Truth, Lies, and Data: Credibility Representation in Data Analysis

James Schaffer*, Tarek Abdelzaher†, Debra Jones‡, Tobias Höllerer*, Cleotilde Gonzalez§ and John O'Donovan*
*Department of Computer Science, University of California Santa Barbara, Santa Barbara, California 93106
Email: {james_schaffer, holl, jod}@cs.ucsb.edu
†Department of Computer Science, University of Illinois at Urbana Champaign, Urbana, IL 61801
Email: zaher@cs.uiuc.edu
‡Department of Social and Decision Sciences
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213
Email: coty@cmu.edu
§SA Technologies Inc, 751 Hebron Parkway Suite 310, Lewisville, Texas 75057
Email: debra@satechnologies.com

***Abstract*—The web has evolved in a scale free manner, with available information about different entities developing in different forms, different locations, and at massive scales. This paper addresses the cognitive limitations information analysts typically experience as they approach the boundaries where automated analysis algorithms are sorely needed. We introduce *Fluo*, a hybrid graph/spreadsheet approach for exploring information from disparate sources. An experiment is conducted to explore information analysts' interactions with recommendations from an automated fact-finder algorithm, during the task of answering questions in a humanitarian aid delivery scenario. Our experiment (N=285) is performed using three increasingly complex user interfaces, with and without the presence of the automated recommendations. Results show that in the best performing group, interaction with the fact-finder recommendations was 47 percent greater than the worst performing group.**

I. Introduction

Current web technology provides rapid generation and collection of information from a vast array of networked sensors. The accessibility of social media allows for easy exchange of information, and, as a result, the amount of information available to decision makers has become too large to efficiently and effectively analyze. Culling through the plethora of information available to find the most relevant, reliable, and credible information on which to base a decision can be a daunting, time consuming task. Interactive visual interfaces can be employed to help the decision maker create an understanding of data sets and establish parameters that tailor the information to a set of criteria. However, practical limitations of an analyst's attention and the increasing size of available data reinforce the need for automated assessments and recommendations about the information. Unfortunately, automated processing of data often leaves the user out of the loop and inhibits data understanding due to the difficulty in fully understanding the processes of complex algorithms. Finding optimal combinations of automated and human analyses of information remains a key challenge in recommender systems [14][15][7], human-computer interaction [4][5], and visual data mining [6].

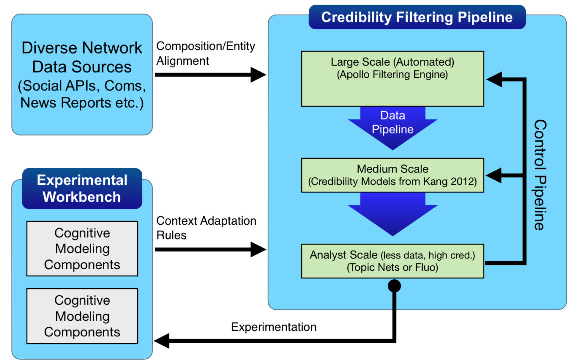

Fig. 2. Overview of a credibility-based recommender system with cognitive modeling component.

A. Research Questions

- *What are the scientific challenges that arise from modeling credibility in networks of different sizes?*
- *What are the cognitive limitations of human analysts that can inform where automated algorithms should take over?*
- *How can we leverage the theoretical and practical boundaries of different types of credibility modeling to improve/optimize a credibility filtering pipeline?*
- *How do we leverage cognitive and human-factor models to discover rules that help analysts to better adapt to specific contexts/missions?*

B. Preview of Study

This research study evaluates several methods for how to visually incorporate results from data mining algorithms into a visual tool and explores the limitations, potential synergies, and other theoretical boundaries between automated credibility analysis algorithms and credibility assessments made by

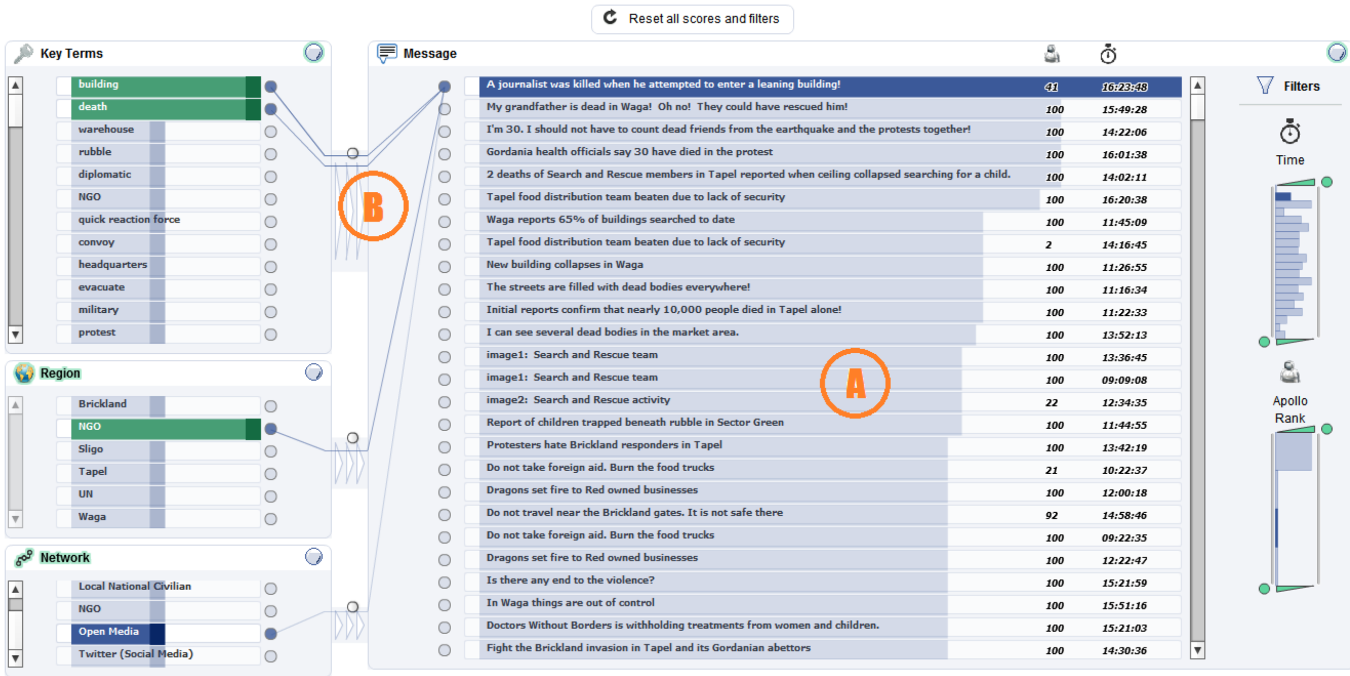

Fig. 1. Fluo is a system for interacting and visualizing relational data, such as those found in most modern databases. Data is hierarchically and semantically organized into different groups called 'tubes' (A) and related objects (e.g. by foreign key) are connected by pipes (B). Queries are performed by specifying relevance to upstream data which flows to downstream data through pipes and tubes. Each tube sorts its contents by this flow of relevance.

human analysts. Amazon Mechanical Turk is used to collect experiment results. Online workers are presented with an interactive visual interface of varying complexity and tasked with answering questions about a hypothetical relief scenario dataset. For half of the participants, the interface incorporated automated results from a highly scalable credibility analysis engine based on the Apollo system [17]. The hypothesis is that there is a 'sweet spot' between complexity of visual and interactive features and the inclusion of recommendations from complicated automated tools.

II. Related Work

The *Fluo* experimental workbench integrates components from multiple research areas. Our discussion of related work covers user interface research, cognitive modeling, recommender systems and relevant research on mining information credibility.

A. Interactive Interfaces for Data Analysis

Effective user interface design can overcome limitations in the user's attention and working memory [3][2]. Additionally, by increasing visual explanation of computational processes, user interface design can facilitate the correct perception of trust and data provenance in scientific information analysis [9]. Developers and interface designers must deal with the challenge of combining and representing large amounts of data at the right time in the right format. Unfortunately, there are finite limits to an analyst's ability to efficiently and effectively summarize large amounts of data, especially when tasks are time-sensitive.

B. Cognitive Models

In Intelligence tasks humans often go through a detective process of "search and relate" to determine who is doing what to and with whom. Intelligence officers often gather information by field observation or consulting public records and confidential information sources. Information helps them make connections and finally answer important questions and make decisions. Cognitive models of a user can help improve computation, filtering, visualization and comprehension of credible information. Understanding of how to filter and visualize network data matched to individuals cognitive states is necessary in order to improve inspectability, control, and situation awareness. Figure 2 shows how cognitive models are leveraged in our experimental data analysis framework. Cognitive models are representations of human behavior that rely on mathematical and computational mechanisms. These mechanisms represent cognitive processes and effects that influence human behavior, such as memory retrieval, forgetting, recognition, judgment and decision making. In military intelligence tasks cognitive models play an important role in explaining, predicting, and supporting the collection of intelligence that can serve a mission.

C. Recommender Systems

There have been numerous studies that have addressed optimizing the synergy between human analysts and expert agents. The visual analytics community has leveraged automated algorithms to intelligently limit the subset of displayed information when viewing multivariate data [16]. Recent

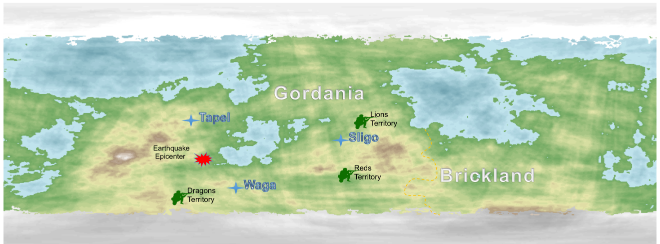

Fig. 3. The map of the Brickland-Gordania region shown to participants to increase their immersion in the scenario.

studies in recommender systems have noted the importance of user-recommender trust in improving satisfactions with recommendations [11][10]. The importance of system transparency and explanation of recommendation algorithms has also been shown to increase the effectiveness of user adoption of recommendations [7].

D. *Presenting Credibility Recommendations*

A large number of studies spanning multiple disciplines have focused on information credibility in networks. However, there has been a lack of focus on the importance of the interface in communicating credibility information to end users. In many cases the problem of data sparsity is unavoidable and there is simply not enough information available to reliably estimate credibility of information in a network. Insights include the following: in online settings, the window of data available to assess credibility of a piece of information is small compared with real world scenarios [11]. In information networks, credibility models can focus on node content, node connectivity, information flow around a node, or some combination of these. Inspectability and Control both independently improve the perceived credibility of information in a network. [7]. Many visualization tools lack a careful consideration of cognitive phenomena. Cognitive Models can be applied to model interaction behavior with the information system [8], and resulting insights can be used to refine a system design.

III. System Description

This section describes the scalable fact-finder algorithm, Apollo [17], and the Fluo user interface, which were used in the experimental setup. Apollo is a fact-finding tool designed to jointly assess both the credibility of information and the reliability of sources. The underlying algorithm performs maximum-likelihood estimation. Given a set of sources and their claims (e.g., statements, tweets, blogs, or other assertions), Apollo iteratively computes the credibility of those claims given their degree of corroboration, and the credibility of sources given credibility of their claims. Apollo also considers non-independence relations between sources to discount rumors that are corroborated only within one social group. Once credibility values are computed, Apollo can rank the information based on credibility.

Fluo (Figure 1) displays multivariate data with different semantic schemas as connected nodes that are organized into sort-able lists (similar to a traditional spreadsheet), called tubes. Tubes can be placed serially (creating an upstream/downstream relationship) or in parallel. Foreign keys from the relational schema are used to connect the nodes representing the underlying data, and relationships are indicated through dark blue edges on-demand. Numeric data from each schema is displayed on each node, and the prototype in this paper provides basic sorting and filtering functionality for these variables. Users can mark data objects of interest by increasing the object's importance score, which is then propagated to connected downstream objects. A user's visual attention is directed to relevant objects by limiting the nodes shown in each tube only to those with high scores. The query mechanism works best when users perform scoring on tubes that contain a limited number of objects, such as one that enumerates the dimensions of a variable in another schema (for example, analysts can work with a dataset of 'people' by scoring nodes in an upstream 'nationality' tube). When many tubes (and their corresponding schemas) are connected in series and scoring is performed, the relationships between data in each schema can be uncovered through interaction. A full presentation and evaluation of the Fluo methodology is beyond the scope of this paper.

Participants were presented with one of three configurations of the Fluo user interface (Figure 4), which varied in the diversity and complexity of interaction methods and data presentation. The simplest user interface presented to users, spreadsheet Fluo, mimics the functionality and visual layout of a basic spreadsheet by organizing the scenario data in a single tube. Spreadsheet Fluo offered only two methods for interaction: sort (by schema column) and scan (via a scroll bar). In this interface, the Apollo credibility ranking of each message was presented in column format and sort-able by clicking the column header. The second interface, limited Fluo, increases visual and interactive complexity by organizing categorical properties (such as region) of each message into upstream tubes. Limited Fluo allows a user to get on-demand details of all messages that match a certain category (e.g. messages that originated in the region of Brickland), where spreadsheet Fluo required sorting and then scanning. Apollo ranking of messages was still displayed as text on each cell and items were sort-able, as in spreadsheet Fluo. Importance scoring of nodes in upstream columns was not allowed. Finally, advanced Fluo expands on limited Fluo primarily by introducing the scoring mechanism and automatically sorting downstream messages. Users were also given the option to filter messages interactively by specifying a numeric range (right side of Figure 1) in addition to the sorting that was available in limited and spreadsheet Fluo. Given sufficient understanding of the query metrics, advanced Fluo potentially allows a user to off-load some of the work involved in complex queries.

Obtaining the correct answer to each analysis question in the experimental task was always possible regardless of the interface configuration, but the minimum level of time and interaction required to perform the same query varied from

Fig. 4. Spreadsheet Fluo (A) and limited Fluo (B). Advanced Fluo is shown in Figure 1. The answer to the question 'What was the occupation of the person who died when entering a leaning NGO building?' is highlighted in each interface.

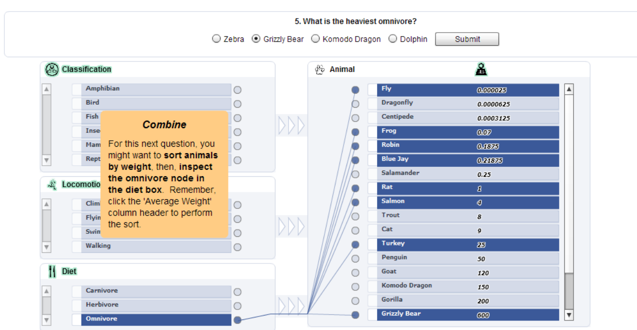

Fig. 5. During the training session, users are required to correctly answer a series of questions designed to verify their understanding of the interface. An easily understood animal dataset is used to help ease users into views and interaction modalities that may be unfamiliar to them. Here, Limited Fluo is shown answering a query related to the diet and weight of the animals present in the dataset.

interface to interface. For instance, comparing messages from different regions in the spreadsheet version of Fluo requires a single click for sorting, followed by scrolling back and forth between the two message groups from each region of interest; in the limited and complex Fluo network views, the relevant region is first selected and then scrolling is used to scroll through messages. Advanced Fluo could also be used to more quickly answer complex queries. For instance, if users want to compare messages between two regions that have high Apollo ranking and match a set of key terms, they can boost the score of the regions and key terms of interest, then adjust the Apollo ranking filter to match the query.

IV. Scenario Description

Before the analysis session, participants were presented with a brief description of a scenario and a map of the fictional Brickland-Gordania region shown in Figure 3. The scenario is presented to users as a collection of 400+ messages broadcast from various fictional regions over different fictional media networks. The hypothetical task was to search through these messages to dispense advice on humanitarian efforts during a crisis situation following an earthquake and involving riots by three insurgent factions. Below is a full description of the scenario, as shown to the participants.

> It is August 2013... You are at your desk in the Operations Center of the main Forward Operating Base (FOB) in the city of Tapel, Gordania, monitoring the flow of data and messages coming in, related to conditions, casualties and relief requirements. The Brickland Army Task Force Commander wants to determine appropriate actions to ensure timely relief operations, mitigate further loss of life due to stalled or poorly prepared search and rescue operations, protect relief assets and personnel including UN and non-government organization (NGO) resources, protect vital remaining infrastructure and key buildings, and reestablish the security of the Gordanian population from looters and militia forces.
>
> You have direct communications with two other FOB locations in the cities of Waga and Sligo, which were likewise affected by the earthquake. These FOBs are monitoring activities and coordinating specific humanitarian relief operations in their respective cities. There are three insurgent cells operating in the region: Reds, Dragons, and Lions. Each one is vying for power with the population, local and national politicians. Each one is seeking to take advantage of the situation to consolidate their political positions and establish local control with their militia forces. The militia forces have access to small arms weapons and limited explosives. The militias are stirring up the local population to protest the presence of Brickland forces and the incompetence of the Gordanian government.
>
> The UN and other NGOs are supporting large food and water distribution operations to prevent mass starvation and cholera outbreaks. UN and NGO medical and Gordanian Red Cross assets are operating in each locale to provide medical support to injured civilians.

Participants were asked to inspect the messages using the interface provided and answer the four Priority Intelligence Requirement questions (PIR), as listed below:

1) *Which insurgent/militia cell is encouraging the most violence against Brickland and the Gordanian government?*
2) *Of the three cities, where are search and rescue efforts most needed?*

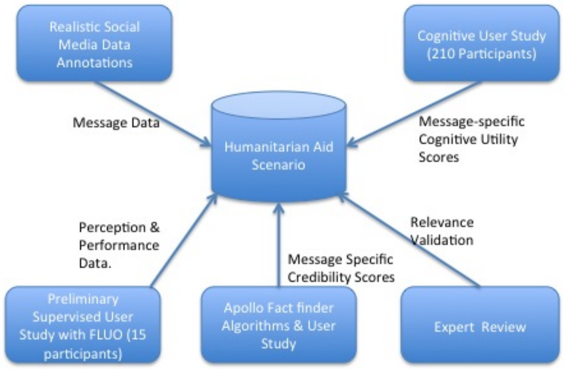

Fig. 6. Elements contributing to the "Brickland" Humanitarian Aid Scenario Data and Metadata.

Experimental Conditions		
Condition Number	*Apollo*	*Fluo*
1	No	Spreadsheet Fluo
2	Yes	Spreadsheet Fluo
3	No	Limited Fluo
4	Yes	Limited Fluo
5	No	Advanced Fluo
6	Yes	Advanced Fluo

3) *Where will protests against Brickland and Gordania most likely occur to disrupt relief operations?*
4) *What degree of risk exists to NGO elements operating in the cities and towns around the cities?*

Figure 6 shows a graph of the different components of the Brickland humanitarian aid delivery scenario. In addition to the core messages, entities and ground truths in the scenario, multiple augmentations were provided to improve the data set. A collection of related social media (mainly Twitter) messages were appended to the scenario to add realism, and to provide more scale to test the Apollo fact-finder tool. Each message was appended with a credibility score from Apollo. This data is revealed to participants in some of our experimental conditions, as detailed in the next section.

V. Experimental Setup

The study described in this paper explored how well parameters that allow the systematic manipulation of data based on criteria related to data credibility and source reliability support the decision maker in efficiently exploring a large data set. Specifically, this study addressed the following hypothesis.

- H1: Increased control over automated credibility filtering mechanisms (i.e., a control pipeline that is regulating the data pipeline) improve human analysts information requirement research process, thereby improving the speed and quality of decision making.

A 3x2 between-subjects design was utilized. The conditions varied the level of functionality available in the user interface, along with presence of credibility information in order to assess which manipulations improve decision speed and quality.

The experimental system was deployed on Amazon Mechanical Turk and data was collected from AMT workers. The AMT web service is attractive for researchers who require large participant pools and inexpensive overhead for their experiments, however, there is valid concern that data collected online may be of low quality and require robust methods for validation. Numerous experiments have been conducted, notably [1] and [13], that have attempted to show the validity of using the service for the collection of data intended for academic and applied research. These studies have generally found that the quality of data collected from AMT is comparable to what would be collected from supervised laboratory experiments, if studies are carefully set up, explained, and controlled.

After accessing the experimental system online, participants were presented with a pre-study questionnaire that collects some basic demographic and expertise information, and required the user to answer three screening questions to test their attention. They were then directed to an interactive training session where the interface was explained while operating on a simple animals (taxonomy) dataset. After the training, participants were prompted to either continue or to re-take the training session. If they were ready to continue, the relief scenario data was loaded into the interface and participants were allowed to explore the data and required to answer some basic comprehension questions before metrics were collected. When all analysis questions were answered and metric collection was complete, the users were directed to a post-study questionnaire where they provided feedback on the user interface and scenario data.

Data was collected from 297 participants, 12 of which were marked as erroneous after closer inspection for a total of 285. Satisficing elimination was done similar to [12] - participants were required to undergo 3 Instructional Manipulation Checks (IMCs) in the pre-study and were called out if they answered the questions incorrectly. Furthermore, an extra level of satisficing detection was added by requiring participants to manually type their answers to the analysis questions which were inspected by hand. Participants were removed either because their answers to the analysis questions were unintelligible or due to timing glitches in our experimental system. Of the 285 use-able data points, participant age ranged from 18 to 67 with an average of 31.42 and a median of 29. 59% (168) of participants were male while 39% (117). were female.

Participants could not be observed as they undertook the study, so the system itself logged detailed results for nearly every possible interaction, including time taken, node clicks, filter manipulations, sorts, and score manipulations.

VI. Results and Discussion

Table I shows a list of the primary metrics recorded in the experiment. These can be classified into accuracy, Interaction and perception based metrics. To test the hypothesis outlined in the introduction, the following key questions were asked about the experimental data:

- How did the user interface condition impact on accuracy of PIR questions? What about analyst score (response

TABLE I
AN OVERVIEW OF THE CRITICAL METRICS RECORDED FOR EACH PARTICIPANT.

Metrics		
Type of Metric	*Specific Metric*	*Measurement Method*
PIR Response	Speed to answer PIR	Time from beginning of scenario until PIR answered.
PIR Response	Accuracy of PIR answer	Compare participant response to ground truth.
Semantic Entity Usage	Quantity/quality of terms used	Specific terms and number of times revisited
Semantic Entity Usage	Nature of term usage	Order of usage, frequency of search, frequency of modification.
Interface Interaction	Utilization of each control	Frequency of manipulation.
Interace Interaction	Patterns of Interaction	Ordering of control usage.
Usability	Usefulness/value of features	Survey administered at the end of the scenario.

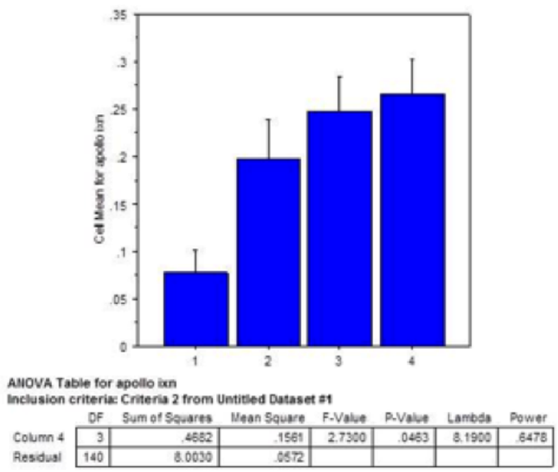

Fig. 7. Number of correct PIRs v/s Apollo Interaction Level. Mean interaction with Apollo was 47.1% higher for correct answer group than incorrect group (p=.0463, ANOVA)

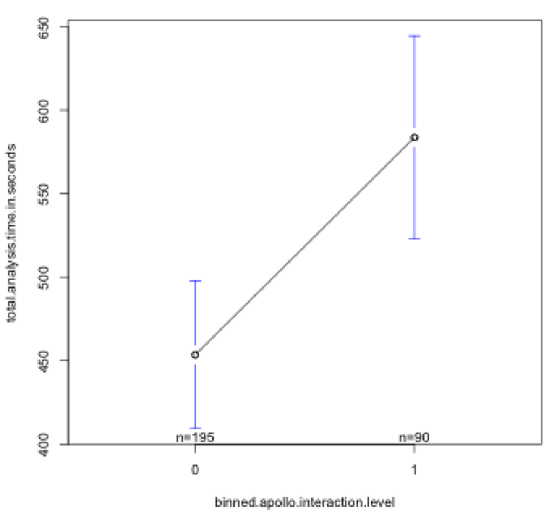

Fig. 8. Response time v/s Apollo Interaction Level (bins: high N=195, low N=90)

time and accuracy?

- How did presence of Apollo impact on accuracy or speed of PIR questions?
- Did participants who interacted with Apollo (when it was available) do better (correctness) overall than people who did not?
- What is the performance difference between participants who showed a high level of interaction overall and participants who showed a high level of interaction with Apollo? This was to validate that observed performance improvements were a result of Apollo interaction, rather than an increase in general interaction with the system.
- For the three previous questions, did the effect size change based on the user-interface (i.e. did participants who used limited Fluo benefit more from the recommendations)?

Before a discussion of results begins, we note that scenario-based experimentation of this type has high variance of participant responses. The experimental subjects were 285 AMT users, all with different propensity and abilities for scenario-based data analysis. While a number of significant effects in the collected data were detected, there were cases where it was not possible to show significant effects, despite a careful power analysis in the initial design phase. For example, no effect was revealed for user interface level on speed or accuracy in this experiment. However, from interviews with participants in pilot studies, the authors believe that user interface level does have an effect on the user experience for certain populations and can impact on time and accuracy of analysis. The authors also believe it is likely that training time with the analysis tool plays a key role and can greatly improve analyst performance. On low-pay crowd sourcing platforms such as AMT, time and participant attention are limited. To address this, a followup study is currently in progress to evaluate the Fluo framework using expert data analysts with experience in aid delivery scenarios.

To answer questions 2 and 3, performance and time were assessed across the conditions with and without the Apollo credibility data. Interestingly, there was no significant difference in correctness for the Apollo present group compared with those who did not have Apollo. However, the best performers (those who answered 3 or 4 of the PIR questions correctly), showed a relative increase of 47 % more interaction with Apollo compared to those who correctly answered 0-2

PIR questions. This is a strong result showing that credibility information is important in the process of data analysis, and also that *analyst interaction* (e.g.: sorting and filtering of messages) based on automated credibility information leads to better performance. Figure 7 shows the results of the experiment along with an ANOVA table (p=0.463).

Results show that Apollo credibility data can lead to increased accuracy, but this does come at a cost. Figure 8 shows a comparison of the mean response times for participants with low and high Apollo interaction scores. Response time was recorded as the number of seconds required to answer all four of the PIR questions (note that the y-axis begins at 400 seconds). Participants in the high-use bin took 27% longer to answer the PIR questions than those in the low-use bin. When considering this result, the limited training time available to participants in this experiment should be kept in mind. In a real-world scenario, where an analyst is well versed in the system functionality, it may not be the case that use of credibility data leads to longer analysis times. This question will be addressed in a follow-up study with expert information analysts.

VII. Future Work

This research has many potential avenues to explore. Specifically, an additional experiment will be devised to better understand the role of user interface complexity on data comprehension. More pronounced paradigms between conditions (spreadsheet, graph views, and hybrid spreadsheet/graph) in terms of data model, visualization, and interaction may show more significant results. The authors believe that analysis skills are a critical factor in the experiment, so a followup study is planned using expert data analysts, sourced through US Army research laboratory. Additionally, a deeper cognitive evaluation of analyst interactions will be performed using Instance Based Learning cognitive models [8].

VIII. Conclusion

A crowd sourced experiment (N=300) was performed and evaluated to assess the cognitive limitations of human analysts in a variety of conditions, particularly with and without presence of credibility information from a large scale fact-finder tool (Apollo). Additionally, this paper has introduced Fluo, a novel data analysis tool to support complex queries over message streams and metadata from disparate sources. Our results showed that the best performers interacted with the interface components related to credibility recommendations, showing a 47% improvement over other participants (p=0.463). However, the group with credibility information exposed did not show a significant accuracy improvement over those who did not. This indicates that viewing credibility information is not enough to impact accuracy, but that the analyst needs to *interact* (e.g.: sort, filter) with the credibility information in order to improve the effectiveness of analysis.

Acknowledgment

The authors would like to thank Laura Marusich at ARL for her valuable advice, and John Hyatt at SA Technologies for his role in constructing the Brickland dataset.

References

[1] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

[2] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[3] Mica R Endsley. *Designing for situation awareness: An approach to user-centered design*. Taylor & Francis US, 2003.

[4] Mica R Endsley and Esin O Kiris. The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(2):381–394, 1995.

[5] Christopher D Hundhausen, Sarah A Douglas, and John T Stasko. A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing*, 13(3):259–290, 2002.

[6] Daniel A Keim. Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):1–8, 2002.

[7] Bart P Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. Inspectability and control in social recommenders. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 43–50. ACM, 2012.

[8] Tomás Lejarraga, Varun Dutt, and Cleotilde Gonzalez. Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*, 25(2):143–153, 2012.

[9] Bertram Ludäscher, Ilkay Altintas, Chad Berkley, Dan Higgins, Efrat Jaeger, Matthew Jones, Edward A Lee, Jing Tao, and Yang Zhao. Scientific workflow management and the kepler system. *Concurrency and Computation: Practice and Experience*, 18(10):1039–1065, 2006.

[10] Paolo Massa and Paolo Avesani. Trust-aware recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 17–24. ACM, 2007.

[11] John O'Donovan and Barry Smyth. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 167–174. ACM, 2005.

[12] Daniel M Oppenheimer, Tom Meyvis, and Nicolas Davidenko. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4):867–872, 2009.

[13] Gabriele Paolacci, Jesse Chandler, and Panagiotis Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419, 2010.

[14] Rashmi Sinha and Kirsten Swearingen. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 830–831. ACM, 2002.

[15] Kirsten Swearingen and Rashmi Sinha. Interaction design for recommender systems. In *Designing Interactive Systems*, volume 6, pages 312–334. Citeseer, 2002.

[16] Andrada Tatu, Georgia Albuquerque, Martin Eisemann, Jörn Schneidewind, Holger Theisel, M Magnork, and Daniel Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 59–66. IEEE, 2009.

[17] Dong Wang, Lance Kaplan, Hieu Le, and Tarek 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*, pages 233–244. ACM, 2012.