ELSEVIER

Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list

Yang Xiang[a,*], Michael Chau[b], Homa Atabakhsh[a], Hsinchun Chen[a]

[a]*Department of Management Information Systems, Eller College of Business and Public Administration, The University of Arizona, Tucson, Arizona 85721, USA*
[b]*School of Business, Faculty of Business and Economics, The University of Hong Kong, Pokfulam, Hong Kong*

Abstract

In crime analysis, law enforcement officials have to process a large amount of criminal data and figure out their relationships. It is important to identify different associations among criminal entities. In this paper, we propose the use of a hyperbolic tree view and a hierarchical list view to visualize criminal relationships. A prototype system called COPLINK Criminal Relationship Visualizer was developed. An experiment was conducted to test the effectiveness and the efficiency of the two views. The results show that the hyperbolic tree view is more effective for an "identify" task and more efficient for an "associate" task. The participants generally thought it was easier to use the hierarchical list, with which they were more familiar. When asked about the usefulness of the two views, about half of the participants thought that the hyperbolic tree was more useful, while the other half thought otherwise. Our results indicate that both views can help in criminal relationship visualization. While the hyperbolic tree view performs better in some tasks, the users' experiences and preferences will impact the decision on choosing the visualization technique.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Information visualization; Hyperbolic tree; Hierarchical list; Criminal data analysis

1. Introduction

Information overload has been a challenge facing the police and justice systems. In crime analysis, information overload comes from the large amount of information relevant to a case, and from the great number of associations that crime analysts need to draw between a new case and any related past incidents to pull out leads for investigation [3]. However, it is difficult for crime analysts to identify such associations from textual records. This comprises a set of knowledge-intensive tasks that include identifying different entities from the text, discerning linkages between the entities which are usually cross-records, and mapping the entities and their relationships into a whole picture,

---

* Corresponding author. Tel.: +1 520 621 2748.
E-mail addresses: yangx@bpa.arizona.edu (Y. Xiang),
mchau@business.hku.hk (M. Chau), homa@bpa.arizona.edu
(H. Atabakhsh), hchen@bpa.arizona.edu (H. Chen).

sometimes onto a physical chart or graph, for better understanding. Such tasks require a great amount of mental effort. Considering the large quantity of incidents to be solved, the restrained number of hours that crime analysts can spend on each case, and the desired timeframe for solving a case, it is crucial to reduce the time spent locating relevant criminal information and their associations.

Visualization techniques have been used to address information overload problems. Card et al. [4] identified the following ways in which visual aids can amplify human cognition:

(a) crystallizing knowledge with good visual representation that may abstract patterns, reveal links, and create a concise overall view of the information accessed;
(b) facilitating interactive manipulation of the representation with a working space for control of data access, selection of different levels of data abstraction, and plotting of visually communicable presentations;
(c) reducing the time needed for searching data, discovering patterns, or doing inferences.

It has been exhibited that visual clues are especially helpful in systems that deal with large amounts of real-time, dynamic data to detect patterns or to gain other insights. One example is the file structure map which supports searching and browsing of document collections generated by electronic meeting systems [21]. Other examples include the three-dimensional perspective views of flight plans for detecting possible aircraft route conflicts in the Air Traffic Control system [2], and the Sammon map designed to visualize data in a college selection application [10]. In all three cases, visualization helps discover knowledge in a more effective and timely manner.

While visualization techniques have been shown to be useful in various domains, they have not been widely studied for applications in crime analysis. In this paper, we describe our experience in applying two visualization techniques, namely hierarchical list and hyperbolic tree, to a crime analysis application. The paper is organized as follows. Section 2 reviews different visualization techniques for criminal record relationships and networked data. In Section 3, we discuss the problems in existing approaches and

propose our research questions. Section 4 describes a prototype system called COPLINK Criminal Relationship Visualizer that we developed in this research. In Section 5 we report on a user study designed to test and compare the performance of a hierarchical list view and a hyperbolic tree view which we adopted. In Section 6, we summarize our findings and propose some future research directions.

2. Research background

2.1. Criminal relationship analysis and visualization

In crime analysis, it is often useful to identify the relationships among different entities such as people, vehicles, addresses, organizations, etc. These relationships are usually distinguished using explicit data (such as father–son, suspect–victim, owner–property) or statistical techniques such as co-occurrence analysis [13]. It is critical for crime analysts to retrieve, understand and analyze these relationships in an efficient and effective way.

Because of its cognitive benefits, information visualization has been employed for representing criminal data and relationships. One of the most commonly used visualization techniques in crime analysis is crime mapping, which helps visually locate incidents on a thematic map, search cases based on selected areas, identify spatial relationships among crimes, or expose "hot spots" (areas of high crime). Crime mapping is more suitable for representing information about location, distance, direction, and patterns relating to both time and geographic space [12]. Fig. 1 shows an example of mapping the changes of drug arrests over time onto a city map. The darkness of the colors indicates the degrees of increase or decrease in those arrests from the year 1997 to 1998. However, the technique is not effective for abstracting the relationships between people, places, incidents, organizations, vehicles, criminal weapons, and stolen properties.

Two-dimensional and three-dimensional networks have been studied for exploring crime associations. The Link Discovery Tool developed jointly by the University of Tennessee and the St. Petersburg Police Department is one of the early three-dimensional graphical representations of criminal relationships. It

Fig. 1. A Crime Stats Map generated by the Omega Group. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

visualizes associations such as relatives, criminal contacts, telephone toll data, financial transfers, flow of resources, etc. The tool was used to display a network of wiretapped telephones with the nodes representing individual phones and the links representing incoming and outgoing calls between the phones [16].

Another example of crime analysis systems is the COPLINK Detect, a system that automatically finds various types of criminal associations from crime data [9,13]. It can identify relationships among five types of entities: person, vehicle, organization, location, and incident/crime type. Information generated by the system is displayed in a tabular format. Fig. 2 shows a screenshot of the result page for a search based on three search terms: (1) person "Matthew Clay"; (2) address "1300 W Silverlake Rd"; and (3) vehicle "HOND/1983". In the table, this first column shows the "Add" buttons which allow users to add an entity to perform a new search. The second column lists all entities found in the database which are related to any of the three search terms. The third column lists the id of the search term(s) to which a found entity is related. This display requires users map the ids back to the individual search terms, which makes it difficult to read. Although the entities in the results are all

locations in our example, it is common that the result include mixed entity types. This display does not show the entity types of the items in the search results.

2.2. Information visualization techniques for networked data

Crime analysis data are essentially networked data because the analysis is based on multiple entities (person, vehicle, location, etc.) and there can be a linkage between any two entities. The networking of the data implies that individual entities interact with other individual entities, which in turn interact with still others [17]. A network is usually built within a distance from a particular entity and includes all relevant interactions and linkages. It can then be used to detect the importance of particular entities in that network.

Intuitively, network charts are suitable for visualizing networked data. Nodes can be designated as entities while edges can represent interactions and linkages between entities/nodes. Network charts can closely represent the real data and their relations. However, they are often too complex for display because of the "clutter" problem (overcrowded display) and the "perception" problem (different values
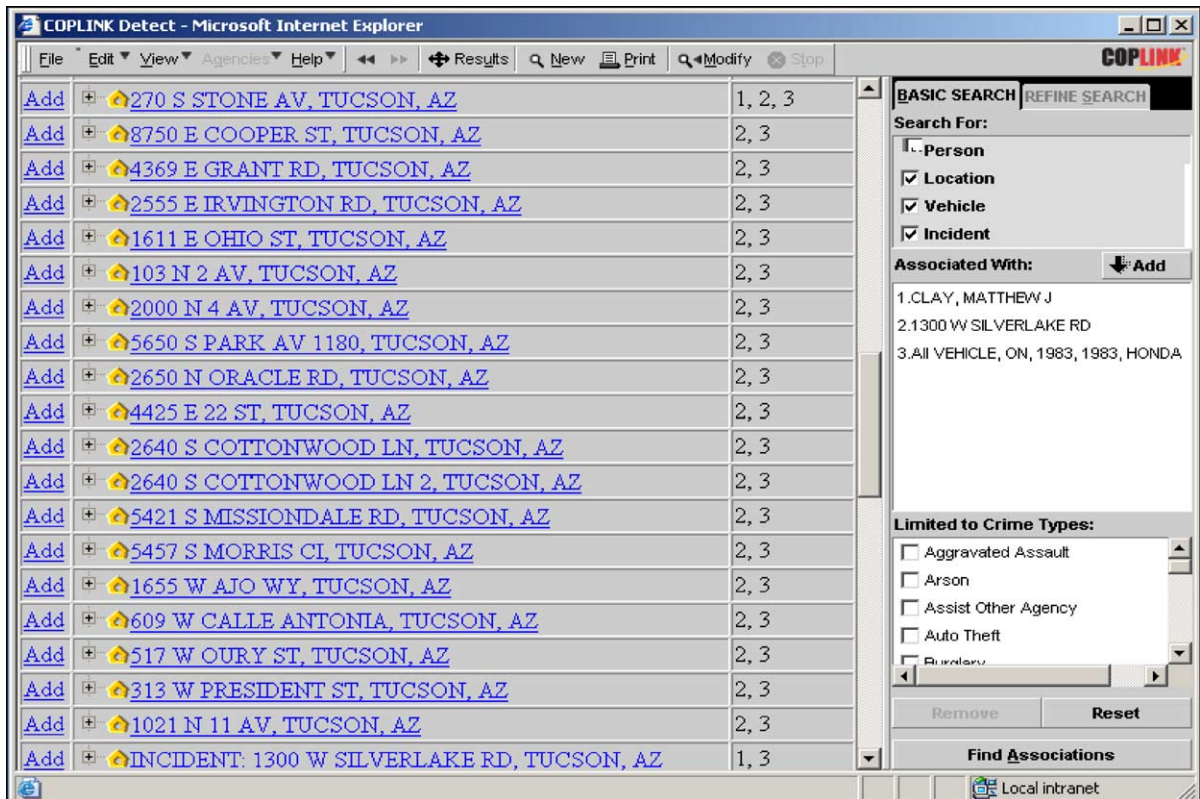
Fig. 2. Entities related to the search terms are shown in a tabular form in COPLINK Detect.

that humans perceive from visual clues) [11]. As a result, complicated technologies, such as clustering or zooming, are often required for effective visual representation [28].

To address these issues, other techniques have been used to render large networks. A tree metaphor is one of the alternatives. Trees have most advantages that networks have especially when cycles are not very important or if we consider trees as simplified networks [4]. The most commonly used tree structure visualization is a hierarchical folder view (e.g., the Microsoft Windows Explorer shown in Fig. 3). The folder/subfolder structure creates a clear sense of hierarchical structure and linkage. Another representation of a tree structure is the hyperbolic tree developed by Xerox PARC [19,20]. Unlike the folder view which extends along horizontal or vertical lines, the hyperbolic tree radiates towards the edge of a circular space (see Fig. 4). Meanwhile, a hyperbolic tree is mostly used to visualize large hierarchies of data, such as Web site structures [14]. Its advantages include increasing the amount of information immediately available to users and placing more information into their attention [25,26]. The hierarchical list view and the hyperbolic tree view have been compared in several previous studies, but no conclusive results have been drawn regarding their performance [20,23,26].

Although the tree metaphor is easy to understand and widely used, it has its weaknesses in visualizing networked data. As a tree cannot represent cycles, it cannot show the link between items in different sub-trees in the hierarchy and users can easily miss the relationships between these items. Also, hierarchical structures are often used to represent "is a" or "composite" types of relationships, which are different from the "degree of separation" or "associate" relationships that often exist in networked data.
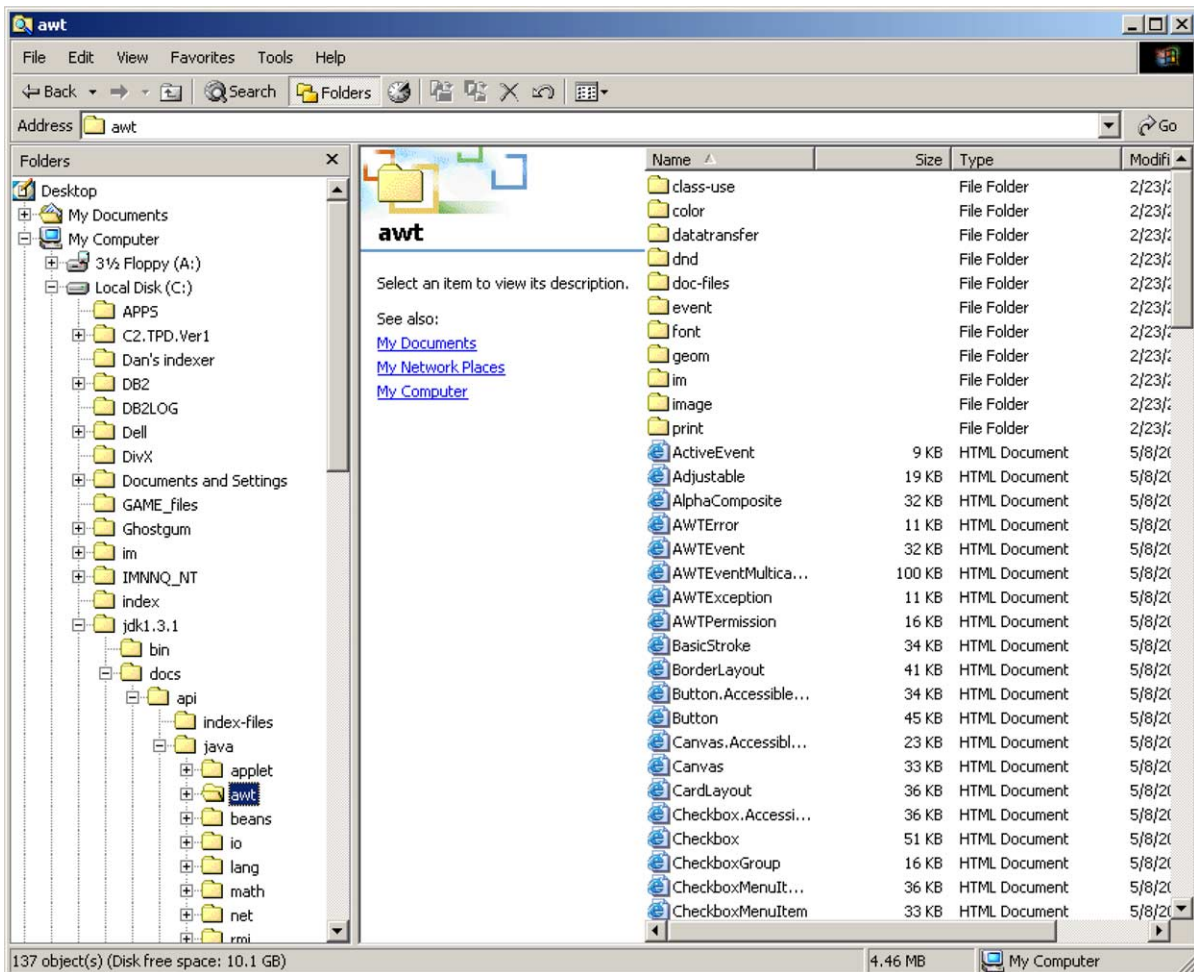
Fig. 3. Microsoft Windows Explorer.

2.3. Problems with existing approaches

There are two major problems with existing crime analysis and visualization applications. First, many tools (e.g., the Link Discovery Tool) only deal with one entity type instead of different entities such as persons, locations, and vehicles. Such tools are not able to identify the link between two persons if they are associated with other entities (e.g., if the two persons have used the same vehicle). The second problem is the insufficient or ineffective visualization support of existing tools. While network charts are most commonly used among crime analysts, they are usually drawn manually. Few software applications

are available for generating network charts for crime analysis purposes. The development of powerful relational databases has enabled crime analysts to search among voluminous records to retrieve relevant criminal data. However, using existing tools, most of these data are either too crowded to be see clearly (e.g., Link Discovery Tool) or can only be viewed in tabular format (e.g., COPLINK Detect), making it difficult for crime analysts to see the big picture and draw connections. A large amount of mental effort is still needed to understand and interpret the relationships among criminal data. Considering the tradeoff between levels of abstraction and amount of information, and thus levels of simplicity and legibility,
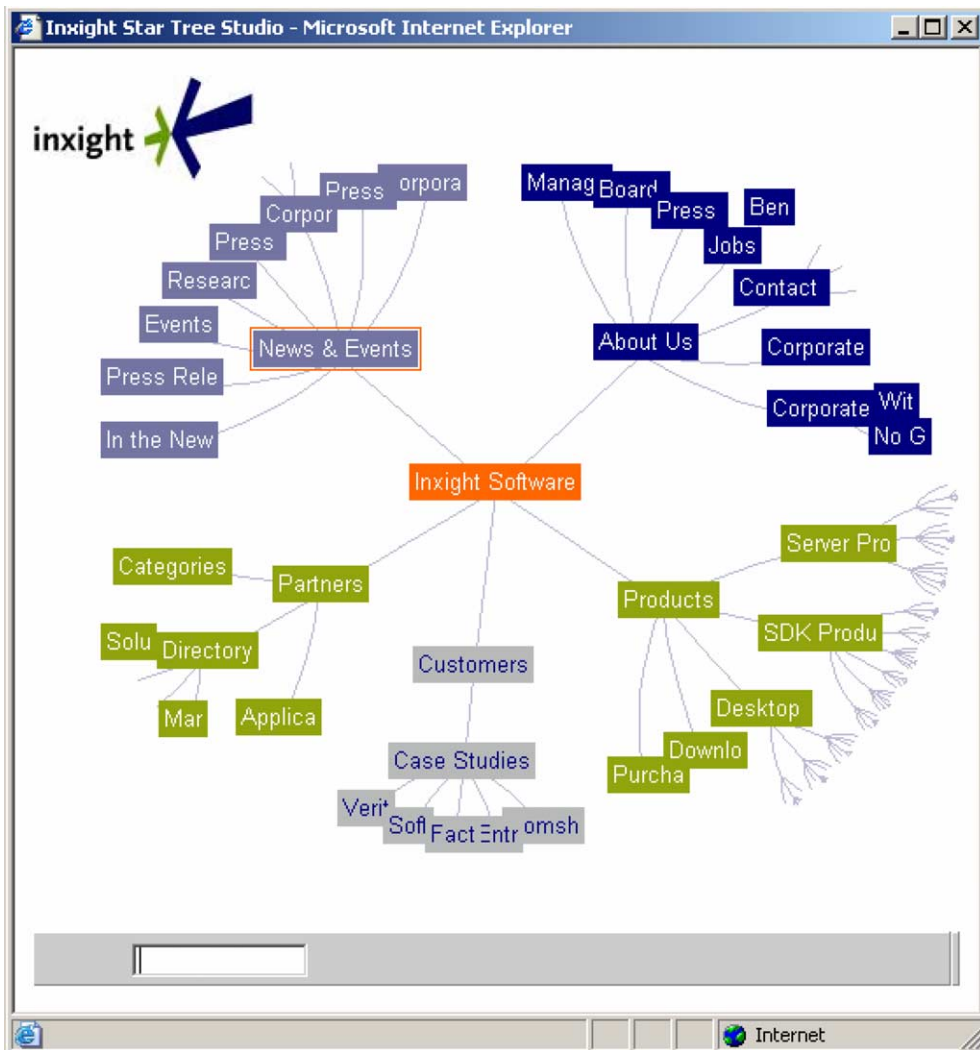
Fig. 4. The Star Tree Site Map of InXight Software.

alternate visualization techniques are examined for visualizing criminal relationships.

3. Research questions

In this study, we focus on the second problem discussed above, i.e., the visualization problem. Based on our review, we have identified the hierarchical list and the hyperbolic tree as two of the options. According to Card et al. [4], nodes and links in a tree can signify relations among objects without the constraint of mapping variables onto multi-dimen- sional axes. Thus, a tree has the potential of using space more effectively than a map. In both hierarchical list and hyperbolic tree, the root (first-level folder) and child nodes (various levels of subfolders and items inside) can be used to denote objects, while the links between them can represent connections. Different colors can be applied to the nodes to distinguish various types of objects. When used to search relations between different entities in crime analysis, they can show all the entities related to a given search term or be further expanded to show multiple levels of associations. Since the relative positions of nodes and links to one another show proximity, the visual properties express some

underlying data relationship such as clustering or partial trend. In crime analysis, these visual cues can signal heavily involved entities (according to the number of their associations), or members of a criminal group (when their individual clusters of associations show great similarity).

Previous experiments comparing a hyperbolic tree browser and other text-based browsers exhibit significantly higher effectiveness in browsing and searching with a hyperbolic tree [23,24,26]. However, the experiment tasks were primarily designed around general retrieval tasks rather than other analysis tasks such as identifying clusters and associations in the data. Our study focuses on comparing the potential of the hyperbolic tree and the hierarchical list in visualizing networked data to support various tasks in law enforcement. In this study, we investigate the following research questions:

- Can we apply the hyperbolic tree and the hierarchical list to criminal relationship visualization?

- Is the hyperbolic tree a better method than the hierarchical list for users to find information? We consider a system "better" if it can help users get the correct results more easily and quickly.
- Does the hyperbolic tree present more information/ knowledge than does the hierarchical list based on the same data set?

4. COPLINK Criminal Relationship Visualizer

To study these issues, we developed the COPLINK Criminal Relationship Visualizer, a prototype user interface for interactively retrieving and visualizing the criminal relationship data stored in the COPLINK Detect database which we developed in our previous research [9,13]. COPLINK is a joint project between the University of Arizona Artificial Intelligence Lab and the Tucson Police Department (TPD), started in 1997 and funded by the National Institute of Justice (NIJ) and the National Science Foundation (NSF).
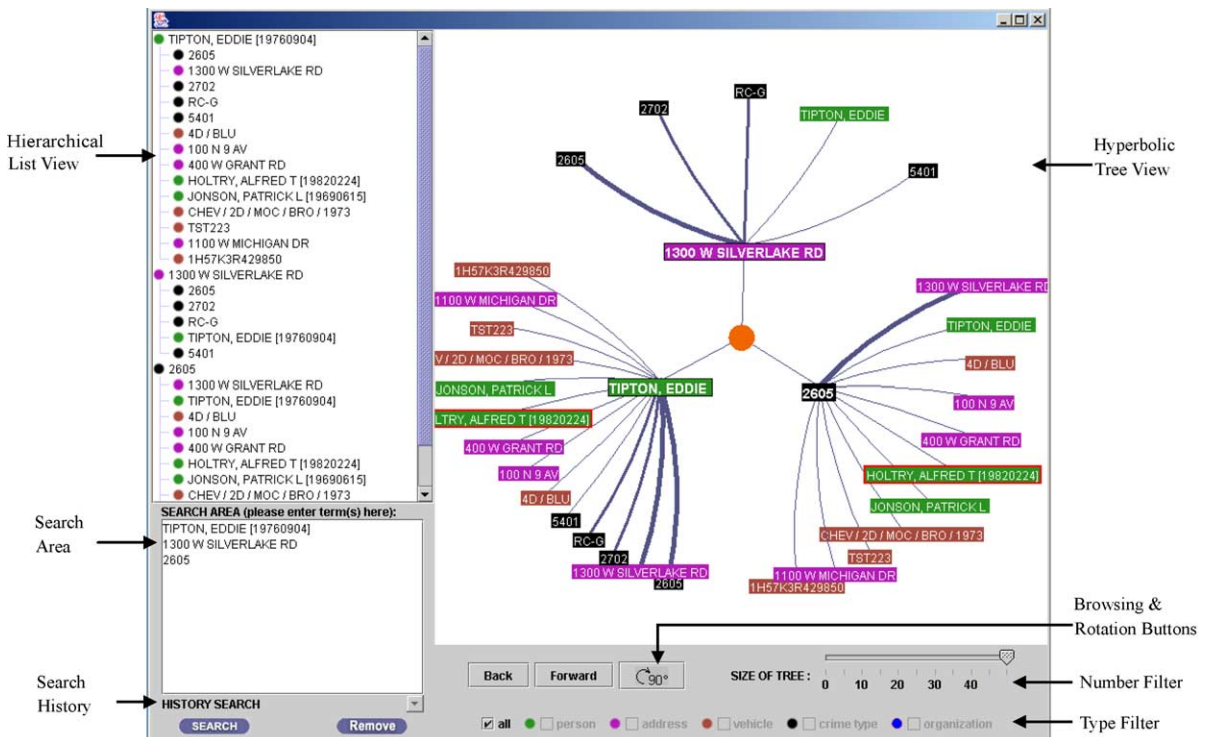

Fig. 5. The COPLINK Visualizer. Both views group search results by search terms.

The project aims to construct an efficient, user-friendly information analysis and sharing system by integrating data from various sources and enabling law enforcement agencies to use data and information more effectively [8]. The COPLINK Detect database currently contains approximately two million criminal case reports from the TPD.

COPLINK Detect analyzes crime data based on co-occurrence analysis [9,13]. Assuming that two entities appearing in the same crimes may have an association, a nonzero co-occurrence weight can indicate the existence of an association. The higher a co-occurrence weight, the more likely that the two entities involved have a strong association. COPLINK Detect calculates the co-occurrence score between each pair of entities in the database. When a search query is submitted to the system, the entities having a high co-

occurrence weights will be retrieved and used to generate the tree in the COPLINK Visualizer.

The COPLINK Visualizer includes a hyperbolic tree view and a hierarchical list view implemented in Java. The hyperbolic tree was developed based on a freely available program written in Java (http://www.soi.city.ac.uk/~livantes/Research.html). Users can submit search terms to the system and the search terms have to be one of the five chosen entities, namely person, vehicle, organization, location, and incident/crime type. Both views group results by search terms (see Fig. 5). The connections between the search term(s) and the search results are displayed using parent–child relationships between nodes. In the hyperbolic tree, the center node is the search term when there is only one, and the nodes immediately surrounding the center node are the search terms when

Fig. 6. In the hyperbolic tree view, users can focus on one part of the tree enlarged in the center while still viewing the rest of the tree that is smaller and further away from the center.

there are multiple search terms; results are displayed as the children of these nodes. In the hierarchical list, the search term(s) is/are the first-level item(s); results are displayed with horizontal indents. On both views, different colors are used to denote the five entity types, making it obvious to which types of entities the search term(s) is/are related. Currently, the COPLINK Visualizer only shows the associations among different entities; it does not handle supertype/subtype relationships.

The hyperbolic tree view, with its "focus+context" features, provides more advantages to users' need for visualization. First, users usually need both context and focus for viewing related data. For our application, the context refers to the overall patterns and the focus refers to detailed information about the object of interest for criminal relationship detection. Second, users prefer less scrolling for maintaining both global structure and local details at the same time [20]. Fig. 6 shows that using the hyperbolic tree, users can focus on one part of the tree and view that section enlarged and put in the center, while still viewing the rest of the tree that is smaller and farther away from the center. The hierarchical list is limited on this feature as the information space grows larger and requires more physical space for display, both horizontally and vertically [15]. However, the hierarchical list is easy to understand and many users are strongly rooted in this type of conventional linear displays [6]. Both views are included in the COPLINK Visualizer to accommodate users' varied preferences. The views also share the same function area, which provides the following functionalities:

- *Search Area* allows users to enter new search term(s) and start a new basic search.
- *Search History* records users' searching history.
- *Browsing Buttons* help users to go back or forward one step at a time in their search history list and redisplay a basic search.
- *Number Filter* and *Type Filter* enable users to reduce the number of nodes to show by electing how many nodes and/or what type or combinations of types of entities they want to view. The *Number Filter* filters out nodes having a low co-occurrence weight with the search terms, while the *Type Filter* allows user to remove any chosen entity types from the display.

- *Rotation Button* allows users to rotate the graph clockwise by 90° for better layout of the tree.

5. Empirical study

We designed an empirical study to examine the effectiveness and efficiency of the two views in the COPLINK Visualizer in presenting to users the relations between entities in crime data. The study looked at how different users' performances were in information searching using either the hyperbolic tree view or the hierarchical list view. The two views are readily comparable as they share the following similarities: (1) both views connect to the same relational database of scrubbed criminal relationships from the TPD; (2) they both group information by search terms; (3) under each search term, both views align search results according to their closeness to the search term; and (4) both provide similar functions such as searching and filtering.

5.1. Task design

A set of low-level, domain-independent visual tasks were used to examine general and fundamental steps users perform using an interface when trying to retrieve information [22,29]. Examples of these low-level tasks include *locate*, *identify*, *distinguish*, *emphasize*, *reveal*, *categorize*, *cluster*, *distribution*, *rank*, *compare*, *associate*, and *correlate* [22,29,31]. By using these low-level visual tasks, we can eliminate features that are specific to the domain of a visualization application and focus only on the visualization components in system evaluation. This approach is often known as the "defeaturing" approach [22,32].

In crime investigation and analysis, it is often necessary to visually identify important objects and patterns, compare different objects, search for specific objects and groups, or identify the associations between objects [13]. Any of these tasks can be broken down into a sequence of low-level tasks. For example, to find the relationships between two people, a crime analyst would need to *locate* or *identify* the two people, *compare* their attributes, and *associate* them based on the attributes. These low-level tasks often enable users to search or verify information on the display [31].

After consulting with our domain expert, a detective at TPD who has been working in law enforcement for more than 20 years, the following five basic task types were selected for our experiment:

- *Associate:* form relationships between objects in a display.
  ◦ Sample question: List all entities that link the organizations "QUIZON" and "KOFRON".
- *Cluster*: find the similarity among visual objects that have multiple attributes.
  ◦ Sample question: Identify the person directly associated with "1300 W SILVERLAKE RD" who has a similar display pattern (in terms of related entity types and number of results) to "EMMICK, JEFFERY S [19570407]".
- *Compare*: compare objects based on certain attribute.
  ◦ Sample question: Of the following two addresses: "100 N 9 AV" and "100 W GRANT RD", which one has more types of crimes that occur there?
- *Identify*: find a visual object with a certain feature.
  ◦ Sample question: Of all people directly associated with "TST331", who has/have committed a crime of type "0403"?
- *Rank*: find the extremes (the best and the worst cases).
  ◦ Sample question: Of the five search terms, which one has the largest number of vehicles directly related to it?

Altogether, 20 objective tasks (four for each search task type) were used in the experiment. By "objective tasks" we refer to those whose solutions are independent of individuals' personal experience and are generally agreed upon. Our expert verified that these tasks, though general in nature, would be useful in helping targeted users (i.e., crime analysts) better understand and retrieve criminal relationships.

5.2. Hypotheses

Based on our experience with the two views, the following hypotheses have been proposed.

H1. For an "associate" task, the hyperbolic tree has similar effectiveness to and higher efficiency than the hierarchical list.

The global structure provided by a hyperbolic tree makes it easier to see connections among objects. On the other hand, the hierarchical list can only list items in one dimension (vertically) and there is a limit on the maximum number of nodes that can be displayed on the same screen. It requires users to remember more and slows them in understanding associations. Therefore in the "associate" task, participants were expected to take less time finding connections using the hyperbolic tree than using the hierarchical list.

H2. For a "cluster" task, the hyperbolic tree has similar effectiveness to and higher efficiency than the hierarchical list.

Because a "cluster" task requires recognition of many attributes of an object, the visual clues in the hyperbolic tree view would reduce users' need to memorize the attributes of different objects and thus result in a lower possibility of error and shorter completion time. However, if the visual contrasts among result groups are not obvious, both views would have similar effectiveness and efficiency.

H3. For a "compare" task, the hyperbolic tree has similar effectiveness and efficiency to the hierarchical list.

Since a "compare" task involves a few objects and requires few visual clues, both views would show similar effectiveness and efficiency. Although the hyperbolic view allows users to view and compare more objects on the same screen, it may not be helpful when the contrast between the two designated objects on the specified attribute is not very obvious.

H4. For an "identify" task, the hyperbolic tree has similar effectiveness and efficiency to the hierarchical list.

For an "identify" task, users need to browse through and see the detail of a large number of objects. The visual clues in a hyperbolic tree view would not make much difference in task performance here. Although the hyperbolic tree view would allow users to find the required object without scrolling, it may become crowded as the number of objects grows. Therefore, we hypothesized that the two views perform similarly.

H5. For a "rank" task, the hyperbolic tree has similar effectiveness and efficiency to the hierarchical list.

For a "rank" task, users need to browse through a large number of objects. We hypothesize that both views show similar effectiveness and efficiency, especially when the contrast among the objects is not obvious. The hyperbolic view allows users to see and compare more objects on the same screen, but the view would become less controllable in terms of enlarging an area of interest to see the detail, whereas one can always look at any local detail in a hierarchical list view.

5.3. Experimental setup

Two experimental conditions were created for the experiment, one with only the hierarchical list view, and the other with only the hyperbolic tree view. Both views were connected to the same database which contained scrubbed criminal data.

A total of 42 undergraduate students at the University of Arizona were recruited as participants. Each participant went through four experimental steps: (1) introduction and training, (2) demographic information survey, (3) system testing, and (4) post-test questionnaire survey. During system testing, each participant performed five tasks using each view, namely the hierarchical list view and the hyperbolic tree view. For each view, each participant was randomly assigned to one of the categories: single-term search or multiple-term search. Under the assigned category, each participant was randomly assigned one of the two task sets (with five tasks in each set). Thus 20 tasks were designed in total (two

sets of single-term search tasks and two sets of multiple-term search tasks). In the post-test questionnaire, we asked the subjects about their perceived ease of use and usefulness of the two views. Because of incomplete or missing data, the data of three participants were discarded. The data of the remaining 39 participants were used throughout our analysis.

5.4. Results and discussion

To measure users' performance of the two views, we adopted measures proposed and used in previous visualization research [18,32]. We use the percentage of accurate answers as the measure of effectiveness and length of response time as the measure of efficiency. Users' perceived evaluations are also examined to see if the visualizer has cognitive overload or disorientation problems [14,30].

5.4.1. Quantitative analysis

Since each participant provided one data point for each task comparing the two views, we obtained in total 39 observations for each task. A two-way t-test was run to compare the performance differences between the Hyperbolic Tree view and the Hierarchical List view for each task. The results are shown in Table 1.

5.4.1.1. Associate—H1 supported. For the "associate" task, the hyperbolic tree view was faster than the hierarchical list, and the difference was statistically significant (at $p < 0.05$). The hyperbolic tree view also had a slightly higher accuracy rate, but the difference was not statistically significant. Hyperbolic tree provides the global structure even with an expanded

Table 1
Experimental results

Task	Measure	Hierarchical list	Hyperbolic tree	n	t-test p-value
Associate	effectiveness (percentage of accuracy)	84.6%	87.2%	39	0.744
	efficiency (seconds)	130.8	99.9		0.043
Cluster	effectiveness (percentage of accuracy)	84.6%	82.1%	39	0.786
	efficiency (seconds)	74.8	83.1		0.613
Compare	effectiveness (percentage of accuracy)	84.6%	74.4%	39	0.291
	efficiency (seconds)	75.4	81.5		0.588
Identify	effectiveness (percentage of accuracy)	69.2%	87.2%	39	0.070
	efficiency (seconds)	80.7	74.9		0.536
Rank	effectiveness (percentage of accuracy)	97.4%	97.4%	39	1.000
	efficiency (seconds)	120.5	136.4		0.395

search, thus making it easier to associate different entities. However, the subjects felt it was hard to manipulate the movement of the hyperbolic tree graph to exhibit the text of the nodes. Browsing through the hierarchical list took longer, but subjects felt more comfortable controlling the view. Although it is cognitively more demanding to find relations in this fashion, the subjects still yielded a high accuracy rate but with significantly more time.

5.4.1.2. Cluster—H2 not supported. The hyperbolic tree view provides visual clues for the distribution of entity types and the convenience of comparing each required item to the benchmark at the same time. However, since the "cluster" task requires subjects to browse through many attributes and visual objects, the filters provided became unusable and no other methods were available for reducing the crowding in the hyperbolic tree view. Under this condition, the hierarchical list view, with its clear text and manipulative display, yielded better results although we did not observe statistically significant difference.

5.4.1.3. Compare—H3 supported. A "compare" task involves a few objects and does not require a lot of visual clues. As a result, the hierarchical list and the hyperbolic tree views did not show significant difference in effectiveness or efficiency. In fact, the hierarchical list view showed a slightly better result because the view is less crowded, making it easier for the participants to count the results.

5.4.1.4. Identify—H4 not supported. For the "identify" task, the hyperbolic tree was slightly faster than the hierarchical list, but not statistically significant. What was unexpected was the statistically higher accuracy rate of the hyperbolic tree. Since the hierarchical list view displays text more clearly and systematically, we hypothesized that the hierarchical list should have shown similar effectiveness if not higher. One possible reason for the unexpected result might be the way subjects interacted with the system. We found that subjects used filters to reduce information overload when adopting the hyperbolic tree view since the view would be overcrowded otherwise. On the other hand, when using the hierarchical list view, they trusted their own counting and did not bother to use the type filter. As a result, many of the users were

overwhelmed by the large amount of information that needed to be remembered.

5.4.1.5. Rank—H5 supported. In a "rank" task, subjects need to browse through a large number of objects. Although the density of links in the hyperbolic tree view can help identify the extreme case, it is not very useful when the contrast is not very obvious or more than one sub-tree has a large number of child nodes. In these situations, the clear layout of the hierarchical list view makes browsing faster than in the hyperbolic tree view.

5.4.2. Qualitative analysis

None of our undergraduate student participants knew much about crime data or crime analysis systems; they might have heard about such systems but had never used them before. However, most of them were able to understand the requirements of the test questions after the short training session. When asked after the experiment which view was preferred in terms of ease of use, 76.9% of the participants voted for the hierarchical list view. Their reasons include: (1) similarity to Windows Explorer-like interface reduces stress over information search in a new application and new domain; and (2) clearer text details enable easy accessibility—reading, following links, and counting. The major weaknesses of the hyperbolic tree views that the participants pointed out include: (1) the graph becomes difficult to read when a large number of result nodes are returned; (2) when the visual objects overlap one another, the view becomes less trustworthy since some nodes might be hidden behind; and (3) there is a longer learning curve for controlling the hyperbolic tree view.

When asked which view was preferred in terms of usefulness, the votes are almost equally split between the two views (48.7% for the hierarchical list vs. 51.3% for the hyperbolic tree). Those who prefer the hierarchical list emphasized that the list view organizes information in a familiar, orderly way and makes the connections understandable. Those who prefer the hyperbolic tree listed their reasons as: (1) visual clues are helpful for seeing links and dependencies between entities; (2) it is useful for identifying patterns; (3) it is efficient for comparing and contrasting number of entities related to certain nodes; (4) it is good for gaining an overall view of the information searched;

(5) when one is particularly interested in a node, one can view direct associations of the node while still maintaining a picture of where one comes from.

Some participants exhibited hesitation when expanded searches were needed for finding the answers. The observation reflected that some participants were not sure where to find information although they knew that the information existed somewhere in the system. It is a sign that users need more training in order to fully grasp the necessary functions of the COPLINK Visualizer.

In order to see whether the COPLINK Visualizer is useful to law enforcement personnel, we also showed it to our domain expert (the detective) and let him perform the search tasks. Overall, he was pleased with the system and he believed the system could be useful in real-world situations in the following ways:

- to considerably reduce the time to draw connections between certain given entities;
- to visualize a group of associations for effective presentation of a case;
- to visually track the search paths in crime analysis and help find hidden links that are not recorded in the database but are known to crime analysts;
- to help detect the significant entities which have a large number of associations.

6. Conclusions and future directions

Our experiment has shown that the hyperbolic tree view has higher efficiency than the hierarchical list in the "identify task" and higher effectiveness in the "associate task". Although the results do not show significantly superior performance for the hyperbolic tree in other tasks, we did find that the hyperbolic tree performed well even for first-time users, and that it allowed users to draw conclusions more easily. We also observed that for visualization applications, users' familiarities with the views often decide their preferences. It suggests that training will be necessary when deploying our application to law enforcement agencies.

Our study also provides some insights to the design of information retrieval systems. Because of the growth of the Internet and the lowering cost of computer hardware and software, more and more information is available. Networked data besides criminal relationship data, such as similarities between customers or relationships between products, are often available but difficult to retrieve. Our findings indicate that effective visualization techniques would be necessary and valuable in information retrieval systems involving such networked data.

One future direction is to investigate the effect of the size of the data on the performance of the two views. It would be interesting to study whether the difference between the two views would be more significant for search results with more nodes. We are also exploring the use of other visualization techniques for other tasks that are currently performed manually by crime analysts. For example, we are investigating the use of an interactive timeline builder [27] and a periodic-pattern tool [5] to visualize the temporal dimension of crime data. We are also developing a geomapping tool for visualization of crime data and relations [1,7]. We are currently deploying some of our systems to the Tucson Police Department and we hope these tools will help law enforcement agencies to fight crimes more effectively.

References

[1] N. Andrienko, G. Andrienko, P. Gatalsky, Visualization of spatio-temporal information in the Internet, The 11th International Workshop on Database and Expert Systems Applications Proc., 2000 (September 4–8), pp. 577–585, London, UK.

[2] R. Azuma, H. Neely, M. Daily, R. Geiss, Visualization tools for free flight air-traffic management, IEEE Computer Graphics and Applications 20 (5) (2000) 32–36.

[3] D.E. Brown, S.C. Hagen, Correlation analysis for decision support with applications to law enforcement, Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 1999, pp. 1074–1078, Tokyo, Japan.

[4] S.K. Card, J.D. MacKinlay, B. Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, San Francisco, 1999.

[5] J.V. Carlis, J.A. Konstan, Interactive visualization of serial periodic data, User Interface Software and Technology (UIST) '98 Conference Proceedings, 1998, pp. 29–38, New York, USA.

[6] H. Chen, A.L. Houston, R.R. Sewell, B.R. Schatz, Internet browsing and searching: user evaluation of category map and concept space techniques, Journal of the American Society for Information Science 49 (7) (1998) 582–603.

[7] H. Chen, H. Atabakhsh, D. Zeng, J. Schroeder, T. Petersen, D. Casey, M. Chen, Y. Xiang, D. Daspit, S. Nandiraju, S. Fu, COPLINK: visualization and collaboration for law enforcement, Proceedings of the National Conference for Digital Government Research, 2002 (May), Los Angeles, USA.

[8] H. Chen, J. Schroeder, R.V. Hauck, L. Ridgeway, H. Atabakhsh, H. Gupta, C. Boarman, K. Rasmussen, A.W. Clements, COPLINK connect: information and knowledge management for law enforcement, Decision Support Systems 34 (3) (2003) 271–285.

[9] H. Chen, D. Zeng, H. Atabakhsh, W. Wyzga, J. Schroeder, COPLINK: managing law enforcement data and knowledge, Communications of the ACM 46 (1) (2003) 28–34.

[10] E. Condon, B. Golden, S. Lele, S. Raghavan, E. Wasil, A visualization model based on adjacency data, Decision Support Systems 33 (4) (2002) 349–362.

[11] S.T. Eick, Aspects of network visualization, IEEE Computer Graphics and Applications 16 (2) (1996) 69–72.

[12] K. Harries, Context and Concepts, Mapping Crime: Principle and Practice, U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC, 1999.

[13] R.V. Hauck, H. Atabakhsh, P. Ongvasith, H. Gupta, H. Chen, Using COPLINK to analyze criminal-justice data, IEEE Computer 35 (3) (2002) 30–37.

[14] M. Heo, S.C. Hirtle, An empirical comparison of visualization tools to assist information retrieval on the web, Journal of the American Society for Information Science and Technology 52 (8) (2001) 666–675.

[15] R.P. Holley, R.E. Killheffer, Is there an answer to the subject access crisis? Cataloging and Classification Quarterly 1 (2) (1982) 125–133.

[16] R.D. Horn, J.D. Birdwell, Link Discovery Tool, Office of National Drug Control Policy/Counterdrug Technology Assessment Center 1997 International Symposium, 1997 (August 18–22), Chicago, Illinois.

[17] F.A.J. Ianni, E. Reuss-Ianni, Network analysis, in: P. Andrews, M.B. Peterson (Eds.), Criminal Intelligence Analysis, Palmer Enterprises, California, USA, 1990, pp. 67–84.

[18] P.W. Jordan, An Introduction to Usability, Taylor & Francis, Bristol, Pennsylvania, USA, 1998.

[19] J. Lamping, R. Rao, Laying out and visualizing large trees using a hyperbolic space, Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '94), 1994 (November), pp. 13–14, Marina del Rey, California.

[20] J. Lamping, R. Rao, P. Pirolli, A focus+context technique based on hyperbolic geometry for visualizing large hierarchies, ACM CHI '95 Conference on Human Factors in Computing Systems, 1995 (May), Denver, Colorado.

[21] M.J. McQuaid, T.H. Ong, H. Chen, J.F. Nunamaker, Multi-dimensional scaling for group memory visualization, Decision Support Systems 27 (1–2) (1999) 163–176.

[22] E. Morse, M. Lewis, K.A. Olsen, Evaluating visualizations: using a taxonomic guide, International Journal of Human–Computer Studies 53 (2000) 637–662.

[23] K. Mullet, C. Fry, D. Schiano, On your marks, get set, browse! (The Great CHI'97 Browse Off), Proceedings of the ACM CHI'97 Conference on Human Factors in Computing Systems, 1997 (March 22–27), Atlanta, Georgia.

[24] P. Pirolli, S.K. Card, M.M. van der Wege, The effect of information scent on searching information visualizations of large tree structures, Advanced Visual Interfaces International Working Conference, 2000 (May 23–26), Palermo, Italy.

[25] P. Pirolli, S.K. Card, M.M. van der Wege, Visual information foraging in a focus+context visualization, Proceedings of the ACM CHI 2001 Conference on Human Factors in Computing Systems, 2001 (March 31–April 5), Seattle, Washington, USA.

[26] P. Pirolli, S.K. Card, M.M. van der Wege, The effects of information scent on visual search in the hyperbolic tree browser, ACM Transactions on Computer–Human Interaction 10 (1) (2003) 20–53.

[27] C. Plaisant, B. Milash, A. Rose, S. Widoff, B. Shneiderman, LifeLines: visualizing personal histories, Proceedings of the ACM CHI '96 Conference on Human Factors in Computing Systems, 1996 (April 13–18), pp. 221–227, Vancouver, Canada.

[28] D. Schaffer, Z. Zuo, S. Greenberg, L. Bartram, J. Dill, S. Dubs, M. Roseman, Navigating hierarchically clustered networks through fisheye and full-zoom methods, ACM Transactions on Computer–Human Interaction 3 (2) (1996) 162–188.

[29] S. Wehrend, C. Lewis, A problem-oriented classification of visualization techniques, Proceedings of the 1st IEEE Conference on Visualization, 1990, pp. 139–143, San Francisco, California, USA.

[30] R.S. Wurman, Information Anxiety, Doubleday, New York, USA, 1989.

[31] M.X. Zhou, S.K. Feiner, Visual task characterization for automated visual discourse synthesis, Proceedings of the ACM CHI '98 Conference on Human Factors in Computing Systems, 1998, pp. 392–399, Los Angeles, California, USA.

[32] B. Zhu, H. Chen, Social visualization for computer-mediated communication: a knowledge management perspective, Proceedings of Workshop On Information Technologies and Systems (WITS'01) Conference, 2001 (December 15–16), New Orleans, Louisiana, USA.

Yang Xiang is currently an assistant professor of multimedia at the College of Lake County, Grayslake, Illinois. She was a master's student in Management Information Systems at the University of Arizona and worked as a research assistant at the Artificial Intelligence Lab from 2000 to 2002. Her research interests include information retrieval and visualization, human–computer interface, and Web-based database development. She received an MA degree in Media Arts (2000) and an MS degree in Management Information Systems (2002) from the University of Arizona.

Michael Chau is a research assistant professor in the School of Business at the University of Hong Kong. He received his PhD degree in Management Information Systems from the University of Arizona and a Bachelor degree in Computer Science (Information Systems) from the University of Hong Kong. When at the University of Arizona, he was an active researcher in the Artificial Intelligence Lab, where he participated in several research projects funded by NSF, NIH, NIJ, and DARPA. His current research interests include information retrieval, natural language processing, Web mining, and multi-agent systems.

Homa Atabakhsh is currently Associate Director of the Coplink Center for Excellence. She is a Principal Research Specialist at the University of Arizona Management Information Systems Department. She received her BSc, MSc (1984), and PhD (December 1987) in Computer Science from the University of Toulouse in France. She was Assistant Professor in Toulouse from January 1988–January 1989. She was employed by the National Research Council of Canada, as Research Scientist, from January 1989–1996 and worked in areas such as knowledge-based systems, object-oriented design and programming, GUI, applications in manufacturing and business.

Hsinchun Chen is McClelland Professor of MIS and Andersen Professor of MIS at the University of Arizona, where he is the director of the Artificial Intelligence Lab and the director of the Hoffman E-Commerce Lab. His articles have appeared in Communications of the ACM, IEEE Computer, Journal of the American Society for Information Science and Technology, Decision Support Systems, and many other publications. Professor Chen has received grant awards from NSF, DARPA, NASA, NIH, NIJ, NLM, NCSA, HP, SAP, 3COM, and AT and T. He serves on the editorial board of Decision Support Systems, the Journal of the American Society for Information Science and Technology, and the ACM Transactions on Information Systems.