Facilitating the Reuse Process in Distributed Collaboration: A Distributed Cognition Approach

Syavash Nobarany[1], Mona Haraty[1], Brian Fisher[2]

[1]Department of Computer Science
University of British Columbia, Vancouver, Canada
{nobarany, haraty}@cs.ubc.ca

[2]School of Interactive Arts and Technology
Simon Fraser University, Surrey, Canada
bfisher@sfu.ca

ABSTRACT

Facilitating the reuse process and enabling unanticipated reuse can improve efficiency of distributed collaboration. However, supporting the reuse process in complex and dynamic contexts, where future use of information is difficult to predict, is challenging. Collaborative analytics exemplifies such a context. We employed distributed cognition theory to design a collaborative visual analytics system, called AnalyticStream, for facilitating the reuse of analysis outcomes. In contrast with the commonly used detail-oriented approach to applying distributed cognition, we performed a high level analysis of the design situation and we identified the cognitive processes that could be distributed over people to facilitate their collaboration. We examined some of the ideas derived from the theoretical analysis, by designing a simple reminding process through recommending relevant pieces of analysis, as well as a mechanism for attention management through allowing users greater control over their shared activity streams. A mixed-methods study of AnalyticStream showed that suggesting relevant artefacts facilitated discovering and consequently reusing them, and provided context-relevant awareness of other analysts' activities.

Author Keywords

Distributed cognition; reuse; collaborative visual analytics; distributed collaboration; reusability.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Computer-supported cooperative work.

General Terms

Human Factors; Design.

INTRODUCTION

Distributed collaboration heavily relies on reuse of shared artefacts. Artefacts generated by some users may become an input for other users to perform their tasks and to build new artefacts based on them. Sometimes, especially in organizational settings, the reuse process is planned and the artefacts generated for reuse are expected to be well documented and hence reusable by other users. Moreover, in a planned reuse scenario, collaborators know which resources are available to them. However, in large-scale collaboration often opportunities for reuse are hard to plan for, and these unanticipated opportunities are eventually missed. Unanticipated reuse plays a prominent role in collaborative systems [15,20]; however, knowing which artefacts should be shared with others is challenging, especially when people are not aware of how their contributions may be helpful to others [15,21].

In collaborative analytics, human reasoning is the valuable resource that is shared and reused. However, only a fraction of reasoning turns into products that are widely distributed. The absence of support for collaboration and reuse of the analysis process and products inhibits analysts from efficient cooperation and engaging in collaborative sensemaking processes [13,31]. Supporting collaboration in visual analytics tools have been a subject of study over the last few years [11,13,32,33]. However, there has been little emphasis on supporting reuse of reasoning artefacts (e.g. hypotheses, evidence and causal relations).

We approached this problem by analyzing distributed collaboration through the lens of distributed cognition theory, and we investigated how the distribution of memory and attention processes can provide a platform for supporting unanticipated reuse in large-scale collaborative systems. We applied the proposed theoretical perspective in the design of a collaborative visual analytics system called AnalyticStream. We designed a simple socially distributed reminding process through recommending pieces of analysis that are relevant to the user's workspace, as well as a mechanism for attention management through allowing users greater control over their shared activity streams. The specific contributions of this work are as follows:

1. Introducing a non-traditional approach to applying distributed cognition theory to the design of CSCW systems.
2. Applying the proposed approach to the design of a collaborative visual analytics system called AnalyticStream.
3. Collecting qualitative and quantitative data about under- or unexplored design ideas resulted from adopting the proposed approach.

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

In the following sections, first, we provide an overview of the collaborative analysis process to describe the context of our design. Then, we briefly explain how distributed cognition theory shaped the design of AnalyticStream, and we conclude with a mixed-methods study of AnalyticStream.

COLLABORATIVE VISUAL ANALYTICS

Visual Analytics is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [31]. A typical use case of visual analytics systems in the context of intelligence analysis is supporting analysts "to piece together information buried in disparate data—including immigration records; patterns of travel; telephone calls; and names, affiliations, and locations of suspected terrorists—to enable them to spot an emerging attack before it can be executed" [31]. Similar use cases can be found in other analytical processes such as scientific research where researchers piece together theories, previous findings, and new data to explain a phenomenon or to make predictions.

Collaborative analytics systems are designed to provide the required infrastructure for efficient analytical collaborations and collaborative sensemaking processes. In the collaborative analytics process, analysts engage in the intertwined processes of sensemaking and sensegiving to address a problem more accurately and more comprehensively. Analysis processes and products should be externalized to enable collaborative processes. Traditionally, analysts use paper notes, concept maps, and other reasoning and argumentation diagrams for this externalization [18]. Recently, reasoning support systems such as Analyst's Notebook [35], Scalable Reasoning System [25], and Oculus SandBox [34] support an extensive set of analytical methods. However, these systems do not provide specific mechanisms for facilitating reuse of analysis' outcomes.

The analytics process involves seeking relevant resources followed by extraction, marshalling and summarizing relevant pieces of information and analysis. This process often involves the reuse process in the form of building upon the previous findings. The cycle of reuse starts with sharing of the analysis findings, and continues with the discovery of them by other analysts or the same analyst at another time, understanding them, integrating them in the current analysis and ultimately sharing the outcome. Every step of the reuse cycle presents challenges and needs to be supported by collaborative analytics systems.

ADOPTING A DISTRIBUTED COGNITION APPROACH

As was stated by Simon [29], "...solving a problem simply means representing it so as to make the solution transparent". Looking through a theoretical lens enables perceiving and representing the problem space differently. We applied distributed cognition theory [16] to investigate how analysts, reasoning artefacts, and user interface elements take part in various cognitive processes to form a distributed cognitive system. Cognitive processes are the processes by which the cognitive system's input is transformed, reduced, elaborated, stored, recovered, or used [22]. Learning, remembering, attention, reasoning and problem solving are some of the cognitive processes that may happen as a cognitive system operates.

From the distributed cognition perspective, a cognitive process is delimited by functional relationships among the elements that participate in it [16]. A cognitive process can be distributed across people, through time, or across internal and external representations. In this section, we demonstrate how distributed cognition theory can inform and shape the design of collaborative analytics systems. Our approach to applying distributed cognition theory is slightly different from the detail-oriented approach that is commonly used in HCI and CSCW research [10,14]. In the detail-oriented approach, a researcher performs a detailed analysis of how information is propagated through the cognitive artefacts and actors, and how its representation is transformed in the process. This analysis helps the researcher to identify breakdowns in the system and design new representations or workflows to address it [10]. Our approach, while similar in nature, focuses on higher-level analysis of collaborative workflows through identifying various cognitive processes that can be distributed over people, time and artefacts to facilitate the collaboration. In our approach, the collaborative system is considered as a cognitive system, which needs to perform various cognitive processes such as remembering, attention, reasoning and learning to achieve its goals. Therefore, the design of a collaborative system involves identifying the cognitive processes that can help the cognitive system succeed, and designing the infrastructure required for an efficient distribution of these processes.

A collaborative analytics system can be seen as a cognitive system composed of analysts, reasoning artefacts and tools that are used to represent, communicate and manipulate the artefacts. Analysts' mental resources normally constitute a significant part of a distributed cognitive system and are responsible for complex or creative computations and coordination of cognitive resources (including themselves). They also act as part of the distributed memory of the system. However, they externalize the information stored in their memory to be able to cope with their complexity and volume. The externalized reasoning artefacts (e.g. evidence, notes, comments, causal networks, hypotheses, etc.) can be stored, shared, retrieved, and reused by other actors in the system.

As discussed earlier, first we need to identify the cognitive processes that should be supported to achieve the goal of the system, which is facilitating collaborative reasoning and reuse of reasoning artefacts in collaborative analytics systems. One approach is to identify such cognitive processes based on how an individual achieves this goal. If we consider reuse of one's own knowledge or work, memory processes are the most apparent cognitive processes that should be explored. We are constantly reusing our knowledge or work, because we remember the

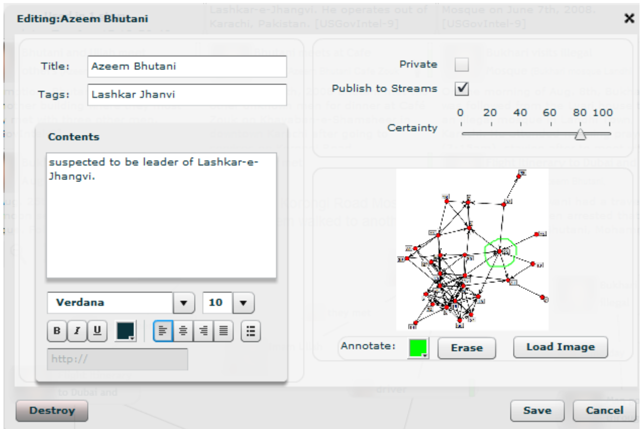

Figure 1. UI for creating and editing a piece of analysis

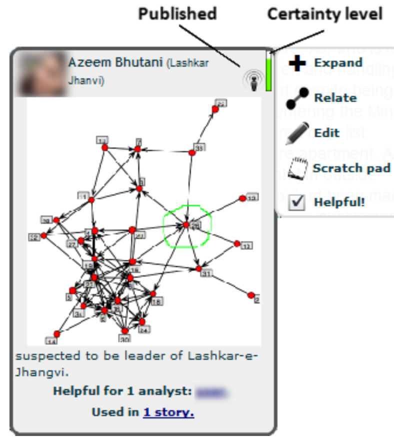

Figure 2. Visual representation of a piece of analysis

relevant ones, intentionally or unintentionally, when we need them.

Looking closely, we see that some of the knowledge that is reused has been intentionally memorized or kept carefully, because we had anticipated the future reuse. We pay attention to the pieces of information that are deemed to be helpful in future, and we can only remember and reuse what we have noticed or attended to. After remembering the relevant experience or piece of information, we need to find a way to apply the previously acquired knowledge to the current situation. Decontextualization and recontextualization of the knowledge require reasoning [1]. In the context of collaborative analytics, reuse of analysis strategies, methods, and heuristics requires the process of recontextualization. In this work, for the sake of simplicity, we focused on facilitating the reuse of products of analysis, and in the next sections, after a brief overview of AnalyticStream, we describe the social distribution of attention and memory processes in AnalyticStream.

OVERVIEW OF ANALYTICSTREAM

AnalyticStream is a collaborative visual analytics system that provides an environment for evidence filing and marshalling. Analysts can share and schematize their pieces of analysis, build cases and tell stories to communicate their analysis. AnalyticStream is designed to support early-stage sharing and reuse of analysis outcomes. Often, sharing the outcomes of the analysis process is postponed to the end of the project or milestones and most of the intermediate outcomes of the analysis do not reach a publishable state. Several benefits are associated with the sharing of ideas and outcomes in the early stages of analysis and research, including early identification of flaws in the analysis process, refining the ideas and incomplete hypotheses, accelerating the review of available information, and avoiding missing of the important sources of information.

Using AnalyticStream, analysts can create sequences of pieces of analysis that are generated by various analysts to convey their line of reasoning. The *piece of analysis* is the basic unit of analysis in AnalyticStream; it includes a title, rich text contents and possibly an image/visualization that can be annotated (Figure 1 and 2). A piece of analysis can be used to represent a fact, an assumption, an entity, a relation, or any other type of artefact that analysts may want to externalize or record. A piece of analysis can be expanded to a sequence of pieces of analysis. We call this sequence the *story* of the piece of analysis; however, the *story* may be just a group of pieces related to the expanded piece and the semantic relations between the pieces and the story is arbitrary. For example, Figure 3 shows the story of "Azeem Bhutani", a person that is represented by a piece of analysis. In this example, the story is a sequence of events and other information related to the person's activities. Analysts can also create relations between pieces; these explicit relations are shown in the Graph view (Figure 3, bottom). In addition, analysts can edit the pieces, add them to their personal scratchpad (Figure 4, bottom-left), or mark them as helpful (Figure 2) to praise or thank the producer of the piece, as a basic incentive mechanism.

The main workspace is the tabbed area shown in Figure 3. Each tab represents the story of a piece of analysis using a multiple-coordinated view approach to enable both sequential marshalling using Narrative view (Figure 3, top) and relation-oriented interaction using the Graph view (Figure 3, bottom). Document-Viewer (Figure 4) is a permanent tab that provides an environment for reading documents during the analysis process. The user interface components that contribute to the social distribution of attention and memory processes are explained in the following sections.

SOCIAL DISTRIBUTION OF MEMORY PROCESSES

The externalized artefacts should be accessible by the other analysts to enable the reuse process. Rudimentary support for sharing and reuse of artefacts can be provided by supporting storage and retrieval of them. Looking through the lens of distributed cognition, to facilitate the reuse process the system should support various processes that are associated with memory such as memorizing, remembering, reminding, and perhaps even forgetting.

As discussed earlier, reusing one's own work and knowledge is only possible because we remember the

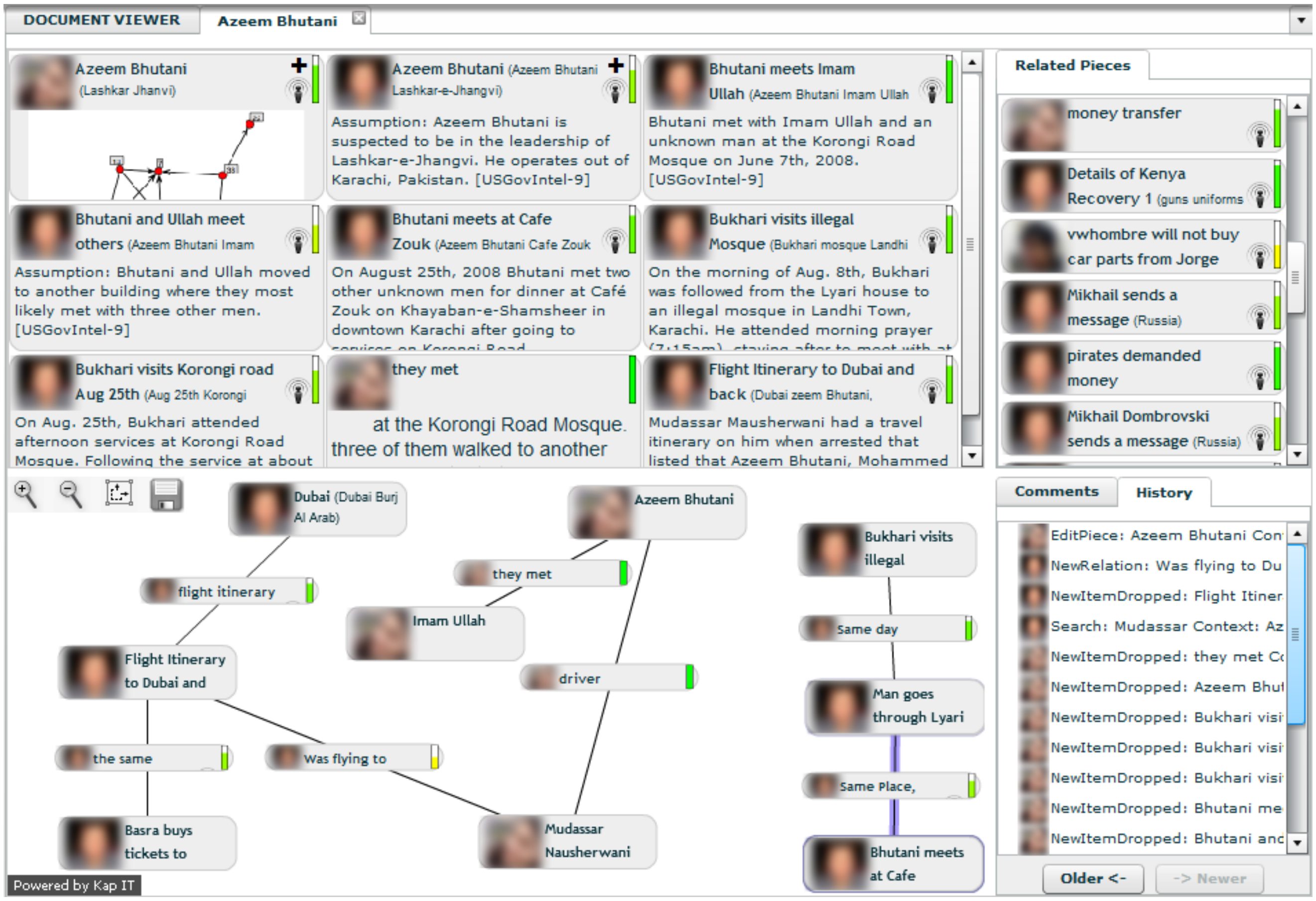

Figure 3. Representation of the story of a piece of analysis ("Azeem Bhutani") using Narrative view (top), Graph view (bottom), Related-Pieces panel (top-right), History panel and Comments panel (bottom-right).

relevant previous experiences and works. Supporting the aforementioned memory processes can facilitate various collaborative workflows. Memorizing (storage) and active remembering (retrieval through search or browsing) processes are supported by all collaborative systems; however, there has been little work on supporting the forgetting and the reminding processes, mainly because the need for them is less apparent. For example, recording (memorize) the entire user interactions can be desirable, because it enables users to navigate through their interaction histories to find and reuse sequences of interactions. Forgetting mechanisms can be designed to handle the unbounded interaction logs efficiently. The simple strategy that is often used is to eliminate the old data, which is not desirable when designing for unanticipated reuse. Cognitive theories of human memory architecture and processes can inform the design of these mechanisms. For example, based on the Schema theory of cognition, forgetting occurs when certain pieces of information get integrated and lose their individual identity [3]. In a collaborative system, this integration may occur over several users' interaction histories over a period of time. Looking at the design space through the lens of distributed cognition theory enables us to transform the design problem by bringing it to the cognition domain and subsequently enables applying other theories of human cognition to solve it.

Various history recording and navigation mechanisms are proposed to enable the reuse of one's own previous works [12,17,26]. Allowing users to replay, manipulate, or simply continue the steps taken by other users enables a greater level of reuse in collaborative systems. However, finding and managing activity sequences is challenging, and various related privacy issues should be addressed. Efficient recording and automatic integration of activity history are exciting but challenging research problems that we postpone for future research.

AnalyticStream provides the basic mechanisms for interacting with activity history. The History panel (Figure 3, bottom-right) shows the activities related to the pieces that are visible in the main workspace (the Narrative or the Graph view). Each item presents the name of the user, the related piece, and the name of the action performed (e.g. adding pieces to the story, creation of the piece, edits, etc.). Moreover, clicking on the user name or profile picture displays the activities of the user. Despite providing basic history navigation mechanisms, we decided to focus on the less explored mechanisms for supporting the distribution of memory and attention processes. The Activity-Stream panel

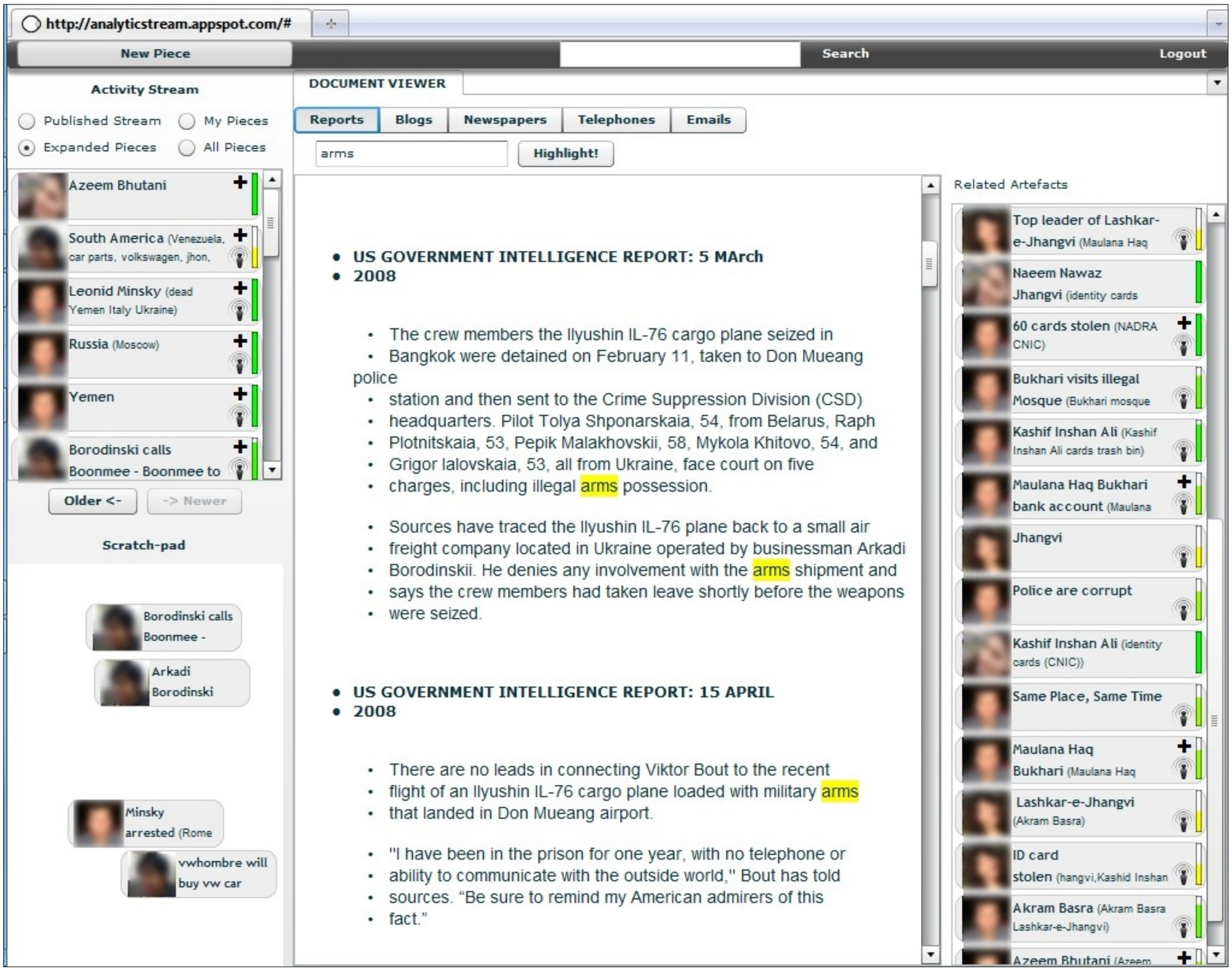

Figure 4. Initial AnalyticStream user interface: The left sidebar includes the Activity-Stream panel (top) and the Scratchpad (bottom-left). The Document-Viewer includes a reading area and its right sidebar shows the relevant pieces.

(Figure 4, top-left) provides a coarse-grained history and is considered as an awareness mechanism, which will be discussed in the section on the distribution of attention.

In a collaborative analytics system, the reuse process starts with searching for other analysts' findings. This process may be ineffective when the analysts are not aware of the items that others have stored. To generate a query, analysts must know enough to know what is not known and which of the many unknowns can be understood from the other analysts' findings. Finding out about all relevant abstractions and entities is challenging when the analysts do not know what to ask [8]. Due to the exploratory nature of the analysis process, the discoverability is a prominent problem. The analytics process deals with vague situation models, which makes it near impossible to ask the right questions.

Looking through the lens of distributed cognition, this challenge can be addressed by extending its analogous process of reusing one's own experiences and prior knowledge. Storing an item in a shared repository can be seen as the process of memorization of an item by a group; therefore, reusing it requires remembering. Information can be retrieved from memory in two different ways: The first one is actively trying to remember something that we are aware of having in our memory, and the second one is being reminded about the artefacts and processes that are related to the current context, which happens frequently in our daily lives. Reminding process often happens when a situation or a change in the external world, triggers a related memory; thus it is often a process distributed over the artefacts, times and one's (internal) memory. Schank's theory of reminding [27] posits that when we process events as they happen, we need to find specific episodes in memory that are closely related to the current input; however we do not consciously look for them, because we are not explicitly aware of their existence. In a collaborative system, the system can assist users by reminding them about other users' products that are related to the user's current task. We hypothesize facilitating the social distribution of reminding process through automatic recommendations creates opportunities for reuse.

Related-Pieces panel (Figure 3, top-right) automatically suggests the pieces of analysis that are related to the analyst's workspace. Previous research has shown that content similarity and collaborative filtering algorithms can be effective for similar purposes [5]. In this study for the sake of simplicity and due to the small number of subjects, we only used content similarity as a measure of relevance for generating suggestions. To assess content similarity, AnalyticStream generates a query based on the contents of the pieces of analysis in an analyst's workspace and searches the repository of pieces to find related pieces. The search is automatically performed whenever the workspace is updated; for example, adding a piece to the workspace or removing or editing it triggers the search process. We used Stanford Named Entity Recognizer [7] to extract the entities from the pieces of analysis and give them more weight in the search process to improve the relevance assessment. Similarly, in the Document-Viewer (Figure 4) the pieces that are related to the reading area are suggested to the user. In this case, the query is generated based on the visible text in the Document-Viewer and scrolling the page triggers the generation of suggestions. We hypothesize that these suggestions improve the discoverability of the pieces of analysis, which can facilitate the reuse process. During the course of this study, we found out that a similar approach has been evaluated in a case study with one analyst, which showed a high rate of reuse in the single-user scenario [28].

Additionally, AnalyticStream provides a scratchpad (Figure 4, bottom-left) to drag and drop pieces to and from other components such as the Related-Pieces panel, the Narrative and the Graph views, etc. The scratchpad supports organizing the pieces spatially and was aimed at providing a personal space like a desktop. Through the lens of distributed cognition theory, the scratchpad acts as an extension of short-term memory that helps a user to keep track of his analysis during a session as well as between sessions. However, due to its personal (non-collaborative) nature, analysing its role is out of the scope of this paper.

SOCIAL DISTRIBUTION OF ATTENTION PROCESSES

Attention is a cognitive process that serves various purposes. It helps us to select a target item from distracters and enables the ongoing awareness of the environment [19]. Attention can be socially distributed in a collaborative analytics system. Socially distributed attention mechanisms have been subject of research under the name of awareness. User awareness in a collaborative system is defined as the way users perceive their collaborators and what they are doing, without direct communication [6].

From the point of view of distributed cognition theory, awareness is a distributed attention mechanism in which users' behaviour attracts other users' attention to the extent that makes them aware of their behaviour. In a collaborative analytics system, sharing of attention among several analysts can facilitate the coordination between analysts. Analysts should be able to share their attention, i.e. to let others know of their foci of attention, and actively attract other analysts' attention to tasks, artefacts and people of interest. Analysts as elements of the distributed cognitive system should be able to find out about the foci of attention of other elements of the system as well as the foci of attention of the cognitive system (or subsystem) that they are part of. Awareness helps analysts to coordinate their activities based on the common or related goals, for example, by devoting their cognitive resources to the common foci of attention. Supporting awareness is essential for facilitating collaborative sensemaking processes [24]. However, recognizing awareness as one of the attention mechanisms that can enhance the flow of information in a complex cognitive system helps to explore and examine new interactions and workflows in collaborative systems. In AnalyticStream, the Activity-Stream panel (figure 4, top-left) provides awareness about recently created or edited pieces. Additionally, it serves as a browsing mechanism for finding pieces of analysis. Similar to a social network stream, the Activity-Stream panel makes analysts aware of the foci of attention of other analysts and helps them to better allocate their cognitive resources.

Privacy and Attention

The concept of privacy has been defined in various ways in social science literature. Parent [23] reviewed several previously proposed definitions and based on them, he proposed a new definition that describes privacy as "the condition of a person's not having undocumented personal information about himself known by others" or in other words "the absence of undocumented personal knowledge about a person". However, this definition does not capture nuances of this concept. The definition suggests that if personal information is documented somewhere, it does not matter if others know it or not. This issue is challenged by recent privacy researchers. Boyd introduced the concept of privacy and security through obscurity [4], which implies that one can put her/his personal information in a public or shared space, while maintaining some level of privacy. Boyd underscored the difference between something being public versus something being announced to others.

The concept of privacy by obscurity suggests that privacy is not just about the existence of information in a public place, but also strongly interrelated with the availability of information to others, awareness of the availability, and the amount of attention that it receives. The design of collaborative systems should aim at providing mechanisms that enhance discoverability and availability of information and foster sharing of the potentially useful information while maintaining collaborators' privacy.

In AnalyticStream, a piece of analysis may have one of the following privacy states: private, public, or published. Private pieces are only visible to the analyst who produced them. Public pieces are visible to all analysts, but they do not show up in the default view of the Activity-Stream panel (hence unpublished). Finally, published pieces are visible to all the analysts and show up in the Activity-Stream panel (Figure 4, top-left). The Activity-Stream panel provides 4 different modes for browsing through pieces of analysis: browsing through one's own pieces, expanded pieces, published pieces, and all pieces. The creators of pieces of analysis mark them as published, when they want to make other analysts aware of them. The default mode of the Activity-Stream shows the published pieces so that analysts become aware of the latest important activities of other analysts. An important cost of information sharing is the cost incurred by the recipients of the information which is their attention. Therefore, devising new attention management mechanisms can help to balance the trade-offs in sharing [2,30].

METHODS

As part of the evaluation of AnalyticStream, we were interested in understanding how recommendations affect participants' reuse behaviour and how participants compare it with search. Moreover, as a secondary goal we were interested to see how the shared activity stream and the enhanced privacy settings affect sharing behaviours.

We used a mixed methods design to collect and analyze qualitative and quantitative data. The qualitative data, collected through interviews and participants' diaries, are the core of the study, which were complemented by quantitative data derived from Likert-scale questions. A controlled lab study could allow for methods such as think-aloud that can provide a clear sense of user behaviours; however, we decided to allow participants to conduct analysis in their comfort zones and at times that work for them to enable them to engage in the analysis process. Moreover, the collaboration in natural settings cannot happen in a one or two-hour session and we wanted the participants to get involved in analyzing the dataset, so that the collaboration and the reuse process occur naturally. Although we did consider using simple datasets that can be analyzed by untrained users, we decided to use a complex realistic dataset to enhance the ecological validity of the study.

Task and Dataset

Participants were given a text corpus including five large document sets related to illegal arms dealing. The dataset was acquired from the Visual Analytics Science and Technology (VAST) contest 2010 [9] and contained intelligence reports, emails, blog posts, telephone transcripts, and news articles. The participants were asked to provide a forensic analysis of illegal arms dealing. In particular, we asked them to find the hidden plot in the data, which involved summarizing and making sense of the activities that happened in each country with respect to illegal arms dealing and the associations among the involved players, based on a synthesis of the information from the different sources (reports, emails, etc.).

Participants

A total of 6 participants (3 females), between 25 and 38 years of age, with science and engineering backgrounds were recruited by calling for volunteers through emails to visual analytics research groups and researchers of a North American university. The participants should have been able to participate in a collaborative analysis of a complex realistic dataset and deal with uncertain or missing data, in multiple analysis sessions.We did not have access to intelligence analysts, but we had the opportunity of recruiting a group of visual analytics researchers who had experience performing analytical tasks in the VAST challenge and/or in visual analytics courses. As a result, our subjects were familiar with typical visual analytics tools, tasks, and processes.

Procedure

Participants were asked to use AnalyticStream for multiple sessions to make sense of the dataset. Initially, we asked them to use the system for a specific amount of time in every session, but later we noticed that because of not being in a lab environment, we cannot dictate the details of their participation. Therefore, we did not specify any timing for their participation, and some of them started their analysis process several days later than others, which is normal for a distributed collaboration. Especially because we primarily focused on collecting qualitative data, it was desirable for us to have participants in different situations, i.e. different levels of experience with analytics tasks and starting analysis with different numbers of pieces of analysis available to them. Participants were asked to fill out a semi-structured diary during and after each analysis session. After the experiment period, participants were asked to fill out a post-test questionnaire, to reflect on the overall experience and evaluate various features of the system. Finally, based on the collected data, we interviewed them to disambiguate or validate our uncertain interpretations of diaries and answers to open-ended questions.

RESULTS AND FINDINGS

Memory Distribution Mechanisms

Just being exposed to the artefacts produced by self or others can remind people of related information. One participant mentioned using this strategy actively by browsing other participants' products: *"I tried to remember some parts using [one of the participant's] stories and some pieces created by a new guy. But still after these 5 sessions, the topic and the written reports are hard to absorb for me and I can't remember events very well..."*.

Participants were asked to assess the usefulness of the two main memory distribution mechanisms: Active retrieval using search and passive retrieval (or reminding) using recommendations. All participants found recommendations to be useful for finding other analysts' pieces of analysis and 4 (out of 6) of them found search to be useful for this purpose (Figure 5). Since we anticipated the helpfulness of both mechanisms, additionally, we asked the participants to compare the helpfulness of the two mechanisms. Overall, most of the participants (4/6) found recommendations more useful than search for finding other analysts' pieces of analysis (Figure 6).

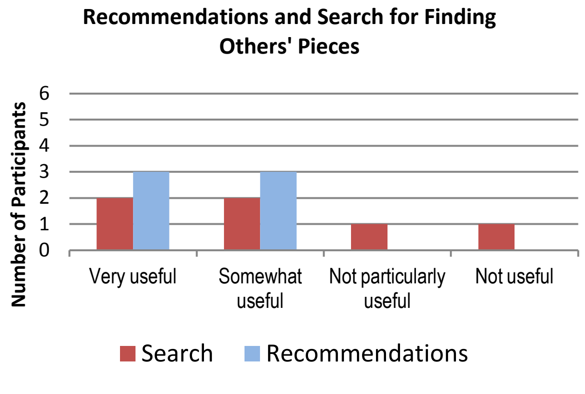

Figure 5. Usefulness of and search for finding other participants' pieces of analysis

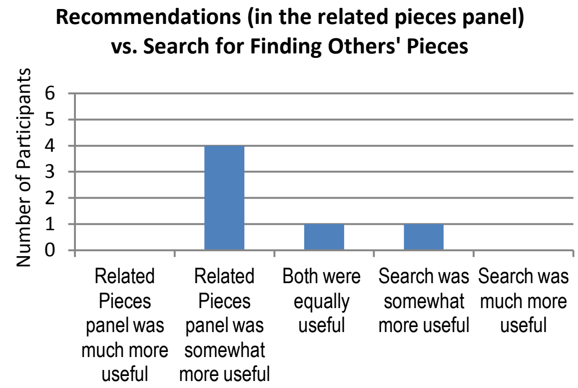

Figure 6. Comparing usefulness of recommendations vs. search for finding other participants' pieces of analysis

One of the participants explained her preference for recommendations, as follows: *"I wasn't looking for a specific thing; I read related artefacts during reading documents; I guess search would be more useful later in my analysis process"*. The dataset was designed to be analyzed by several analysts over a long period; therefore, during this study participants were mostly in the initial phases of analyzing the dataset, where they had an ambiguous model of the problem space, and they could not distinguish the key entities in the dataset. The recommendations that referred to the entities in the workspace, as one participant said, helped to *"relate some events that the reports didn't connect"* and identify the entities that require further investigation. This benefit of recommendations can be realized in a variety of collaborative systems that deal with complex problems where collaborators have to deal with constantly changing data and situations.

Another participant explained his usage of search, as follows: *"I use the search [...] to find pieces I know exist."* Even in these situations, sometimes as he reported, recommendations worked as a substitute for search: *"I was going to search for [a piece] but it turned out it was already listed [in the Related-Pieces panel]"*. The only participant who found recommendations less helpful than search was the least active participant of the study (based on the number of interactions with the system). This suggests that the more analysts engage in their analysis, the more recommendations become helpful; because when analysts are focused on their analysis, they may spend less time exploring others' analysis products. Investigating this hypothesis is left for future work.

One of the participants found recommending relevant artefacts useful for connecting subplots when analysts work on different subplots, and the subplots had something in common. This was one of the main anticipated benefits of recommendations. In distributed collaboration, often, implicit or explicit division of labour makes collaborators focus on a specific set of resources and the resources connected to them (similar to exploring research literature through citations). In such situations missing relevant resources is easy and unsolicited recommendations can help to solve this problem.

According to usage logs, all participants browsed through the recommendations and read many of the recommended pieces. This can be considered as an indicator of reuse; every instance of reading relevant pieces has potentially helped an analyst to develop his understanding of the information space. For example, one of the participants mentioned that by seeing other analysts' relevant pieces, he noticed: *"We had common hypotheses and there were pieces that added new info to what I knew."* This statement also shows that the recommendations can be employed as a context-based awareness mechanism, as another participant said: *"[I] used related artefacts to know what others analyze and what they discovered."* Additionally, we reviewed the usage logs to find if any explicit instance of reuse has occurred, and we found that the more active half of the participants (3/6) explicitly reused the recommended pieces of analysis in building stories.

We also asked the same questions regarding the reuse of *own* pieces of analysis. However, participants' responses were equally divided between search and recommendations. One explanation is that the analysis process was not long enough for participants to forget about the pieces that they have created earlier. The major advantage of recommendations over search is evident when users cannot ask the right questions. Being aware of the existence of the pieces and remembering specific information or keywords related to them allows for effective search.

In sum, the study showed that the recommendations were particularly helpful when the participants did not know what to look for or even were not aware of the existence of helpful pieces of information.

Attention Distribution Mechanisms

Usage logs showed that all participants browsed through the activity stream right after logging in. Interviews indicated that browsing through the shared activity stream served two purposes. First, it helped remembering their own activities, which exemplifies the distribution of refocusing attention and remembering over time. Second, browsing through the activity stream helped them become aware of others' activities, which exemplifies the distribution of attracting and paying attention across people and over time.

As discussed earlier, understanding the effects of enhanced privacy settings requires a focused study in ecologically valid settings where analysts care about the consequences of sharing. Nevertheless, we looked at how participants used the privacy settings, and we found that the two most active participants used the public (unpublished) privacy level. Content analysis of the public pieces revealed two usage patterns: first, creating helper pieces or relations that did not have an independent meaning, but were generated to explain or comment on other pieces, and second, creating pieces that were considered to be less important or less relevant to the participants' main line of reasoning. These pieces may be helpful to other analysts or may become helpful later and sharing them creates opportunities for unanticipated reuse.

All participants used the published mode of sharing pieces of analysis, which was the default mode and only one participant created private pieces. In realistic settings, people are more concerned about when and with whom they share pieces of information as well as about consequences of sharing them [2]. This suggests that in-depth investigation of the effects of the enhanced privacy/attention settings requires a higher level of ecological validity, which is left for future research.

In addition, allowing participants to mark helpful pieces worked as another mechanism for attracting other participants' attention. Originally, this feature was designed to provide an incentive for creating and sharing valuable pieces of analysis. As a side benefit, as one of the participants said: *"It helps in emphasizing the better posts"*, which suggests that marking as helpful was used to signal valuable pieces and attract other analysts' attention.

In sum, the study showed that the attention distribution mechanisms can help users to make their products available without necessarily attracting other analysts' attention, and to attract attention only to helpful pieces of analysis.

SUMMARY AND CONCLUSION

Looking through the lens of distributed cognition theory helped us gain a different understanding of the design space of collaborative systems. Rather than focusing on explicit requirements for a collaborative analytics system, we analyzed how various cognitive processes can be distributed over collaborators and how the system can facilitate the distribution of these processes.

One approach to employing distributed cognition theory for designing a collaborative system is to consider it as a cognitive system and identify the cognitive processes that can be socially distributed to support the goals of the collaborative system. Using this approach, we identified several possibilities for facilitating the reuse process through supporting the social distribution of memory and attention processes. Then we analyzed how user interface components can augment and support the social distribution of memory and attention processes. Implicit division of labour can be facilitated through enhancing shared attention and awareness mechanisms. Awareness of others' efforts can help collaborators to reuse each other's products and avoid redoing. Additionally, collaborators should be able to attract each other's attention to their products and processes as a method for initiating or enhancing collaborative work.

We focused on the use of recommendations as a distributed reminding process, through which analysts are reminded of other analysts' products that are related to the current workspace. This study showed that suggesting the relevant artefacts facilitated discovering them and helped participants to identify the key entities based on their occurrence in multiple pieces of analysis. Additionally, the Relevant-Artefacts panel provided context-based awareness of other analysts' activities. The study showed that public (unpublished) privacy level was used for sharing seemingly unimportant pieces or supplementary pieces. However, further research in realistic settings is needed to understand how enhancing control over attention and awareness affects sharing and reuse.

In sum, AnalyticStream enables the reuse process when users are not aware of the availability of useful shared artefacts or do not have a clear model of their information space to ask right questions. Moreover, AnalyticStream provides attention management mechanisms to allow collaborators share their products and decide about attracting others' attention to them. The design process and evaluation of AnalyticStream helped to develop an understanding of how the proposed approach to applying distributed cognition theory can be implemented in the design of collaborative systems. Future research is needed to further test the generalisability of our findings and the applicability of the proposed approach in different design situations.

ACKNOWLEDGMENTS

This work was supported by NSERC, and the GRAND NCE.

REFERENCES

1. Ackerman, M.S. and Halverson, C. Organizational Memory as Objects, Processes, and Trajectories: An Examination of Organizational Memory in Use. *Computer Supported Cooperative Work (CSCW) 13*, 2 (2004), 155-189.
2. Ackerman, M.S. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human–Computer Interaction 15*, 2 (2000), 179.
3. Ausubel, D.P., Novak, J.D., and Hanesian, H. *Educational psychology: A cognitive view*. Holt Rinehart and Winston, 1978.

4. Boyd, D. Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. *Convergence 14*, 1 (2008), 13-20.
5. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. *Proceedings of the 12th international conference on Intelligent user interfaces*, ACM (2007), 32-41.
6. Dourish, P. and Bellotti, V. Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, ACM (1992), 107-114.
7. Finkel, J.R., Grenager, T., and Manning, C. Incorporating non-local information into information extraction systems by Gibbs sampling. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics (2005), 363-370.
8. Fischer, G. Cognitive View of Reuse and Redesign. *Software, IEEE 4*, 4 (1987), 60-72.
9. Grinstein, G., Plaisant, C., Scholtz, J., and Whiting, M. IEEE VAST 2010 Challenge. 2010. http://hcil.cs.umd.edu/localphp/hcil/vast10/index.php.
10. Halverson, C.A. Activity Theory and Distributed Cognition: Or What Does CSCW Need to DO with Theories? *Comput. Supported Coop. Work 11*, 1-2 (2002), 243-267.
11. Heer, J. and Agrawala, M. Design Considerations for Collaborative Visual Analytics. *Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology*, IEEE (2007).
12. Heer, J., Mackinlay, J., Stolte, C., and Agrawala, M. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1189-1196.
13. Heer, J., Viégas, F.B., and Wattenberg, M. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM 52*, 1 (2009), 87-97.
14. Hollan, J., Hutchins, E., and Kirsh, D. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact. 7*, 2 (2000), 174-196.
15. Holmes, R. Unanticipated reuse of large-scale software features. *Proceedings of the 28th international conference on Software engineering*, (2006), 961–964.
16. Hutchins, E. *Cognition in the Wild*. The MIT Press, 1996.
17. Kadivar, N., Chen, V., Dunsmuir, D., et al. Capturing and supporting the analysis process. *IEEE Symposium on Visual Analytics Science and Technology*, (2009), 131-138.
18. Kirschner, P.A., Buckingham-Shum, S.J., and Carr, C.S. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer, 2003.
19. LaBerge, D. Attention, Awareness, and the Triangular Circuit. *Consciousness and Cognition 6*, 2-3 (1997), 149-181.
20. Lutters, W.G. and Ackerman, M.S. Beyond Boundary Objects: Collaborative Reuse in Aircraft Technical Support. *Computer Supported Cooperative Work (CSCW) 16*, 3 (2006), 341-372.
21. Marshall, C.C. Toward an ecology of hypertext annotation. *Proceedings of the ninth ACM conference on Hypertext and Hypermedia*, ACM (1998), 40–49.
22. Neisser, U. *Cognitive Psychology*. Prentice Hall, 1967.
23. Parent, W.A. Recent Work on the Concept of Privacy. *American Philosophical Quarterly 20*, 4 (1983), 341-355.
24. Paul, S.A. and Morris, M.R. CoSense: enhancing sensemaking for collaborative web search. *Proceedings of the 27th conference on Human factors in computing systems*, (2009), 1771–1780.
25. Pike, W., Bruce, J., Baddeley, B., et al. The Scalable Reasoning System: Lightweight visualization for distributed analytics. *Information Visualization 8*, 1 (2009), 71-84.
26. Robinson, A.C. and Weaver, C. Re-Visualization: Interactive Visualization of the Process of Visual Analysis. *Workshop on Visualization, Analytics & Spatial Decision Support*, (2006).
27. Schank, R.C. *Dynamic Memory Revisited*. Cambridge University Press, 1999.
28. Shrinivasan, Y.B., Gotzy, D., and Lu, J. Connecting the dots in visual analysis. *IEEE Symposium on Visual Analytics Science and Technology*, (2009), 123-130.
29. Simon, H.A. *The Sciences of the Artificial - 3rd Edition*. The MIT Press, 1996.
30. Simon, H. Designing Organizations for an Information-Rich World. In *Computers, communications, and the public interest*. 1971, 37-72.
31. Thomas, J.J. and Cook, K.A. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.
32. Viegas, F. and Wattenberg, M. Communication-Minded Visualization: A Call to Action. *IBM Systems Journal 45*, 4 (2006).
33. Viegas, F.B., Wattenberg, M., van Ham, F., Kriss, J., and McKeon, M. ManyEyes: a Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics 13*, (2007), 1121–1128.
34. Wright, W., Schroh, D., Proulx, P., Skaburskis, A., and Cort, B. The Sandbox for analysis: concepts and methods. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2006), 801-810.
35. i2 Analyst's Notebook. http://www.i2inc.com/products/analysts_notebook/.