TopicLens: An Interactive Recommender System based on Topical and Social Connections

Laura Devendorf
University of California
Berkeley
CA 94720
ldevendorf@berkeley.edu

John O'Donovan
University of California
Santa Barbara
CA 93106, USA
jod@cs.ucsb.edu

Tobias Höllerer
University of California
Santa Barbara
CA 93106, USA
holl@cs.ucsb.edu

ABSTRACT

This paper describes TopicLens, an interactive tool for exploring and recommending items within large corpora, based on both social metadata and topical associations. The system uses a hybrid visualization model that represents topics and content items side by side, allowing the user to actively explore recommendations rather than passively viewing them. The approach provides insight into the composition of relevant topics as they relate to the meta-data of underlying texts. We describe a novel approach to sorting and filtering, which can be topic or document-driven, and two novel interaction styles termed "view inversion" and "human-review", each of which enable novel perspectives on topic modeled sets of documents. To evaluate the system, three use cases are presented to highlight interesting insights across three different data sets using our novel recommendation interface.

1. INTRODUCTION

Recommender systems attempt to ease the information overload problem by providing the right information to the right person at the right time [31, 19, 33]. However, presentation *mechanisms* for these systems are becoming increasingly important, as they are applied to increasingly more diverse data on the social web. For example, Herlocker's early experiments on the value of explaining recommendations [19] have informed and influenced many of today's recommender system designs. Tintarev and Masthoff [38] survey the role of explanation as an integral part of the recommendation process and outline seven distinct advantages of providing explanation. More recent efforts to analyse the effect of "inspectability and control" [21], interactive visual feedback [6], and dynamic critiquing [30, 10] clearly show that the interface components play an important role in a user's acceptance and overall trust in a recommendation.

In this paper we focus on one specific interface design (Figure 1 for exploration of recommendations which have been derived from a topic modeling algorithm. Topic modeling is a statistical method for extracting relevant topics from a large corpus of text. Visualization of connections formed through topic modeling can enable users to quickly identify trends and other insightful details from a large data set. Successful visualizations are especially effective at highlighting patterns within high dimensional data. Such visualizations may also allow the user to navigate and dynamically filter information in order to extract specific and relevant items. Example use cases are:

- To augment the users ability, beyond keyword based search and navigation, to discover topical composition and interrelationships in texts (i.e. recommendation via topic associations).
- To highlight popular trends and conversations within social networks.
- To compare bodies of text, visually exploring similarities, differences and patterns in the underlying texts for better personalized result sets.

The focus of this paper is largely on the UI design and on novel interaction techniques to represent connections formed over large text datasets using topic modeling or other automated text analysis algorithms. The key elements in our visual representations include:

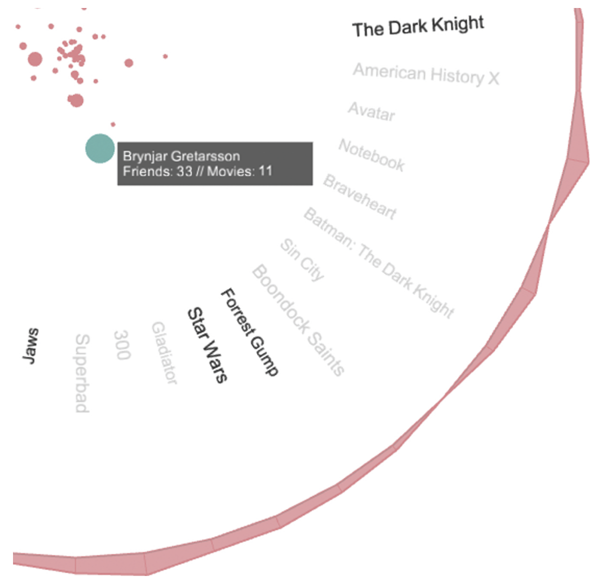

Figure 1: A snapshot of TopicLens interactively recommending movies from Facebook API. The segment shows the popularity of each item among friend groups on the outer ring, and highlights recommendations in bold on the inner ring. This view is highly dynamic and changes based on mouseover interactions

- *Recommendable Item*: An abstract entity which can translate to either a text document or a user within a social network. These are conceptually grouped because they are both represented by collections of terms. For example, in the Twitter data set, a user is represented as a collection of Tweets.

- *Topic*: Multinomial distributions over a set of terms, which can be associated with content items.

While established representations, such as word clouds and tree maps [35] can be useful for visualizing frequency in topic-item relationships, we describe a model that also preserves and represents relationships at the meta-data level. This allows users not only to see which topics arise, but also how they arose and under what conditions. The approach enables more informed reasoning about documents a user wishes to investigate, while highlighting trends over a number of different types of networks with respect to a particular investigation.

Microsoft's "Twahpic" [29] approach to visualizing topics in conjunction with meta-data leverages a composite view that optimizes its visualization strategy for each different facet of the data. This strategy is effective for illustrating and highlighting the multifaceted nature of the data, but is difficult to navigate due to the separation of each frame and the segregation of the data networks. In short, the interaction model helps a user form impressions of the data rather than supporting investigations into the data.

Work by Cao et al. in [8] shows a benefit of using multiple approaches to visualizing the different facets of the data, and in this paper, we will present a model that takes a hybrid approach rather than a segregated approach in order to facilitate navigation and interaction with the data. The key features of the proposed technique are as follows:

- Presents a choice of view modes, sorting parameters and controls for navigation and dynamic filtering.

- Enables a user to filter topics in relation to the pre-existing networks in the data.

- Allows for human oversight of algorithmically generated results.

- Enables exploration of dataset as a map, traversing and isolating regions of particular interest in order to extract relevant items.

- Caters to diverse topic modeling scenarios, including additional data such as social and information networks.

In the remaining sections, we will discuss the related research and provide a brief background of topic modeling before describing in detail the design decisions made when developing the TopicLens interface. The design decisions include those related to overall structure and the mapping of formal elements to relational information. Novel aspects of the interface are also discussed, particularly new techniques that we have termed ï£¡view inversionï£¡ and ï£¡human review.ï£¡ We will then present three applications of the system, one of which uses data that does not contain topic-based relations, thus highlighting a more generalized application of the design.

2. RELATED WORK

Due to the proliferation of data available on the web, there is an increasing need for better techniques for exploration of large amounts of text data. This is commonly known as addressing an information overload problem [20]. Ongoing research has produced proactive, query-based solutions in the fields of search [12] and reactive or filter-based approaches in the field of recommendation [20, 7]. In the context of this work, we are especially interested in approaches that employ visual and interactive methods to tailor an information space to a user's individual needs. The novel approach presented in this paper employs a statistical method known as Latent Dirichlet Analysis (LDA) or "topic modeling" [4, 3] to discover useful linkages between documents upon which visualizations are built.

Mathematical Theory	theorem lemma proof follow constant bound exist definition
Software Engineering	software process tool project development design system developer
Gene Expression	protein genes expression network motif interaction pathway genome
Politics and Society	political social policy economic china law government national
Business and IT	business firm services customer technology management market product
Fluid Dynamics	flow velocity wall fluid turbulence reynold pressure channel

Table 1: Examples of LDA topics learned on a corpus of research papers

While there has been a significant amount of research in this domain from a variety of perspectives, from early approaches such as [27, 40, 18] to more recent work in [36, 37, 39, 22, 25], visual techniques for exploring large sets of documents have not yet been widely adopted.

2.1 Topic Modeling

LDA or "topic modeling" is a statistical technique introduced by Blei et al. [4] that computes focused probability distributions over the words in a set of documents. The algorithm functions by mapping documents onto a smaller number of "topics". In this sense, a topic consists of a multinomial distribution over words or stemmed terms in a document set. For example, as $p(w|t)$, for $t \in 1 \ldots T$, where T is the number of topics [4, 16]. In many cases, topics are displayed as a list of the top n words with the highest probability in the set. Table 1 from [15] shows some example topics produced by an LDA algorithm. In this case, the words "theorem, lemma, proof, follow, constant..." seem to relate to the topic "Mathematical Theory". Recent research in [9, 26] has shown that although LDA topics can be misinterpreted, they are generally well understood by users. Techniques for the automatic labeling of topics have been presented in [23].

In TopicLens, topics are leveraged to form associations among items in a large corpus, and these associations are used to produce informative and highly flexible representations of the broader content item space, using novel layout and interaction techniques. Before describing our approach to visualizing a topic space, we now present a discussion of existing approaches to visualization of large document sets.

Many approaches in the literature dealing with the representation of large text collections, ranging from traditional static representations, e.g. [18], to more recent and highly interactive representations which use advanced methods to relate documents together, e.g. [**?**]. They can rely on pre-existing meta data, or can compute relations on the fly. In this paper, we present a novel interactive design and layout for exploring topic based and social network relations in large document sets. Before presenting the prototype system in detail, the following section provides a brief account of the design choices for using a combination of river and graph-like visual representations in the system.

2.2 The Need for a Hybrid Model

As shown in Figure 2, we are supporting exploration of multi-faceted data in a variety of ways. Specifically, examples are demonstrated on three different network types: social network data with

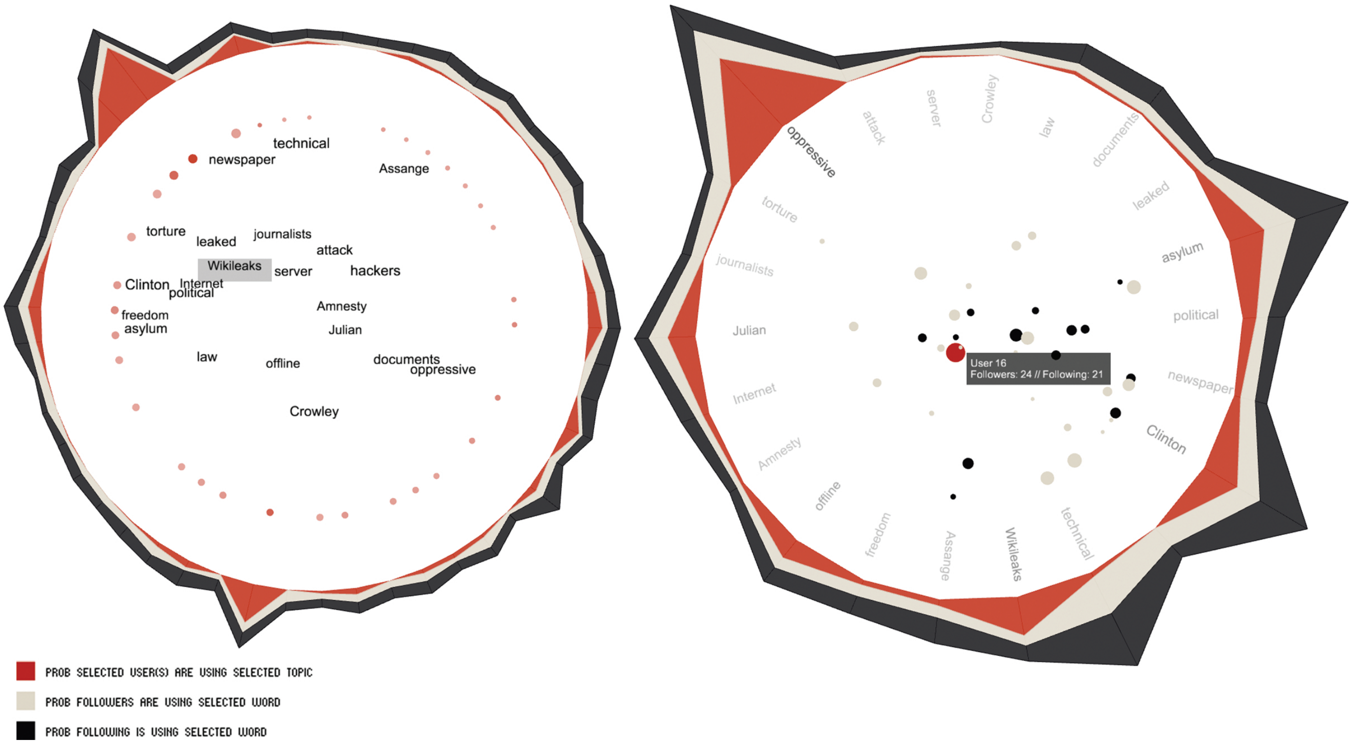

Figure 2: Two detail views of the TopicLens visualization, each showing connections between items and related topics. In this case the data is from the Twitter social network, so our generic "items" represent Twitter users. Frequency measures are shown on the outer river-like component. The two views are of the same data, with items and topics inverted.

unidirectional edges (followers and followees) from Twitter; augmented with topic relations, and a topic modeled network of news articles from the New York Times; and social network data with bi-directional connections from Facebook. Across all examples, the goal is to use simple interaction and novel layouts to facilitate user comprehension of complex data, particularly to communicate the "credibility' factor of peers in a network with respect to particular topics of interest. This complexity would be inherently difficult to communicate with a single visualization technique such as a river or graph visualization. Accordingly, we have opted for a hybrid approach which uses a graph-like mechanism similar to TopicNets [15] for highlighting relations between document and topic nodes, and a river-like view similar to ThemeRiver [17] overlaid to communicate frequency or "credibility" of different sets of peers within the context of a topic selection. This approach has been successful in applications such as Freire's ManyNets [13].

3. DESIGN CONSIDERATIONS

At the core, the TopicLens interface seeks to empower the user to explore a large datasets based on a number of factors. We designed the interface with the idea that potential users would benefit most from learning and engaging in the system rather than making sense of the data at a glance. We see applications of our system being beneficial for any researcher who is looking to glean insights into a large body of text. This includes analysts of social networks as well as scholars in the humanities who may want to use TopicLens to explore trends in the bodies of work by a single author or works belonging to a single or set of genres. We provide functionality with the goal of avoiding a crowded interface and we took great measures to ensure clarity and consistency across multiple view modes. In our informal tests and observations of interactions with the system, we have found it easy to learn and that users take quickly to the dynamic filtering and sorting tools we provide.

4. VISUALIZATION DESIGN

The most prominent feature of the visualization, shown in Figure 2, is its use of the wheel to structure information. Using a circular structure allows us accommodate variability in the size of the datasets. The wheel dynamically expands to fit the data and contracts upon filtering. Zooming and font size are adjusted in order to keep information present within the visualization space, regardless of how much there is to display.

The visualization is designed to fit within a rectangular window with width larger than height. The exact dimensions can vary and in our examples, we found it most effective to use a full screen view on a high-resolution display (1280x1024 and higher), especially when dealing with large sets. The left side of the screen contains the controls and legends and the wheel rotates on an axis in the center of the screen. A static camera is also positioned at the center of the screen, allowing the user to zoom in towards and away from the center. The river is positioned along the outer edge of the wheel and protrudes in different directions depending on the current data selection.

4.1 Organization

In order to support the user in exploring the data at varying levels of detail, the organization of the visualization needs to clearly distinguish the different relationships that are represented. We classify those relationships into three types: primary, secondary and ternary.

The data we collect has pre-existing relationships as formed through meta-data (primary relations), the topic modeling algorithm provides information about relationships between items and topics (secondary relationships), and we found it helpful to further analyze the topics in relation to items and item meta-data (ternary relationships). By dividing the wheel into three concentric regions, we were able to map each type relationship to its own location on the wheel. As you travel from the center out, the information represented reflects a increasing number of factors. The wheel, combined with zooming, was intended to give the user the idea that zooming out will provide them with a big picture, birds-eye overview of the data and zooming in closer will focus on the finer detailed relationships. The following paragraphs provide a detailed explanation of the relationship types and the regions they map to.

4.1.1 Primary Relations: Center

Primary relations are formed though associations in item meta-data. In the analysis of Twitter networks, a single item represents a Twitter user. Item meta-data includes, but is not limited to, a list of followers of this user and a list of other Twitter users that this user is following. In the case of topic modeling run over New York Times articles, primary relationships would be formed between two or more articles that share the same author formed by two articles. Primary relationships are mapped to the center so these relationships can be viewed in a local space. Figure 2 shows primary relations through coloring in the view on the right. In the view on the left, topics are featured in the center. Since primary relations don't exist within topics, no explicit color mapping is represented.

4.1.2 Secondary Relations: Center & Inner Ring

Secondary relationships occur as a result of the topic modeling and define the relationships between topic and item nodes. Each of these relationships occurs with a given probability as defined by the LDA algorithm. These relationships as well as their respective probabilities are represented by interactions between the center and inner ring. While the nodes in the center are not bound to any axis or predetermined path, the nodes in the inner ring are equidistantly laid out in a circle. This is primarily because the inner ring also functions as the axis points for the river visualization but also reinforces simplicity by defining only one type of data to be related spatially. On the left side of Figure 2, highlighting Wikileaks changed the opacity of the nodes on the inner ring in order to indicate how related each item is to this topic. On the right, highlighting User 16 changed the opacity of the topics in the inner ring, similarly showing the strength of the connection.

4.1.3 Ternary Relations: Outer Ring / River

Ternary relationships are formed between the topic modeled results and the meta-information of the items related to those results. Using the river visualization to graph these relationships allows us to see an overall frequency of the node in addition to the meta-information frequencies within the same space. Depending on the data and filtering, the river model can be customized to show any particular facet of the meta-information. Figure 2 is showing average probabilities over each facet of item meta-data in relation to the selected item. The colors in the river match the colors of the meta-data in the center, reinforcing this relationship.

4.2 Visual Mappings

Because the TopicLens visualization needs to encode a rich variety of data, we took care to make the visual encoding of different relationships and concepts distinct. In order to maintain simplicity we map objects and relationships to specific formal elements. Depending on the underlying dataset, visual features may be turned on and off in order to keep the visual complexity to a minimum.

At the root, our information display consists of two basic entities: topics and content items. Items are mapped to circles and topics are mapped to rectangles with the text label of the topic in the center. We made these entities distinct in order to visually and conceptually separate them. The topic text is always visible but the item text is only present on demand. Similarly, the circular shape of the item is always visible but the rectangular shape of the topic is only visible on selection.

Color is used to visually group items based on meta-data. For instance, if there is meta-information about item categories, each category type would map to a unique color. This mapping was chosen partly because it enables a quick visual grouping of items and extends to a large number of categorizations. Another reason for choosing color, was its ability to support a visual connection between the meta-data of the individual item and the corresponding meta-data represented in the river. This offers the user two levels of understanding by illustrating how the meta information is connected to the item as well as the topic.

Opacity is used to illustrate secondary relationships, relationships between topics and content items. These relationships occur with a probability specified by the LDA algorithm. Opacity is an effective means of illustrating these connections as it indicates relative strength. Darker nodes have strong probabilities of relation, lighter have weaker ones. If a node is unrelated, it is removed from the space. Secondary relations are highlighted upon interaction as the user must specify a single item or topic in order to view its connections. If multiple items or topics are selected, then the opacity value is determined by the average probability from all nodes in the selected set.

Position and order are used in conjunction to highlight patterns in the data. Patterns are exposed by using the ordering of the items or topics on the inner ring to position the items or topics in the center. Each value begins in the center of the circle and is pulled towards all of its related nodes in the inner ring. The strength of attraction depends on the probability of the connection between the item and the topic. The result is a spatial grouping of items or topics that share similar relationships. A number of interaction techniques for positioning items on the inner ring will be discussed in the following sections.

Size is used to illustrate measures of numerical magnitude such as frequency or number of relations. Similar to position and ordering, some mappings of attributes to size can be more informative than others. For this reason, we allow the user to indentify the node attribute that determines node size.

5. IMPLEMENTATION

This visualization evolved through a number of design iterations. Using Processing to program the design and interaction allowed us to easily explore changes in the design and instantly see the results. The Processing framework also made it simple to program animations and transitions between states. A number of libraries were used to extend the scope and flexibility ofProcessing. The PeasyCam library provided the basic virtual viewpoint control, the ControlP5 library was used to implement text boxes, range sliders and list boxes and an OpenGL library was used to add custom functionality into the system such as smoothing and alpha blending.

The TopicLens application creates node and edge objects by parsing configuration and data files on load. During the execution of the program, nodes and edge objects are referenced in order to create dynamic links. Links are the elements that are drawn to the canvas and much of the code is devoted to maintaining those links and

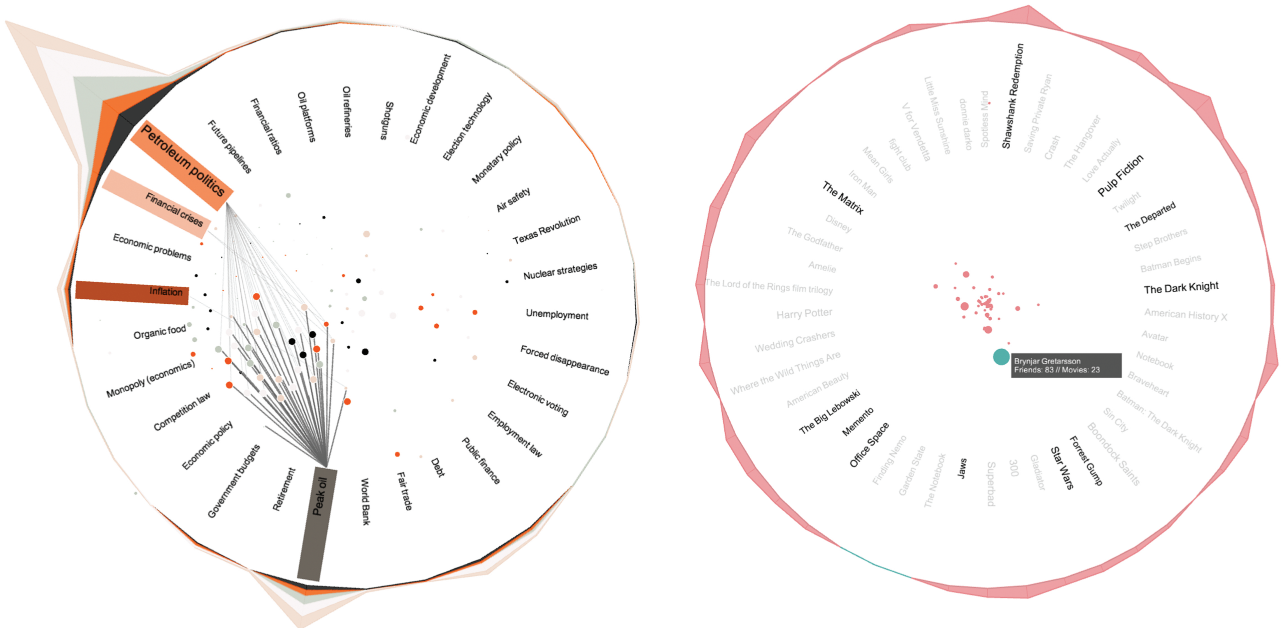

Figure 3: TopicLens view details for varying data sets. Left: TopicLens view of news articles from the New York Times showing topics on the inner ring and articles/items in the center. The view shows selection of an individual topic (peak oil) and edges linking to related articles. Right: TopicLens view of a Facebook social network showing an individual's friend network and associated item preferences.

dynamically updating their values to indicate relationships. The smooth transitions were created using an integrator class that allows the user to specify characteristics such as mass, position, damping and attraction. When a link targets a given position, the integrator dynamically updates its position depending on its physical characteristics.

6. USE CASES AND DISCUSSION

In order to showcase the flexible applicability of our visual model, we present three use cases that explore different dynamic datasets. Each use-case will discuss the design decision made to cater to the specific data domain as well as a usage scenario to illustrate its potential for a variety of applications.

6.1 Recommending Credible Information In Twitter

Preserving social network relations in topic modeled systems allows us to glean insights into the networks and salient topics therein. This example is catered specifically as an attempt to visualize credibility in Twitter networks. Our definition of "credibility" relates to the probability by which a user is connected with a particular topic, based on LDA analysis over a bag of words representation of all of that user's tweets. In analyzing credibility, we also examine that user's followers and followees and their respective associations with the given topic.

In this visualization, which is represented earlier in Figure 2, each topic node contains a label that represents the list of words in a mined topic. Primary relationships are formed between a user and their followers and followees. Secondary relationships occur between users and extracted keywords and the ternary relationships represent probabilities over the meta data. In this example, the user meta data contains a list of other users following this user and a list of the users this user is following. One type of ternary relationship in this example is the relationship between a topic and the average probability that the users friends are discussing that topic.

As noted in Section 4.1, ternary relations are mapped to the river and the nodes in the inner ring form the axis points. The river displays specific information depending on the organization of the nodes within the space. This approach affords the user an opportunity to uncover potentially interesting relations in the following 6 view configurations.

With topic nodes in the center, and user nodes on the outer ring:

- *Upon selection of an individual user*, the river view shows that user, their friends and their followers' probabilistic association with each topic on the outer ring.
- *When no user is selected*, the river shows the average probability for each topic across all users.
- *When a topic is selected*, the river shows each user's association with that topic.

With user/item nodes in the center, and topic nodes on the outer ring:

- *When a topic is selected*, the user's friends and follower's opacity is varied to represent association with that topic.
- *When a user is selected*, their association with each topic on the outer ring is shown in the river view.
- *When no user is selected*, the probability of each topic in the global space is shown on the outer ring

For this scenario, the river represents three probabilities for each node, the average probability of the user using the topic, the average probability of the user's followers using the topic and the average probability that the people following this user are using the topic. Since topics are represented along the inner ring, this information is available for every topic. Each of the probabilities is represented on the river, using color matching to indicate the group or single user it applies to. To further explain what the river is visualizing, a legend on the bottom left of the interface dynamically updates, explaining the current model. In this case mode visibility is of particular importance as the river maps different values through the life of the visualization.

When a Twitter user is highlighted in the space, interactions take place at each of the three levels. In the center, the primary relationships are presented through colors. All users who don't belong to this user's network are removed and the remaining users are color coordinated to indicate whether they are a follower of the selected user, or someone the selected user is following. Spatially, each user is attracted to the topic nodes in the inner ring by the positioning algorithm mentioned above. Topics related to the selected node vary by opacity in order to indicate the strength of connection.

The probability mappings were specifically designed to investigate credibility or trust. The top left of Figure 2 shows a network with two people selected. All of the nodes in the set represent both of the selected people's networks. On the outer river, one can see the probability distributes for this network over each topic in the network. From the river you can conclude that these two users are using the topic "Crowley" quite a bit, however their friends and followers are not. For this reason, they may not be a trusted source for this topic since their followers do not appear to be interested in similar topics. On the other side of the visualization is the topic "asylum" which is being used largely by the network and not so much by these users.

Drawing firm conclusions at this level is not necessarily reliable but better information can be introduced by selecting the right network. For instance, if you know the terms "Assange", "Julian" and "Wikileaks" are all terms related to Wikileaks, then you could select those terms from the visualization and view the results over the given network of users associated with those terms. By investigating the probability of these three words occurring together across the social network you may be able to visualize trends about who is followed, by whom and for what reasons.

6.2 Recommending New York Times Articles

In the example shown on the left of Figure 3, topic modeling was performed on a set of New York Times articles and is used for investigation and discovery of related articles that may not have been discovered through traditional search models. Each document node represents an article and topic nodes represent the topics extracted from those articles. Each article contains information about the section of the paper which it belonged to, such as opinion, world or national news.

Two unique design features were included in this interface to improve the functionality in relation to the underlying data. The first one is colored rectangles on topics. These colors are used to reinforce ternary connections though the use of color averaging. The color of the rectangle is determined by the category of each of the articles associated with it. Should a color tend heavily towards a single category's color, on could deduce that the topic tends to appear most frequently within that category. The actual distribution of the categories is explicitly represented in the river.

The second unique feature is the use of lines. When hovering over a topic, darkened lines extend form the topic itself to all related documents. Lighter lines then extend from each of those related documents to all of the other topics they are related to. This conveys information to the user about other topics related to their selection. The user is able to specifically locate the documents that contributed to this relationship by following the lines or selecting multiple topics and browsing the filtered document space.

The lines are particularly useful for illustrating how two topics are related to each other and upon what criteria. This is helpful when browsing for articles associated with a given theme. Let's say a researcher is looking for references on "peak oil." Searching for and selecting "peak oil" from the space would show the researcher other related topics as well as articles specifically contain the relation. If one of the related articles contains a topic that is also of interest to him or presents a particularly interesting comparison, he or she can easily isolate and obtain information about the articles containing both topics by filtering the space and hovering over the document node, revealing information about the article such as title, date and author.

TopicLens could also be used visualize trends associated with a subset of articles. Say a user read a few articles in the Times Opinion section and they would like to find other articles about similar and related subjects. This could be accomplished by typing each article name into the search field. This would in turn select each of the corresponding articles in the space and illustrate the topics associated with them. In order to remove outliers, they would adjust the slider to specify the amount of documents that need to be associate a topic in order for it to appear in the space. After this filtering step, they are presented with a number of related topics, the most popular being the largest and darkest. By selecting that topic, the space is reorganized to show all the the articles related to that topic. The user can now visually browse these articles and quickly identify which one appeared in the opinion section, based on color. Hovering the mouse over a document node would reveal its specific information and provide access to the full text..

6.3 Recommending Movies via Facebook API

In this example, shown in the right of Figure 3, we use the proposed framework to visualize data that is not topic modeled in order to show how the interface also operates on similarly structured datasets. Reinterpreting the definitions of recommendable item and topic allows us to use the existing visual model for this dataset. In this example, a single Facebook user takes the place of a recommendable item and a movie takes the place of a topic. Since movies can be related to any number of Facebook users and Facebook users can be associated with any number of movies, this dataset can function similarly to the topic model examples. Each item-topic, or rather user-movie, combination is assigned the probability of 1 since the user has specified explicitly that they like the given movie. This visualization is able to provide exploratory views of the most popular movies within a Facebook friend network as well as the least popular movies. It can also isolate pockets of users that are fans of these most or least popular movies. Essentially this view is a visual representation of a social collaborative filtering process, since items which are popular among Facebook friends are promoted for a single target user receiving the recommendation.

7. CONCLUSION

In summary, this paper has presented a novel interactive interface for recommending interesting topics and documents from within a large corpus. The design is a hybrid which combines river and graph-like representations of recommended items and can be easily adapted and customized by the end user for different use cases. We have also introduced novel interaction methods that support human skills in the exploration of topic modeled data sets. In doing so, we have extended the efficacy of both the system and the algorithm, allowing the user to navigate large datasets and uncover patterns. Details of our design choices and methodology have been discussed, and demonstrated over three example applications, including social network data from Twitter augmented with topic modeling over users' tweets, a topic modeled set of New York Times news articles, and social network data from Facebook, including item preferences. In each example case, we have discussed ways in which the approach facilitates discovery of relevant information which may go undiscovered in traditional analysis tools. We have also demonstrated TopicLens' ability to act as a flexible interaction layer, supporting exploration of multiple application domains.

8. ACKNOWLEDGMENT

The authors would like to thank Peter Pirolli and Bongwon Suh for use of their Twitter data set. This work was partially supported by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053; by NSF grant IIS-1058132; and by the U.S. Army Research Laboratory under MURI grant No. W911NF-09-1-0553; The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL, NSF, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

9. REFERENCES

[1] M. Analyzer. Matheo analyzer, database analysis and information mapping. http://www.matheo-analyzer.com/, 2010.

[2] K. Andrews, W. Kienreich, V. Sabol, J. Becker, G. Droschl, F. Kappe, M. Granitzer, P. Auer, and K. Tochtermann. The infosky visual explorer: exploiting hierarchical structure and document similarities. *Information Visualization*, 1:166–181, December 2002.

[3] D. Blei and J. Lafferty. Correlated topic models. In *Advances in NIPS 18*, pages 147–154. MIT Press, Cambridge, MA, 2006.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[5] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.

[6] S. Bostandjiev, J. O'Donovan, , and T. Hollerer. Tasteweights: An interactive hybrid recommender system. In *6th ACM Conference on Recommender Systems, Dublin, Ireland, September 9 to 13th*, 2012.

[7] R. Burke. Knowledge-based recommender systems. In *Encyclopedia of Library and Information Systems*, volume 69, 2000.

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

[9] J. Chang. Reading Tea Leaves: How Humans Interpret Topic Models. 2009.

[10] L. Chen and P. Pu. Critiquing-based recommenders: survey and emerging trends. *User Model. User-Adapt. Interact.*, 22(1-2):125–150, 2012.

[11] D. R. Cutting, D. R. Karger, and J. O. Pedersen. Constant interaction-time scatter/gather browsing of very large document collections. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '93, pages 126–134, New York, NY, USA, 1993. ACM.

[12] B. D. Davison, T. Suel, N. Craswell, and B. Liu, editors. *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010*. ACM, 2010.

[13] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck. Manynets: an interface for multiple network analysis and visualization. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 213–222, New York, NY, USA, 2010. ACM.

[14] B. Fry. *Visualizing data - exploring and explaining data with the processing environment*. O'Reilly, 2008.

[15] B. Gretarsson, J. O'Donovan, A. Asuncion, D. Newman, S. Bostandjiev, T. Hï£¡llerer, and P. Smyth. Topicnets: Visual analysis of large text corpora with topic modeling. *ACM Transactions on the Web: Special Issue on Intelligent Text Visualization*, 16:1172–1181, November 2011.

[16] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(Suppl 1):5228–5235, 2004.

[17] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.

[18] M. A. Hearst. Tilebars: visualization of term distribution information in full text information access. In *CHI '95: Proc. of the SIGCHI Conf.*, pages 59–66, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[19] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *CSCW*, pages 241–250, 2000.

[20] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22:5–53, January 2004.

[21] B. Knijnenburg, J. O'Donovan, S. Bostandjiev, and A. Kobsa. Inspectability and control in social recommenders. In *6th ACM Conference on Recommender Systems, Dublin, Ireland, September 9 to 13th*, 2012.

[22] S. Koch, H. Bosch, and T. Ertl. Towards content-oriented patent document processing. *IEEE Symposium on Visual Analytics, Science and Technology*, pages 203–210, 2009.

[23] Q. Mei, X. Shen, and C. Zhai. Automatic labeling of multinomial topic models. In *KDD*, 2007.

[24] N. E. Miller, P. Chung Wong, M. Brewster, and H. Foote. Topic islands - a wavelet-based text visualization system. In *Proceedings of the conference on Visualization '98*, VIS '98, pages 189–196, Los Alamitos, CA, USA, 1998. IEEE Computer Society Press.

[25] D. Newman, T. Baldwin, L. Cavedon, E. Huang, S. Karimi, D. Martinez, F. Scholer, and J. Zobel. Visualizing search results and document collections using topic maps. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(2-3):169–175, July 2010.

[26] D. Newman, Y. Noh, E. Talley, S. Karimi, and T. Baldwin. Evaluating topic models for digital libraries. In *JCDL '10: Proceedings of the 10th annual joint conference on Digital libraries*, pages 215–224, New York, NY, USA, 2010. ACM.

[27] K. A. Olsen, R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams. Visualization of a document collection: the vibe system. *Inf. Process. Manage.*, 29(1):69–81, 1993.

[28] Questel. Qpat, intellectual property patent and trademark searching. http://www.qpat.com/, 2010.

[29] D. Ramage, S. Dumais, and D. Liebling. Characterizing microblogs with topic models. In A. Cohn, editor, *Proceedings of the Fourth International on Weblogs and Social Media*, pages 130–137. AAAI Press, 23–26 May 2010.

[30] J. Reilly, K. McCarthy, L. McGinty, and B. Smyth. Dynamic critiquing. In *ECCBR*, pages 763–777, 2004.

[31] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *CSCW*, pages 175–186, 1994.

[32] D. A. Rushall and M. R. Ilgen. Depict: Documents evaluated as pictures. visualizing information using context vectors and self-organizing maps. In *Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS '96)*, INFOVIS '96, pages 100–, Washington, DC, USA, 1996. IEEE Computer Society.

[33] U. Shardanand and P. Maes. Social information filtering: Algorithms for automating "word of mouth". In *CHI*, pages 210–217, 1995.

[34] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations, 1996.

[35] B. Shneiderman and M. Wattenberg. Ordered treemap layouts. *Information Visualization, IEEE Symposium on*, 0:73, 2001.

[36] W. Spangler, J. Kreulen, and J. Lessler. Mindmap: Utilizing multiple taxonomies and visualization to understand a document collection. *Hawaii International Conference on System Sciences*, 4:102, 2002.

[37] J. T. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[38] N. Tintarev and J. Masthoff. A survey of explanations in recommender systems. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, pages 801–810. IEEE, 2007.

[39] Wanner. Towards content-oriented patent document processing. *World Patent Information*, 30(1):21–33, 2008.

[40] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In N. D. Gershon and S. Eick, editors, *IEEE Information Visualization '95*, pages 51–58. IEEE Computer Soc. Press, 30–31 Oct. 1995.

[41] P. C. Wong, B. Hetzler, C. Posse, M. Whiting, S. Havre, N. Cramer, A. Shah, M. Singhal, A. Turner, and J. Thomas. In-spire infovis 2004 contest entry. *Information Visualization, IEEE Symposium on*, 0:r2, 2004.