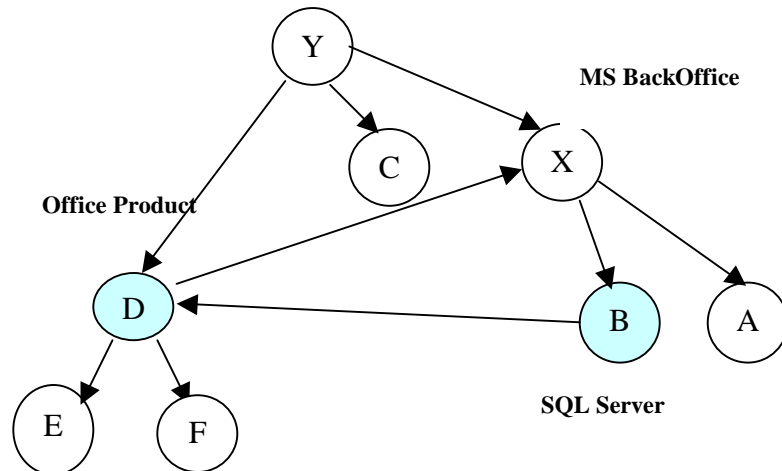
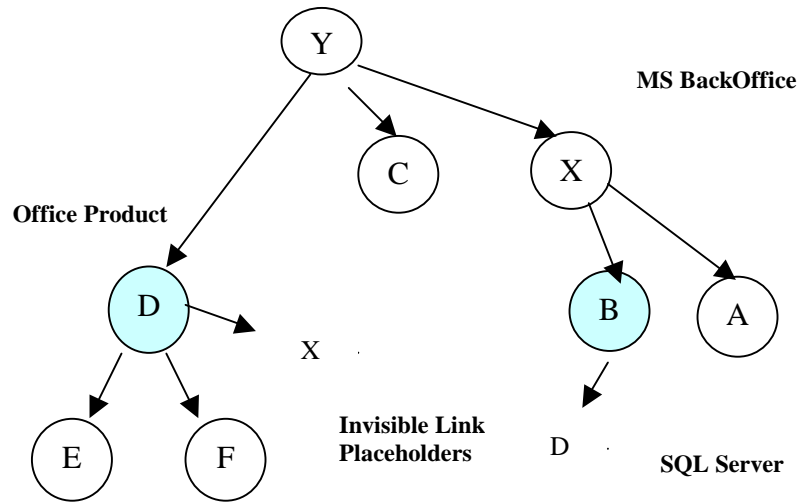


Figure 2: A Definition of Cyclic Hierarchical Space

The above graph displays both the primary and the secondary paths. Note that (1) Node X has 2 parents: primary parent Y and secondary parent D; thus D has a secondary child, X. (2) Node D has 2 parents: primary parent Y and secondary parent B; thus B has a secondary child D. (3) The link (or path) from D to X is a secondary path; so is B to D. (4) This directed graph contains a cycle (X, B, D). (5) Node D, X's child node B has the invisible link property; i.e. it has links to secondary child nodes. In order to represent this cyclic relationship, we need to introduce two extra lines: one from Node D to X, one from Node B to D.



The same graph, showing primary links only, this appears as a tree. This is the view a user is presented with when he starts navigating. Nodes B and D with the invisible link properties are highlighted with color.

- (1) Node D has 2 parents: primary parent Y and invisible secondary parent B. One placeholder to node X.
- (2) Node X has 2 parents: primary parent Y and invisible secondary parent D
- (3) Node B has one parent X and one placeholder to D.

Instead of drawing extra lines (illustrated at Figure 2), this new technique uses an Invisible Link Placeholder residing in the invisible link node X and in D to represent the secondary path to node D and node X.

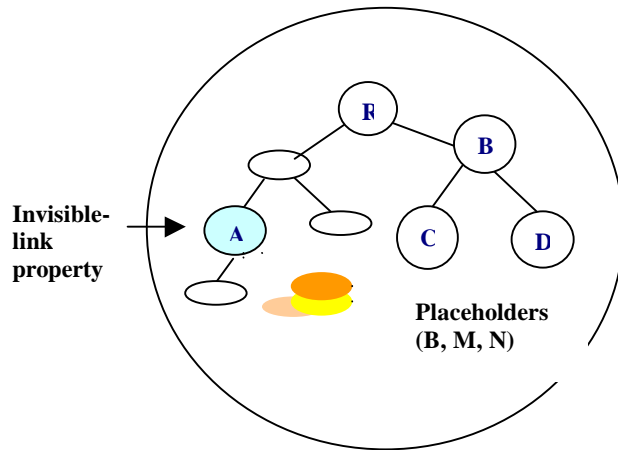
Figure 3: A Directed Hyperbolic Space with Cycles

2.2 Invisible link States and Processing Flow

The invisible link technique is built on a web-based client-server model with multi-threaded parallelism. This technique provides instantaneous mapping and unmapping of secondary linked nodes for navigation in a hyperbolic space. This technique employs placeholders to maintain link relationships. An invisible-link processor is used to manage the following four different processing states:

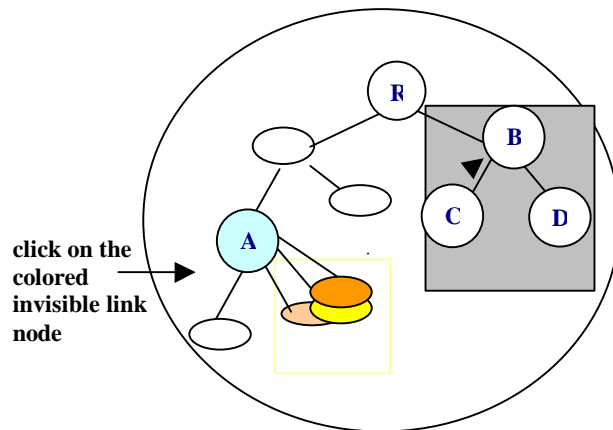
State 1: idle state

A hyperbolic space with invisible link has the same layout as it has without invisible links. There are no extra lines and intersections added in the graph. An invisible link node A is identified with a unique color. It contains invisible placeholders for each secondary path.



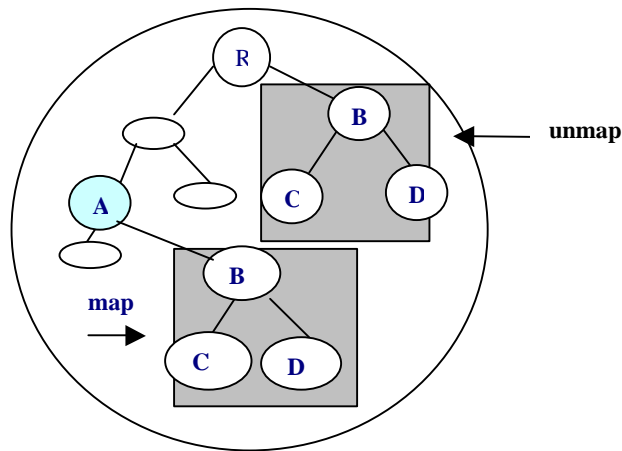
State 2: activate state

When a user is clicking on a node that has invisible links, a number of placeholders will be displayed from the node. Each placeholder contains an invisible “secondary path”. The user can dynamically select a “secondary path” for navigation.



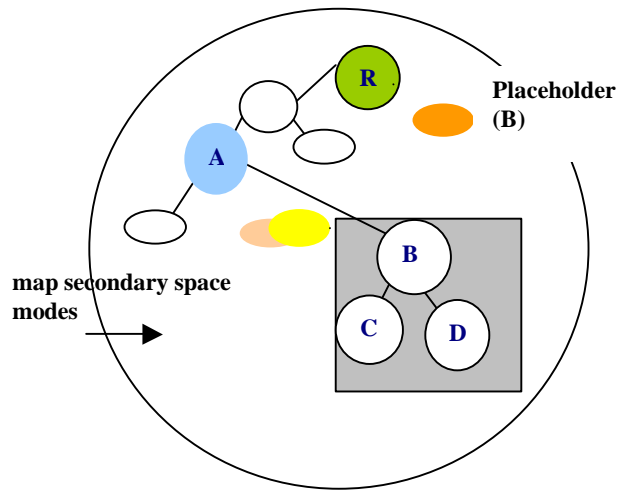
State 3: map/unmap (move) state

The invisible-link processor uses parallel mapping and unmapping methods to move the primary sub-space nodes, the selected node with its children, from their original structure to the current focused node. A placeholder will be placed in the original structure to retain the previous path for mapping.



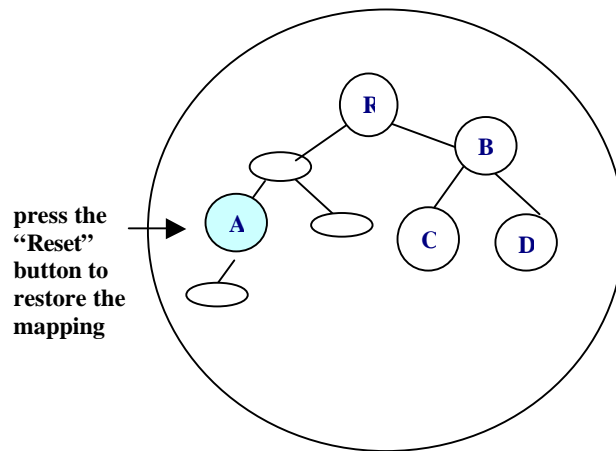
State 4: Navigation State

After the selected secondary sub-space node and its children are mapped under the hidden-link node, the user can start his navigation.



State 5: Reset

At end of navigation, the hidden-link processor dynamically unmaps the secondary sub-space nodes B, C and D from node A, and maps the primary sub-space node B with its child nodes C and D back to the original structure. The hyperbolic space with secondary sub-space and its child nodes only existing during the secondary path processing. The invisible link node's ancestors are checked to continue cyclic mapping of the secondary path.



3. WEB DATA MINING APPLICATIONS

This paper describes the structure of an invisible link visualization server and a client on the Internet for data mining. We use a web browser with Java activator to dynamically create a large hyperbolic space on the web. The visualization Web interfaces are standard HTML and a Java applet, which are used to explore relationships and to retrieve data within a region of interest. The server is integrated with the data warehouse and mining engine. The user is put in the driver seat at the client side to mine the knowledge results. The system allows the user to access large hierarchies with complex links through HTML pages in a Web browser.

There are many data mining applications with large hierarchical cyclic and non-cyclic structures that can be mapped into hyperbolic spaces with invisible links. Figure 2 illustrates a cycle that is formed by OfficeProduct, Microsoft BackOffice, and SQL Server. A user can reach BackOffice products from OfficeProduct, or from SQL Server to find other office products. The invisible link technique enables the user to easily navigate through different links without being overwhelmed with a large number of nodes and paths. Whether

users are able to navigate through answers to link to another set of question and answer group, which does not directly belong to the primary path of the hyperbolic space. The user can easily follow the knowledge values to search for recent patches, technical tips, and editions.

There are many different kinds of data mining information like usage statistics, areas of new knowledge requests, or even search result matches, that can be mapped into the hyperbolic space. The user can analyze this additional knowledge through navigation.

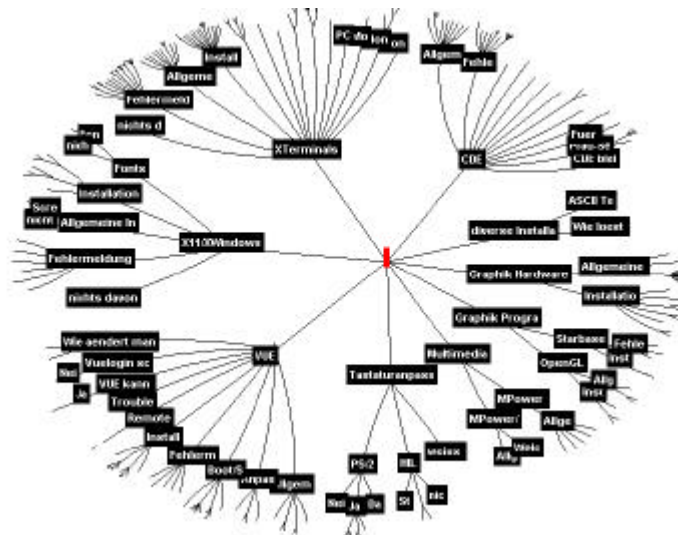


Figure 5: An Interview Application Sample

4. CONCLUSION

Data mining applications in the 90's continuously face difficulties for visual mining of massive large highly connected data sets on the Internet. To date, many practical applications have shown the usefulness of hyperbolic space visualizations [4, 7]. The invisible link technique described in this paper enables the mining of large hierarchical graph with much better visual clarity.

Invisible link enables the user to easily navigate through different links without being overwhelmed with a large number of nodes and paths. Our technique should be extensible to a 3D graph hyperbolic space. In addition, the invisible link technique has been implemented in several prototypes at Hewlett Packard Laboratories.

ACKNOWLEDGEMENT

Thanks to Sharon Beach, Martin Griss, and Graham Pollock from HP Software Technology Laboratory for their advice and suggestions. Also thanks to Jeff Holmbeck, Harold Shinsato from “Inxight Software, Inc.” for their technical support and for allowing us to use their toolkits in our experiments.

REFERENCES

1. Ming C. Hao, Meichun Hsu, Umesh Dayal, Adrian Krug, “Visual Mining Large Web-based Hyperbolic Space Using Hidden Links “, The Third International Conference on The Practical Application of Knowledge Discovery and Data Mining, PADD’99, April.
2. John Lamping and Ramana Rao, “Laying out and Visualizing Large Trees Using a Hyperbolic Space”. ACM /UIST’94, 1994
3. MindMan is a visualization and organizational tool to help individuals to organize, generate and learn ideas and information. 1998
4. Tamara Munster, “Exploring Large Graphs in 3D Hyperbolic Space” IEEE Computer Graphics. Vol. 18, Number 4. 1998
5. Joe C. Pinheiro, Don X Sun, “Methods for Linking and Mining Massive Heterogeneous Databases”, KDD98.
6. George G. Roberson, Jack D. Mackinlay, and Stuart K. Card, “Cone Trees: Animated 3D Visualizations of Hierarchical Information”, 1991 ACM
7. The Hyperbolic Tree Toolkit is a product from Inxight Software for exploiting large amounts of information. 1998

