Making Sense of Large Network Data: Combining Rich User Interaction and Machine Learning

Duen Horng "Polo" Chau, Aniket Kittur, Jason I. Hong, Christos Faloutsos
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{dchau, nkittur, jasonh, christos}@cs.cmu.edu

ABSTRACT

Extracting useful knowledge from large network datasets has become a fundamental challenge in many domains, from scientific literature to social networks and the web. We introduce Apolo, a system that uses a mixed-initiative approach—combining visualization, rich user interaction and machine learning—to guide the user to incrementally and interactively explore large network data and make sense of it. Apolo engages the user in bottom-up sensemaking to gradually build up an understanding over time by starting small, rather than starting big and drilling down. Apolo also helps users find relevant information by specifying exemplars, and then using a machine learning method called Belief Propagation to infer which other nodes may be of interest. We evaluated Apolo with twelve participants in a between-subjects study, with the task being to find relevant new papers to update an existing survey paper. Using expert judges, participants using Apolo found significantly more relevant papers. Subjective feedback of Apolo was also very positive.

Author Keywords

Sensemaking, large network, Belief Propagation

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces; H.3.3 Information Storage and Retrieval: Information Search and Retrieval

INTRODUCTION

Making sense of large networks is an increasingly important problem in domains ranging from citation networks of scientific literature; social networks of friends and colleagues; links between web pages in the World Wide Web; and personal information networks of emails, contacts, and appointments. Theories of sensemaking provide a way to characterize and address the challenges faced by people trying to organize and understand large amounts of network-based data. Sensemaking refers to the iterative process of building up a representation or schema of an information space that is useful for achieving the user's goal [31]. For example, a scientist interested in connecting her work to a new domain must build up a mental representation of the existing literature in the new domain to understand and contribute to it.

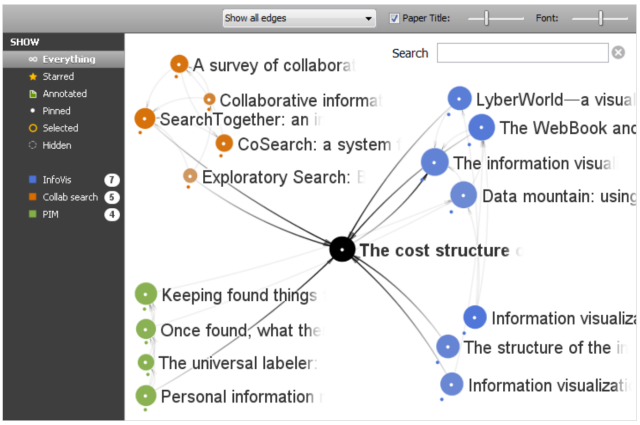


Figure 1. Apolo displaying citation network data around the article *The Cost Structure of Sensemaking*. The user gradually builds up a mental model of the research areas around the article by manually inspecting some neighboring articles in the visualization and specifying them as exemplar articles (with colored dots underneath) for some ad hoc groups, and instructs Apolo to find more articles relevant to them.

For the above scientist, she may forage to find papers that she thinks are relevant, and build up a representation of how those papers relate to each other. As she continues to read more papers and realizes her mental model may not well fit the data she may engage in representational shifts to alter her mental model to better match the data [31]. Such representational shifts is a hallmark of insight and problem solving, in which re-representing a problem in a different form can lead to previously unseen connections and solutions [11]. The practical importance of organizing and re-representing information in the sensemaking process of knowledge workers has significant empirical and theoretical support [28].

We focus on helping people develop and evolve externalized representations of their internal mental models to support sensemaking in large network data. Finding, filtering, and extracting information have already been the subjects of significant research, involving both specific applications [7] and a rich variety of general-purpose tools, including search engines, recommendation systems, and summarization and extraction algorithms. However, just finding and filtering or even reading items is not sufficient. Much of

Submitted for review to CHI 2011.

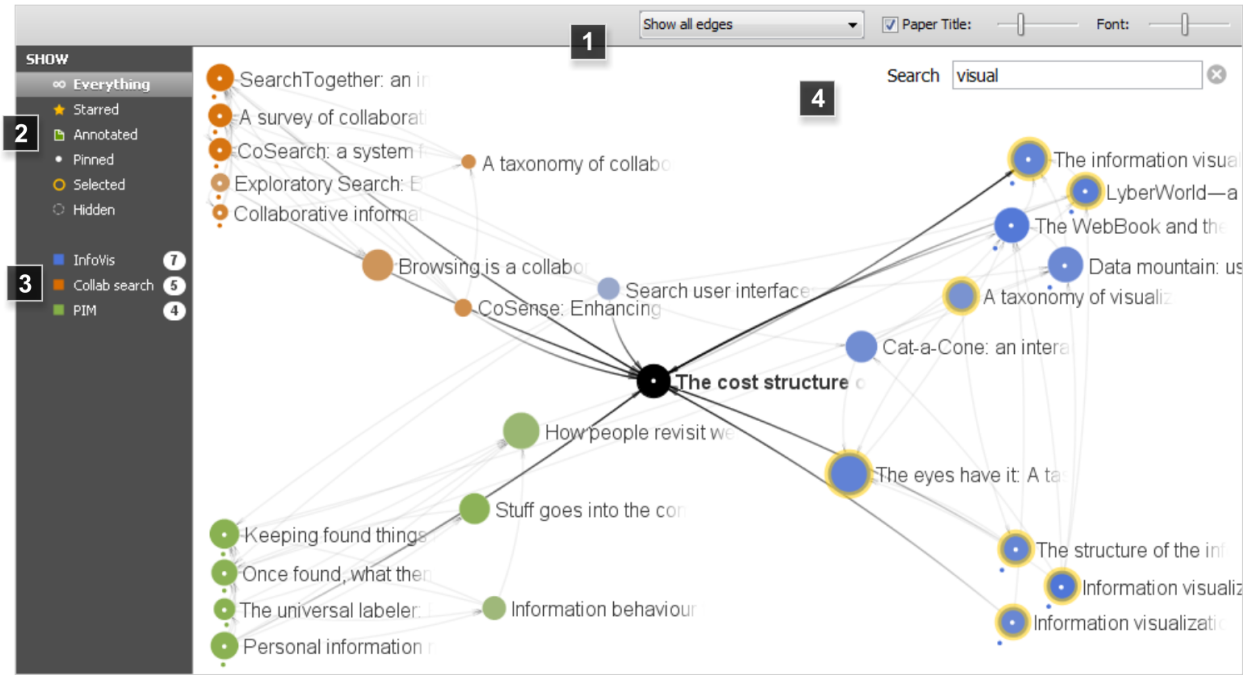


Figure 2. The Apolo user interface displaying citation data around the article *The Cost Structure of Sensemaking*. The user has created three groups: *Information Visualization* (InfoVis, blue) , *Collaborative Search* (orange), and *Personal Information Management* (PIM, green). A color is automatically assigned to each group, and each node is assigned the color of the group it most likely belongs to; saturation indicates "belongness". Each exemplar has a colored dot below it. 1: The *Configuration Panel* for enhancing visualization readability, by setting visibility of citation edges and article titles, varying font size and visible length of titles via sliders. 2: The *Filter Panel* provides filters to control the type of nodes to show; the default filter is "Everything", showing all types of nodes, except hidden ones. Other filters show nodes that are starred, annotated, pinned, selected, or hidden. 3: The *Group Panel* lets the user create, rename, and delete groups. 4: The *Visualization Panel* where the user incrementally and interactively builds up an understanding and a highly personalized visualization of the network. Articles whose titles contain "visual" are highlighted with yellow halos.

the heavy lifting in sensemaking is done as people create, modify, and evaluate schemas of relations between items. Few tools aimed at helping users evolve representations and schemas. We build on initial work such as Sensemaker [3] and studies by Russell et al. [30] aimed at supporting the representation process. We view this as an opportunity to support flexible, ad-hoc sensemaking through intelligent interfaces and algorithms.

These challenges in making sense of large network data motivated us to create Apolo, a system that uses a mixed-initiative approach—combining rich user interaction and machine learning—to guide the user to incrementally and interactively explore large network data and make sense of it. Our main contributions are

- We aptly select, adapt, and integrate work in machine learning and graph visualization in a novel way to help users make sense of large graphs using a mixed-initiative approach. Apolo goes beyond just graph exploration, and enables users to externalize, construct, and evolve their mental models of the graph in a bottom-up manner.

- Apolo offers a novel way of leveraging a complex machine learning algorithm, called Belief Propagation (BP) [40], to intuitively support sensemaking tailored to an individual's goals and experiences; BP was never applied to sensemaking or interactive graph visualization.

- We explain the rationale behind the design of Apolo's novel techniques and how they advance over the state of the art. Through a usability evaluation using real citation network data obtained from Google Scholar, we demonstrate Apolo's ease of use and its ability to help researchers find significantly more relevant papers.

RELATED WORK

Apolo intersects multiple research areas, each having a large amount of relevant work. However, less work explores how to better support graph sensemaking, like the way Apolo does, by combining powerful methods from *machine learning*, *visualization*, and *interaction*. Here, we briefly describe relevant work in some of the related areas.

Sensemaking

Sensemaking refers to the iterative process of building up a representation of an information space that is useful for achieving the user's goal [31]. Numerous models have been proposed, including Russell et al.'s *cost structure* view [31], Dervin's sensemaking methodology [8], and the notional model by Pirolli and Card [28]. Consistent with this dynamic task structure, studies of how people mentally learn and represent concepts highlight that they are often flexible, ad-hoc, and theory-driven rather than determined by static features of the data [4]. Furthermore, categories that emerge in users' minds are often shifting and ad-hoc, evolving to match the changing goals in their environment [14]. These results highlight the importance of a "human in the loop" approach for organizing and making sense of information, rather than fully unsupervised approaches that result in a common structure for all users. Several systems aim to support interactive sensemaking, like SenseMaker [3], Scat-

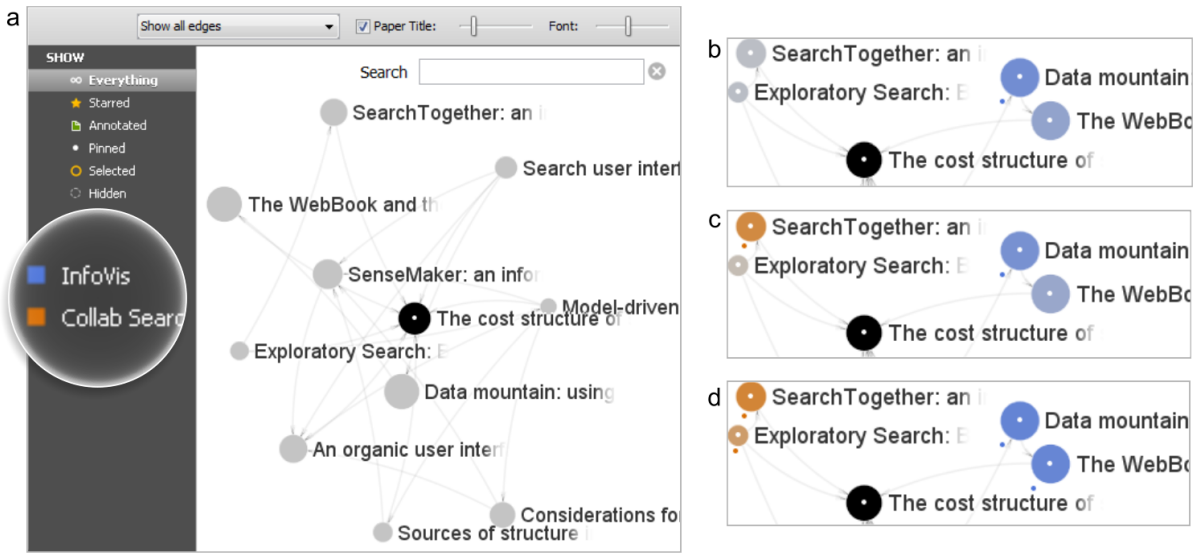


Figure 3. a) The Apolo user interface at start up, showing our starting article *The Cost Structure of Sensemaking* highlighted in black. Its top 10 most relevant articles, having the highest proximity relative to our article, are also shown (in gray). We have selected our paper; its incident edges representing either "citing" (pointing away from it) or "cited-by" (pointing at it) relationships are in dark gray. All other edges are in light gray, to maximize contrast and reduce visual clutter. Node size is proportional to an article's citation count. Our user has created two groups: *Collab Search* and *InfoVis* (magnified). b-d) Our user first spatially separates the two groups of articles, *Collab Search* on the left and *InfoVis* on the right (which also pins the nodes, indicated by white dots at center), then assigns them to the groups, making them *exemplars* and causing Apolo to compute the relevance of all other nodes, which is indicated by color saturation; the more saturated the color, the more likely it belongs to the group. The image sequence shows how the node color changes as more exemplars are added.

ter/Gather [7], Russell's sensemaking systems for large document collections [30], Jigsaw [34], and Analyst's Notebook [16]. Other relevant approaches include [35] and [1] which investigated how to construct, organize and visualize topically related web resources.

Visualization and exploration

Apolo builds on a large body of research aimed at understanding and supporting how people can gain insights through visualization [17]. A number of tools have been developed to support "landscape" views of information. These include WebBook and WebForager [6], which use a book metaphor to find, collect, and manage web pages; Butterfly [21] aimed at accessing articles in citation networks; and Webcutter, which collects and presents URL collections in tree, star, and fisheye views [20]. For a more focused review on research visualizing bibliographic data, see [23].

In contrast to many of these systems which focus on providing overviews of information landscapes, Apolo adopts a bottom-up sensemaking approach [31] aimed at helping users construct their own landscapes of information. This is because people may have very different mental representations of information depending on their individual goals and prior experiences. To address this problem, some research has explored local exploration of graphs, including Treeplus [19], Vizster [10], and the degree-of-interest approach proposed in [37]. These approaches generally support the idea of starting with a small subgraph and expanding nodes to show their neighborhoods (and in the case of [37], help identify useful neighborhoods to expand). One key difference with these works is that Apolo changes the very structure of the expanded neighborhoods based on users' interactions, rather than assuming the same neighborhoods for all users. We will describe this approach in more detail when discussing Apolo's core design factors.

Graph mining algorithms and tools

A lot of research in graph mining studies how to automatically discover clusters (or groupings) in graphs, e.g., Graphcut [18], METIS [13]. Much work has also been done on developing methods to compute relevance between two nodes in a network; many of them belong to the class of spreading-activation [2] or propagation-based algorithms, e.g., HITS [15], PageRank [5], and *random walk with restart* [36]. Used by Apolo, Belief Propagation [40] is a message passing algorithm over link structures similar to spreading activation, but it is uniquely suited for graph sensemaking because it offers *simultaneous* support for: multiple user-specified exemplars (unlike [37]); any number of groups (unlike [2, 12, 37]); linear scalability with the number of edges (best possible for most graph algorithms); and soft clustering, supporting membership in multiple groups (unlike [18]). Few tools have integrated graph algorithms to interactively help people make sense of network information [12, 27, 37], and they often only support some of the sensemaking features offered by Apolo, e.g., [37] supports one group and a single exemplar.

INTRODUCING APOLO

The user interface

The Apolo user interface is composed of three main areas (Figure 2). The *Configuration Panel* at the top (1) provides several means for the user to reduce visual clutter and enhance readability of the visualization. Citation edges and article titles (i.e., node names) can be made invisible, and the font size and the visible length of the titles can be varied via sliders. On the left is the *Filter Panel* (2) and *Group Panel*

(3). The *Filter Panel* provides filters to control the types of nodes to show; the default filter is "Everything", showing all types of nodes, except hidden ones. Other filters show nodes that are starred, annotated, pinned, selected, or hidden. The *Group Panel* lets the user create, rename, and delete groups. Group names are displayed with the automatically assigned color (to the left of the name) and the exemplar count (right side). Each group label also doubles as a filter, which causes only the exemplars of that group to be visible. The *Visualization Panel* (4) is the primary space where the user interacts with Apolo to incrementally and interactively build up a personalized visualization of the data.

Apolo in action

We show how Apolo works in action through a sensemaking scenario of exploring and understanding the landscape of the related research areas around the seminal article *The Cost Structure of Sensemaking* by Russell et al. (Figure 2 shows final results). This example uses real citation network data mined from Google Scholar using a breadth-first strategy to crawl all articles within three degrees from the above paper. The dataset contains about 83,000 articles (nodes) and 150,000 citations relationships (edges, each represents either a "citing" or "cited-by" relationships). This scenario will touch upon the interactions and major features of Apolo, and highlight how they work together to support sensemaking of large network data. We will describe the features in greater depth in the next section.

We begin with a single source article highlighted in black in the center of the interface (Figure 3a) and the ten most relevant articles as determined by the built-in BP algorithm. Articles are shown as circles with sizes proportional to their citation count. Citation relationships are represented by directed edges.

After viewing details of an article by mousing over it (Figure 4), the user moves it to a place on the landscape he thinks appropriate, where it remains pinned (as shown by the white dot at the center). The user can also star, annotate, unpin, or hide the node if so desired. After spatially arranging a few articles the user begins to visually infer the presence of two clusters: articles about *information visualization* (*InfoVis*) and *collaborative search* (*Collab Search*). After creating the labels for these two groups (Figure 3a), the user selects a good example article about *InfoVis* and clicks the *InfoVis* label, as shown in Figure 4, which puts the article

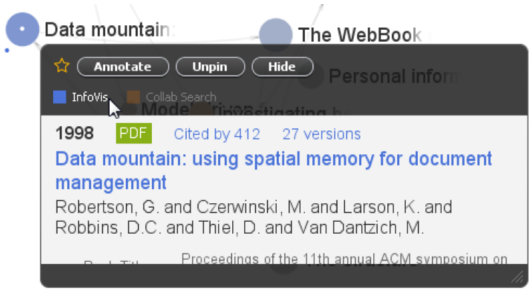


Figure 4. The *Details Panel*, shown when mousing over an article, displays article details obtained from Google Scholar. The user has made the article an *exemplar* the *InfoVis* group by clicking the group's label.

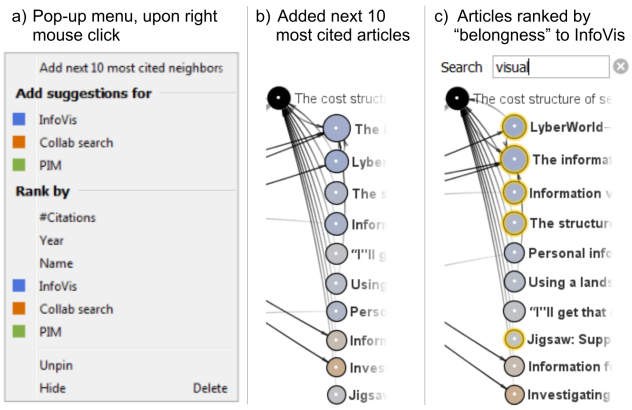


Figure 5. a) Pop-up menu shown upon right click; the user can choose to show more articles, add suggested articles for each group, rank sets of nodes in-place within the visualization, hide or unpin nodes. b) Newly added neighboring nodes ranked by citation count; they are selected by default to allow the user to easily relocate them all to a desired location. c) The same nodes, ranked by their "belongness" to the (blue) InfoVis group; articles whose titles contain"visual" are highlighted with yellow halos.

into that group. A small blue dot (the group's color) appears below the article to indicate it is now an *exemplar* of that group. Changing an article's group membership causes BP to execute and infer the relevance of all other nodes in the network relative to the exemplar(s). A node's relevance is indicated by its color saturation; the more saturated the color, the more likely BP considers the node to belong to the group. Figure 3b-d show how the node color changes as more exemplars are added.

Our user now would like to find more articles for each group to further his understanding of the two research areas. The user right-clicks on the starting paper and selects "Add next 10 most cited neighbors" from the pop-up menu (Figure 5a). By default, new nodes added this way are ranked by citation count (proportional to node size, as shown in Figure 5b) and initially organized in a vertical list to make them easy to identify and process. To see how relevant these new nodes are, he uses Apolo's *rank-in-place* feature to rank articles by their computed relevance to the InfoVis group. To quickly locate the papers about visualization, our user types "visual" in the search box at the top-right corner (Figure 5c) to highlight all articles with "visual" in their titles.

Going further down the list of ranked articles, our users found more InfoVis articles and put them all into that group. Within it, our user further creates two subgroups spatially, as shown in Figure 6, the one on top containing articles about visualization applications (e.g., *Data mountain: using spa-*

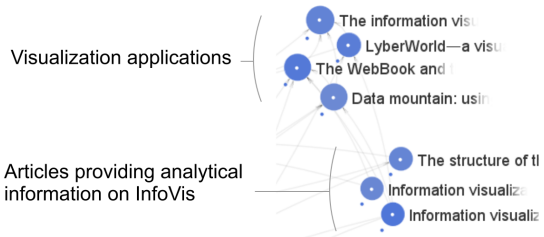


Figure 6. Two spatial subgroups

tial memory for document management), and the lower subgroup contains articles that seem to provide analytical type of information (e.g., *The structure of the information visualization design space*).

Following this work flow, our user can iteratively refine the groups, create new ones, move articles between them, and spatially rearrange nodes in the visualization. The user's landscape of the areas related to Sensemaking following further iterations is shown in Figure 2.

CORE DESIGN RATIONALE

Below we discuss core factors that demonstrate Apolo's contributions in supporting sensemaking.

Guided, personalized sensemaking and exploration

A key factor in the design of Apolo was having exploration and sensemaking be user-driven rather than data-driven—using structure in the data to support the user's evolving mental model rather than forcing the user to organize their mental model according to the structure of the data. This led to several interrelated features and design decisions. First, it drove our decision to support *exploration through construction*, where users create a mental map of an information space. By allowing users to define the map by pinning nodes to the layout, the system provides stability: familiar landmarks do not shift, unless the user decides to shift them. Contrast this to a pure force-directed layout algorithm, which may place items in a different location every time or shift all items when one is moved. Apolo's support for hybrid layout, mixing user-driven and automatic layout, is also different from work on semi-automatic layout (e.g., [32]) that uses constraints to improve a final layout, whereas Apolo supports constraints (fixing node positions) to help users evolve and externalize their mental models.

Second, instead of using an unsupervised graph clustering algorithm that uses the same similarity space for every user (as in [27]), we adapted a semi-supervised algorithm (BP) that would fundamentally change the structure of the similarity space based on user-labeled exemplars. Apolo uses this algorithm to find relevant nodes when starting up or when the user asks for group- or paper-relevant nodes, and to quantify relevance for use in ranking-in-place or indicating relevance through color. This means that even if two users' landscapes included the same nodes, those landscapes could be very different based on their goals and prior experience.

Multi-group Sensemaking of Network Data

Another important factor was the ability to support multiple dynamic, example-based groups. Theories of categorization suggest that people represent categories and schemas through examples or prototypes [29], as opposed to what are typical way of interacting with collections of information online such as search queries or tags. Furthermore, items may and often do belong to multiple groups, leading to a need for "soft" clustering.

Apolo was designed to support multiple groups both in the interface and algorithmically. In the interface, users can easily create multiple groups and move nodes into and out of one or more groups (via the *Details Panel*, Figure 4). Users can also see the degree to which the algorithm predicts items to be in a group through color. The use of the BP algorithm is instrumental as it can support fast, soft clustering on an arbitrary number of groups; many graph-based spreading activation-style algorithms are limited to one or two groups (such as PageRank [5], and random walk with restart [36]).

Evaluating exploration and sensemaking progress

Sensemaking involves evolution and re-representation of the user's mental model. Users typically continue evolving their models until they stop encountering new information; when the new information they encounter confirms rather than changes their existing mental representations; and when their representations are sufficiently developed to meet their goals. To assist in this evaluation process, Apolo surfaces change-relevant information in a number of ways. It helps the user keep track of which items have been seen or hidden. New nodes are added in a systematic way (Figure 5), to avoid disorienting the user. Pinned nodes serve as fixed landmarks in the users' mental map of the information space and can only be moved through direct action by the user. When saving and loading the information space, all pinned nodes remain where the user put them. Apolo uses all these features together to preserve the mental map that the user has developed about the graph structure over time [24].

As a node's group membership can be "toggled" easily, the user can experiment with moving a node in and out of a group to see how the relevance of the other nodes change, as visualized through the nodes' color changes; thus the effect of adding more exemplars (or removing them) is easily apparent. Also, color changes diminishing over time can help indicate the user's representations stabilizing.

Rank-in-place: adding meaning to node placement

Typically, layout algorithms for graphs (e.g., force-based layouts) try to layout nodes to minimize occlusion of nodes and edges, or to attain certain aesthetic requirements [9]. These layouts usually offer little help with the sensemaking process. Approaches to address this problem include "linked

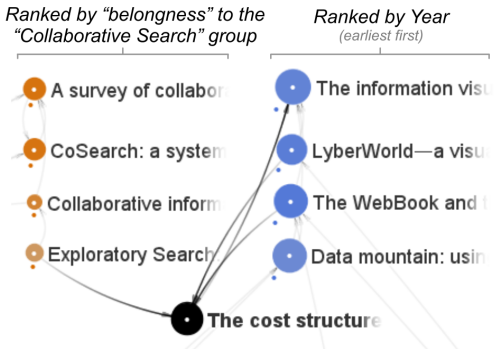


Figure 8. Our user applies two *rank-in-place* arrangements to the exemplars of *Collaborative Search* group (left, ranked by their "belongness" to the group) and the *InfoVis* group (right, ranked by year). These two side-by-side arrangements reveal several insights: 1) the *Collaborative Search* articles are less cited (smaller node sizes), as the research is more recent; 2) only the article *Exploratory Search* in the *Collaborative Search* group cited the source paper, while three papers in the *InfoVis* has either cited or been cited by the source paper.

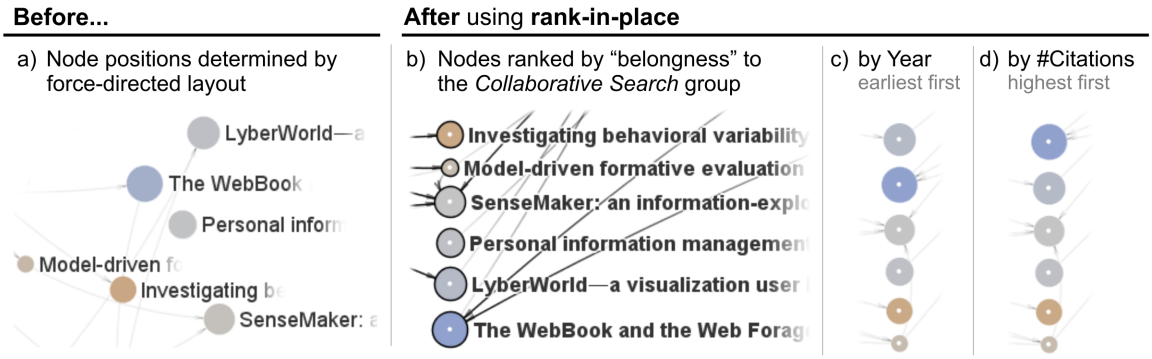


Figure 7. Illustrating how the user can learn more about a set of nodes using the rank-in-place feature, which imparts meaning to the nodes' locations, by vertically aligning and ordering them by a specified node attribute. (Node names have been hidden in figures c and d to reduce clutter.) a) Without using rank-in-place, nodes' positions are determined by a force-directed layout algorithm, which strives for aesthetic pleasantness, but offers little to help with sensemaking. b) Using rank-in-place, our user ranks the articles by their "belongingness" to the *Collaborative Search* group. c) Ranking articles by year reveals a different structure to the data. d) Ranking by citation count shows the "WebBook" article appears at the top as most cited.

views", which use a separate view of the data (e.g., scatter plot) to aid node comparison; and imparting meaning to the node locations (e.g., in NodeXL [33] and PivotGraph [39]). However, the above approaches are often global in nature: all nodes in the network are repositioned, or a new, dedicated visualization created.

Apolo's main difference is that it offers a rank-in-place feature that can rank local subsets of nodes in meaningful ways. Figure 7 shows one such example. Furthermore, the user can create multiple, simultaneous arrangements (through ranking) for different sets of nodes, which can have independent arrangement criteria (e.g., one ranks by year, the other ranks by citation count). We designed for this flexibility to allow other characteristics of the nodes (e.g., their edges, node sizes) to still be readily available, which may then be used in tandem with the node arrangement across multiple rankings (see Figure 8). Recently, researchers are exploring techniques similar to rank-in-place, to create scatterplots within a network visualization [38].

IMPLEMENTATION & DEVELOPMENT

Informed design through iterations

Sensemaking for large network data is an important problem which will undoubtedly take years of research to address. As such, our Apolo system only solves part of it; however, the system's current design is the result of over two years' investigation and development effort through many iterations and two major revisions.

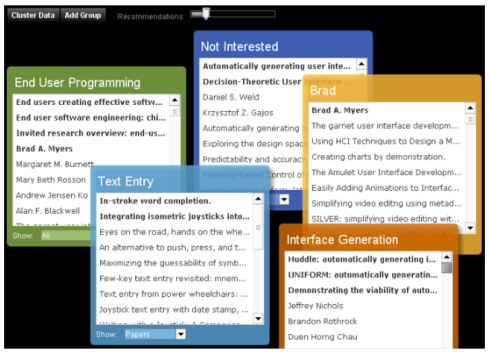


Figure 9. The first version of Apolo

The first version of Apolo presented suggestions in ranked lists without a visualization component (Figure 9), one list for each group in a floating window. Users' exemplars showed up at the top of the lists with the most relevant items in ranked order below them . We initially thought that the high data density and ease of comprehension of the list format might lead to better performance than a spatial layout. However, our users quickly pointed out the lack of a spatial layout, both as a workspace to externalize their mental representations as well as to understand the relations between items and between groups. This finding prompted us to add a network visualization component to the second revision.

While working towards the second revision, we conducted contextual inquiries with six graduate students to better understand how they make sense of unfamiliar research topics through literature searches. We learned that they often started with some familiar articles, then tried to find works relevant to them, typically first considering the articles that cited or were cited by their familiar articles. Next, they considered those new articles' citation lists. And they would repeat this process until they had found enough relevant articles. This finding prompted us to add support for incremental, link-based exploration of the graph data.

We studied the usability of the second revision through a pilot study, where we let a few researchers use Apolo to make sense of the literature around new topics that they recently came across. We learned that (1) they had a strong preference in using spatial arrangement to manage their exploration context, and to temporarily organize articles into approximate (spatial) groups; (2) they used the visualization directly most of the time, to see relations between articles and groups, and they only used the list for ranking articles. These findings prompted us to rethink Apolo's interaction design, and inspired us to come up with the *rank-in-place* technique that offers benefits of both a list-based approach and a spatial layout. *Rank-in-place* lays out nodes at a greater density, while keeping them quick to read and their relations easy to trace. With this new technique, we no longer needed the suggestion lists, and the visualization became the primary workspace in Apolo.

System Implementation

The Apolo system is written in Java 1.6. It uses the JUNG library [26] for visualizing the network. The network data is stored in an SQLite[1] embedded database, for its cross-platform portability and scalability up to tens of gigabytes. One of our goals is to offer Apolo as a sensemaking tool that work on a wide range of network data, so we designed the network database schema independently from the Apolo system, so that Apolo can readily be used on different network datasets that follow the schema.

We implemented BP as described in [22]. The key settings of the algorithm include: (1) a *node potential* function that represents how likely a node belongs to each group (a value closer to 1 means more likely), e.g., if we have two groups, then we assign (0.99, 0.01) to exemplars of group 1, and (0.5, 0.5) to all other nodes; (2) an *edge potential* function that governs to what extent an exemplar would convert its neighbors into the same group as the exemplar (a value of 1 means immediate conversion; we used 0.58).

EVALUATION

To evaluate Apolo, we conducted a laboratory study to assess how well people could use Apolo on a sensemaking task on citation data of scientific literature. At a high-level, we asked participants to find papers that could be used to update the related work section of a highly cited survey paper describing software tools for HCI [25]. We considered using other datasets such as movie data, but we felt that evaluation could be too subjective (genres are hard to define). In contrast, scientific literature provides a good balance between objectivity (some well-known research areas) and subjectivity (subject's different research experience). Another reason we chose scientific literature was because it was easier to assess "ground-truth" for evaluating the study results. More specifically, we used literature from computer science research areas of HCI (i.e., UIST), and had experts at our institution help establish "ground truth."

Participants

We recruited twelve participants from our university through a recruitment web site managed by our institution and through advertisements posted to a university message board. All participants were either research staff or students, and all had backgrounds in computer science or related fields, so they would be comfortable with the technical computer-related terms mentioned in the study materials. Our participants' average age was 24, and 9 were male and 3 were female. All participants were screened to make sure they had (1) participated in research activities, (2) were not familiar with user interface (UI) research, and (3) had conducted literature search before using Google Scholar. Seven of them have used other citation managers/websites, such as PubMed or JSTOR. Each study lasted for about 90 minutes, and the participants were paid $15 for their time.

We also had two judges who were experts with HCI research help evaluate the results of the participants. These judges have taught classes related to the UIST topics that the participants were exploring.

[1] www.sqlite.org

Apparatus

All participants used the same laptop computer that we provided. It was connected to an external 24" LCD monitor, with a mouse and keyboard. The computer was running Windows 7 and had Internet Explorer installed, which was used for all web-browsing-related activities. For the Apolo condition, we created a web crawler that downloaded citation data from Google Scholar using a breadth-first-search within three degrees of separation using the survey paper as the starting point. There was a total of about 61,000 articles and 110,000 citations among them.

Experiment Design & Procedure

We used a between-subjects design with two conditions: the Apolo condition and the Scholar condition, where participants used Google Scholar to search for papers. We considered using a within-subjects design, where the participants would be asked to find related work for two different survey papers from different domains using the two tools; however that would require the participants to simultaneously have a background in both domains while not being knowledgeable about both. We believed these constraints would make the scenarios used in the study overly artificial, and that qualified participants would be much harder to come across. However, we still wanted to elicit subjective feedback from the participants, especially for their thoughts on how the two tools compare to each other for the given task. To do this, we augmented each study with a second half where the participants used the other tool that they did not use in the first half. None of the data collected during these second tasks were used in the quantitative analysis of the results.

We asked participants to imagine themselves as researchers new to research in user interfaces (UI) who were tasked with updating an existing survey paper published in 2000. The participants were asked to find potentially relevant papers published since then, where relevant was defined as papers that they would want to include in an updated version. We felt that defining "relevant" was necessary and would be understandable by researchers. Given that finding relevant papers for the entire survey paper would be a very extensive task, both for participants and for the judges, we asked participants to focus on only two of the themes presented in the survey paper: *automatically generating user interfaces based on models* and *rapid prototyping tools for physical devices, not just software* (both paraphrased), which were in fact section titles in the paper.

In each condition, the participants spent 25 minutes on the literature search task. They were to spend the first 20 minutes to collect relevant articles, then the remaining five to select, for each category, 10 that they thought were most relevant. They did not have to rank them. We limited the time to 25 minutes to simulate a quick first pass filter on papers. In the Google Scholar condition, we set the "year filter" to the year 2000 so it would only return articles that were published on or after that year. In the Apolo condition, we started people with the survey paper. Participants in

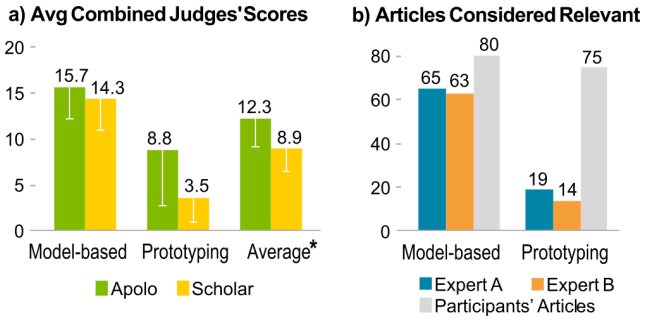


Figure 10. a) Average combined judges' scores; an article from a participant receives a score of 1 when it is considered relevant by an expert (min is 0, max is 2). The average score across both categories was significantly higher in the Apolo condition. Error bars represent -1 stdev, * indicates statistically significant. b) The number of articles considered relevant by the two expert judges (blue and orange). The total number of articles found by participants are in gray. The two experts strongly agree with each other in their judgment.

the Apolo condition were given an overview of the different parts of Apolo's user interface and interaction techniques, and a sheet of paper describing its main features.

Results

We examined our results using both a quantitative approach as well as subject measures. We pooled together all the articles that the participants found in the study and divided them into two stacks, "model-based" and "prototyping", according to how they were specified by the participants. For each article, we located a soft copy (usually a PDF) and printed out the paper title, abstract and author information. We collected these printouts and sorted them alphabetically. These printouts were used by the expert judges. We had the two judges select papers relevant to the topic. We represented each expert's judgment as a vector of 0s and 1s. An article was marked with "1" if the expert considered it relevant, and "0" otherwise. We used cosine similarity to evaluate the similarity between the two experts. A score of 1 means complete agreement, and 0 means complete disagreement. For the "Model-based" articles, the cosine similarity between the experts' judgement was 0.97. For the "Prototyping" articles, despite few papers being considered relevant by the judges, the cosine similarity was 0.86.

In our evaluation, the independent factor is "participant", and the dependent factor is the relevance scores assigned by the expert judges. Using a two-tailed t-test, we found the average score across both categories was significantly higher in the Apolo condition ($t(9.7) = 2.32, p < 0.022$), as shown in Figure 10a. We also note that participants in both conditions did well for finding papers related to model-based interfaces. One reason for this is that papers in this category tended to have the word "automatic" or "model" in the title. However, the same was not true for papers in the prototyping category.

Subjective Results

Overall, participants felt that Apolo improved their sensemaking experience. Based on a post-test survey, Apolo seemed to be easy to use, as shown in Figure 11 and 12. These results are encouraging given that Apolo was not designed as a walk-up-and-use system.

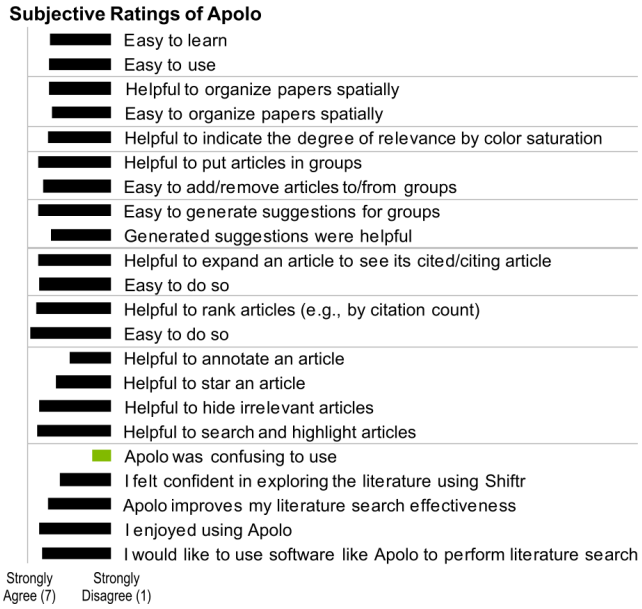


Figure 11. Subjective ratings of Apolo. Participants rated Apolo favorably. Features for grouping, suggesting and ranking nodes, were highly rated. Apolo's usage was very clear to most participants (green bar).

We organized qualitative feedback from participants into three categories. The first relates to general sensemaking and organization. One participant said that Apolo "helped much more in organizing search[...] Google was a bit random." The graph visualization was also helpful. One person said that it "helped me see the citation network in which articles cited which ones". Another said that the "indication of degree relevance by color was extremely helpful and easy to use and it also helped to make sense of the articles."

Being able to spatially organize papers was also a useful feature and participants were happy about using it. One person said that "arrangement of articles in my order of preference & my preferred location was easy." We do note that the task that we had participants do was somewhat limited in scale, but for the current task the screen real estate available was sufficient. Seeing connections between papers was also helpful. Participants said that "it helped me see the citation network in which articles cited which ones", and that it was an "easy way to list & relate citations graphically." Participants also had constructive feedback for improving Apolo. One comment was to have abstracts be more readily available in Apolo (a constraint of the crawling software, since Google Scholar does not have full abstracts in an accessible format). Another comment was that there would often be too many edges, a problem common to graph visualizations.

Discussion and Limitations

While the results of our evaluation was positive, there are also several limitations. For example, we only had participants examine two themes. Having more themes would stress the screen real estate. Apolo currently has minimal features for managing screen real estate.

We also did not evaluate category creation and removal; participants were given two categories that correspond to two

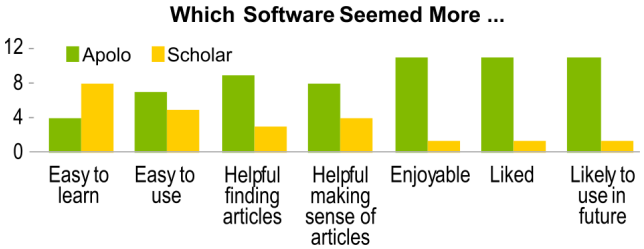


Figure 12. Participants liked and enjoyed using Apolo, and would like to use it again in the future. They generally found Apolo more useful than Google Scholar in helping them find relevant articles and make sense of them.

sections in the given overview paper, which may not be the most natural categories that they would like to create. However, if we were to allow the participants to create any group they wanted, the great variety of possible groups created would make our evaluation extremely difficult. Moreover, a pilot study found that more categories required more prior knowledge than participants would have. Two categories were in fact already challenging, as indicated by the few relevant articles found for the "Prototyping" group.

The need for the participants to categorize articles was created by the tasks; however, in real-world scenarios, such needs would be ad hoc. We plan to study such needs, as well as how Apolo can handle those kinds of tasks, in less controlled situations. For example, how many groups do people create? Do the groupings evolve over time, such as through merging or subdivision. How well do the findings of sensemaking literature apply to large graph data?

DISCUSSION & FUTURE WORK

We believe mixed-initiative approaches for network exploration are becoming increasingly important as more datasets with millions or even billions of nodes and edges are becoming available. These numbers vastly exceed people's limited cognitive capacities. By combining a person's reasoning skills with computer algorithms for processing large amounts of data, we hope to reduce the disparity in gaining access to and comprehending the vast amount of graph data that have become increasingly available. Automated analysis methods and algorithms for extracting useful information, such as patterns and anomalies, are active research topics. However, these methods can rarely give the perfect solutions, or even if they can, it is necessary to convey such information to the user. Our work seeks to reach the goal of supporting sensemaking through a mix of rich user interaction and machine learning.

Visually managing many nodes is also an open problem in graph sensemaking; Apolo currently focuses on filtering, e.g. enabling users to remove nodes from the visualization, re-add them, or focus only on selected ones. By integrating BP, we hope to help users find relevant nodes quickly and remove unnecessary nodes from the visualization more easily than a manual approach. We have been experimenting with methods that will help further, such as semantic zooming, and reversibly collapsing nodes into meta nodes.

Apolo currently relies exclusively on the link-structure of a graph to make relevance judgments. In the future, we would like to integrate other types of information as well, such as the textual content of the nodes and edges (e.g., the content of an article), edge weights, and temporal information (particularly important for analyzing dynamic networks).

We are also developing features in Apolo to support data from many more sources. Some of these include other citation data sets, such as from the ACM and the IEEE digital libraries, CiteSeerX, and using one's own personal citation libraries (in software like Zotero, Mendeley, and EndNote) to discover new papers. However, Apolo can be used on any network data; for example, we have conducted successful proof of concepts integrations of tagging data from delicious, friendship networks from Facebook, and recipe-ingredient networks. As we move forward it will be important to understand the degree to which context-specific features are necessary in order to support sensemaking across a variety of domains. Finally, we plan to release Apolo for general use and collect feedback and ideas from the community about its usefulness and how to improve it.

CONCLUSIONS

This paper introduces Apolo, a mixed-initiative system for helping users make sense of large network data. Apolo tightly couples large scale machine learning with rich interaction and visualization features to help people explore graphs through constructing personalized information landscapes. We demonstrate the utility of Apolo through a scenario and evaluation of sensemaking in scientific literature. The results of the evaluation suggest the system provides significant benefits to sensemaking and was viewed positively by users. This work focuses on the scientific literature domain; a recent work [12] showed that approaches similar to ours (based on spreading activation) could also work well for graph data of websites and tags. We believe that the ideas in Apolo can be helpful for many other kinds of data intensive domains, by aiding analysts in sifting through large amounts of data and directing users' focus to interesting items.

ACKNOWLEDGEMENT

We thank Scott Hudson and Brad Myers for their help with the user study. This material is based upon work supported by the National Science Foundation under Grants No. CNS-0721736, IIS-0968484, and OCI-0943148. This work is also partially supported by an IBM Faculty Award.

REFERENCES

1. B. Amento, W. Hill, L. Terveen, D. Hix, and P. Ju. An empirical evaluation of user interfaces for topic management of web sites. In *Proc. CHI*, page 559. ACM, 1999.
2. J. Anderson. A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3):261–95, 1983.
3. M. Q. W. Baldonado and T. Winograd. Sensemaker: an information-exploration interface supporting the contextual evolution of a user's interests. In *Proc. CHI*, pages 11–18. ACM Press New York, NY, USA, 1997.
4. L. Barsalou. Ad hoc categories. *Memory & Cognition*, 11(3):211–227, May 1983.
5. S. Brin and L. Page. The anatomy of a large-scale

hypertextual Web search engine* 1. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.
6. S. Card, G. Robertson, and W. York. The WebBook and the Web Forager: an information workspace for the World-Wide Web. In *Proc. CHI*. ACM, 1996.
7. D. Cutting, D. Karger, J. Pedersen, and J. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. In *Proc. SIGIR*, pages 318–329. ACM Press, 1992.
8. B. Dervin. An overview of sense-making research: concepts, methods and results to date. In *International Communications Association Annual Meeting*, 1983.
9. G. Di Battista, P. Eades, R. Tamassia, and I. Tollis. *Graph drawing: algorithms for the visualization of graphs*. Prentice Hall PTR Upper Saddle River, NJ, USA, 1998.
10. J. Heer and D. Boyd. Vizster: Visualizing online social networks. *Proc. InfoVis*, pages 33–40, 2005.
11. K. Holyoak and P. Thagard. *Mental leaps: Analogy in creative thought*. The MIT Press, 1996.
12. Y. Kammerer, R. Nairn, P. Pirolli, and E. H. Chi. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proc. CHI*, 2009.
13. G. Karypis and V. Kumar. METIS: Unstructured graph partitioning and sparse matrix ordering system. *The University of Minnesota*, 2.
14. F. Keil. *Concepts, kinds, and cognitive development*. The MIT Press, 1989.
15. J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.
16. P. Klerks. The network paradigm applied to criminal organisations. *Connections*, 24(3):53–65, 2001.
17. B. Kules. From keyword search to exploration: How result visualization aids discovery on the web.
18. V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics*, 22(3):277–286, 2003.
19. B. Lee, C. Parr, C. Plaisant, B. Bederson, V. Veksler, W. Gray, and C. Kotfila. Treeplus: Interactive exploration of networks with enhanced tree layouts. *TVCG*, pages 1414–1426, 2006.
20. Y. Maarek, M. Jacovi, M. Shtalhaim, S. Ur, D. Zernik, and I. Ben-Shaul. WebCutter: a system for dynamic and tailorable site mapping. *Computer Networks and ISDN Systems*, 29(8-13):1269–1279, 1997.
21. J. Mackinlay, R. Rao, and S. Card. An organic user interface for searching citation links. In *Proc. CHI*, page 73. ACM, 1995.
22. M. McGlohon, S. Bay, M. G. Anderle, D. M. Steier, and C. Faloutsos. Snare: a link analytic system for graph labeling and risk detection. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 1265–1274, New York, NY, USA, 2009. ACM.
23. G. McKiernan. New Age Navigation. *The Serials Librarian*, 45(2):87–123, 2003.
24. K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of Visual Languages and Computing*, 6(2):183–210, 1995.
25. B. Myers, S. Hudson, and R. Pausch. Past, present, and future of user interface software tools. *TOCHI*, 7(1):3–28, 2000.
26. J. OMadadhain, D. Fisher, P. Smyth, S. White, and Y. Boey. Analysis and visualization of network data using JUNG. *Journal of Statistical Software*, 10:1–35, 2005.
27. A. Perer and B. Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proc. CHI*, pages 265–274, New York, NY, USA, 2008. ACM.
28. P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. IA*, 2005.
29. E. Rosch and C. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605, 1975.
30. D. M. Russell, M. Slaney, Y. Qu, and M. Houston. Being literate with large document collections: Observational studies and cost structure tradeoffs. In *HICSS*, page 55. IEEE, 2006.
31. D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proc. CHI*, pages 269–276. ACM Press, 1993.
32. K. Ryall, J. Marks, and S. Shieber. An interactive constraint-based system for drawing graphs. In *Proc. UIST*, pages 97–104. ACM, 1997.
33. M. Smith, B. Shneiderman, N. Milic-Frayling, E. Mendes Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave. Analyzing (social media) networks with NodeXL. In *Proc. C&T*, pages 255–264. ACM, 2009.
34. J. Stasko, C. Gorg, Z. Liu, and K. Singhal. Jigsaw: supporting investigative analysis through interactive visualization. *InfoVis*, 7(2):118, 2008.
35. L. Terveen, W. Hill, and B. Amento. Constructing, organizing, and visualizing collections of topically related web resources. *TOCHI*, 6(1):94, 1999.
36. H. Tong, C. Faloutsos, and J. Pan. Fast random walk with restart and its applications. In *ICDM*, pages 613–622, 2006.
37. F. van Ham and A. Perer. "Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest. *IEEE TVCG*, 15(6).
38. C. Viau, M. J. McGuffin, Y. Chiricota, and I. Jurisica. The flowvizmenu and parallel scatterplot matrix: Hybrid multidimensional visualizations for network exploration. *IEEE Transactions on Visualization and Computer Graphics*, 16:1100–1108, 2010.
39. M. Wattenberg. Visual exploration of multivariate graphs. In *Proc. CHI*, page 819. ACM, 2006.
40. J. Yedidia, W. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.