The 5th International Workshop on Information Quality and Quality of Service for Pervasive Computing 2013, San Diego (22 March 2013)

Interactive Interfaces for Complex Network Analysis: An Information Credibility Perspective

(Invited Paper)

James Schaffer*, Byungku Kang*, Tobias Höllerer*, Hengchang Liu†, Chenji Pan†, Siyu Giyu†, John O'Donovan*
*Department of Computer Science, University of California Santa Barbara, Santa Barbara, CA 93106
{james_schaffer, bkang, holl, jod}@cs.ucsb.edu
†Department of Computer Science, University of Illinois at Urbana Champaign, Urbana, IL 61801
{hl4d, cpan8, siyugu2}@illinois.edu

***Abstract*—This paper discusses and evaluates the impact of visualization and interaction strategies for extracting quality information from data in complex networks such as microblogs. Two different approaches to interactive visual representations of data are discussed: an interactive node-link graph and a novel approach where content is separated into interactive lists based on data properties. To assess the two approaches in terms of information credibility, the TopicNets system is compared with "Fluo", a novel system. An analysis scenario is performed through each system on a set of big data filtered from the Twitter message service. The exposure of content, trade-offs between algorithmic power and interaction complexity, methods for content filtering, and strategies for recommending new content are assessed for each system. Fluo is found to improve on TopicNets ability to efficiently find relevant content primarily by providing a more structured content view, however, TopicNets is more customizable and boasts features which are critical for an expert analyst. The paper concludes with general insights on interface design for information filtering systems to maximize perceived quality of information.**

***Index Terms*—information filtering; visualization; interaction; credibility, information networks**

I. Introduction

Increasing availability of networked data is making the task of detecting relevant and credible information ever more challenging and resource intensive. This paper addresses the problem of recommending credible content across composite networks of social media and a range of other communications data. Practical limitations of an information analyst's attention reinforce the need for automated assessments of credibility. Finding optimal combinations of automated and human analyses of network data remains a key challenge since data volume and credibility-determining factors vary greatly across domains, contexts, and missions.

To address this challenge, a scalable credibility analysis toolchain is presented that explores the limitations, potential synergies, and other theoretical boundaries between credibility analysis algorithms and credibility assessments made by human analysts. The toolchain starts with the transfer of data from a credibility analysis engine based on the Apollo system [1] and progresses to a second layer of algorithms that infuse additional modeling, such as the social and content-based credibility models described in [2]. Results are then represented in an interactive visual interface for human analysis. Both the visualization and interaction designs play a key role in an information analyst's perception of Quality of Information in a system.

To explore the role of interaction design in depth, two distinct UI designs with different levels of visual and interaction complexity are analyzed in this paper. Both interaction and visualization approaches that are discussed in the context of a set of Twitter messages filtered by the the prototype toolchain. Before being visualized in either interface, messages are first annotated with information, such as credibility, generated by the Apollo system and subsequent algorithms[1].

Both systems are asked complex questions such as "What is the current difference in sentiment of tweets about #missile between the US and North Korea?" or "What are Twitter users in California saying about #Obama?" The first UI design uses the TopicNets interface from [3], which is a complex graph visualization of messages connected by topic associations. The second approach is a novel interface that organizes a graph view into several columns of ranked and truncated message lists, with a variety of filtering and sorting algorithms that are executed by interacting with data items in each column. In this paper the end goal is to assess interactive mechanisms for analysts that aid in comprehension of the data, the data model, and the underlying filtering algorithms. The longer term goal is to cognitively assess analysts' ability to provide informed feedback that improves underlying filtering models, the interface itself, and most importantly, the credibility-based filtering pipeline as a whole.

II. Architecture

The filtering pipeline in Figure 2 shows three components, which iteratively reduce and refine data to be presented to a human analyst. Data flows through each filter, and the analyst can control each via the UI. The cognitive evaluation mechanisms serve as a second feedback to enable the filtering algorithms to adapt to changes in context. The filtering mechanisms are defined as follows:

A. Large-scale, automated credibility models

The first filtering mechanism in the pipeline focuses on highly scalable models (many millions of nodes). Clearly, human analysts are limited in the scope of data that they can manually assess. Further, many of today's APIs limit data access, making it difficult to run credibility modeling algorithms that rely on many complex queries. Considering these limits, an example of a "large-scale" method might be one that that focuses only on a single network node and associated metadata (available from a source through a

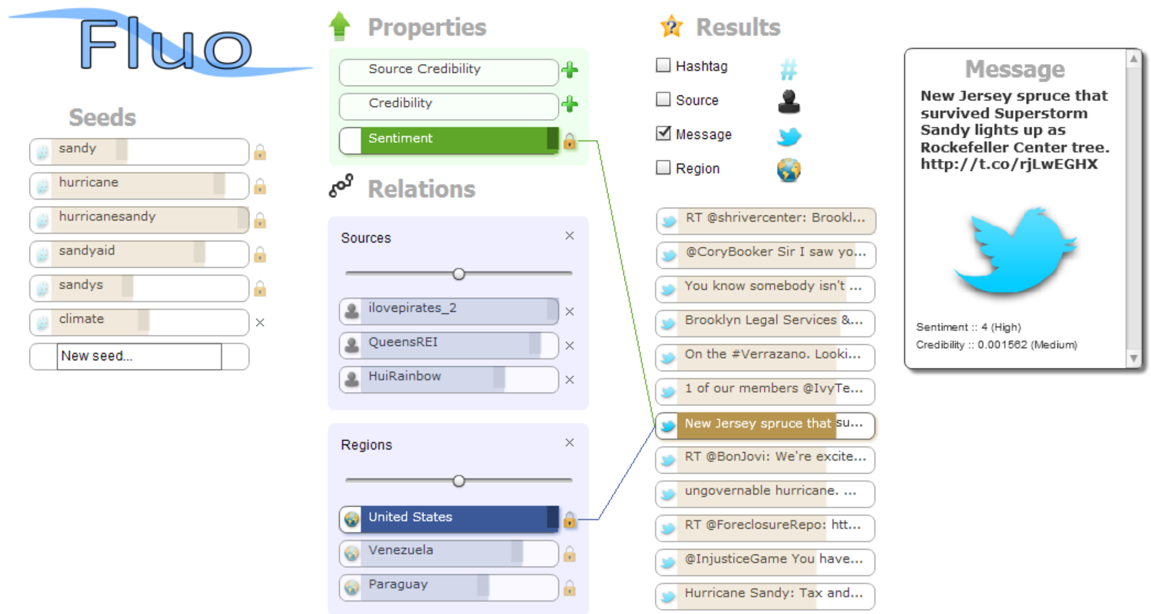

Fig. 1. Screen Shot of the FLUO Interface showing data filtered through Apollo for a query based around "hurricane sandy".

simple query) to arrive at a decision based on credibility. For instance, an analyst might want to query for a user Bob's social network status update. Of course, "Bob" need not necessarily be a human, and simply represents an arbitrary node in an information network. Large-scale credibility-based filtering is performed on network data sources to generate seed data for algorithms that require more complex querying in the subsequent steps of the pipeline.

Apollo [1] is a data distilling service sitting between noisy raw social sensory data and the more sophisticated credibility/topic modeling tools. It aims at distilling large amounts of noisy social sensing data into smaller amounts of more credible information. More concretely, Apollo extracts reliable observations from multitudes of possibly unreliable data sources, which may be unknown to the application in advance, by utilizing a scheme where the credibility of both sources (e.g., users) and claims (e.g., collections of microblog messages about a specific event) are assessed jointly. Apollo converts the collected data into an application-independent form suitable for distillation by representing the reported observations by a graph of sources and claims that is called the source-claim network. The credibility assessment (i.e. ranking) process is performed on both the source and the claim sides, based on the expectation maximization (EM) algorithm [1]. The data source used for the study in this paper is a collection of many millions of tweets that were generated in real-time from Twitter Streaming API, based around the query #sandy, and collected through Apollo.

B. Medium-scale automated models

In the context of the credibility filtering pipeline discussed earlier, the mid-scale approaches are those that apply iterative queries over network(s) to arrive at credibility assertions about a single node. As an example, one might query for status updates from Bob's friends to make a credibility assertion about Bob. This approach enables an analyst to reach various information and its metadata surrounding the desired node that needs to be assessed. Context-sensitive "social" credibility models [4], [2] will be applied to generate more refined assertions about the credibility of information warranting the attention of a human analyst.

In the current prototype of the framework, a subset of these algorithms are implemented. Metadata such as author information and tweet text are gathered and converted to a graph structure which is discussed in more detail later. As a representative sample of the available filtering algorithms from [5], sentiment scores of each claim are computed and passed to the UI. This allows the user to ask interesting questions such as: "Show me the geographic regions that portray a positive sentiment towards the recent North Korean missile launch".

C. Analyst-scale models

Despite recent progress with automated methods for recommending/filtering information, they generally remain far behind human experts [6], especially when a credibility decision is based on a previously unseen or subtle data instance. The proposed integration of scaled credibility modeling algorithms enables an information analyst to focus her efforts on relevant data from a far broader catchment than previously possible. For example, an analyst might want to interactively explore content surrounding the user "Bob" from a composition of previously filtered data sources, probabilistic representations of aligned entities, credibility, and provenance data. Existing data visualization tools [3] are built upon to produce a novel interactive interface that allows analysts to control algorithms at every level in the toolchain. This allows for fast context-based adaptation of credibility models. In preliminary research of the filtering framework, the authors have developed a working prototype called "Fluo" (shown in Figure 1) which combines filtering mechanisms from [1] and [3]. This pro-

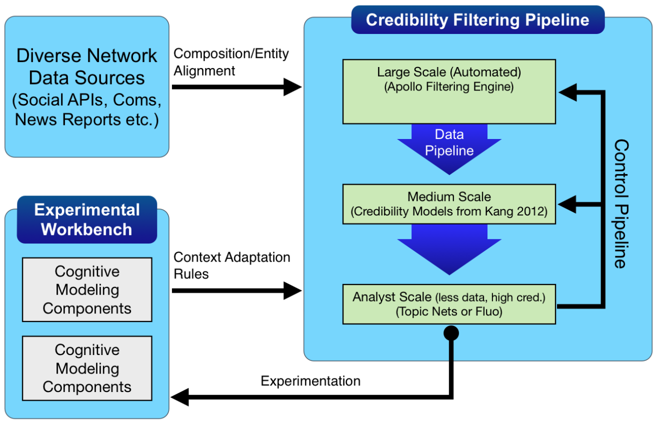

Fig. 2. Overview of the proposed credibility filtering pipeline and the associated experimental workbench.

totype system is compared with "TopicNets"[3], a system for interactively exploring relations between documents via automatically mined common topics.

III. User Interfaces

TopicNets provides tools for an experienced analyst to get both overview and detail views of content while providing interaction and visualization algorithms. Fluo does not offer the breadth of options that TopicNets does, but improves on TopicNets ability to recommend new content principally by providing a more structured content view and interaction techniques that are more focused on rating content to find new nodes. Below, the features of each system that affect an analyst's ability to explore and draw conclusions about sets of big data are discussed.

The TopicNets system (Figure 4) displays content as a node-edge graph and groups similar nodes together on a two-dimensional plane. TopicNets allows an analyst to explore arbitrarily structured information by mapping each data entity onto a node. Properties of each node can either be visualized in a details panel on the right hand side of the screen or discrete properties can be represented as additional nodes with edges to their corresponding data entities. Visualization of content can be customized based on the current state of the system (e.g. which node is selected). New content can be discovered by an analyst by "drilling down" into each node to uncover its properties and relations with other nodes. Collections of related nodes can be ordered spatially to visualize selected properties. An analyst can query over the labels of the nodes in the graph and create a new view of the graph's sub-network representative of the query. Despite supporting the above interaction techniques, the view in TopicNets is frequently cluttered with edges, making it difficult for an analyst to find relevant content. Furthermore, since the layout of nodes is unstructured, it can be difficult to separate query from results. Heavy use of attribute color mapping, spatial layout manipulation, node highlighting, and node labeling only do so much to solve TopicNet's inherent content visibility problem. Expert use of the system is required to produce informed feedback on the mechanisms that generated the data, as well as discoveries within the data itself.

Fluo (Figure 1) is similar to TopicNets in that the system displays content as a node-edge graph and that all content appears as a node. Fluo reduces the complexity of TopicNets' content visualization by showing nodes in one of three columns and only showing a subset of the graph at a time which is determined by a small set of seed nodes that are chosen by the analyst and are known to be credible or interesting. Seed nodes are shown in the leftmost column and their immediate neighbors are shown in the middle "Relation" column. The neighbors of the middle column are then shown in the rightmost "Results" column; in this way, the number of nodes potentially shown in each column increases dramatically from left to right. This problem is solved by assigning a score to each node and only showing the top N nodes in each column. Scoring methods can be customized based on the underlying data model and scores flow from the leftmost column (where the seed nodes are located) along the edges until they reach the rightmost column. Additionally, properties of content that are not represented as nodes in the data model are also shown in the middle column and can be controlled to find nodes with similar properties in the rightmost column. Since the analyst only has to inspect and score the seeds that were manually chosen and the neighbors of those seeds, the number of overall interactions is minimized while still scoring the potentially enormous amount of data in the right hand column. Once nodes have been scored, the analyst can quickly scan down the list of ranked nodes in the rightmost column to find interesting content; if the results are unsatisfactory the leftmost and middle columns can be re-scored until desired results are achieved.

IV. Use Case

Here, the methodology for assessing the TopicNets and Fluo interfaces for information analysis tasks is discussed. A task that includes location, credibility, and sentiment analysis is used for assessment: *"what are the different perspectives on Hurricane Sandy both within the US and abroad?"*. Through "cognitive walkthrough" exercises, these questions are addressed in both systems by inspecting a corpus of data crawled and processed from Twitter API through the Apollo system. This section and the next describe step by step processes for answering the challenge question in each system. Note that these lists do not represent a comprehensive list of functionality for each system and are simply example workflows supported by each system.

A. Data Model

Data is extracted from the Twitter messaging service using the previously described data filtering toolchain. The original dataset consists of nearly two million tweets tagged with metadata such as author, region, timestamps, and associated hashtags. In the example, a small subset of these two million tweets which are further filtered, clustered, and then processed by Apollo into a smaller dataset of "claims". In Apollo, a "claim" is a cluster of tweets on a particular topic, deemed credible by the system. In the output data, claims usually have a single representative tweet. The medium-scale system tags each representative tweet with a sentiment value and processes the information into a source-claim-region-hashtag network for use in the UIs. Rather than having every node in the graph be a particular claim, the model's power to recommend content is maximized by modeling the properties of each claim as its own node and connecting these nodes to the original claim to form a subgraph. In the example, each subgraph contains nodes corresponding to region, hashtag, and source (user) that are related to the central claim node. The subgraphs are iteratively merged based on common nodes (Figure 3), and delete duplicate nodes in the final graph structure, shown in Figure 4. The result is a node-edge graph of the original content where all related content is linked so that recommendations can be generated.

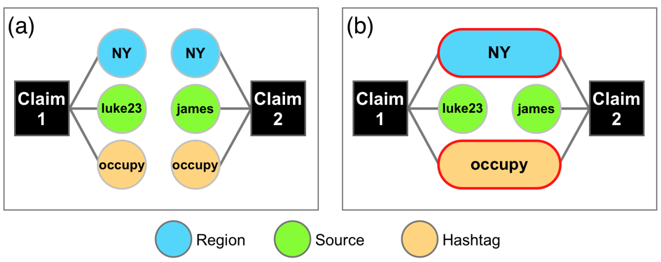

Fig. 3. Connecting subgraphs in Apollo data by finding common nodes.

B. *TopicNets Workflow*

1) Begin with overview of entire corpus (force directed graph layout in Fig 4). In the example, the graph contains a large body of nodes connected to the US region, and a second group of nodes with unknown location. Other nodes form smaller groups around the less represented regions in the dataset.
2) Search for nodes by location (e.g: USA) by clicking on the corresponding node.
3) Interact with highlighted results to better visualize the nodes; in the example the United States node was separated via a dragging motion and the related nodes became a tree with United States as the root.
4) Visualize only the highlighted nodes and related content (topics, users etc) by toggling an option to hide the unrelated nodes.
5) Refine the topic relations by running an LDA algorithm over node contents, and re-running graph layout.
6) Drill down to contents for interesting (central) nodes, the nodes were filtered based on sentiment and credibility to find interesting nodes such as the node selected in Figure 1.
7) Repeat for a search for non-us regions.

C. *Fluo Workflow*

1) Find hashtags (through recommendation) related to initial keyword search. In the example, the analyst starts with #sandy and the system recommends related hashtags until the final seed list is produced as in the left hand column of Figure 1.
2) Rate the relevance of each hashtag in the seed column by dragging the sliders on each node.
3) Weight connections in graph to view only key regions (e.g. USA) as shown in the region box in the second column of Figure 1
4) Examine content recommendations in right column to determine which refinements to be performed.
5) Re-score results by adjusting property weights. For example "Prefer messages with high sentiment and high credibility", the final view of the UI is shown in Figure 1 and interesting messages such as "New Jersey spruce that survived Superstorm Sandy lights up as Rockefeller Center tree" are found.
6) Repeat above for non-US regions.

V. ANALYSIS

Table I shows a breakdown of the key elements that support visual inspection and interactive control in both interfaces. Table II provides a further breakdown of the advantages and disadvantages of each technique, based on a simple cognitive walkthrough with expert users for the use case. Currently, a larger scale automated study is being set up to empirically evaluate both systems with data from large numbers of users. To summarize, TopicNets supports a far more diverse set of features, supporting multiple possible workflows to arrive at an answer to the task question, while Fluo has a more constrained set of functions, but in turn is far more efficient at answering specific questions and consequently is also easier for novice users to understand. TopicNets requires a longer learning-curve to be used to its full extent, making it less suitable for regular website users in an application such as Twitter, and more relevant for trained information analysts. One of the key differences between these systems is the primary modality for visualizing graph data: node-link graph with layout options (TopicNets) versus constrained column layout (Fluo). The graph view has a clear benefit for providing an overview of the entire corpus of data, which is not possible in Fluo's truncated list views. However, as edge complexity increases, the graph view becomes more cluttered and thus less effective for answering questions. Fluo overcomes this problem through ranking and truncating based on inherent data properties such as credibility, sentiment, location and other data scores provided interactively by the user during the analysis task.

As previously discussed, the data filtering toolchain's algorithms append credibility and sentiment scores to each group of messages or "claim". These values are used to guide the analyst towards potentially relevant information in both systems, however the data is utilized in different ways. Fluo presents a slider for credibility data, shown in the center column of Figure 1. This slider affects the ranking of nodes to bias towards messages that have a certain credibility score. For example, by dragging the slider to the maximum value, the analyst is telling the system to boost the ranking of messages with higher credibility scores. This is done through a simple ranking and weighting mechanism over the results in the right column of Figure 1. TopicNets (Figure 4) incorporates credibility data in a different way. Once again, a slider is presented, but this time to control a threshold value over connected entities in the node-link graph. For example, by placing the slider to the maximum value, only nodes with very high credibility will have edges drawn to topics in the document-topic graph. By default, nodes without any edges are not visualized. Both approaches have different benefits and limitations. The filtering approach in TopicNets is better at providing an overview of how the thresholding affects the entire information network and is much more effective when an analyst is searching for a particular node. Fluo's scoring method, however, does not throw out nodes that were highly ranked by other mechanisms in the interface, increasing the quality and diversity of the results. This is especially useful when an analyst is looking for informative feedback on the algorithms that originally generated the credibility values or in the presence of noisy or erroneous metadata.

VI. RELATED WORK

Modeling of trust and credibility information has received research attention from the AI community [6], social and cognitive sciences, e.g. [7] and psychology disciplines, e.g. [8]. Researchers analyzing social web data feeds such as Twitter combine facets from all of these disciplines to develop novel methods for filtering credible information from an abundance of noisy or nonsensical data. For example, [9], [2], [4] all define models that iteratively query over content and underlying network data to make assertions about credibility of a single source or message. Other researchers such as [1] and [3] have focused on scalable techniques, by employing mappings

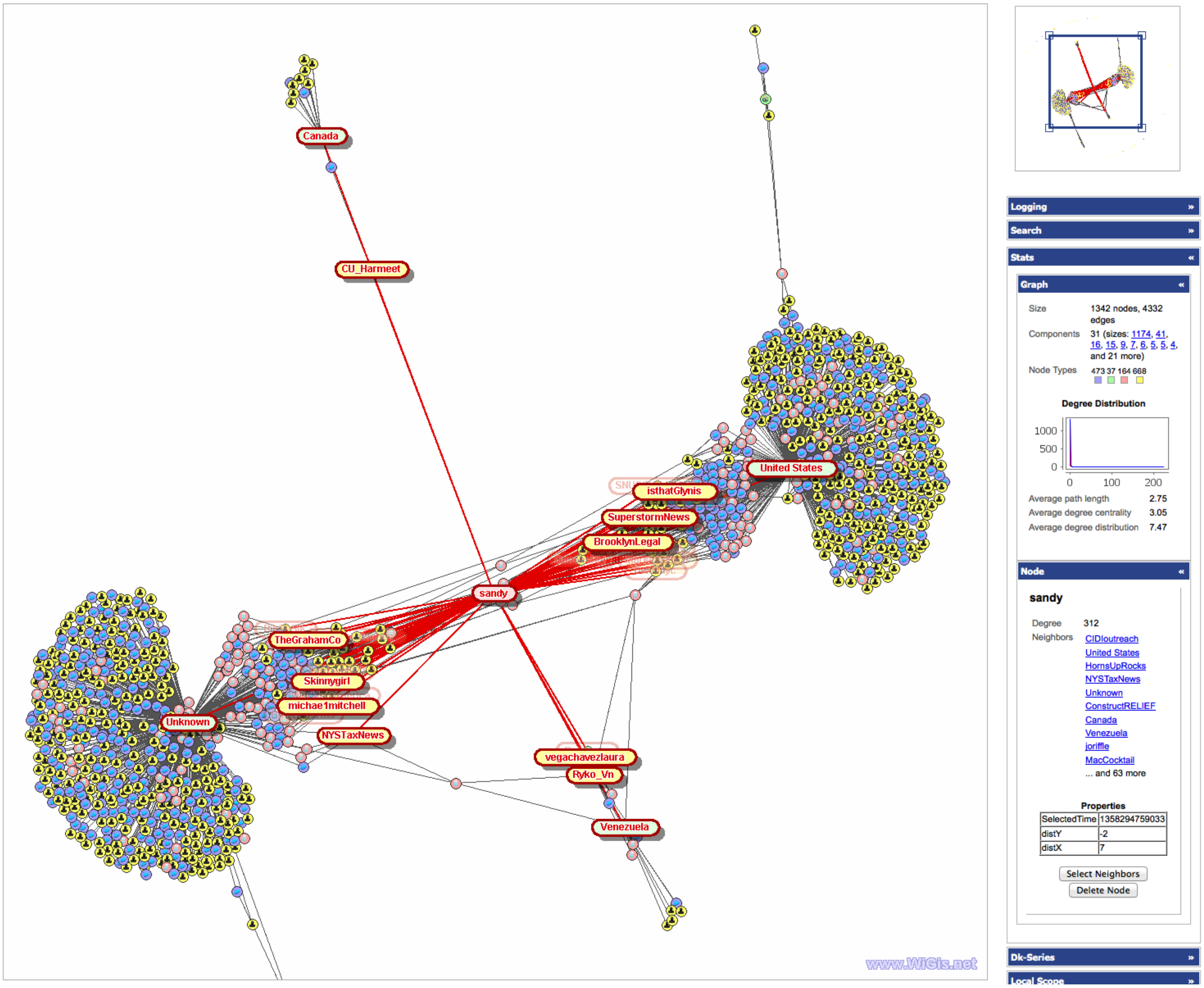

Fig. 4. Screen Shot of the TopicNets Interface showing data filtered through Apollo for the topic "Hurricane Sandy"

such as topic modeling, by reducing query complexity or by leveraging better query facilities (cloud-computing).

At the interactive analysis level, Fluo and TopicNets are related to many published approaches to communicating and explaining credibility information, from early studies such as [10], [11] to more recent experimentally validated approaches in [12] and [13]. Additionally, there are hundreds web tools for analyzing Twitter data. In terms of this research, the most relevant tools include Tweet Archivist [14], Mention Map [15], and TwitterMap [16]. Tweet Archivist is notable for its ability to break down a large group of tweets based on a certain keyword, however, the model must be constantly updated on the server side which makes it unsuitable for streaming, it does not perform any analysis based on sentiment, credibility, or geolocation of the tweets nor does it specialize in recommending topics that are similar to the topic keyword. Mention Map lets you explore Twitter in real-time as a node-edge graph but is user-centric, which can make finding content related to a specific hashtag or keyword very difficult. Finally TwitterMap monitors the current stream of tweets about some hashtag and allows an analyst to visualize them on a world map. Unfortunately, the visualization is tied to the geography and the data model does not allow for sentiment and credibility of tweets. None of the systems that were surveyed were able to answer the complex questions posed in the introduction.

VII. Conclusion

In this workshop paper an exploratory framework for scalable pipelining of different algorithms for filtering network data based on credibility was described. The framework ranges from highly scalable automated algorithms to smaller scale analyst-in-the-loop procedures that require data to be presented through interactively controlled visualizations. As an initial experiment, data was collected through a scalable credibility filter [1] and presented through two different user interfaces for analysis. A cognitive walkthrough of both systems is presented using the same overall task on the same data for both systems, which differ primarily in complexity of the interface and interaction capabilities. The key finding is that both systems

System	*Inspection Elements*	*Control Elements*
TopicNets	Interactive Node-Link Graph with Text and Tabular Elements	Node Dragging, Node Selection, Right Clicking, Control Panels
Fluo	Interactive List View with Text Elements	Node Selection, Slider List

TABLE I
BREAKDOWN OF INSPECTABILITY AND CONTROL ELEMENTS IN BOTH INTERFACES

Mechanism	*Type*	*Advantages*	*Disadvantages*
Node-Link Graph (TopicNets)	Inspection	Good provenance. Easy to inspect paths, neighbor links etc.	Scales badly, gets cluttered quickly (abstraction / clustering can help).
List View (Fluo)	Inspection	Simple, can be reranked with provenance annotations.	Hard to display connectivity.
Interactive Interpolation (TopicNets/Fluo)	Inspection	Can handle lots of information. Creates a "game-like" feel to keep user interested.	Hidden functionality, usually has a learning curve, requires good annotation/help tools.
Tabular View (TopicNets)	Inspection	Easier to understand than a graph.	Hard to display complex connectivity/provenance.
Text-based (TopicNets/Fluo)	Inspection	Simple, lots of detail available.	Does not take full advantage of visual elements, does not scale well.
Node Dragging (TopicNets)	Control	Communicates impact of user input very well.	Not initially intuitive, difficult to re-rank vertically (crossed edges).
Node Selection (TopicNets/Fluo)	Control	Very useful for highlighting subset from a general overview.	Edges cause clutter quickly especially for large graphs.
Slider List View (Fluo)	Control	Clean look, most users familiar with slider input, can be reranked easily with provenance data shown.	Difficult to resize, less freedom.
Right-click (TopicNets)	Control	Useful for node-specific functionality.	Hidden functionality, has small learning curve.
Control Panels (TopicNets)	Control	Easier to understand than a graph, can be labeled more easily.	Can get cluttered quickly depending on the number and complexity of actions.

TABLE II
ADVANTAGES AND DISADVANTAGES OF INSPECTABILITY AND CONTROL ELEMENTS

are capable of recommending useful data by filtering based on credibility. The more feature-rich graph-based system (TopicNets) requires a greater familiarization period with the tradeoff that it can produce additional perspectives on the underlying data, perhaps making it more suitable for trained information analysts than general web users. Fluo does not contain many of the features of TopicNets, but is found to improve on TopicNets ability to efficiently find relevant content primarily by providing a more structured content view. Additionally, since content is discovered through a scoring process rather than filters, it is also capable of producing more diverse results, enabling it to potentially correct for analyst or data error.

ACKNOWLEDGMENT

The authors would like to thank Svetlin Bostandjiev for his innovative research that inspired aspects of the Fluo system. 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; by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory contract number FA8650-10-C-7060; 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, AFRL, IARPA, 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.

REFERENCES

[1] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, "On truth discovery in social sensing: a maximum likelihood estimation approach," in *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, ser. IPSN '12. New York, NY, USA: ACM, 2012, pp. 233–244. [Online]. Available: http://doi.acm.org/10.1145/2185677.2185737

[2] B. Kang, J. O'Donovan, and T. Höllerer, "Modeling topic specific credibility on twitter," in *IUI*, C. Duarte, L. Carriço, J. A. Jorge, S. L. Oviatt, and D. Gonçalves, Eds. ACM, 2012, pp. 179–188.

[3] B. Gretarsson, J. O'Donovan, S. Bostandjiev, T. Höllerer, A. Asuncion, D. Newman, and P. Smyth, "Topicnets: Visual analysis of large text corpora with topic modeling," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 2, pp. 23:1–23:26, Feb. 2012. [Online]. Available: http://doi.acm.org/10.1145/2089094.2089099

[4] J. ODonovan, B. Kang, G. Meyer, and T. Höllerer, "Credibility in context: An analysis of feature distributions in twitter," in *Proceedings at the 2012 ASE/IEEE International Conference on Social Computing*, ser. SocialCom '12. Amsterdam, The Netherlands: ACM, 2012.

[5] B. Kang, J. O'Donovan, and T. Höllerer, "Modeling topic specific credibility on twitter," in *IUI*, 2012, pp. 179–188.

[6] J. O'Donovan and B. Smyth, "Trust in recommender systems," in *Proceedings of the 10th international conference on Intelligent user interfaces*, ser. IUI '05. New York, NY, USA: ACM, 2005, pp. 167–174. [Online]. Available: http://doi.acm.org/10.1145/1040830.1040870

[7] C. Gonzalez, J. F. Lerch, and C. Lebiere, "Instance-based learning in dynamic decision making," *Cognitive Science*, vol. 27, pp. 591–635, 2003.

[8] N. Luhmann, *Trust and power: two works*, ser. UMI Books on Demand. Wiley, 1979, no. pts. 1-2. [Online]. Available: http://books.google.com/books?id=WZu4AAAAIAAJ

[9] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 675–684. [Online]. Available: http://doi.acm.org/10.1145/1963405.1963500

[10] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ser. CSCW '00. New York, NY, USA: ACM, 2000, pp. 241–250. [Online]. Available: http://doi.acm.org/10.1145/358916.358995

[11] N. Tintarev and J. Masthoff, "Effective explanations of recommendations: user-centered design," in *Proceedings of the 2007 ACM conference on Recommender systems*, ser. RecSys '07. New York, NY, USA: ACM, 2007, pp. 153–156. [Online]. Available: http://doi.acm.org/10.1145/1297231.1297259

[12] B. P. Knijnenburg, S. Bostandjiev, J. O'Donovan, and A. Kobsa, "Inspectability and control in social recommenders," in *Proceedings of the sixth ACM conference on Recommender systems*, ser. RecSys '12. New York, NY, USA: ACM, 2012, pp. 43–50. [Online]. Available: http://doi.acm.org/10.1145/2365952.2365966

[13] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "Tasteweights: a visual interactive hybrid recommender system," in *Proceedings of the sixth ACM conference on Recommender systems*, ser. RecSys '12. New York, NY, USA: ACM, 2012, pp. 35–42. [Online]. Available: http://doi.acm.org/10.1145/2365952.2365964

[14] T. Archivist, "Tweet archivist website," Accessed January 7, 2013. [Online]. Available: \url{http://www.tweetarchivist.com/}

[15] M. Mapp, "Mention mapp," Accessed January 7, 2013. [Online]. Available: \url{http://mentionmapp.com/}

[16] TwitterMap, "Twittermap," Accessed January 7, 2013, http://twittermap.appspot.com/.