Supporting Collaborative Information Analysis: A Classroom Study

1st Author Name
Affiliation
City, Country
e-mail address

2nd Author Name
Affiliation
City, Country
e-mail address

3rd Author Name
Affiliation
City, Country
e-mail address

ABSTRACT

Collaborative information analysis involves strategic analysis and representation of complex information space through synchronous and asynchronous team interactions over extended time periods. Chances are rare to observe how technology mediates team analysis because such complex scenarios are difficult to model in lab studies, and analysts are often limited to access in real world. Classroom study provides a testbed in which students are trained to become professional analysts and course projects are designed to simulate real world tasks. We deployed our tool, a collaborative visual analytic tool informed by earlier work, in class and observed the usage of the tool by teams in their project. Particularly, this paper reports how participants used the tool in evidence collection, evidence schematization, hypothesis generation, and overall coordination, four critical parts in the task of collaborative information analysis. The paper ends with discussion over several theoretical and design implications.

Author Keywords

Collaborative information analysis; visualization; classroom study;

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation (e.g. HCI): Computer-supported cooperative work

INTRODUCTION

Collaborative information analysis is a form of sensemaking wherein a team analyzes a complex information space of facts and relationships to identify and evaluate causal hypotheses. A common example of collaborative information analysis is crime investigation; a variety of putative facts are assembled, including financial records, witness observations and interviews, and social connections of various sorts among persons of interest, from which investigators collaboratively assess means, motives, and opportunities, articulate and investigate further hypotheses and deductions, and develop one or more theories of the crime. Other examples include intelligence

analysis, business intelligence, scientific research, and social constructivist learning.

A critical challenge for information analysts is building an adequate preliminary data model from textual documents, and insuring that the data model is employed effectively in hypothesis development and evaluation. This is an open challenge [1]. Standard methods often do not support it at all; for example, Analysis of Competing Hypotheses (ACH) assumes that data has been modeled, and that relevant evidence can be adduced appropriately to various hypotheses, but provides no structured support for either. Other techniques, such as Information Extraction and Weighting (IEW), help structure modeling of evidence, but do not extend utilization of evidence to hypothesis generation. We therefore are motivated to develop an integrated workspace in which analysts can model and analyze data in one place, and investigate how that will affect analytic process.

Any work of information analysis at a non-trivial scale is fundamentally collaborative. A key enabler for effective collaboration is *activity awareness*, defined as team's awareness of its own sustained collaborative activity [5]. Derived from Activity Theory, activity awareness transcends synchronous awareness of who collaborators are, where a collaborator is looking, etc. It encompasses issues of many different kinds of information covering all aspects of an activity, such as events, tasks, goals, mediating artifacts, social interactions, and group values and norms, which becomes higher demanding as the activity becomes more complicated. Awareness support in such a complex activity of information analysis is perhaps also more challenging than many other situations (e.g. collaborative writing) or casual social settings (e.g. [10]). Teammates could be working with much more complex data structure, mediating through multiple analytic artifacts, and making sense of different levels of analysis, assumptions, and hypotheses, both synchronously and asynchronously throughout a long-term course of collaborative interaction. Hence we will investigate how technology can mediate team collaboration in a complex analytic task over extended period of time.

We situate our study in classroom learning of information analysis. Classroom study provides a natural environment in which participants engage in long term, complex class projects. The students are in fact learning to be information analysts, and are graded in those courses on their ability to understand and enact professional practices of information analysis. This strong normative emphasis on problem solv-

ing practices is a great evaluation context for new interactive tools: Tools are only valuable to the students insofar as they actually support better practices and better outcomes.

The study also provides us an opportunity to examine learning experience with collaborative tools in education of information analytics. Learning information analytics is challenging because it requires students to not only get familiar with analytic techniques but also to be able to apply them in an effective manner. Research has called for enhancement of information analysis training and education in the context of globalization characterized by an increasing degree of complexity and unprecedented acceleration of change, especially the need for incorporating innovative technology to transform exploding data into meaningful and actionable information [15].

We thus are motivated to investigate the feasibility, effectiveness and consequence of supporting integrated data modeling and analysis, as well as supporting activity awareness in complex information analysis, in the context of classroom study. We have developed a tool that includes annotation for data modeling, interactive visualization for data analysis, and a set of awareness features. For the balance of this paper, we describe the tool we have developed, the classroom settings, and our observations. We conclude with design implications derived from the study as well as future work

RELATED WORK

Many researches have been conducted to investigate design features and tools to support collaborative information analysis. For example, Goyal and Fussell [8] studied the effect of hypotheses sharing on sensemaking. Mahyar and Tory [**Mahyar2013**] designed a visualization to connect collaborators' common findings and evaluated its support for team performance. Hajizadeh et al. [11] explored how sharing teammate's interactions affects awareness. These studies report interesting result of controlled lab studies to validate hypotheses of specific design features. However, they do not provide insight on how teams would collaborate in the real world over extended period of time.

Several field studies were conducted aimed to understand design requirements of collaborative information analysis in more realistic setting. Chin et al. [6] observed and analyzed the analytic strategies, work practices, tools ad collaboration norms of professional intelligence analysts. Kang and Stasko [14] studied how student analysts, as in our study, completed in-class intelligence projects. Carroll et al. [4] attempted to model a complex analytic task scenario in a lab setting, and examined the development of team awareness in a four-hour-long task. These studies helped improve understanding of current work practice with commercial tools or no supporting tools at all. However, it remains unknown how teams would behave with technology that is explicitly designed for collaborative information analysis.

Before the study with our tool, we observed how teams of students accomplished a course project in the class. Students practiced applying IEW to extract and model data from documents. They then replicated key facts into analytic artifacts such as an ACH Matrix in PARC ACH, a timeline and a network graph in Analyst's Notebook. However, these tools lack serious collaboration support. Analysts were unable to contribute simultaneously. The analysis work was often divided by tools: each individual created and analyzed an artifact with a tool on their own. This had the consequence that findings and hypotheses be made without integrating collective efforts and diverse knowledge. Analysts must coordinate work by manually sharing notebooks or graphs, resulting in a scattered placement of results, requiring repeated manual resynchronizing to identify redundant or missing pieces of information, analysis of information, and analytic hypotheses. The instructor and students in our study were aware of the shortcomings of available tools with respect to support of collaboration.

CANALYTICS FEATURES

We developed a collaborative information analysis tool, CAnalytics, to support teams of analysts in identifying, visualizing, integrating and assessing facts from multiple sources. The design is informed by our paper prototype studies, where we examined communication patterns and team's spontaneously created artifacts when teams were engaged in a complex crime scenario. We also take into account findings from empirical studies conducted by Chin et al. [6] and Kang and Stasko [14] when making design decisions.

CAnalytics supports evidence modeling through annotation. In the document view users can select and highlight snippets of information and annotate them as a type of entity such as a person, location, events, etc., or as a relationship between entities. Unlike in other entity-based systems such as [18, 3], in which natural language processing (NLP) techniques are employed to extract and model entities automatically, we use annotation to allow analysts to manually create evidence objects of interest. While algorithms for named entity recognition have improved significantly, developers of the entity-based systems (e.g. [7]) admitted that the accuracy of automatic entity extraction is not sufficient to support human analysis. Besides, relationships identified by algorithms are limited to co-occurrence of entities in the same document, whereas identifying meaningful relationships *implicit* in documents that are of more importance to analysts is not possible. And perhaps the most problematic is the fact that the NLP treats all pieces of information equally with no regard to the problem analysts have at hand. We posit that information analysis is a progressive process, wherein analysts place varying priority on evidence depending on how they frame the problem [13]. Manual annotation allows for greater user control, allows more integrated source data objects to be identified, and avoids the user problems associated with automatic identification of disaggregated people, locations and times [2].Users can decide their own information of interest and granularity that best suits their ad-hoc analytic needs.

Users can add attributes to the annotated object, e.g. adding time attribute in an event, and placing coordinate in a location. Users can also make reference to other objects in the attribute; for example, users can add people objects to an event indicating that these people were involved in the event. In this case, a relationship between the people and the event is
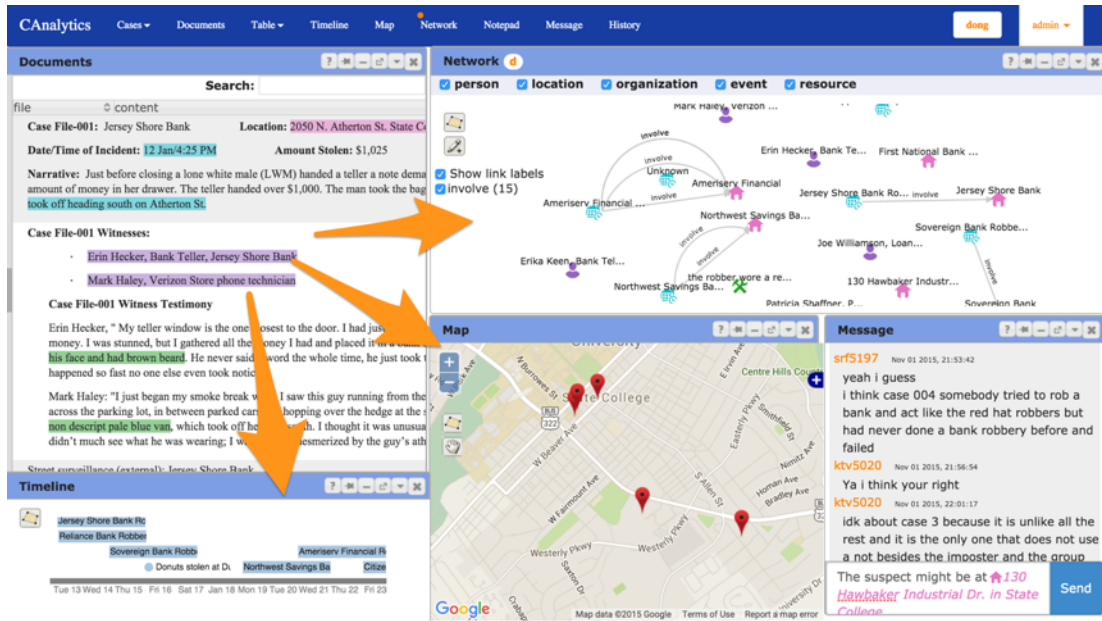
Figure 1. CAnalytics user interface

automatically created. Users can also explicitly create a relationship between two entities by selecting a source entity and a target entity and labeling the relation name.

When users are explicitly creating an annotation, they are also implicitly creating a provenance of the modeled entity— annotation records the source where a data object was created. As observed in Carroll et al.'s [4] empirical study, while integrating information in a visual artifact helps sharing and pooling information with teammates, the action also removes problem information from its original context. Participants often forget what an entity refers to and why it is such positioned in an artifact. Entities in CAnalytics are linked back to the location of documents where they were created through annotation. Users can always re-access the data objects for provenance, a critical requirement emphasized by Chin et al. [6].

The data modeled through annotation is then displayed in multiple coordinated views in the same workspace, including table, timeline, map, and network graph—artifacts frequently constructed to hold attribute data, temporal data, spatial data and relational data respectively [4]. Figure 2 shows an example of the tool interface: when an annotation is created in the document view with information about time, location, participants, and their relationships, a new event is created in the timeline view, a new location is created in the map view, and new people are added to the network graph with a labeled edge representing the relationship (or new edges are added to existing nodes). Hovering mouse over an entity will activate an entity detail window that displays attributes in detail, and analysts can modify, or re-model the entity in situ.

The views are coordinated and afford brushing and linking interaction; that is, when users apply a graphic filter in one view, related information is displayed in other views. Thus

the analyst can retrieve entities within a time range using timeline filter, make a spatial query with map filter, or select a cluster of entities by drawing a filter area in the network view.

CAnalytics supports real-time collaborative editing, similar to the Google Tools. Users can open several concurrent editors to collaboratively edit multiple annotations. Annotations are immediately shared within a team and are automatically added to teammates' document views and other visualizations. As far as we know, this tool is the first to support multi-concurrent editing, which we think can improve fine grain coordination and awareness in complex collaborative work.

In addition to real-time data sharing, CAnalytics is built in with several other awareness features, including a notification system, a feature we named "tool coordinator"', a message tool, and a history tool, and a collaborative editor. A notification system sends individual's actions to the team, in the form of a text box in the top right corner of the workspace. The tool coordinator is an iconic indicator on top of a tool window, suggesting who is working on the tool. The message tool is a real time chat window that enables team communication with persistent message history. The system also maintains a traceable log of time-stamped individual activities in a history tool. Users can learn team activity about who did what to which object at when. Entities and relationships mentioned in the message tool and history tool are hyperlinks that will trigger pop-up detail window when being moused over. With these awareness features, users who work synchronously can be informed of others' activity continuingly; users who work asynchronously will be able to use the history to reconstruct their work status and become aware of changes beyond the point of their last interaction.

We also included a simple notepad to support collaborative hypothesis development. We integrated Etherpad [1], an open-source collaborative editor similar to Google Doc, for teams to compose their hypotheses. Users can insert tables (e.g. an ACH matrix) and images (e.g. screenshots of the tool views).

CLASSROOM STUDY SETTINGS

The context for this study was an undergraduate course in an intelligence training program in a US university. The program was designed to train students to become professional intelligence analysts. A key requirement of the course is to emphasize hands-on practice on team-based intelligence analysis. During the first nine weeks, students learned strategic knowledge (e.g. bottom-up analysis and top-down analysis) and structured analytic techniques, such as IEW, ACH and network analysis, and practiced to apply these techniques to solving two small projects with state-of-the-art tools including PARC ACH and IBM Analyst's Notebook.

Our study began from the 10th week of the course and lasted for one week. The task was to investigate a series of bank robberies fabricated by the course instructor. Teams were provided with a set of documents pertaining to seven robberies, including police reports, witnesses reports, video records, and news media. The task was designed open-ended, which meant that there was no single answer to the task. The instructor explained that the task was to simulate real world scenarios, in which analysts always reasoned in the circumstances of uncertainty, ambiguity, and complexity. The instructor told the students that 6 hours was expected to complete the project, including in-class and outside-class work. In the end of the project students were required to submit a team report, describing their hypotheses, assumptions, conclusions and supporting evidence.

Students were given a tutorial on CAnalytics a week before the project began. One of the authors walked through features of CAnalytics and then let students accomplish a small case analysis on their own pace. During the study week, one author was always available to help with any technical issues. Although students were encouraged to make full use of CAnalytics, to ensure a naturalist environment students were always free to employ any other tools that they believed useful.

Of the 98 students enrolled in the course (from two sections), 73 consented to participate in the study. Students were randomly assigned into 25 teams (23 three-person teams and 2 two-person teams). Research suggested that group size be an important factor in group collaboration, thus two-person teams might behave differently from three-person teams. We thus excluded data from the two-person teams in our analysis in this paper. Also, from the log (and confirmed by their questionnaire), we found that one team made little use of CAnalytics and opted for other tools (Google Doc). Hence their data was also excluded. Thus in this paper we reported the result from 22 teams.

All the students held major in the program of Security and Risk Analysis. Most (75%) of them were in the third academic year (3.05 years in average), indicating that partic-

ipants in our study had relatively advanced experience and knowledge in intelligence analysis. Participants' age ranged from 19 to 28 (20.3 in average). 77% of the participants were male.

We employed several data collection approaches. We administrated a post-study questionnaire, which included several 7-index questions measuring individual's self-reported awareness (adapted from [**Convertino2011**]), team communication (adapted from [**Convertino2011**]), collective efficacy (adapted from [**Convertino2011**]), perceived performance (adapted from [9]), and cognitive load [12]. The questionnaire also included open-ended questions asking how the tool helped or impeded their work. We captured user interactions with system logs. Instead of simply logging low-level events like mouse click and keyboard strokes, we recorded actions such as making an annotation and deleting an entity. Finally, we reviewed team reports and graded them as an indicator of team performance. We designed an assessment rubric together with the course instructor.

RESULT

We first report descriptive data about usage and use experience of CAnalytics. We then report details on the aspects of team behavior, awareness, and learning experience, through methods of artifact analysis, log analysis, and qualitative analysis of questionnaire result.

Usage overview

Over the week, teams created 1805 entities and 1529 relationships in total. The number of entities teams created ranged from 24 to 223 (M=82, SD=39.9), and the number of relationships ranged from 7 to 237 (M=69.5, SD=51.0). The large variety of modeled data was related to team strategy, which will be detailed later.

Overview of the survey items indicates that students rated positive on CAnalytics overall, as shown in Figure [**fig:survey**]. CAnalytics were ranked favorably in all aspects except cognitive load. They had a close to neutral feeling towards cognitive load, which suggests that the task is still cognitive heavy.

Behavior of data modeling and data analysis

We examined the pattern of data modeling and data analysis by looking at a visualization of the entire interaction log (e.g. Figure [**fig:sequence**] shows the interaction log of one team). All teams started with data modeling as they got themselves familiar with the documents. However, teams did not wait to start analysis till they finished data modeling; instead, the activity of data modeling and data analysis were intertwined. Participants switched from one activity to the other activity frequently. This is also demonstrated in the state transition diagram [**fig:transition**], in which we encode the number of switches into width of the link.

To look at the consequence of intertwined data modeling and data analysis, we examined the analytic product teams created, the network graph in particular because social relationships played the most critical role in this specific case and teams spent most time on network analysis (as reflected from
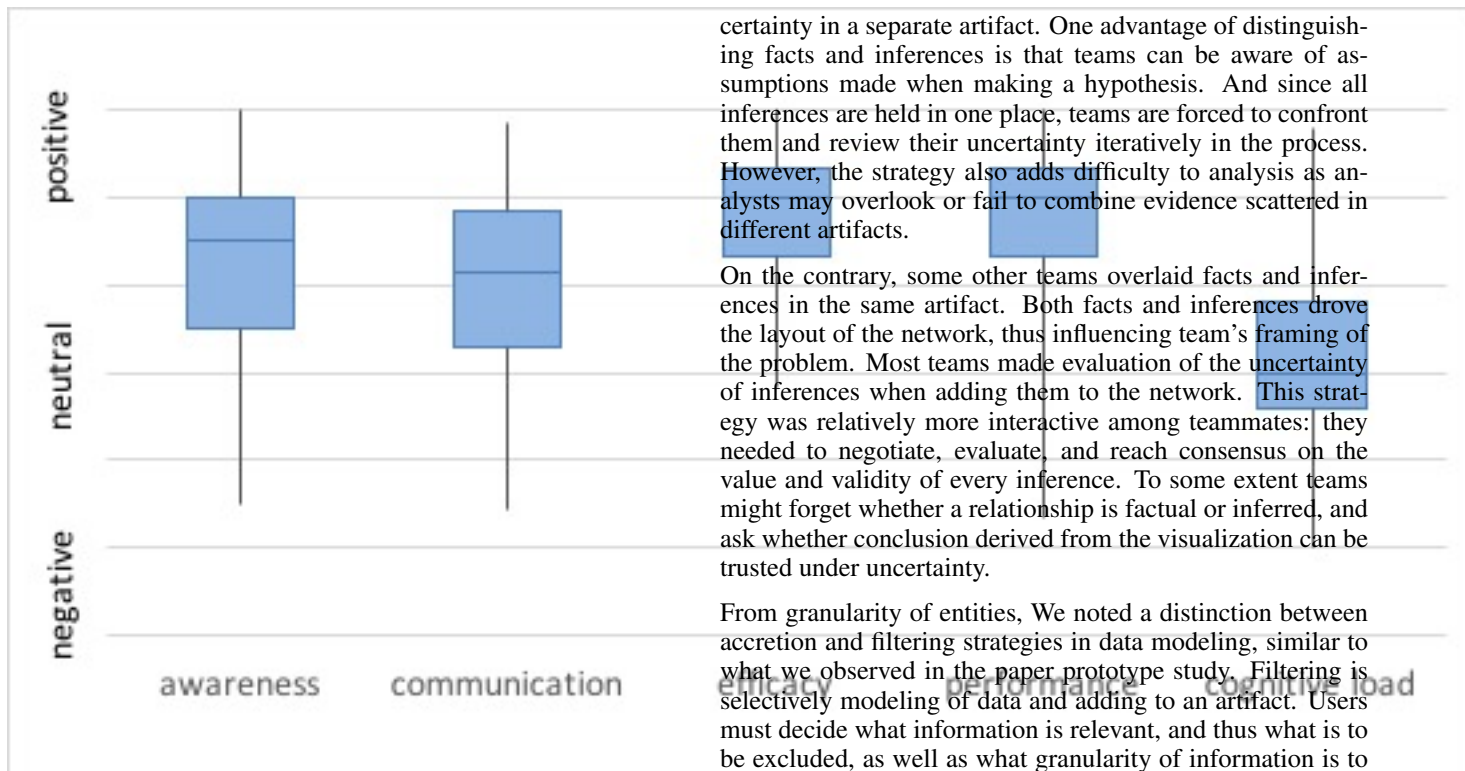

Figure 2. Survey responses (box shows Q1-Q3 and median; whiskers show maximum and minimum)

certainty in a separate artifact. One advantage of distinguishing facts and inferences is that teams can be aware of assumptions made when making a hypothesis. And since all inferences are held in one place, teams are forced to confront them and review their uncertainty iteratively in the process. However, the strategy also adds difficulty to analysis as analysts may overlook or fail to combine evidence scattered in different artifacts.

On the contrary, some other teams overlaid facts and inferences in the same artifact. Both facts and inferences drove the layout of the network, thus influencing team's framing of the problem. Most teams made evaluation of the uncertainty of inferences when adding them to the network. This strategy was relatively more interactive among teammates: they needed to negotiate, evaluate, and reach consensus on the value and validity of every inference. To some extent teams might forget whether a relationship is factual or inferred, and ask whether conclusion derived from the visualization can be trusted under uncertainty.

From granularity of entities, We noted a distinction between accretion and filtering strategies in data modeling, similar to what we observed in the paper prototype study. Filtering is selectively modeling of data and adding to an artifact. Users must decide what information is relevant, and thus what is to be excluded, as well as what granularity of information is to model. Filtering requires more team coordination, because teammates must reach a common ground of the current problem as well as information needed to answer the problem. Figure [**fig:network**]a is an example of filtering, highlighting only the key information of each robbery and how robberies are connected.

Accretion is an attempt to comprehensively represent the problem by adding all information to an artifact. Users extract every fact from the document, regardless of its immediate relevance to the problem. Accretion costs less coordination as it is relatively mechanical note taking. A disadvantage of accretion is that it could be time consuming to model all details and the produced artifact could be fairly complex. An example is Team 108, who modeled every step the suspects took, which resulted in many more entities than the average and much more cluttered network view (Figure [**fig:network**]d). Users reported that they spent too much time in details that they lost the bigger picture:

> We would find ourselves glued to our computer screens, and spent too much time on intelligence gathering rather than analysis (User 135)

Different from the result in paper prototype study, however, participants seemed to be more tempted to accretively add information with CAnalytics. Students reflected that many annotations did not help them solve the problem at all because those entities were unrelated to their problem.

> I felt that after we were done annotating, we hadn't really accomplished anything and that we were no closer to solving the case than when we had started. In the end it didn't really help that we had annotated the data,

the log). We found the network views fell into one of two categories: the networks consisted of 1) separate clusters, or 2) connected clusters. For example, the network in Figure [**fig:network**]b consists of separate clusters. Networks from 8 teams (36%, mean performance=7.8) fall into this category. Nodes within a cluster are connected, representing information space of a robbery case. Nodes between clusters are nonetheless not connected. However, these teams did not just claim that no relationship existed between robberies; instead, we found that they documented possible relationships between robberies in the notepad tool.

In contrast, 6 other teams (27%, mean performance=8.3) created networks composed of connected clusters. While a cluster is still a representation of a robbery, some of them are connected through an evidence node. An example is Figure [**fig:network**]c, in which we mark four *connectors* that link the clusters. These connectors were key evidence that led the teams to hypothesize that those robberies were related and might be committed by the same criminal group.

While many causes might account for the different network views, we attempt to interpret the difference from a perspective of uncertainty. For instance, links within a cluster are factual relationships literally modeled from raw documents (e.g. a white van was witnessed at a location), but links between clusters are often inferences beyond literally documented (e.g. a white van at location A is the same van witnessed at location B). Teams creating separate clusters only represented facts in the network and held evidence with un-

Why did this happen? We guess both the context of classroom study and the system design contributed. Unlike in the lab study where teams were temporarily assembled, teams in a class evaluated peers either consciously or unconsciously. Such social pressure motivated individuals to make contributions, and to make *visible* contributions more than valuable contributions. The awareness features in our system unfortunately made some contributions more *visible* than others. For example, creating an annotation would be immediately broadcast to the team, whereas writing a hypotheses on a notepad produced no notification, although the text was also shared. The selective awareness seemed to also exert bias to recognition of contribution.

Finally, a third category of network consists of few links. These networks seemed incomplete. These teams appeared not to devote much time on the project. However, it was worth noting that their average performance was significantly higher than the other types of teams. We guess these teams employed a different strategy in making use of the tool. Instead of comprehensively representing the problem first, the teams started early off to generate hypotheses. This can be demonstrated from their log: these teams started writing report on the notepad in their first class whereas most other teams waited till second class or later. Examining their network view, the nodes were critical entities in their report. The teams seemed to use the tool as external memory of important information, rather than tool for analysis.

Collaboration and awareness

Participants rated positively of the collaborative support. One participant thought of CAnalytics as a combination of Analyst's Notebook and Google Doc. made an analogy to Google Doc. Similar to being able to collaborate on text editing in Google Doc, they were able to collaboratively work with more complex data structure. Participants could contribute simultaneously without getting blocked.

> CAnalytics seemed like the beta-version of an analyst's notebook that multiple people could work on at once... It was like having an analyst's version of a google doc... [Previously] all of us could not edit a single document at the same time. With CAnalytics, we were all able to work on everything at once with ease

Participants reported that they could easily integrate team efforts.

> they [teammates] were able to work directly off something that I had also created. **This ability to work off of each other's own work allowed us to all contribute to each element of the analytic process**. An example of this would be when I was creating an association chart for one of the bank robberies. I was able to cover my bases and include in the chart all the information that I thought was relevant. At a later time, one of my group members received the document of the same bank robbery and found information he thought should be included in the association chart that I had not included. **He was able to work off of my initial diagram and**

add and manipulate it in order to include additional relevant information I had missed. (U31)

The awareness features were received well. When asked what features helped them stay aware of team activities, 28 participants mentioned the tool coordinator, 24 mentioned the notification system, 19 mentioned the history tool, 14 mentioned the real-time update of user-generated data 12 mentioned the collaborative editor, and 7 mentioned the message tool. While the number of mentions does not simply indicate tool usefulness, it suggests users appropriate these features and are explicitly aware of their support.

We categorized students' feedback based on the element of awareness, or the essential problem of awareness of *what* [16], into social awareness, information awareness, action awareness, history awareness, and intention awareness, as shown in Table XX.

By analyzing the interaction, we also found indicators of high awareness. For example, we measured the number of entities accessed by collaborators versus by the author only, and the time lapse when the entity was first accessed by collaborator since created. While data generated by users is automatically shared, it is up to collaborators to choose to read the shared information or ignore information altogether. A high awareness team would keep updated with collaborators' generated information and read information soon after it is shared; whereas a low awareness team might experience a significant delay or even never access it. We found that most teams shared a high proportion of entities (mean=77.6%). We found that in average, 77.6% of the created entities were accessed by at least one teammate

One major critique is the lack of sharing of intermediate analytic insight for close collaboration. When individuals are exploring visualizations (e.g. zoom, filter, pan, drag, highlight), an insight pops up with a specific visualization state. Students complaint that they could not easily communicate that insight together with the associate views to the team. The team could "be looking at the same information but arranged in completely different ways" (P131). Unable to share immediate insights in the graphic context seems to make communication out of context.

Learning experience

Students reported that being aware that teammates were contributing motivated themselves to contribute.

> the notifications every time you saw someone annotated something kept you peace at mind that your teammates are also working efficiently on this project. (U64)

> During class I wasn't sure if my teammates were doing work for that class or another thing but then seeing their dot switch between applications on the software and updates pop up on my screen I knew they were doing work for 231. (U141)

> CAnalytics allowed us to see how much each person had contributed to their work (142)

The fact that you can see what other teammates are doing and they can see what you are doing creates a sense of accountability in terms of separating the work load. (User 51)

DISCUSSION

Reflections on method

The goal of the study is to explore design opportunities to support collaborative information analysis by evaluating tool usage in a natural environment over extended period of time. This is to complement research that only tests tools in short term lab studies (e.g. e.g. [**Convertino2012**, 8]). Due to the constraint of time (usually about one hour), these studies had to employ a simplified task with significantly reduced content and complexity. Teams would then not create complex information artifacts and had less difficulty balancing limited cognition between problem solving and team coordination. More complex information artifacts and higher cost of coordination would have provided more insights into team process of combining information and tool usage for integrating efforts.

Besides, the limited time may prevent teams from developing sufficient awareness to work properly. Participants in lab studies face a fresh new setting: a new formed team, a new collaborative tool, a new task, and a new environment. It takes time for teams to establish common ground, and to learn to appropriate the tool to best serve their team. Allowing teams for more time to explore and make trials would have provided an opportunity to observe how team awareness have developed and how teams have appropriated the tool to best solve the problem.

Our classroom study attempts to gain deeper insights on collaborative information analysis behavior by simulating real world settings in at least two aspects:

1. The study was one-week long. Teams were able to explore multiple strategies to solve the problem, and to change a strategy if they encountered a problem. For example, two teams decided to change the use of the tool halfway in their analysis. One team started with dividing work by case documents, but later decided members should annotate different entity types. Another team started with an accretion strategy by annotating all entities. Later they discovered that this strategy brought too much noise instead of useful information, and decided to clean out irrelevant entities (filtering strategy). Such change occurs as a consequence of increased awareness of team functions and tool capabilities, which takes time to develop.

2. Participants in this study are being trained to be professional analysts. They already have knowledge of analytic techniques and tools to implement those techniques. In the study, participants often compared CAnalytics to those tools, as well as teamwork with CAnalytics to practice with those tools. Their feedback provides more insight into strength and weakness of CAnalytics.

Reflections on result

A misconception about information analysis is that data modeling and data analysis are two staged activities. Most analytic tools assume data has already been modeled and ready to be visualized and analyzed. This is akin to the waterfall software development model, which features a sequential process that flows downwards through the phases of requirement conception, software design, software implementation, testing and maintenance. Critics have pointed out that the staged approach may not work properly, because clients may not know exactly what their requirements are before they see the working software and designers may not be fully aware of future difficulties in a new software product. Instead, an iterative design process is often required that leads to reframe user requirements, redesign, redevelopment, and retesting.

Similarly, relying only on information that has already been modeled and delivered to analysts will probably not solve all analytical problems [13]. It will probably be necessary to look elsewhere, re-model the data, and dig for more information. As suggested by the Data-Frame theory, constructing a frame is essential to determine how to model the data. Contemporary hypotheses help frame the problem, identify the gap, and determine what data to model. Re-modeling of the data could lead to a different picture of the problem (e.g. adding a link between two clusters changes the layout of the network), leading to completely different analytic path. The interaction log we captured demonstrated such an iterative analysis process and the positive subject feedback confirmed data modeling as an integral part of analysis.

We noted the importance of representing uncertainty. We observed teams in our study spontaneously employed two different approaches to deal with team uncertainty given that the tool did not include specific support: to hold facts and inferences in separate artifacts, and place facts and inferences of high certainty in an artifact. This demonstrates both challenge and opportunity to design for uncertainty support. We propose that a richer graphic language be designed so that analysts can encode uncertainty into the network view. Links and entities with different uncertainty are visualized in different colors. Users can *filter by uncertainty* so that users can choose to have only facts to take into account or confront all inferences to make a review.

We found that participants tended to make more visible contributions than valuable contributions, and collaborative technology only made some activities more visible, thus unintentionally encouraged participants to do more these activities, although they might not be useful at all. This is especially true when team practice has not been established and thus can be easily shaped by outside incentives, such as the awareness features in our system. We should be cautious to distinguish between awareness system and contribution system. A contribution system should only include factors that bear value to the task, such as hypotheses created and validated in the context of information analysis, and should be explicitly displayed to users to highlight the value of these factors. Awareness system, on the other hand, should share all relevant activities (and perhaps highlight information most relevant to the current user, for example, when teammates edit your en-

tity). The information, while valuable to the task goal, can motivate teammates to contribute in the same direction, and can remind teammates to pull you back when it deviates from the team goal (e.g. when one user created too many low-level entities in Team 108)

One major critic was the lack of view sharing support in the tool. In addition to data sharing, we find that views of data should become shareable resources as well. With the identical data pool, analysts often have different views of data. For example, analysts can apply a filter to have a reduced data view, highlight an area to sharpen analytic focus, and re-layout the node-link graph to cluster relevant entities. While the data pool represents the information the team have available, individual views of the data reflect analyst's *interpretation* toward the information. The views together with the underlying data embody user's intermediate analytic status. Therefore we propose that just like data, views, which embodies interpretation of data, should be shareable.

Views as resources should also be extensible and reusable. For example, several participants reflected that there were situations when they found a collaborator's view useful and wanted to build their own work upon that view without manually reproducing the view. With views as resources, individuals can take the views to their need. They can also deliberatively share their own view when they feel other collaborators will be interested. Shared views are interactive rather than static images, so that analysts can still perform full functions including filtering and highlighting, and are able to evolve the view with collective team efforts, a critical requirement emphasized in [4]

Our study suggests that the instrument plays an important role in shaping students' behavior towards more collaborative learning. With traditional single-user tools, students often employ a divide-and-conquer strategy; they divide their job responsibility, work individually on their own part, and put the results together in the end for report submission. In our study, we observed that students spontaneously conducted closer collaboration and enjoyed being able to contribute simultaneously.

Anecdotal reflections from the instructor suggested that the system can include support for instructor intervention. During the study, the instructor would go over to students and check their computer screen about how they were doing, and provide guidance if necessary. The instructor commented that he valued students' thinking and reasoning process, and believed that monitoring and guiding students' ongoing performance would be a valuable supplement to classroom instruction. As claimed by Heuer [13], training will be more effective if supplemented with ongoing advice and assistance. CAnalytics could do more in supporting the instructor. CAnalytics already provides an integrated workspace for data modeling, information analysis and hypothesis generation, and thus makes it easier to monitor the whole analytic process. Besides, students' interaction logs are already captured (for team awareness and research purpose only now), and could be streamed to the instructor for performance monitoring. The process data provides the instructor a new window to assess students' performance and to provide intervention when necessary, as suggested by learning analytics techniques [17].

References

[1] Richard V Badalamente and Frank L Greitzer. "Top Ten Needs for Intelligence Analysis Tool Development". In: *First Annual Conference on Intelligence Analysis Methods and Tools* May (2005).

[2] Eric a. Bier, Stuart K. Card, and John W. Bodnar. "Entity-based collaboration tools for intelligence analysis". In: *2008 IEEE Symposium on Visual Analytics Science and Technology*. Ieee, Oct. 2008, pp. 99–106. ISBN: 978-1-4244-2935-6. DOI: 10.1109/VAST.2008.4677362. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4677362.

[3] Eric a Bier, Stuart K Card, and John W Bodnar. "Principles and tools for collaborative entity-based intelligence analysis." In: *IEEE transactions on visualization and computer graphics* 16.2 (2010), pp. 178–91. ISSN: 1077-2626. DOI: 10.1109/TVCG.2009.104. URL: http://www.ncbi.nlm.nih.gov/pubmed/20075480.

[4] John M. Carroll, Marcela Borge, and SI Shih. "Cognitive artifacts as a window on design". In: *Journal of Visual Languages & Computing* 24.4 (2013), pp. 248–261. URL: http://www.sciencedirect.com/science/article/pii/S1045926X13000207.

[5] John M. Carroll et al. "Awareness and teamwork in computer-supported collaborations". In: *Interacting with Computers* 18.1 (Jan. 2006), pp. 21–46. ISSN: 09535438. DOI: 10.1016/j.intcom.2005.05.005. URL: http://iwc.oxfordjournals.org/cgi/doi/10.1016/j.intcom.2005.05.005.

[6] George Chin Jr, Olga A. Kuchar, and Katherine E. Wolf. "Exploring the analytical processes of intelligence analysts". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2009, pp. 11–20. ISBN: 9781605582467. URL: http://dl.acm.org/citation.cfm?id=1518704.

[7] C. Gorg, Z. Liu, and J. Stasko. "Reflections on the evolution of the Jigsaw visual analytics system". In: *Information Visualization* 13.4 (July 2014), pp. 336–345. ISSN: 1473-8716. DOI: 10.1177/1473871613495674. URL: http://ivi.sagepub.com/lookup/doi/10.1177/1473871613495674.

[8] Nitesh Goyal and Susan R Fussell. "Effects of Sensemaking Translucence on Distributed Collaborative Analysis". In: *CSCW '16 Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 2016.

[9] Nitesh Goyal et al. "Effects of Implicit Sharing in Collaborative Analysis". In: *Proceedings of the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 129–138. ISBN: 9781450324731. URL: http://www.cs.cornell.edu/%7B~%7Dngoyal/1470%7B%5C_%7DChi2014.pdf.

[10] Saul Greenberg and Michael Rounding. "The Notification Collage: Posting Information to Public and Personal Displays". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)* 3.3 (2001), pp. 514–521. DOI: `10.1145/365024.365339`. URL: `http://portal.acm.org/citation.cfm?doid=365024.365339`.

[11] Amir Hossein Hajizadeh, Melanie Tory, and Rock Leung. "Supporting awareness through collaborative brushing and linking of tabular data." In: *IEEE transactions on visualization and computer graphics* 19.12 (Dec. 2013), pp. 2189–97. ISSN: 1941-0506. DOI: `10.1109/TVCG.2013.197`. URL: `http://www.ncbi.nlm.nih.gov/pubmed/24051785`.

[12] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". In: *Advances in Psychology* 52.C (1988), pp. 139–183. ISSN: 01664115. DOI: `10.1016/S0166-4115(08)62386-9`. arXiv: `arXiv:1011.1669v3`.

[13] Richards J Heuer. *Psychology of intelligence analysis*. Lulu.com, 1999.

[14] Youn-ah Kang and John Stasko. "Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study". In: *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Ieee, Oct. 2011, pp. 21–30. ISBN: 978-1-4673-0014-8. DOI: `10.1109/VAST.2011.6102438`. URL: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6102438`.

[15] Iulian Martin. "Current training approaches in intelligence analysis". In: *The International Annual Scientific Session Strategies XXI 2*. 2014.

[16] Kjeld Schmidt. "The Problem with fffdfffdfffd Awareness '". In: *Computer Supported Cooperative Work (CSCW)*. 2002, pp. 285–298.

[17] George Siemens and Phil Long. "Penetrating the fog: Analytics in learning and education." In: *EDUCAUSE review* 46.5 (2011), p. 30.

[18] John Stasko, Carsten Görg, and Robert Spence. "Jigsaw: supporting investigative analysis through interactive visualization". In: *Information Visualization* 7.2 (2008), pp. 118–132. ISSN: 1473-8716. DOI: `10.1057/palgrave.ivs.9500180`. URL: `http://ivi.sagepub.com/lookup/doi/10.1057/palgrave.ivs.9500180`.