Multimedia Summary Generation from Online Conversations: Current Approaches and Future Directions

Enamul Hoque and **Giuseppe Carenini**
University of British Columbia, Canada
{enamul,carenini}@cs.ubc.ca

Abstract

With the proliferation of Web-based social media, asynchronous conversations have become very common for supporting online communication and collaboration. Yet the increasing volume and complexity of conversational data often make it very difficult to get insights about the discussions. We consider combining textual summary with visual representation of conversational data as a promising way of supporting the user in exploring conversations. In this paper, we report our current work on developing visual interfaces that present multimedia summary combining text and visualization for online conversations and how our solutions have been tailored for a variety of domain problems. We then discuss the key challenges and opportunities for future work in this research space.

1 Introduction

Since the rise of social-media, an ever-increasing amount of conversations are generated every day. People engaged in asynchronous conversations such as blogs to exchange ideas, ask questions, and comment on daily life events. Often many people contribute to the discussion, which become very long with hundreds of comments, making it difficult for users to get insights about the discussion (Jones et al., 2004).

To support the user in making sense of human conversations, both the natural language processing (NLP) and information visualization (InfoVis) communities have independently developed different techniques. For example, earlier works on visualizing asynchronous conversations primarily investigated how to reveal the thread structure of a conversation using tree visualization techniques, such as using a mixed-model visualization to show both chronological sequence and reply relationships (Venolia and Neustaedter, 2003), thumbnail metaphor using a sequence of rectangles (Wattenberg and Millen, 2003; Kerr, 2003), and radial tree layout (Pascual-Cid and Kaltenbrunner, 2009). However, such visualizations did not focus on analysing the actual content (i.e., the text) of the conversations.

On the other hand, text mining and summarization methods for conversations perform content analysis of the conversations, such as what topics are covered in a given text conversation (Joty et al., 2013b), along with what opinions the conversation participants have expressed on such topics (Taboada et al., 2011). Once the topics, opinions and conversation structure (e.g., reply-relationships between comments) are extracted, they can be used to summarize the conversations (Carenini et al., 2011).

However, presenting a static/non-interactive textual summary alone is often not sufficient to satisfy the user information needs. Instead, generating a multimedia output that combines text and visualizations can be more effective, because the two can play complementary roles: while visualization can help the user to discover trends and relationship, text can convey key points about the results, by focusing on temporal, causal and evaluative aspects.

In this paper, we present a visual text analytics approach that combines both text and visualization to helps users in understanding and analyzing online conversations. We provide an overview of our approach to multimedia summization of online conversations followed by how our generic solutions have been tailored to specific domain problems (e.g., supporting users of a community question answering forum) . We then discuss further
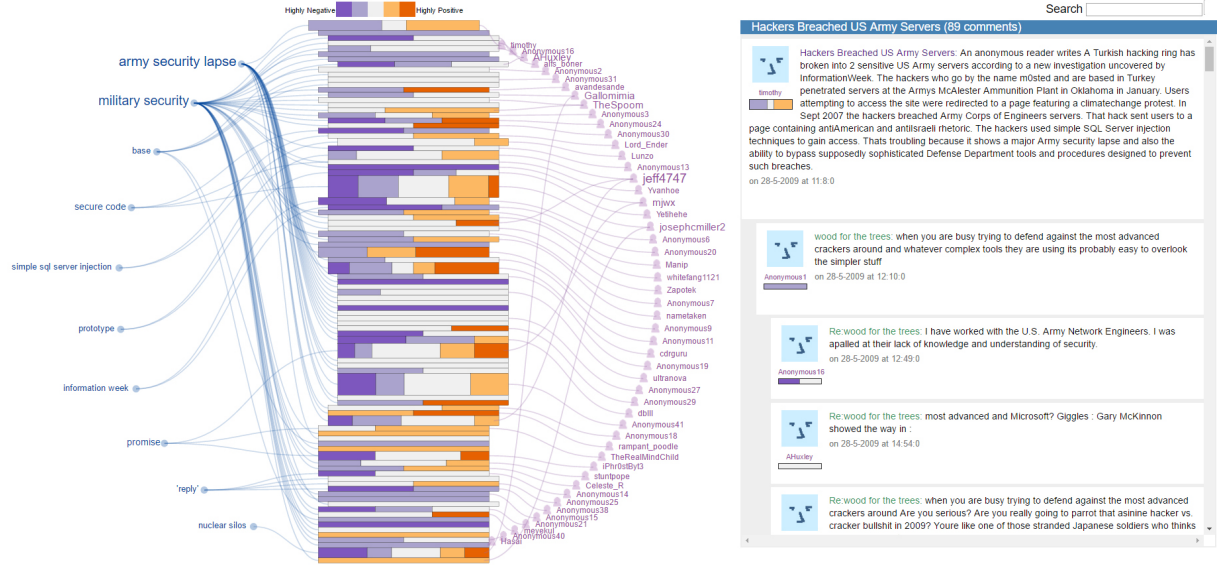
Figure 1: The ConVis interface: The Thread Overview visually represents the whole conversation encoding the thread structure and how the sentiment is ex-pressed for each comment(middle); The topics and authors are arranged circularly around the Thread Overview; and the Conversation View presents the detailed comments in a scrollable list (right).

challenges, open questions, and ideas for future work in the research area of multimedia summarization for online conversations.

2 Multimedia Summarization of Online Conversations

2.1 Our Approach

To generate multimedia summary for online conversation, our primary approach was to apply human-centered design methodologies from the InfoVis literature (Munzner, 2009; Sedlmair et al., 2012) to identify the type of information that needs to be extracted from the conversation as well as to inform the design of the visual encodings and interaction techniques.

Following this approach, we proposed a system that creates a multimedia summary and supports users in exploring a single asynchronous conversation (Hoque and Carenini, 2014, 2015). The underlying topic modeling approach groups the sentences of a blog conversation into a set of topical segments. Then, representative key phrases are assigned to each of these segments (labeling). We adopt a novel topic modeling approach that captures finer level conversation structure in the form of a graph called Fragment Quotation Graph (FQG) (Joty et al., 2013b). All the distinct fragments (both new and quoted) within a conversation are extracted as the nodes of the FQG. Then

the edges are created to represent the replying relationship between fragments. If a comment does not contain any quotation, then its fragments are linked to the fragments of the comment to which it replies, capturing the original 'reply-to' relation.

The FQG is exploited in both topic segmentation and labeling. In segmentation, each path of the FQG is considered as a separate conversation that is independently segmented (Morris and Hirst, 1991). Then, all the resulting segmentation decisions are consolidated in a final segmentation for the whole conversation. After that, topic labeling generates keyphrases to describe each topic segment in the conversation. A novel graph based ranking model is applied that intuitively boosts the rank of keyphrases that appear in the initial sentences of the segment, and/or also appear in text fragments that are central in the FQG (see (Joty et al., 2013b) for details).

While developing the system, we started with a user requirement analysis for the domain of blog conversations to derive a set of design principles. Based on these principles, we designed an overview+detail interface, named ConVis that provides a visual overview of a conversation by presenting topics, authors and the thread structure of a conversation (see Figure 1). Furthermore, it provides various interaction techniques such as brushing and highlighting based on multiple facets to

support the user in exploring and navigating the conversation.

We performed an informal user evaluation, which provides anecdotal evidence about the effectiveness of ConVis as well as directions for further design. The participants' feedback from the evaluation suggests that ConVis can help the user to identify the topics and opinions expressed in the conversation; supporting the user in finding comments of interest, even if they are buried near the end of the thread. The informal evaluation also reveals that in few cases the extracted topics and opinions are incorrect and/or may not match the mental model and information needs of the user.

In subsequent work, we focused on supporting readers in exploring a collection of conversations related to a given query (Hoque and Carenini, 2016). Exploring topics of interest that are potentially discussed over multiple conversations is a challenging problem, as the volume and complexity of the data increases. To address this challenge, we devised a novel hierarchical topic modeling technique that organizes the topics within a set of conversations into multiple levels, based on their semantic similarity. For this purpose, we extended the topic modeling approach for a single conversation to generate a topic hierarchy from multiple conversations by considering the specific features of conversations. We then designed a visual interface, named MultiConVis that presents the topic hierarchy along with other conversational data, as shown Figure 2. The user can explore the data, starting from a possibly large set of conversations, then narrowing it down to the subset of conversations, and eventually drilling-down to the set of comments belonging to a single conversation.

We evaluated MultiConVis through both case studies with domain experts and a formal user study with regular blog readers. Our case studies demonstrate that the system can be useful in a variety of contexts of use, while the formal user study provides evidence that the MultiConVis interface supports the user's tasks more effectively than traditional interfaces. In particular, all our participants, both in the case studies and in the user study, appear to benefit from the topic hierarchy as well as the high-level overview of the conversations. The user study also shows that the MultiConVis interface is significantly more useful than the traditional interface, enabling the user to find insightful comments from thousands of comments, even when they were scattered across multiple conversations, often buried down near the end of the threads. More importantly, MultiConVis was preferred by the majority of the participants over the traditional interface, suggesting the potential value of our approach for combining NLP and InfoVis.

2.2 Applications

Since our visual text analytics systems have been made publicly available, they have been applied and tailored for a variety of domain problems, both in our own work as well as in other research projects. For example, we conducted a design study in the domain of community question answering (CQA) forums, where our generic solutions for combining NLP and InfoVis were simplified and tailored to support information seeking tasks for a user population possibly having low visualization expertise (Hoque et al., 2017). In addition to our work, several other researchers have applied or partially adopted the data abstractions and visual encodings of MultiConVis and ConVis in a variety of domains, ranging from news comments (Riccardi et al., 2015), to online health forums (Kwon et al., 2015), to educational forums (Fu et al., 2017). We now analyze these recent works and discuss similarities and differences with our systems.

News comments: SENSEI[1] is a research project that was funded by the European Union and was conducted in collaboration with four leading universities and two industry partners in Europe. The main goal of this project was to develop summarization and analytics technology to help users make sense of human conversation streams from diverse media channels, ranging from comments generated for news articles to customer-support conversations in call centers.

After the research work on developing ConVis was published and the tool was made publicly available, the SENSEI project researchers expressed their interest in adopting our system. Their primary objective was to evaluate their text summarization and analytics technology by visualizing the results with ConVis, with the final goal of detecting end-user improvements in task performance and productivity.

In their version of the interface[2], they kept the main features of ConVis, namely the topics, au-

---

[1]www.sensei-conversation.eu

[2]A video demo of their version of the interface is available
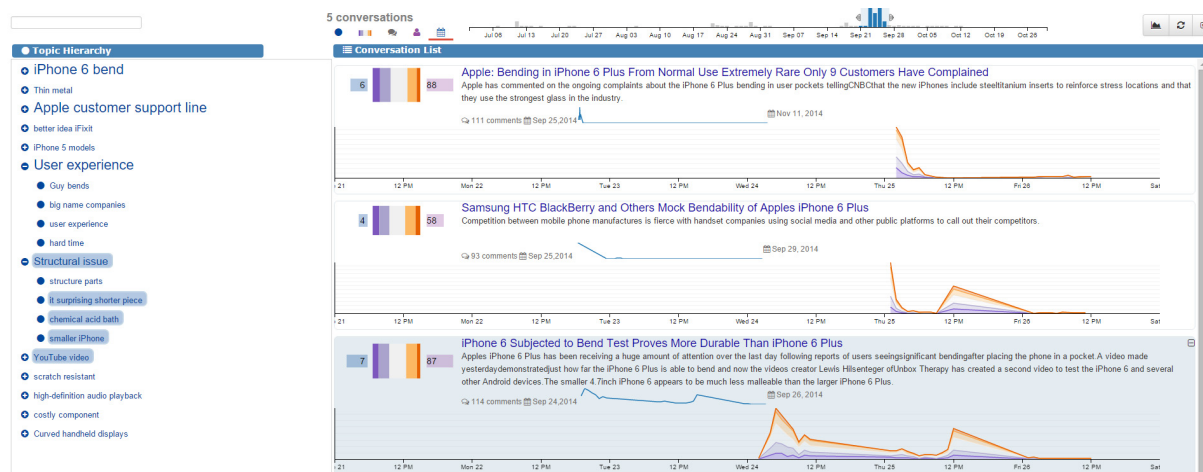
Figure 2: The MultiConVis interface. Here, the user filtered out some conversations from the list using the Timeline located at the top, and then hovered on a conversation item (highlighted row in the right). As a consequence, the related topics from the Topic Hierarchy were highlighted (left).

thors, and thread overview; and then added some new features to show text analytics results specific to their application, as shown in Figure 3 (Riccardi et al., 2015). In particular, within the thread overview, for each comment they encoded how much this comment agrees or disagrees with the original article, instead of showing the sentiment distribution of that comment. Another interactive feature they introduced was that clicking on an author element results in showing the predicted mood of that author (using five different mode types, i.e., amused, satisfied, sad, indignant, and disappointed). Furthermore, they added a summary view that shows a textual summary of the whole conversation in addition to the detailed comments. Finally, they introduced some new interactive features, such as zooming and filtering to deal with conversations that are very long with several hundreds of comments.

Online health forums: Kwon et al. developed VisOHC (Kwon et al., 2015), a visual analytics system designed for administrators of online health communities (OHCs). In this paper, they discuss similarities and differences between VisOHC and ConVis. For instance, similar to the thread overview in ConVis, they represented the comments of a conversation using a sequence of rectangles and used the color encoding within those rectangles to represent sentiment (see Figure 4). However, they encoded additional data in order to support the specific domain goals and tasks of OHC administrators. For instance, they

at www.youtube.com/watch?v=XIMP0cuiZIQ

used a scatter plot to encode the similarities between discussion threads and a histogram view to encode various statistical measures regarding the selected threads, as shown in Figure 4.

Mamykina et al. analyzed how users in online health communities collectively make sense of the vast amount of information and opinions within an online diabetes forum, called TuDiabetes (Mamykina et al., 2015). Their study found that members of TuDiabetes often value a multiplicity of opinions rather than consensus. From their study, they concluded that in order to facilitate the collective sensemaking of such diversity of opinions, a visual text analytics tool like ConVis could be very effective. They also mentioned that in addition to topic modeling and sentiment analysis, some other text analysis methods related to their health forum under study, such as detection of agreement and topic shift in conversation, should be devised and incorporated into tools like ConVis.

Educational forums: More recently, Fu et al. presented iForum, an interactive visual analytics system for helping instructors in understanding the temporal patterns of student activities and discussion topics in a MOOC forum (Fu et al., 2017). They mentioned that while the design of iForum has been inspired by tools such as ConVis, they have tailored their interface to the domain-specific problems of MOOC forums. For instance, like ConVis, their system provides an overview of topics and discussion threads, however, they focused more on temporal trends of an entire forum, as op-
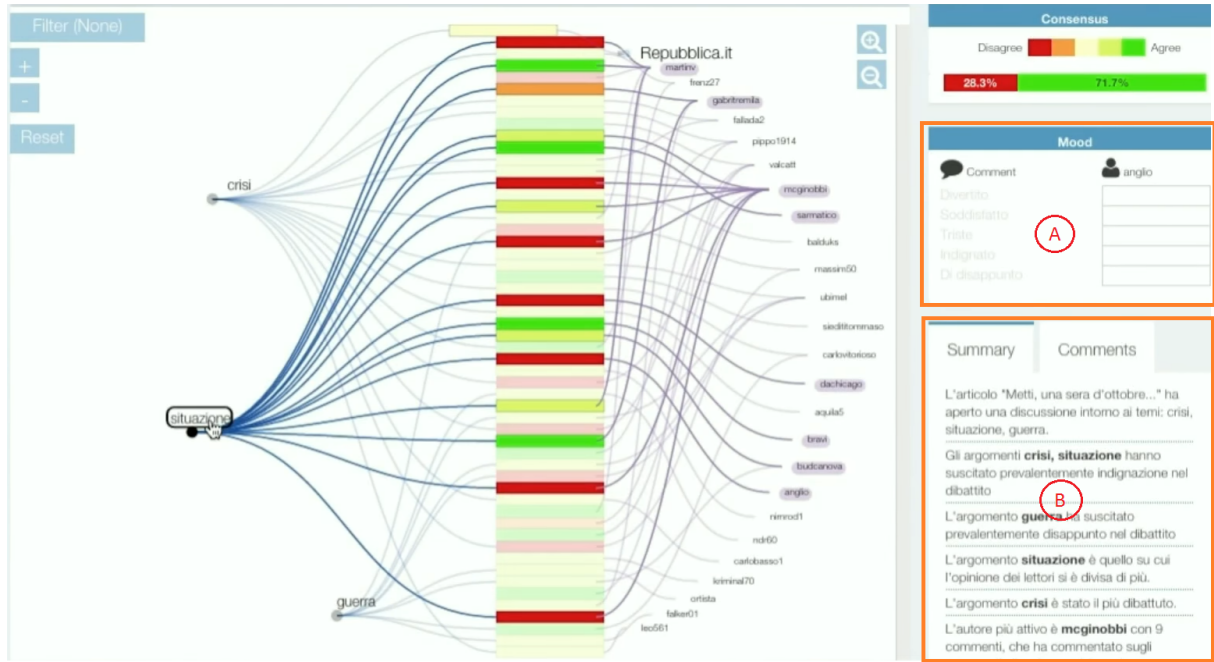
Figure 3: A screenshot of the modified ConVis interface used in the SENSEI project. The interface shows the results of some additional text analysis methods, namely the degree of agreement/disagreement between a comment and the original article (within the thread overview), the predicted mood of the corresponding author (A), and the textual summary of the conversation (B) (Riccardi et al., 2015).
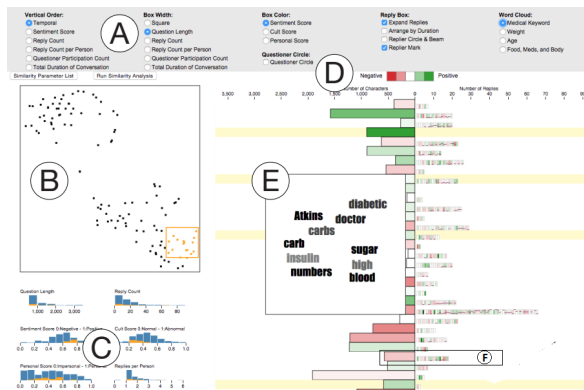

Figure 4: VisOHC visually represents the comments of a conversation using a sequence of rectangles (F), where color within each rectangle represents sentiment expressed in a comment. Additionally it shows a scatter plot (B), and a histogram view (C) (The figure is adapted from (Kwon et al., 2015)).

posed to an individual conversation or a set of conversations related to a specific query.

3 Challenges and Future Directions

While our approach to combining NLP and Info-Vis to generate multimedia summaries has made some significant progress in supporting the ex-

ploration and analysis of online conversations, it also raises further challenges, open questions, and ideas for future work. Here we discuss the key challenges and opportunities for future research.

How can we provide more high-level summary to users? In our current systems, we used the results from topic modeling which can be viewed as crud summary of conversations, because each topic is simply summarized by a phrase label and the labels are not combined in a coherent discourse. Based on the tasks of real users we identified the need for higher level summarization. For instance, users may benefit from a more high-level abstract human-like summary of conversations, where the content extracted from the conversations is organized in a sequence of coherent sentences.

Similarly, during our evaluations some users found the current sentiment analysis insufficient in revealing whether a comment is supporting/opposing a preceding one. It seems that opinion seeking tasks (e.g., 'why people were supporting or opposing an opinion?') would require the reader to know the argumentation flow within the conversation, namely the rhetorical structure of each comment (Joty et al., 2013a) and how these structures are linked to each other.

An early work (Yee and Hearst, 2005) attempted to organize the comments using a treemap like layout, where the parent comment is placed on top as a text block and the space below the parent node is divided between supporting and opposing statements. We plan to follow this idea in ConVis, but incorporating a higher level discourse relation analysis of the conversations along with the detection of controversial topics (Allen et al., 2014).

How can we scale up our systems for big data? As social media conversational data is growing in size and complexity at an unprecedented rate, new challenges have emerged from both the computational and the visualization perspectives. In particular, we need to address the following aspects of big data, while designing visual text analytics for online conversations.

Volume: Most of the existing visualizations are inadequate to handle very large amounts of raw conversational data. For example, ConVis scales with conversations with hundreds of comments; however, it is unable to deal with a very long conversation consisting of more than a thousand comments. To tackle the scalability issue, we will investigate computational methods for filtering and aggregating comments, as well as devise interactive visualization techniques such as zooming to progressively disclose the data from a high-level overview to low-level details.

Velocity: The systems that we have developed do not process streaming conversations. Yet in many real-world scenarios, conversational data is constantly produced at a high rate, which poses enormous challenges for mining and visualization methods. For instance, immediately after a product is released a business analyst may want to analyze text streams in social media to identify problems or issues, such as whether customers are complaining about a feature of the product. In these cases, timely analysis of the streaming text can be critical for the company's reputation. For this purpose, we aim to investigate how to efficiently mine and summarize streaming conversations (tre, 2017) and how to visualize the extracted information in real time to the user (Keim et al., 2013).

How can we leverage text summarization and visualization techniques to develop advanced storytelling tools for online conversations? Data storytelling has become increasingly popular among InfoVis practitioners such as journalists, who may want to create a visualization from social media conversations and integrate it into their narratives to convey critical insights. Unfortunately, even sophisticated visualization tools like Tableau [3] offer only limited support for authoring data stories, requiring users to manually create textual annotations and organize the sequence of visualizations. More importantly, they do not provide methods for processing the unstructured or semi-structured data generated in online conversations.

In this context, we aim to investigate how to leverage NLP and InfoVis techniques for online conversations to create effective semi-automatic authoring tools for data storytelling. More specifically, we need to devise methods for generating and organizing the summary content from online conversations and choosing the sequence in which such content is delivered to users. To this end, a starting point could be to investigate current research on narrative visualization (Segel and Heer, 2010; Hullman and Diakopoulos, 2011).

How can we support the user in tailoring our systems to a specify conversational genre, a specific domain, or tasks? In the previous section, we already discussed how our current visual text analytics systems have been applied and tailored to various domains. However, in these systems, the user does not have flexibility in terms of the choice of the datasets and the available interaction techniques. Therefore, it may take a significant amount of programming effort to re-design the interface for a specific conversational domain. For example, when we tailored our system to a community question answering forum with a specific user population in mind, we had to spend a considerable amount of time modifying the existing code in order to re-design the interface for the new conversational genre.

In this context, can we enable a large number of users - not just those who have strong programming skills to author visual interfaces for exploring conversations in a new domain? To answer this question, we need to research how to construct an interactive environment that supports custom visualization design for different domains without requiring the user to write any code. Such interactive environment would allow the user to have more control over the data to be represented and the interactive techniques to be supported.

---

[3]www.tableau.com

To this end, we will investigate current research on general purpose visual authoring tools such as Lyra (Satyanarayan and Heer, 2014) and IVisDesigner (Ren et al., 2014), which provide custom visualization authoring environments, to understand how we can build a similar tool, but specifically for conversational data.

How can the system adapt to a diverse range of users? A critical challenge of introducing a new visualization is that the effectiveness of visualization techniques can be impacted by different user characteristics, such as visualization expertise, cognitive abilities, and personality traits (Conati et al., 2014). Unfortunately, most previous work has focused on finding individual differences for simple visualizations only, such as bar and radar graphs (Toker et al., 2012). It is still unknown how individual differences might impact the reading ability of multimedia summary that requires coordinations between text and visualization. In this regard, we need to examine what aspects of a multimedia output are impacted by user characteristics and how to dynamically adapt the system to such characteristics.

4 Conclusions

Multimedia summarization of online conversations is a promising approach for supporting the exploration of online conversations. In this paper, we present our current work on generating multimedia summaries combining text and visualization. We also discuss how our research has influenced the subsequent work in this research space. We believe that by addressing the critical challenges and research questions posed in the paper, we will able to support users in understanding online conversations more efficiently and effectively.

References

2017. TREC real-time summarization track (accessed June 05, 2017). http://trecrts.github.io/.

Kelsey Allen, Giuseppe Carenini, and Raymond T Ng. 2014. Detecting disagreement in conversations using pseudo-monologic rhetorical structure. In *Proceedings of the Empirical Methods on Natural Language Processing (EMNLP)*.

Giuseppe Carenini, Gabriel Murray, and Raymond Ng. 2011. *Methods for Mining and Summarizing Text Conversations*. Morgan Claypool.

C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker. 2014. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Computer Graphics Forum (Proceedings of EuroVis)* 33(3):371–380.

Siwei Fu, Jian Zhao, Weiwei Cui, and Huamin Qu. 2017. Visual analysis of MOOC forums with iForum. *IEEE Transactions on Visualization and Computer Graphics (Prooceedings of VAST)* 23(1):201–210.

Enamul Hoque and Giuseppe Carenini. 2014. ConVis: A visual text analytic system for exploring blog conversations. *Computer Graphics Forum (Proceedings EuroVis)* 33(3):221–230.

Enamul Hoque and Giuseppe Carenini. 2015. ConVisIT: Interactive topic modeling for exploring asynchronous online conversations. In *Proceedings ACM conference on Intelligent User Interfaces (IUI)*. pages 169–180.

Enamul Hoque and Giuseppe Carenini. 2016. MultiConVis: A visual text analytics system for exploring a collection of online conversations. In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI)*. pages 96–107.

Enamul Hoque, Shafiq Joty, Màrquez Lluís, and Giuseppe Carenini. 2017. CQAVis: Visual text analytics for community question answering. In *Proceedings of the ACM conference on Intelligent User Interfaces (IUI)*. pages 161–172.

Jessica Hullman and Nick Diakopoulos. 2011. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 17(12):2231–2240.

Quentin Jones, Gilad Ravid, and Sheizaf Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research* 15(2):194–210.

Shafiq Joty, Giuseppe Carenini, Raymond Ng, and Yashar Mehdad. 2013a. Combining intra-and multi-sentential rhetorical parsing for document-level discourse analysis. In *Proceedings of the Annual Meeting on Association for Computational Linguistics*.

Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. 2013b. Topic segmentation and labeling in asynchronous conversations. *Journal of Artificial Intelligence Research* 47:521–573.

Daniel A Keim, Miloš Krstajic, Christian Rohrdantz, and Tobias Schreck. 2013. Real-time visual analytics of text data streams. *IEEE Computer* 46(7):47–55.

Bernard Kerr. 2003. Thread arcs: An email thread visualization. In *IEEE Symposium on Information Visualization*. pages 211–218.

Bum Kwon, Sung-Hee Kim, Sukwon Lee, Jaegul Choo, and Ji Yi Jina Huh. 2015. Visohc: Designing visual analytics for online health communities. *IEEE Transactions on Visualization and Computer Graphics* .

Lena Mamykina, Drashko Nakikj, and Noemie Elhadad. 2015. Collective sensemaking in online health forums. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pages 3217–3226.

Jane Morris and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics* 17(1):21–48.

Tamara Munzner. 2009. A nested model for visualization design and validation. *Transactions on Visualization and Computer Graphics (Proocedings of InfoVis)* 15(6):921–928.

Víctor Pascual-Cid and Andreas Kaltenbrunner. 2009. Exploring asynchronous online discussions through hierarchical visualisation. In *IEEE Conference on Information Visualization*. pages 191–196.

Donghao Ren, Tobias Höllerer, and Xiaoru Yuan. 2014. iVisDesigner: Expressive interactive design of information visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 20(12):2092–2101.

Giuseppe Riccardi, A R Celli Balamurali, Favre Benoit Fabio, Ferrante Carmelo, Adam Funk, Rob Gaizauskas, and Vincenzo Lanzolla. 2015. Report on the summarization views of the sensei prototype. In *Technical report*.

Arvind Satyanarayan and Jeffrey Heer. 2014. Lyra: An interactive visualization design environment 33(3):351–360.

Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design study methodology: reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics* 18(12):2431–2440.

Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics* 16(6):1139–1148.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics* 37(2):267–307.

Dereck Toker, Cristina Conati, Giuseppe Carenini, and Mona Haraty. 2012. Towards adaptive information visualization: on the influence of user characteristics. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, pages 274–285.

Gina Danielle Venolia and Carman Neustaedter. 2003. Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pages 361–368.

Martin Wattenberg and David Millen. 2003. Conversation thumbnails for large-scale discussions. In *Extended Abstract Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pages 742–743.

Ka-Ping Yee and Marti Hearst. 2005. Content-centered discussion mapping. *Online Deliberation 2005/DIAC-2005* .