Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets

Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan
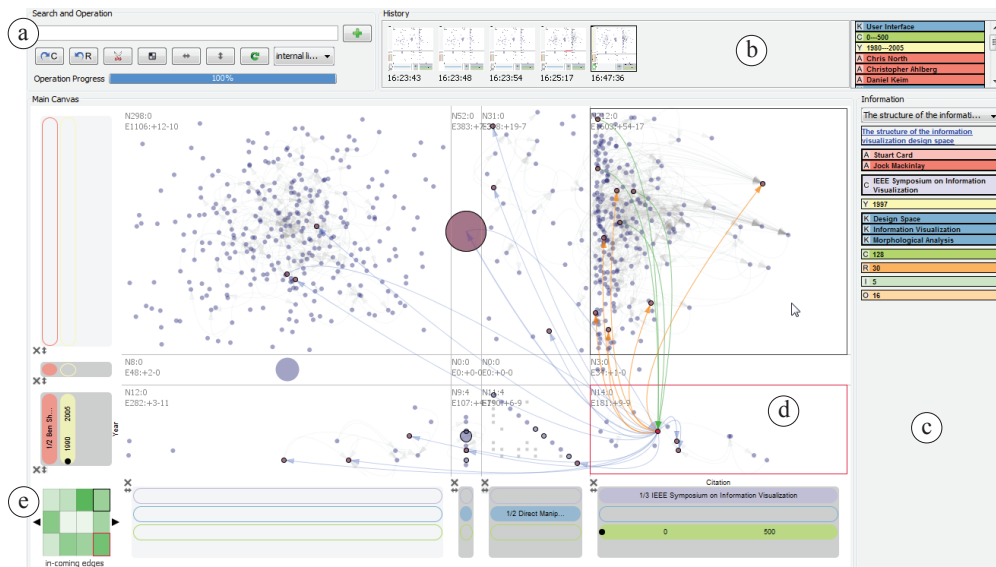
Fig. 1. The main PivotSlice interface: (a) Search and Operation Panel, (b) History Panel, (c) Information Panel, (d) Main Canvas, and (e) Cell Relation Panel. The user is conducting faceted exploration by sub-dividing the entire data into 3-by-4 regions and discovering relationships among different sub-datasets, using the a table of queries constructed by two sets of filters on both axes.

Abstract—Many datasets, such as scientific literature collections, contain multiple heterogeneous facets which derive implicit relations, as well as explicit relational references between data items. The exploration of this data is challenging not only because of large data scales but also the complexity of resource structures and semantics. In this paper, we present PivotSlice, an interactive visualization technique which provides efficient faceted browsing as well as flexible capabilities to discover data relationships. With the metaphor of direct manipulation, PivotSlice allows the user to visually and logically construct a series of dynamic queries over the data, based on a multi-focus and multi-scale tabular view that subdivides the entire dataset into several meaningful parts with customized semantics. PivotSlice further facilitates the visual exploration and sensemaking process through features including live search and integration of online data, graphical interaction histories and smoothly animated visual state transitions. We evaluated PivotSlice through a qualitative lab study with university researchers and report the findings from our observations and interviews. We also demonstrate the effectiveness of PivotSlice using a scenario of exploring a repository of information visualization literature.

Index Terms—Faceted browsing, network exploration, dynamic query, interaction, information visualization, visual analytics

◆

1 INTRODUCTION

In many application domains, the datasets people want to explore and analyze are becoming not only larger in size but also richer in structure and more complicated in semantics. Datasets can have *explicit* relations between entities, consisting of links between data items that

- *Jian Zhao is with University of Toronto, Canada. E-mail: jianzhao@dgp.toronto.edu.*
- *Christopher Collins is with University of Ontario Institute of Technology, Canada. E-mail: christopher.collins@uoit.ca.*
- *Fanny Chevalier is with University of Toronto, Canada; and OCAD University, Canada. E-mail: fanny@dgp.toronto.edu.*
- *Ravin Balakrishnan is with University of Toronto, Canada. E-mail: ravin@dgp.toronto.edu.*

are intrinsically defined (e.g., friendship between actors in a social network, or railways between cities in a transportation network); but also *implicit* relations, which are correlations between entities that can be derived from their attributes (e.g., actors in a social network that share the same language and profession, cities in a transportation network that have more than one airport and exhibit a population increase).

To illustrate, let us consider online academic databases. These datasets consist of a large number of data items (e.g., publications) and a variety of attributes associated with each publication (e.g., authors and years). References and citations make up the explicit relations between articles, while implicit relations can be formulated by queries on specific properties, such as "share the same author(s) and be published before 2000", "contain the same keyword(s) and be widely referenced in cognitive science after a few years", or any other combination of data facets. Academic publication data offers a complex dataset with a variety of relation types, making it useful for research into visualization techniques [4, 11, 12, 24]. Improved interfaces for analysis and exploration of academic publication data could lead to better discovery of relevant background research, a holistic understanding of emerging

topics in research to inform policy and funding, and the establishment of new collaborations between researchers. In this paper we use the academic publication domain as a testbed for a new technique for visualizing faceted datasets which also contain referential relationships.

The size scale of this type of data suggests a manner of *multi-focus* and *multi-scale* visual exploration, providing progressive, dynamic and flexible access of various parts of data at different levels of detail [34, 35, 37]. Moreover, the structural aspects of the dataset are often expressed in *heterogeneous facets*—attributes that are grouped into multiple orthogonal classifications, in which the attributes could be numerical, categorical, or ordinal. Hence, in addition to the traditional direct and navigational search, facet-based exploration is necessary, allowing users to freely narrow down and view data across various dimensions [11, 24, 38].

In this work we focus on supporting visual analytics of datasets which include two types of relationships—the *explicit* referential relationships of the core data items (e.g., paper citations) and the *implicit* correlations and trends between different subsets or certain attributes of data (e.g., publications by authors versus years). However, few existing visualization systems have adequately addressed all the above issues. There are two main large bodies of research—faceted classification systems and multivariate network visualizations. The former focuses on mechanisms of pivoting and dynamic filtering but lacks visualization of the structural overview and explicit relationships in datasets. The latter provides many efficient means of displaying and manipulating the topological layout of data, but they usually have limited exploration abilities for implicit relations, where only small numbers of attributes or a single type of facet (e.g., numerical) are allowed, often through visual encoding (e.g., coloring) the nodes of a graph.

In this paper, we present a novel interactive visualization system, PivotSlice, equipped with flexible multidimensional query mechanisms over complex data, which seamlessly provides efficient faceted exploration capabilities while enabling to reveal the relationships and correlations. Our core visualization design is based on a dynamic tabular view which represents a set of logical visual queries performed by a user progressively filtering, pivoting and slicing the information space, supporting a multi-focus access of the whole dataset with different scales and layout methods. With the metaphor of direct manipulation [29], individual or groups of data attributes can be dragged, split, combined, and pivoted to construct a customized tabular subdivision of the whole dataset, supporting a free-from faceted exploration with heterogeneous attributes. Based on the node-link view and flexible graph layout, PivotSlice provides the visual representations of both explicit referential relationships across all data items and implicit trends and correlations between different facet attributes.

We developed and evaluated PivotSlice with a seed dataset of academic publications augmented with the Microsoft Academic Search API [2] which was integrated with the tool to make the dataset live and dynamic, where users can grow, modify and tailor the dataset according to their needs. We chose scientific literature data because: 1) it is representative that includes all the common features of faceted datasets, and 2) expert users are readily available in academic settings to participate in evaluation experiments. We conducted a qualitative laboratory study with 6 university researchers and report our observations, user feedback, and usage scenarios derived from our findings.

2 RELATED WORK

2.1 Faceted Exploration

There has been extensive research on exploring large faceted collections through visualization. Several early systems including FOCUS [33] and InfoZoom [32] allow users to perform dynamic queries by arranging filtered data objects in a large table where the attributes on the vertical facets axis can be pivoted, expanded, and collapsed; multi-scale aggregations and focus+context techniques are applied to display the highly compressed table content. Along the same lines, Yee *et al*. [38] proposed an image searching tool that allows the user to progressively narrow down the query using a facet metadata panel displaying item distributions. However, these techniques assume that the exploration process is linear, forcing the user to drill down into the

data through increasing the query specifications step by step, which may limit serendipitous discoveries that would otherwise be possible by comparing different query results. TimeSlice [39] employs a multi-focus approach allowing users to browse faceted temporal events with a set of related queries simultaneously. But none of the above tools displays relationships and trends of data items across different facets.

Several techniques have been proposed to visualize the correlations and trends among faceted data items. Polaris [34], which later became Tableau [1], provides a tabular representation of relational databases where the columns and rows form a set of dynamic queries with ordinal or numerical facets to divide the whole dataset. However, the number of attributes in each facet are limited. PaperLens [23] and NetLens [21] use multiple coordinated views of bar charts and textual lists to support the exploration of trends and connections for large academic repositories. Similarly, FacetLens [24] introduces linear facets (e.g., publication years) and integrates richer faceted navigation techniques to further facilitate sense-making tasks over faceted datasets. GraphTrail [12] records the exploration history of queries conducted on a canvas,supporting users in tracing back, recalling and modifying their interactions. While these examples focus on the implicit relations, the topology of the referential structures (e.g., citations) across the entire or different subsets of data is not adequately visualized. FacetAtlas [7], though it expresses relations between text documents via topological layout, is limited in visualizing rich faceted information. Most of these systems also force the user to explore linearly within the information space rather than in a flexible multi-focus manner.

2.2 Multivariate Network Visualization

Also related to our research is the area of network visualization techniques for exploring referential (explicit) relations between data items. For example, referential relationships can be aggregated with algorithms to support the analysis of scientific research [9] or social networks [28] on a higher level. Focus+context techniques have been proposed to direct the user's exploration of large graphs with degree-of-interest functions [35] and semantic lenses [17]. To further reveal the genealogy of citation patterns, Citeology [26] connects papers titles organized in a chronological layout. In addition to a similar timeline view of nodes, EdgeMaps [10] integrates another layout which is based on multidimensional scaling to expose implicit similarities of entities. Apolo [8], on the other hand, relies on machine learning algorithms and interactions to guide the user to incrementally explore large network data. However, one significant limitation of these systems is that faceted exploration interactions such as pivoting are not permitted. PivotPath [11], while providing several efficient faceted exploration mechanisms, limits the ability to compare different query results in a multi-focus manner to just one or two attribute values, and like the above techniques, it does not emphasize the overall relationships or trends between different data attributes.

The Parallel Scatterplot Matrix [36] presents the correlation of every attribute-pair in a matrix layout and GraphDice [4] further incorporates smooth animations into the transition of different view perspectives. While such techniques provide a clear overview of all the potential implicit relationships, a large number of attributes could cause the matrix view to be too small and crowded to interact with. In addition to duplicating all the nodes in every matrix cell, another approach is to subdivide the whole graph based on attributes and present the correlations accordingly. ManyNets [15] applies a computational approach to extract subnetworks and visualizes their statistical properties side-by-side in a table. PivotGraph [37] aggregates nodes and edges based on their attributes and renders the graph on a grid. Semantic Substrates [31] place graph nodes in non-overlapping regions according to user-defined attribute semantics. Ploceus [25], though designed for relational databases, presents the connections of data items in node-link graphs by slicing the visualization into a tabular view of subnetworks along user-selected attributes. However, these techniques only investigate the dividing of the information space with one level of orthogonal attributes (i.e., a data point must be placed in one unique attribute region), which is not flexible enough to support the investigation of questions involving multiple types of facet simultaneously.
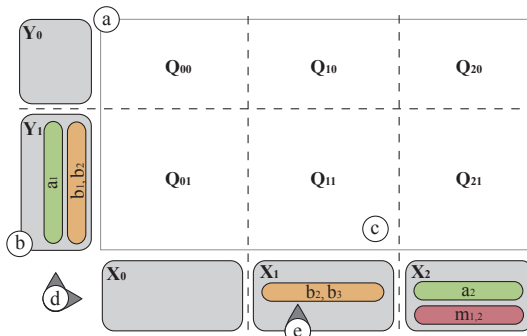
Fig. 2. An illustration of the visual query logic in PivotSlice.

2.3 Visualization Task Models

Beyond existing approaches to implicit and explicit relation analysis, the design of PivotSlice relates to prior work on high-level task models of visual analytic systems in general. Shneiderman proposed seven basic abstract tasks—*Overview, Zoom, Filter, Details-on-demand, Relate, History, and Extract*—summarized in the visual information seeking mantra, which motivates the creation of many visualization tools [30]. Keim later extended this idea in the context of visual analytics—*Analyze first, Show the important, Zoom, Filter and analyze further, Details on demand*, indicating the iterative nature of the visual exploration process [22]. Further, Russell *et al.* described a sense-making loop which concurs with the above model, suggesting that interfaces should guide users progressively search for a representation and encoding of data to answer their specific questions [27]. Our design decisions were guided by these models.

3 INTRODUCING PIVOTSLICE

The main user interface of PivotSlice is composed of five interactively coordinated views (Figure 1): (a) the *Search and Operation Panel* which provides the meta-data searching abilities and essential functions on data management and view configuration; (b) the *History Panel* which shows the recent states of the visualization, allowing users to undo or redo operations; (c) the *Information Panel* presenting the detailed attributes of visualization objects that the user is currently interacting with; (d) the *Main Canvas* displaying the dynamic visual representations of the dataset based on user interactions; and (e) the *Cell Relation Panel* which summarizes the overall relationships among different subsets of the data on the Main Canvas.

In PivotSlice, implicit relations, such as trends and correlations, are revealed through the layout of data items based on user-defined queries. Explicit relations, such as cross-referencing, are revealed by directed edges drawn between items on the Main Canvas. In this section, we will introduce the visual query language which is used to specify a data layout on the canvas, and describe the usage of Pivot-Slice through a scenario. In the following section we will describe the overall design motivation and the individual features in detail.

3.1 Visual Query Language

The Main Canvas is the primary interface component where the user can explore the data by performing dynamic visual queries. As depicted in Figures 1 and 2, PivotSlice logically organizes a set of user-defined queries in a *Query Table* view (Figure 2-a) made of a matrix of *Query Cells* (Figure 2-c). The Query Table layout is built according to the two groups of *Data Filters*, shown as gray rectangles (Figure 2-b), on the horizontal and vertical *Query Axes* (Figure 2-d). A Data Filter contains several *Facet Panels* (Figure 2-e), displayed as color-coded bars, each of which composed of several attribute constraints.

Let us consider a multifaceted dataset $\mathscr{D}$ of data entries d. We note the categorical facets as $A, B, \ldots$, and a_i the ith attribute value of the facet A. For numerical facets, we use the notation $M, N, \ldots$, and $m_{i,j}$ as the interval of values within $[m_i, m_j]$ on facet M. Further, let $A(d)$ be the attribute value of facet A for the entry d. Based on the universal relational model—which states that "one can place all data attributes into a table, which may then be decomposed into smaller

tables as needed" [18], the visual query language in PivotSlice subdivides a logical map of all data entries into a dynamic table according to the criteria (i.e., Data Filters) defined by the user. Figure 2 shows such an example of dividing the information space into a 2-by-3 grid constructed with filters $\{Y_0, Y_1\} \times \{X_0, X_1, X_2\}$.

The query associated to each cell is evaluated in three steps:
1) Each Facet Panel corresponds to a logical *OR* operation on the selected attribute values. For example, the Facet Panel B of Data Filter X_1 corresponds to $\{b_2 \cup b_3\} = \{d \in \mathscr{D}, B(d) \in \{b_2, b_3\}\}$. The numerical Facet Panel M in X_2 corresponds to $\{m_{1,2}\} = \{d \in \mathscr{D}, M(d) \in [m_1, m_2]\}$. We chose the *OR* operation for attributes on a Facet Panel to make it conceptually consistent between numerical and categorical facets: since a query on a numerical facet corresponds to a range of values equivalent to a *OR* relation over all values in the interval, we use the same operation for categorical facets.
2) Each Data Filter corresponds to a logical *AND* operations on the basic queries of its Facet Panels. For instance, the Data Filter Y_1 forms the query $\{a_1 \cap (b_1 \cup b_2)\} = \{d \in \mathscr{D}, A(d) = a_1 \text{ and } B(d) \in \{b_2, b_3\}\}$.
3) Each Query Cell corresponds to a compound query $Q_{i,j}$ with a logical *AND* operation over the two resulting queries of the corresponding Data Filters X_i and Y_j, such as $Q_{2,1} = \{(a_1 \cap (b_1 \cup b_2)) \cap (a_2 \cap m_{1,2})\}$.

Following this process, the whole dataset is sliced into multiple subsets filled with items satisfying the constraints of every Query Cell. When query results overlap, data items are duplicated in multiple cells.

The top-most row and left-most column correspond to the "others" or "remaining" filters, that represent the complement query to all other queries on the corresponding Query Axis. For example, $Y_0 = \mathscr{D} \setminus Y_1 = \overline{\{a_1 \cap (b_1 \cup b_2)\}}$ and $Q_{1,0} = \{\overline{a_1 \cap (b_1 \cup b_2)} \cap (b_2 \cup b_3)\}$. In particular, the top-left cell contains data items that are not compatible with any of the other Query Cells. This allows a better understanding of the Query Table structure since the location of the cells remains consistent when the user dynamically adds or removes Data Filters.

3.2 A Usage Scenario: Exploring Visualization Research

We introduce the use of PivotSlice in visual biographic analytics with a usage scenario by exploring a collection of scientific literature. We use the InfoVis 2004 Contest dataset [13], consisting of a directed network of 591 nodes (publications) and 1866 edges (citations and references). The metadata of publications (gathered using the Microsoft Academic Search API [2]), consists of titles, URLs and a range of facets including authors, keywords, venues, publication years, reference numbers and citation counts. The scenario is derived from the interview results and feedback of participants in the process of our design and study.

Suppose that Sally, a graduate student from the data mining group, is interested in visualization as a future area of research. But she knows very little about the field, so she wants to gain more knowledge about the literature in order to find research opportunities, discover foundational readings, and get a feel for the influential authors. Sally loads the InfoVis dataset into PivotSlice, which presents the entire citation network on the Main Canvas.

3.2.1 Uncovering Entry Points for Exploration

Sally starts by looking for papers that relate to her research expertise. Just after she types "data" in the Search box, a list of matched items drops down, including journals, conferences, keywords and so on. Within the list, she identifies some familiar conferences such as *Knowledge Discovery and Data Mining* as well as the keyword *Data Mining*. She drags the latter keyword onto the horizontal Query Axis, which results in a split of the dataset into a table of two Query Cells with smooth animation: papers containing the keyword move to the right column, and the remaining papers move to the left column. However, there are only a few papers associated to the keyword *Data Mining*. She thinks of other keywords and starts to type "visualization" in the Search Box. This time the drop-down panel suggests a list of attribute values, many of which are unfamiliar. A common keyword *Data Visualization* catches her attention. She drags this keyword onto the existing Data Filter with *Data Mining* (resulting on a *OR* query of the two keywords), and more papers move back from the left column to the Query Cell of interest.
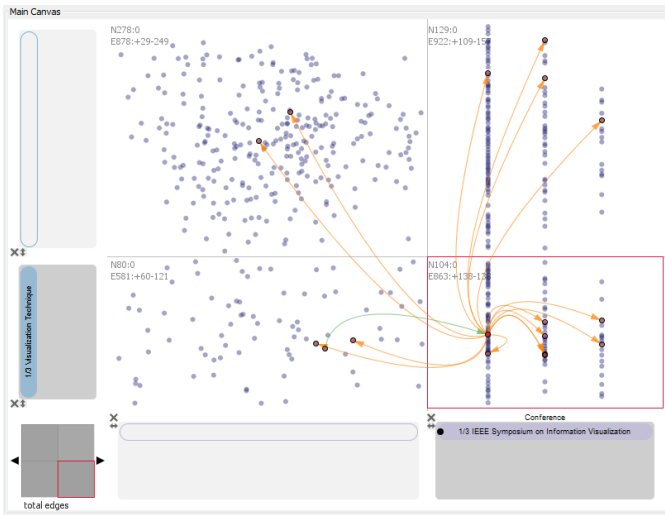
Fig. 3. A 2-by-2 division of the dataset with keywords *Information Visualization*, *User Interface*, and *Visualization Technique* versus conferences *InfoVis*, *CHI* and *UIST*, where the conference facet in the Data Filter is used for aligning the nodes.
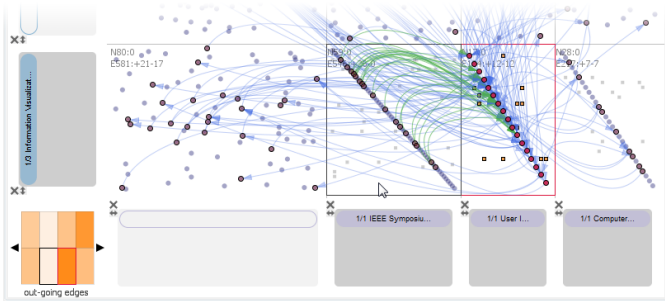

Fig. 4. After selecting the all papers from the *UIST* conference, hovering over the *InfoVis* cell indicates a lot of citations (green) from *InfoVis* and no references to *InfoVis* (orange). This can be further confirmed with the colormap of out-going edges on the Cell Relation Panel).

By hovering over these nodes, Sally accesses the details on the Information Panel, and finds that these articles focus on specific domain applications. So she decides to change her exploration strategy by using more common keywords. With the selection of all the papers in this Query Cell via the context menu, she obtains a rank of frequently mentioned keywords on the Information Panel. In addition to the two keywords she just dragged, she identifies *Information Visualization*, *User Interface*, and *Visualization Technique* to be interesting general keywords for a broader exploration.

3.2.2 Investigating Top Venues

Sally drags all three keywords onto the horizontal Query Axis to create a new Data Filter, and in the meantime clears the previous filter. To obtain the summarized information, she selects all the newly filtered nodes and looks at the ranked metadata attributes associated with these papers on the Information Panel. She finds that *InfoVis*, *CHI* and *UIST* are the top three popular conferences for the selected topics. Similarly, Sally drags these purple conference attributes to create a new Data Filter on the other Query Axis, which further divides the canvas into a 2-by-2 Query Table. To get a better idea of paper distributions over the three conferences, she aligns the nodes according to the *Conference* facet in the focused Query Cell (see Figure 3).

Sally observes that most of the papers matching her current interest are published at *InfoVis*. Wanting to ensure she spends time investigating important papers, she is curious about the citation patterns between the three conferences for the selected keywords. By double-clicking the *Conference* Facet Panel, she splits the original filter into three separate queries, with each conference alone as a the filtering attribute. Using the Cell Relation Panel, which indicates the overall connection
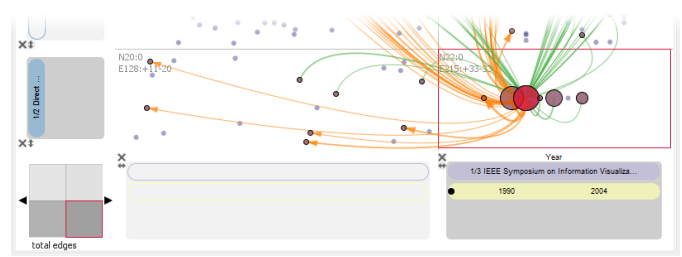

Fig. 5. Viewing the time trend of papers using the keywords *Direct Manipulation* or *Dynamic Query* from conferences *InfoVis*, *CHI* and *UIST*.
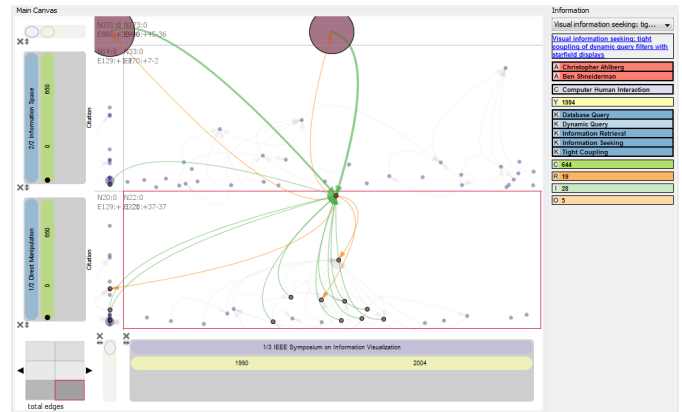

Fig. 6. Identifying highly-cited publications with keywords of *Direct Manipulation* and *Dynamic Query* as well as *Information Space* and *World Wide Web*. The "others" row and column are minimized.

relations between the selected cell and other cells, she gets a rough idea of the citation patterns among the three conferences. For further verification, she changes the layout of the three cells to the Matrix-Style View for more clarity, since it aligns all the nodes diagonally and declutters internal directed links by showing them as tiny squares (Figure 4). Sally conducts some observations of the crossing links by selecting and hovering over several papers. An intriguing insight she uncovers is that *InfoVis* cites a lot of papers from *CHI* and *UIST* but the relationship is not reciprocal, at least in this dataset.

3.2.3 Revealing Specific Hot Keywords and Trends

Next Sally wants to narrow down her exploration by looking for more specific popular topics, so she first goes back to the original 2-by-2 Query Table using the History Panel. By selecting all the nodes in the top-right and bottom-right Query Cell, which are all the papers from the three conferences, Sally identifies a range of top-mentioned keywords, in which *Direct Manipulation* and *Dynamic Query* seem to be interesting. Thus she deletes the original general keywords and adds the two new ones to the Data Filter.

Sally now wonders what are the trends for publications of selected topics along time. So she adds a yellow *Year* facet to the Data Filter on conferences and aligns the nodes with this facet, which results in a distribution of papers over time. To get a better view of the trends, Sally aggregates the nodes vertically in this Query Cell, where she finds that those topics are more active in years 1994 and 1995 (Figure 5).

To search for other more recent research topics, Sally moves her attention to the top-right cell, which contains publications in these conferences but without the two selected keywords. She selects all the papers after year 1996 in that cell, and from the list of top-mentioned keywords, she identifies two more interesting keywords: *Information Space* and *World Wide Web*. She then adds these keywords into a new Data Filter next to the original one on the vertical Query Axis, since she is not sure about which group of keywords to explore further.

3.2.4 Studying Influential Papers and Authors

To gain more knowledge about the selected topics, Sally wants to find the most influential papers in both keyword groups. Hence, she appends the green *Citation Count* facet to the Data Filters and organizes
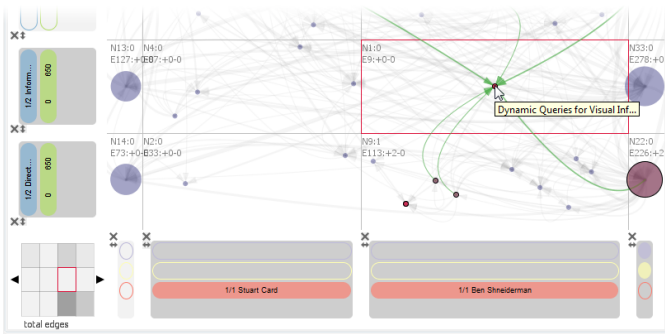
Fig. 7. Investigating publications of authors *B. Shneiderman* and *S. Card* versus two collections of keywords.

Fig. 8. Comparing publication trends between *B. Shneiderman* and *S. Card* by year.

the nodes by this facet, so that more highly cited papers appear higher in the cell. The distribution of nodes clearly indicates the most cited papers. However, Sally realizes that the Main Canvas is a bit crowded as more Query Cells are generated, thus she minimizes the rows and columns that are less important (Figure 6).

Next Sally decides to identify the dominant authors in the these fields by selecting and hovering over a number of highly-cited papers and accessing the paper author attributes on the Information Panel. After a short time of observation, she finds that *B. Shneiderman* coauthors almost all of the influential papers with keywords *Direct Manipulation* or *Dynamic Query*, and similarly *S. Card* is an influential author for the other topic about *Information Space* or *World Wide Web*. So Sally determines to further study the work of these two authors by adding them separately on the horizontal Query Axis, which allows her easily compare their work on the selected keywords.

At the first glance, Sally finds that both authors publish papers in the two selected topics, but *B. Shneiderman* seems to mainly focus on *Direct Manipulation* and *Dynamic Query*, as she identifies only one paper he published on the other topic—*Dynamic Query for Visual Information Seeking*—which is actually a duplicated node (highlighted with a black outline in PivotSlice) that exists in the other Query Cell as well (Figure 7). However, Sally notices that the values of citation and reference counts for a lot of the articles do not equal the actual in-degree and out-degree attributes of the network, indicating that the dataset is quite incomplete.

3.2.5 Digging into Authors of Interest

In order to get better knowledge of the work of these two authors, Sally starts to build her own specific but more complete dataset. She first prunes the data to crop all the rest of the network except their papers, then selects a number of interesting and highly-cited publications and fetches their references and citations into the current dataset with the online search ability of PivotSlice.

After a while, Sally starts to explore this more coherent and thorough data. By placing the two authors in filters on different axes, Sally finds a book coauthored by them: *Readings in Information Visualization: Using Vision to Think*, which might be worthy looking into in the future. To compare the authors, she keeps both author filters side-by-side, then adds the *Year* facet to the other Query Axis and further aligns and aggregates papers by it. She identifies that *B. Shneiderman* has more active publishing records than *S. Card* in recent years (Figure 8). In a similar way, Sally compares their publications by adding the three main conferences explored (i.e., *InfoVis*, *CHI* and *UIST*), and an interesting pattern is that *S. Card* publishes at all of the three venues but *B. Shneiderman* seems to rarely publish at *UIST*. In the end, Sally quits PivotSlice and saves her data into a new file for future reference.

4 DESIGNING PIVOTSLICE

In the design process of PivotSlice, we conducted a 2-hour interview and discussion session with a group of 14 university researchers (including graduate students, post-docs and professors) about how they explore and organize scientific publications and find interesting topics in their fields, in order to determine the essential elements of interactive faceted browsing tools for their tasks. Based on their comments
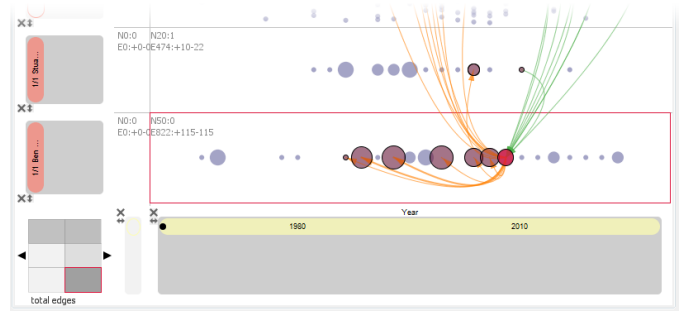
and feedback as well as the knowledge we gained from our literature survey, we present the core design rationales of PivotSlice in this section. These can be taken as general guidelines for designing visualization systems of faceted dataset exploration.

4.1 Multi-Focus and Multi-Scale Exploration

Numerous previous studies have indicated that multi-focus visualization, which allows the user to browse several parts of a dataset simultaneously at higher levels of detail while maintaining the context, is an efficient way to explore graphs, relational databases, faceted datasets and other forms of data [17, 34, 35, 37, 39]. In particular for large scale datasets, integrating multi-focus views and level-of-detail representations of data objects has been shown to facilitate navigation [17, 35, 37].

PivotSlice supports the multi-focus exploration of faceted datasets by allowing users to systematically divide the dataset into meaningful subsets via dynamic queries in a tabular form. This enables convenient comparisons of different parts of a dataset across user-defined semantics by viewing the Query Cells along columns or rows. For example, in Figure 7, the user compares publications of two keyword groups versus the authors using them or not in a easy side-by-side manner. To make best utilization of the screen real estate, the size of each table row or column in the Main Canvas is allocated dynamically based on the number of nodes contained in it.

To view the subsets of data at different scales, the user can minimize rows or columns of the tabular view by clicking the "arrow" button at the corner of each filter, which results in aggregating the nodes in the corresponding cells (Figure 6). Additionally, the aggregation views of a cell can be initiated by using the horizontal and vertical "arrow" buttons on the Search and Operation Panel, which does not minimize the entire row or column.

The form of node aggregation is determined by the current cell layout method (i.e., nodes can be aligned with one of the facets on each axis; see Section 4.2.3 for details) and its visual status (i.e., minimized or not). Specifically, if a facet on the axis to be minimized is configured as the layout-aligning facet, the nodes will be aggregated into buckets corresponding to a single attribute on that facet. For example, along the vertical *Conference* facet in Figure 9-a, and along the horizontal *Year* facet in Figure 9-b; otherwise, all the nodes will be grouped as one aggregated node (Figure 9-c). The size of an aggregated node is mapped to the number of actual publications in it. Depending on the node aggregations, the directed edges connecting the same nodes are aggregated as well, where the line thickness is mapped to the number of merged links.

4.2 Information Seeking with Direct Manipulation

Our design of PivotSlice follows the principles of creating efficient visualization systems suggested in many fundamental works including the well-known visual information seeking mantra [22, 27, 30]. Particularly, we further extend Shneiderman's seven general abstract tasks [30] to the following essential elements supporting the iterative knowledge discovery process in faceted data exploration. Moreover, we design the interaction of these features in PivotSlice with direct manipulations and smooth animations which can greatly ease the use, learnability, and understanding of the system [4, 29].
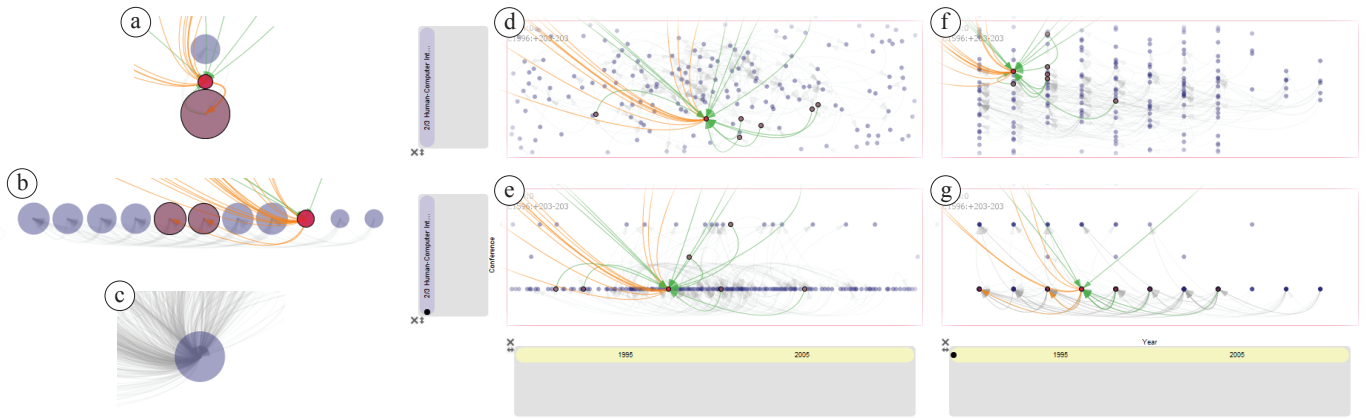
Fig. 9. Aggregated (a–c), force directed (d–f), and facet aligned (e–g) layouts of nodes in a cell (the vertical axis is the *Conference* facet with 3 attributes and the horizontal axis is the *Year* facet with range 1990–2010).
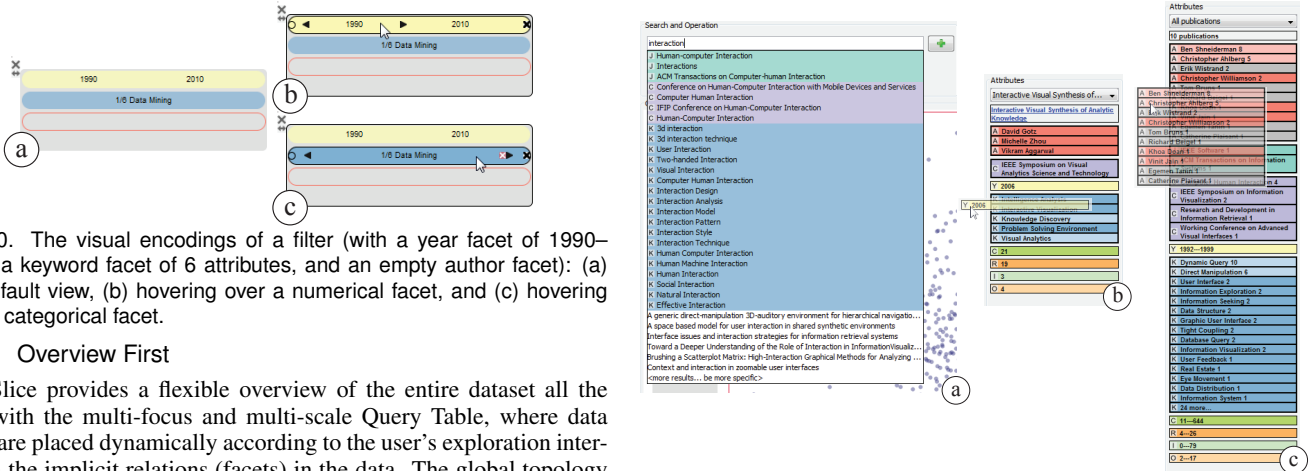

Fig. 10. The visual encodings of a filter (with a year facet of 1990–2010, a keyword facet of 6 attributes, and an empty author facet): (a) the default view, (b) hovering over a numerical facet, and (c) hovering over a categorical facet.

4.2.1 Overview First

PivotSlice provides a flexible overview of the entire dataset all the time with the multi-focus and multi-scale Query Table, where data items are placed dynamically according to the user's exploration interests in the implicit relations (facets) in the data. The global topology of network data referential patterns (i.e., the explicit relations) is presented by combining the visualizations in every Query Cell with: 1) either default node-link views by applying a force-directed layout algorithm [16] or 2) matrix-style views with nodes aligned diagonally (Figure 4). The user can selectively display the network edges using the options on the Search and Operation Panel, including "no links", "cell-internal links", "cell-crossing links" and "all links". At the top-left corner of each cell, a summary of the sub-network information such as the numbers of total nodes and edges is displayed (Figure 1). Further, the aforementioned various forms of node aggregations, such as the node grouping and the minimization of rows and columns, allow the easy access of the overall data at different levels of details.

4.2.2 Pivot and Slice

Pivoting is a widely-used browsing method in many faceted visualizations [11, 12, 24]. PivotSlice extends the traditional pivoting concept into a more general and flexible filtering technique by systemically organizing a set of dynamic queries in a table. As in the aforementioned visual query language, the filters composing the query support complicated logical operations on multiple attributes and facets as well as categorical and numerical values. With the interactive slicing of a dataset in the Query Table, the user can explore and compare the data by building her own semantics as described in Section 3.2.

Visual Query Encodings. In PivotSlice, as shown in Figure 1 and Figure 10, a Data Filter is represented as a gray round rectangle and a Facet Panel is displayed as a color-coded rectangular bar inside the filter (color schemes in [6] were used), in which a hollow bar with only the outline indicates no attribute associated with that facet in the filter. For categorical facets, attributes are rendered as textual labels on the Facet Panel. For numerical facets, a continuous range of values is shown by two separated labels for the starting and ending numbers.

Slicing Interactions. The user can progressively slice the dataset by interactively adding filters to the column or the row on the Query

Fig. 11. Adding new attributes via (a) searching on the Search and Operation Panel and (b)(c) dragging from the Information Panel.

Axes. A new Data Filter can be created if an attribute is drag-and-dropped onto the "others" filter representing the complement query (always the first one on each Query Axis). This newly-created filter adds one more column or row to the existing tabular division of the data. To drill down the data, the user can keep placing attributes of any facet onto existing filters (except the "others" filter), which updates all the queries and data items in cells of the corresponding row or column.

Pivoting Interactions. There are two ways of adding a new attribute to a filter in PivotSlice to further pivot the dataset. First, the user can type keywords in the multi-functional search box to open a drop-down list of matching categorical attributes, which can be selected and added onto a filter by pressing or dragging across the green "plus" button (Figure 11-a). Second, as shown in Figure 11-b, the user can grab single attribute labels from the Information Panel. Multiple categorical attributes can be selected through standard multiple selection interaction then dragging the group and dropping it on the Query Axis to form an OR query between all selected attribute values (Figure 11-c).

4.2.3 Relate and Extract

The ability to view relationships among data objects is a key function in visualization applications to support visual analytics [22, 30]. Especially for faceted datasets, not only does the topology of explicit referential relationships of data items need to be presented, but also the implicit cross-concept relations and trends between different facets.

Revealing Within-Cell Relationships. For one particular subset of data in a Query Cell, PivotSlice supports the discovery of trends and relations between different attributes by allowing flexible layout of data items. As Figure 9-d indicates, the default view is produced by force-directed layout [16]. When the user sets a layout-aligning facet for only one axis (by toggling the black circle on the Facet Panel), the nodes split into rows or columns where each contains the data in-

stances of one specific attribute value (Figure 9-ef). Node positions in the "free" dimension are still determined using the force-directed layout. When both of axes have been specified with a layout-aligning facet (Figure 9-g), the view becomes a regular scatter-plot. Together with the aggregation views (Figure 9-a–c) introduced in Section 4.2.1, PivotSlice provides flexible and powerful analytical abilities of data correlations in categorical-to-categorical, numerical-to-numerical, and categorical-to-numerical facets.

Revealing Between-Cell Relationships. PivotSlice assists the user with discovering relations at the cross-cell level in several ways. First, the Cell Relation Panel displays the inter-connection measurements between the selected table cell (highlighted in the red outline) and every other cell via color-mapping the strength of cell-to-cell connections in four different metrics (Figure 1-e). The connection strength values are computed by simply normalizing the numbers of total crossing edges, in-coming edges, out-going edges and overlapping nodes. Other network metrics could be easily integrated into PivotSlice. Second, the user can read the exact numbers of overlapping nodes, incoming and out-going edges at the top-left corner of each cell. Third, when a cell is selected and another cell is hovered over, only the links between these cells are highlighted in green (out-going) and orange (in-coming), relative to the selected nodes; other links are displayed in a less saturated color (Figure 1).

Extracting and Modifying Queries. PivotSlice offers the ability to modify query parameters through direct manipulations at multiple interface levels, including filters, facets, and attributes. First, the entire Data Filter can be dragged: if dropped on to another filter, the queries of these two filters are merged; if dropped in the empty area between filters, the filter is inserted in place; otherwise, the dragged filter is removed. Second, at the facet level, all the attributes on a Facet Panel can be manipulated as a group in a similar manner. Third, the user can pull-out one attribute by dragging the label and place it onto another filter. Other attribute-level modifications are supported: for numerical facets, the range of values can be adjusted by using the black forward and back arrows (Figure 10-b) or double-clicking to enter the number; for categorical facets, these arrows are used for attributes navigation and clicking the red cross button removes one (Figure 10-c). Moreover, the user can extract part of the visual query at the above three interaction levels without modifying the original one by performing the drag-and-drop operations while holding a modifier key.

4.2.4 Details-on-Demand

In PivotSlice, information details are presented in many forms when the user selects or hovers over a visualization object. For example, tooltips are presented for nodes, filters, and Facet Panels. Moreover, as in Figure 1, when a node is selected or hovered over, which is emphasized in red color, all its neighbor nodes are highlighted in lighter red and the edges connected to them are shown in green or orange (i.e., references or citations). Meanwhile, all the metadata of the selected node appears on the Information Panel: if the node represents a single publication, its title, URL, and all attributes on the facets will be displayed; if the node contains a collection of publications (e.g., an aggregated node), the panel will show a similar summary view containing the top 15 attributes (also with the number of occurrences) of each categorical facet ranked by frequencies as well as the attribute ranges of every numerical facet in these publications (Figure 11-c), and the user can further select and browse an individual publication. In addition to single-node selections, the user can choose a group of nodes using a polygon selection tool, or via the context menu to make special selections, such as all the nodes in a cell and neighboring nodes with the citation/reference relation.

4.2.5 History Available

History mechanisms, which allow the user to undo or "time travel", play a very important role in interface design [30]; especially for visual analytics systems, graphical histories of representation states can effectively facilitate the analysis process [19]. PivotSlice maintains two types of interaction history (Figure 1). First, every time when there is a change to the visualizations on the Main Canvas, such as

adding and removing a filter or a number of attributes, the state of the canvas is recorded and shown as thumbnails on the History Panel. Then the user can navigate back in time and restore to previous visualizations by simply clicking the thumbnails. Second, PivotSlice also stores a list of recently added or removed attributes on the right side of the History Panel, which can be added back to the existing filters via drag-and-drop interactions. The history mechanisms benefit the user in many ways, such as when the user accidentally deletes an object and wants to construct queries with the same attributes.

4.3 Dynamic Dataset

As noted in many studies [11, 17, 35], the data we want to analyze today is often very large but the user may be only interested in part of it. Moreover, the dataset are distributed and continuously evolving. For example, it would be very difficult to collect all the publications into a single academic database for exploration and the data collections themselves are updated all the time. Often, research tasks involving academic papers require up-to-date data.

PivotSlice attempts to support the above nature of datasets by providing several dynamic data management functions. The user can reduce the dataset according to her needs with the "prune data" button on the Search and Operation Panel, which prunes away all data items except selected items or items within a selected table cell. On the other hand, with the integration of web search, the citations or references of selected papers can be retrieved online and added to the current dataset. Furthermore, with the Search Box, the user can initiate a web search with the desired keywords and choose a list of publications in the results to insert into the data. To highlight new data points, recently added publications are displayed in red on the Main Canvas.

5 QUALITATIVE EVALUATION

We conducted a laboratory study to access how well people can use PivotSlice to explore, make sense and discover findings of faceted datasets as well as to find any usability issues for improving the system. The dataset used in this evaluation is as described in Section 3.2.

5.1 Participants and Apparatus

We recruited six participants (two females), aged 24–33, from our university network. All participants were either graduate students or research staff who had computer science or related backgrounds. We pre-screened participants to ensure that they all had at least two years research experience and had ever conducted literature searches in their fields. In order to balance the user expertise levels of our experimental dataset, we selected two participants whose research interests were information visualization and interaction in general, and others were from completely different research areas. The experiment was conducted on a desktop which was running on Windows 7 and connected with a 24-inch LCD display (resolution 1920×1200). Participants used a mouse and keyboard to interact with PivotSlice.

5.2 Tasks and Procedure

Based on the experiments conducted in previous studies with similar datasets [12, 21, 23, 24] and our own experience of doing literature surveys, we developed a list of tasks for evaluating PivotSlice which was guided by the 10 low level components of analytic activities in information visualization proposed by Amar *et al.* [3] (T1–T20 in Table 1). Although classified in those themes, some tasks required the combination and interleaving of multiple low-level analytical components and might be followed with open-ended questions. In general, we presented these tasks to the participants from simple to complicated ones. We also designed three high-level exploratory tasks (T21–T23), where the attributes were not mentioned in the low-level tasks, to evaluate PivotSlice in a more general situations.

The study began with a brief introduction and demonstration (about 10 minutes) of the features of PivotSlice. Then participants were asked to complete a series of tasks shown in Table 1 by exploring the provided dataset. During the process, participants were encouraged to articulate their intentions of performing each operation ("think aloud").

Table 1. Experimental tasks for evaluating PivotSlice, organized by the 10 low level components of analytic activities in information visualization proposed by Amar *et al.* [3], plus three high level open-ended tasks.

Retrieve Value	T1. Who are the authors of Cone trees: Animated 3D visualization of hierarchical information? T2. What are the papers cited by paper H3: Laying out large directed graph in 3D hyperbolic space?
Filter	T3. What are the papers coauthored by John Stasko and Robert Amar? T4. What are the papers with keyword Information Visualization in 2000–2002? How do papers in these years cite each other?
Determine Range	T5. What are the publication years of author Catherine Plaisant? T6. What is the range of citation counts for papers written by Chris North?
Find Extremum	T7. What are the most popular keywords of papers in conferences excluding Computer Human Interaction? T8. Which author publishes the most papers in conference IEEE Symposium on Information Visualization?
Compute Derived Value	T9. Find all the citations of the papers with keyword Information Visualization in conference Computer Human Interaction. T10. Who are the collaborators of author Stuart Card? Are there any patterns in his coauthors?
Sort	T11. Order papers with keywords Information Space or Dynamic Query, by year. Tell us one insight about the citation patterns among papers in different years. T12. Rank papers written by author George Robertson with and without Jock Mackinlay by citation count. What is the most cited paper and how do other papers cite it?
Characterize Distribution	T13. What is the distribution of Stuart Card's papers by conference? How do papers in those conferences refer to each other? T14. Compare the distributions of papers by year between conference Computer Human Interaction and IEEE Symposium on Information Visualization. What can you identify?
Find Anomalies	T15. Are there any exceptions for the conference IEEE Symposium on Information Visualization in terms of paper year trends? T16. Observe papers with keywords Information Seeking or Dynamic Query in this dataset. Is there any incomplete data in terms of citations and references? What are the papers?
Cluster	T17. Identify groups of papers in conference IEEE Symposium on Information Visualization, in terms of similar year and citation count attribute values. T18. Find clusters of papers written by Stuart Card, George Robertson, or Jock Mackinlay. Can you tell more about the collaboration patterns between them?
Correlate	T19. Compare papers containing both keywords Information Visualization and User Interface, and the ones containing neither of them. Who are the authors that publish the most in the two categories? What are the popular keywords in addition to the two above? T20. Compare papers published in years 1990–2000 and 2001–2010. Of keywords Three Dimensional, Information Space and World Wide Web, which one is more popular in the year ranges? How many papers exactly? Can you tell more about the citations across those paper categories?
High Level Tasks	T21. Tell three facts about the author Ben Shneiderman. T22. Give three comments on the conference User Interface Software and Technology. T23. Name three interesting findings for the Direct Manipulation research keyword.

Meanwhile, interaction logs were recorded by the software. Different approaches of finding the answers to the task questions were allowed. Participants could ask for further explanations of features of PivotSlice when necessary. In the end of study, we gave participants questionnaires to rate PivotSlice and conducted semi-structured interviews (about 30 minutes) to collect their feedback. The whole study lasted around 1.5 hours for each participant.

5.3 Results

Based on the interviews conducted during the design of PivotSlice and the results in our study, we derived a usage scenario for demonstrating the application of PivotSlice in practice, which has already been described in Section 3.2. Thus in this section, we report the other findings resulting from the experiment.

5.3.1 Questionnaire

In the post-study survey, participants completed a questionnaire of 10 questions using a 1–7 Likert scale (from strongly-disagree to strongly-agree). Figure 12 indicates the questions and average user ratings. Based on the results of Q1 and Q2, PivotSlice seemed to be easy to use and learn, which was fairly encouraging as novice users were able to conduct complicated querying and filtering operations in PivotSlice with a very brief introduction and minimal assistance during the experiment. For the faceted browsing and analytical capabilities (Q3–Q6), participants ranked PivotSlice with very high scores and the small variances further indicated their agreement on the effectiveness and usefulness of these PivotSlice functions. For other features, participants felt that the animated transitions between two different layouts and the abilities of importing online search results were useful (Q7 and Q9); but the graphical history component received a more diverging opinions (Q8). Some participants thought the thumbnails of the historical views on the History Panel looked too similar to identify useful information about which step to go back. They suggested that "*adding textual labels of the interactions performed under each thumbnail*" and "*associating the thumbnails with the list of historical operations on the right of the panel*" would be helpful. Overall, participants ranked their desire of using PivotSlice to explore faceted data as 6 out of 7, which was very encouraging given that it was only a prototype system.

5.3.2 Observations

Participants were able to complete the low level tasks (T1–T20 in Table 1) relatively quickly, where the average time was about 1.5 minutes per task. Simple tasks that requires only a few steps of operations, such
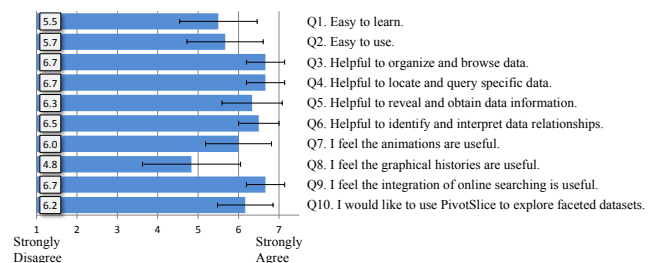

Fig. 12. User satisfaction results from the post-study questionnaire.

as T1–T8, were accomplished roughly within half a minute. Tasks with follow-up open-ended questions took longer. Participants spent much more time (around 2–3 minutes) on complicated tasks consisting of many analytical reasoning operations and comparisons, such as T17–T20. Assistance from the experimenter was barely needed when participants conducted T1–T15. For some tasks demanding complex semantics in facet attribute layout and alignment, participants sometimes needed time for planning to solve the problem or hints on the features they should use. For example, some participants did not think of using the *Reference* and *Out-degree* facets (or *Citation* and *In-degree*) as a way to identify the incompleteness of data in T16, or placing and aligning nodes with *Citation* and *Year* on different axes to solve T17 easily; but they suddenly realized so with a bit more instructions, commenting in the end "*Awesome! I never thought it [PivotSlice] could do that!*". Participants demonstrated various approaches to complete the tasks through pivoting attributes in filters to slice the data and obtain the desired queries, especially for T19 and T20. Participants were impressed by the flexibility and effectiveness of PivotSlice. Some interesting insights could be easily identified by participants in tasks with open questions. For example in T11, one participant found that there was a trend toward fewer cross references between papers with the two keywords *Information Space* and *Dynamic Query*.

For the high level tasks (T21–T23), participants were more familiar with PivotSlice and primed to explore more deeply and use different functions to find insights. The total time spending on these three tasks was around 10–15 minutes for each participant. Generally in T21, participants looked for the year, keyword and citation distributions of *B. Shneiderman*'s publications as well as his collaborators. Participants sometimes imported papers with the online search feature, to further explore and confirm the data patterns. In T22, even after fetching more data, some participants found that more citations appeared from papers of *InfoVis* to those of *UIST* but fewer in the other direction, although

both conferences contained similarly frequent keywords. When conducting the last task, some participants sought deeper findings based on previous layouts, such as investigating publications of the *Direct Manipulation* keyword for different conferences and authors. Other interesting insights found in T23 included: *Dynamic Query* was the most related keyword, and papers of such keyword tended to cite those of *Direct Manipulation* more often than the other way around.

5.3.3 Initial User Feedback

During the post-study interview, all participants indicated that the PivotSlice design was aesthetically pleasing and user-friendly, although a bit of learning time to understand the visual query logic was necessary. They agreed that PivotSlice was capable of flexible, efficient and adequate analytical functionalities for faceted data exploration and relation discovery. One mentioned that PivotSlice "*makes it easy for organizing complicated filters and searches*". Another said that "*it is extremely helpful in uncovering trends and distributions between data attributes [...] relating different parts of data is easy by using the connections and different alignments*". One participant who works in social network modeling research indicated that PivotSlice could be very useful in mining relationships in multivariate social networks and was eager to apply PivotSlice to his data in the future. Some of them thought PivotSlice's flexibility allowed conducting various basic analytical tasks in a unified interface, including correlation analysis, anomaly detection and similarity clustering. However, a few participants commented that too much flexibility made the PivotSlice interface complicated and difficult to learn, since "*it is sometimes confusing that you can perform the same task in many different ways*". But we argue that maintaining useful and adaptable functionalities is more critical in visual analytics.

While some participants mentioned that in the beginning they were confused about whether the logical *AND* or *OR* relations were computed for attributes of a filter, they could easily understand the rules of the query logic after a few manipulations of the dataset. One mentioned that it was difficult to remember the *AND/OR* operations forming the queries, suggesting that "*using similar visual cues from the filed of logical circuit design*". We also identified some usability issues during the study. One participant suggested that functions of merging adjacent cells and freely adjusting cell spaces should be implemented. Another thought that integrating short-cuts for common filtering operations, such as "*add facet and order by it*", as well as for all the drag-and-drop interactions, could speed up the data exploration process. Some commented that text labels should be displayed when the nodes were aligned or aggregated along a facet.

6 DISCUSSION

The core interface component of PivotSlice is the Main Canvas, in which the user can construct dynamic queries organized in a table, thus to build customized semantics of subdividing the whole dataset. This visualization provides not only an efficient multi-focus and multi-scale browsing of the explicit data relation (i.e., the topology), but also a flexible manner of correlating data items based on their facets to reveal implicit trends. PivotSlice supports these kinds of exploration for faceted datasets with both categorical and numerical attributes, which can be further extended to include more general data formats.

In particular, the ordinal facet is a special type of categorical facet, which can be easily supported in PivotSlice by restricting the attribute sequence in a Facet Panel. Although we only demonstrated PivotSlice with numerical facets of integer values (e.g., years and citation counts), it can further support real-value numerical attributes by simply including continuous adjustments of attribute ranges, such as using a slider, and customized binning thresholds when aligning the data items with this facet (i.e., the current binning threshold is always unit one for integer attributes). Although the research prototype was developed with certain data models tied in the Microsoft Academic Search APIs, the concept of PivotSlice design is applicable to many other types of faceted datasets, even for ones not containing referential links, where a different node layout algorithm such as the multidimensional scaling (MDS) [5] could be used instead of force-directed layout. Therefore,

we believe that PivotSlice is flexible and general enough to support various kinds of faceted data existing many application domains.

The ability to analyze large scale datasets is a perennial challenge for designing visual analytics systems. Although PivotSlice was not evaluated with datasets containing a very large number of data items, its facet attribute space is considerably rich: 5 numerical and 4 categorical facets, consisting of 1169 authors, 1163 keywords, 50 conferences, and 63 journals. We designed PivotSlice with several scalability considerations. First, the visualization is based on a multi-focus and multi-scale exploration of the data topology, relationships and trends, combined with different aggregation views and data item layout mechanisms. When the number of nodes is extremely large, the visual clutter of nodes and links will become an issue. However, PivotSlice can be easily extended with more advanced graph layout algorithms and visualization techniques such as edge bundling [20] and focus+context graph views [17]. Second, the flexible data management functions in PivotSlice, such as data cropping and resource extension with online searches, make it capable of visualizing large and dynamic datasets. Users can progressively modify and load data based on their evolving interests during exploration. Lastly, PivotSlice displays summarizations of selected objects on the Information Panel, which is very helpful in identifying points of interests within large scale data.

However, to further support very large datasets and advanced analytic capabilities, PivotSlice should be augmented with smart representation of data in each cell by well-suited informative charts and visualization techniques, such as tag clouds, linked bar charts [12], heatmaps, and linked treemaps [14]. In this way, the advantages of each visualization can be balanced in the flexible dynamic query table thus to overcome disadvantages of specific techniques, for example, occlusion problems of node-link views when the dataset is large. Moreover, proper analytical algorithms such as MDS can be integrated in processing the data before visualizing them in the cell, in order to guide the user's exploration in large scale data following the principle of "Analyze First" in Keim's visual analytics mantra [22].

7 CONCLUSION AND FUTURE WORK

We have presented PivotSlice, an interactive visualization tool for faceted data exploration, which allows the user to systematically and logically manage a collection of dynamic queries in a multi-focus and multi-scale tabular view. PivotSlice not only provides flexible faceted browsing capabilities but also assists the user with revealing both the explicit and implicit data relationships. The visualizations and interactions of PivotSlice are seamlessly integrated with direct manipulation and smoothly animated transitions. PivotSlice was developed and evaluated with a dataset of scientific literature that contains rich faceted structures, and a further incorporation of an online academic search API extends the abilities of PivotSlice with dynamic data management. A laboratory study, consisting of high-level and low-level tasks derived from basic analytical activities, was conducted with the dataset. The result shows that PivotSlice is not only relatively easy to use and learn but also useful, flexible and effective in discovering data relationships during faceted exploration.

In the future, we plan to test PivotSlice with other types of faceted data, such as music and movie databases, as well as with larger scale datasets of more items. We also aim to conduct more thorough comparative user studies to further evaluate PivotSlice with these datasets both in laboratory and practical settings. In terms of the visualization design, we would like to introduce other techniques regarding graph layout and edge rendering, semantic zooming, and focus+context exploration techniques based on the current PivotSlice framework of organizing dynamic queries in the multi-focus and multi-scale view. Finally, we are interested in developing other visualization techniques to further extend the analytical abilities of PivotSlice.

REFERENCES

[1] Tableau Software, 2010. http://www.tableausoftware.com/.

[2] Microsoft Academic Search API, 2013. http://academic.research.microsoft.com/About/Help.htm.

[3] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of the 2005 IEEE Symposium on Information Visualization*, INFOVIS '05, pages 15–, Washington, DC, USA, 2005. IEEE Computer Society.

[4] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist, and J. Fekete. Graphdice: A system for exploring multivariate social networks. *Computer Graphics Forum*, 29(3):863–872, 2010.

[5] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 1997.

[6] C. A. Brewer. http://www.colorbrewer.org, 2009.

[7] N. Cao, J. Sun, Y.-R. Lin, D. Gotz, S. Liu, and H. Qu. Facetatlas: Multifaceted visualization for rich text corpora. *IEEE Trans. on Visualization and Computer Graphics*, 16(6):1172 –1181, 2010.

[8] D. H. Chau, A. Kittur, J. I. Hong, and C. Faloutsos. Apolo: Making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 167–176, 2011.

[9] C. Chen. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.*, 57(3):359–377, 2006.

[10] M. Dörk, S. Carpendale, and C. Williamson. Visualizing explicit and implicit relations of complex information spaces. *Information Visualization*, 11(1):5–21, 2012.

[11] M. Dörk, N. Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through faceted information spaces. *IEEE Trans. on Visualization and Computer Graphics*, 18(12):2709–2718, 2012.

[12] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson. GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1663–1672, 2012.

[13] J.-D. Fekete, G. Grinstein, and C. Plaisant. IEEE InfoVis 2004 Contest, The History of InfoVis. www.cs.umd.edu/hcil/iv04contest, 2004.

[14] J.-D. Fekete, D. Wang, N. Dang, A. Aris, and C. Plaisant. Overlaying graph links on treemaps. In *Posters compendium of InfoVis*, pages 82–83, 2003.

[15] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck. ManyNets: An interface for multiple network analysis and visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 213–222, 2010.

[16] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Softw. Pract. Exper.*, 21(11):1129–1164, Nov. 1991.

[17] S. Hadlak, H.-J. Schulz, and H. Schumann. In situ exploration of large dynamic networks. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2334–2343, 2011.

[18] I. T. Hawryszkiewycz. *Database analysis and design*. Science Research Associates, 1984.

[19] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Trans. on Visualization and Computer Graphics*, 14(6):1189–1196, 2008.

[20] D. Holten and J. J. van Wijk. Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, 28(3):983–990, 2009.

[21] H. Kang, C. Plaisant, B. Lee, and B. Bederson. NetLens: Iterative exploration of content-actor network data. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, pages 91–98, 2006.

[22] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In

A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, chapter 7, pages 154–175. Springer-Verlag, 2008.

[23] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson. Understanding research trends in conferences using PaperLens. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1969–1972, 2005.

[24] B. Lee, G. Smith, G. G. Robertson, M. Czerwinski, and D. S. Tan. FacetLens: Exposing trends and relationships to support sensemaking within faceted datasets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1293–1302, 2009.

[25] Z. Liu, S. B. Navathe, and J. T. Stasko. Network-based visual analysis of tabular data. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 39–48, 2011.

[26] J. Matejka, T. Grossman, and G. Fitzmaurice. Citeology: visualizing paper genealogy. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 181–190, 2012.

[27] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 269–276, 1993.

[28] Z. Shen, K.-L. Ma, and T. Eliassi-Rad. Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1427–1439, 2006.

[29] B. Shneiderman. Direct manipulation: A step beyond programming languages. *Computer*, 16:57–69, 1983.

[30] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[31] B. Shneiderman and A. Aris. Network visualization by semantic substrates. *IEEE Trans. on Visualization and Computer Graphics*, 12(5):733–740, 2006.

[32] M. Spenke and C. Beilken. Infozoom - analysing formula one racing results with an interactive data mining and visualisation tool. In *Proceedings of Data Mining*, pages 455–464, 2000.

[33] M. Spenke, C. Beilken, and T. Berlage. Focus: The interactive table for product comparison and selection. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST '96, pages 41–50, 1996.

[34] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Trans. on Visualization and Computer Graphics*, 8(1):52–65, 2002.

[35] F. van Ham and A. Perer. "search, show context, expand on demand": Supporting large graph exploration with degree-of-interest. *IEEE Trans. on Visualization and Computer Graphics*, 15(6):953–960, 2009.

[36] C. Viau, M. J. McGuffin, Y. Chiricota, and I. Jurisica. The FlowVizMenu and Parallel Scatterplot Matrix: Hybrid multidimensional visualizations for network exploration. *IEEE Trans. on Visualization and Computer Graphics*, 16:1100–1108, 2010.

[37] M. Wattenberg. Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 811–819, 2006.

[38] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 401–408, 2003.

[39] J. Zhao, S. M. Drucker, D. Fisher, and D. Brinkman. TimeSlice: Interactive faceted browsing of timeline data. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 433–436, 2012.