Interactive Assigning of Conference Sessions with Visualization and Topic Modeling

Yun Han[1] * Zhenhuang Wang[1] † Siming Chen[4,5] ‡ Guozheng Li[1] Xiaolong Zhang[6] §
Xiaoru Yuan[1,2,3]

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) National Engineering Laboratory for Big Data Analysis Technology and Application, Beijing, China
3) Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University
4) Fraunhofer IAIS, Sankt Augustin, Germany 5) University of Bonn, Bonn, Germany
6) College of Information Sciences and Technology, Pennsylvania State University, USA

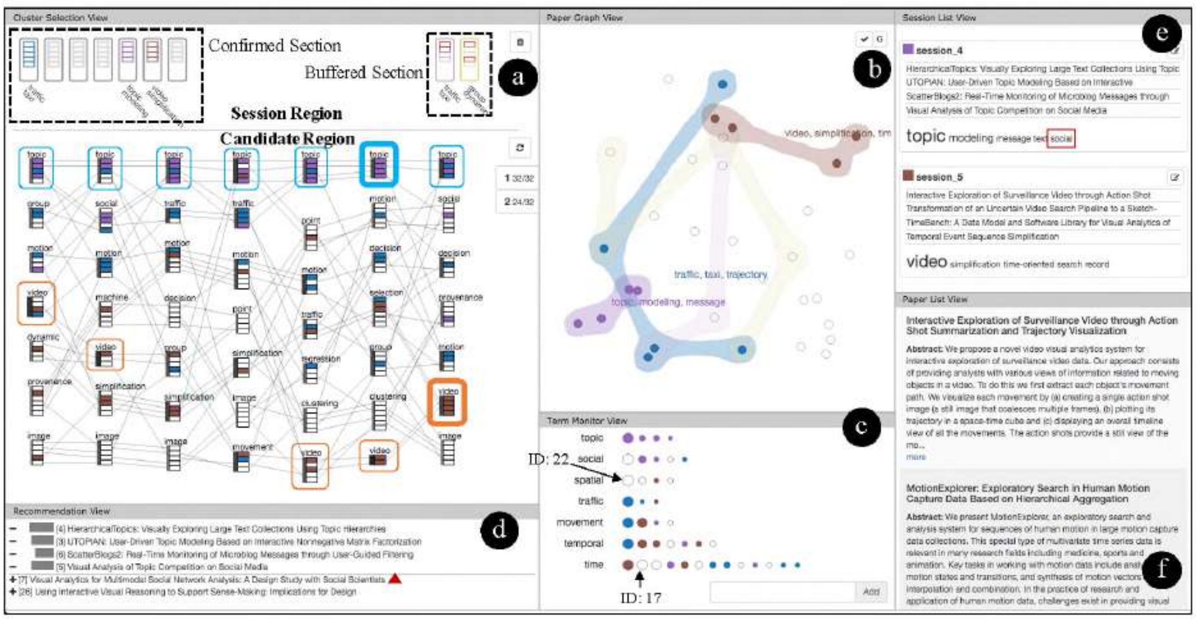

Figure 1: Interface of the system: (a) Cluster Selection View, providing users with the situational awareness of the paper selection process. (b) Paper Graph View, representing an overview of papers with different topics. (c) Term Monitor View, interactive analyzing the papers according to keywords. (d) Recommendation View, recommending related papers. (e) Session List View and (f) Paper List View, showing the raw data of generated sessions and papers.

Abstract

Creating thematic sessions based on accepted papers is important to the success of a conference. Facing a large number of papers from multiple topics, conference organizers need to identify the topics of papers and group them into sessions by considering the constraints on session numbers and paper numbers in individual sessions. In this paper, we present a system using visualization and topic modeling to help the construction of conference sessions. The system provides multiple automatically generated session schemes and allows users to create, evaluate, and manipulate paper sessions with given constraints. A case study based on our system on the VAST papers shows that our method can help users successfully construct coherent conference sessions. In addition to conference session management, our method can be extended to other tasks, such as event and class schedule.

Keywords: Interactive clustering visualization, Visual analytics, Topic modeling, Conference scheduling, Constraint solving

1 Introduction

While organizing academic conferences, program committee chairs need to make decisions on paper sessions. Creating thematic and proper sessions can improve the experience for attendees by making it easier for them to choose topics of talks rather than individual papers, and attend the talks on papers of interest more coherently. Recently, researchers have explored designs to support the construction of conference sessions. For example, tools have been provided to address concerns with scheduling constraints [13] and the social dynamics among program committee members [5].

However, existing research largely leaves the selection of papers to program committee members. For example, when using Frenzy [5] to determine sessions in big conferences like ACM CHI, associate chairs have to rely on their domain knowledge and experience in deciding what topics a paper is about, what candidating

*e-mail: {yunhan, guozheng.li, xiaoru.yuan} @pku.edu.cn. Xiaoru Yuan is the corresponding author.
†e-mail: wangzhenhuang.zeek@gmail.com
‡e-mail: siming.chen@iais.fraunhofer.de
§e-mail: lzhang@ist.psu.edu

2020 IEEE Pacific Visualization Symposium (PacificVis)
978-1-7281-5697-2/20/$31.00 ©2020 IEEE

2020 IEEE Pacific Visualization Symposium (PacificVis) 978-1-7281-5697-2/20/$31.00 ©2020 IEEE 10.1109/PacificVis48177.2020.1027x

sessions the paper may belong to, and what session the paper eventually goes to. For large conferences with hundreds of accepted papers, it could be a daunting task for program committee members to extract topic information of individual papers and then to group them based on topics. Committee members must consider the size constraint on each session, multiple possible ways to set up sessions and the conflicts among sessions.

In this paper, we introduce a system to address these challenges in constructing conference sessions. Our approach uses interactive visualization and topic modeling techniques, which can help users understand and manage the topics of papers and turn topic-based candidate paper groups into paper sessions by considering certain constraints (e.g., the minimum and maximum numbers of papers in each cluster). Our system can also provide users with several options for conference sessions based on various possibilities of paper-topic clusters. It allows users to explore, analyze and adjust paper clusters, i.e., sessions, and interactively decide what conference sessions to have and which papers each session could include. Finally, we present an exploration case to show how our system can help a user iteratively construct conference sessions based on papers accepted by a real conference, and we also compare the result produced by our system with the sessions of the conference.

2 Related Work

Our research concerns several areas, including conference session scheduling systems, visualization of topics generated by topic modeling algorithms, and interactive data clustering. Our review here focuses on these three areas.

2.1 Conference session scheduling systems

In most conferences, grouping accepted papers into sessions still requires the active involvement of people. Generally, program chairs organize discussions with committee members to create a preliminary session partition scheme and then refine it to satisfy constraints. To utilize knowledge of committee members and improve the efficiency of grouping, Weld et al. [5] proposed a system called Frenzy to support collaborative session construction. Frenzy divided the process of session construction into two stages, a meta-data elicitation stage and a global constraint satisfaction stage. It collects suggestions of categories from a large group of experts collaboratively in a crawled-sourced way firstly and then refines sessions to solve the length constraint of sessions. Kim et al. [13] consider the crowd-sourced suggestions from both program committee members and authors. In addition, an Attendee-Sourcing technique [2] is introduced to gather inputs from potential conference attendees and consider them as preferences in the process of creating sessions.

2.2 Visualization of topics in topic modeling

Papers in a session often have similar topics, and our research uses topic modeling to help cluster papers. Topic modeling can be regarded as a soft clustering algorithm and visualization has been used to support topic analysis. The Topic Browser [12] provided an interactive tool to browse the output of topic models. Termite [7] visualizes the relationship between the layers of words and topics, while Serendip [1] allows users to access all cross-layer relationships. To better display the topics, mapping documents to a two-dimension plane is a frequently used method [6, 14]. In this way, similar documents are placed close and the distribution of topics is obvious in the graph. Topics can be organized in a hierarchical structure, Visualization designs have been developed to help people understand and analyze such topic hierarchies [9, 11].

2.3 Interactive visual clustering

Automatic clustering algorithms may not provide satisfying results, so allowing people to involve in data clustering processes could be a solution to produce better results. IVIBRATE [4], an interactive visualization framework was proposed for clustering large datasets, and this framework enabled users to be involved in large-scale clustering processes. Interactive Visual Clustering [10] used force-directed layouts to display data and allow the users to move data nodes between different blocks and to edit the clusters. Recently, as latent dirichlet allocation(LDA) is widely used in large-scale text data sets, some works focus on allowing human users to influence the clustering model. iVisClustering [14] provided a visual analytics system to perform document clustering interactively based on the output of LDA model. iVisClustering offered multiple forms of visualization and rich interaction so that users can conduct an in-depth analysis of the clustering result and interactively refine the model in various ways. For the same purpose, UTOPIAN [6] enabled users to interact with topic modeling methods and improved the result by keyword refinement, topic merging/splitting, and topic creation.

3 Method

In this section, we discuss the design goals based on our problem definition, and present the system pipeline designed to achieve these goals.

3.1 Design Requirements

The target of this research is to help conference program organizers understand the topics of papers, group papers based on topic relevance, and create conference sessions with appropriate numbers of papers in each session, so we identify the following design goals.

- **D1. Helping program chairs understand the topics of papers.** Deciding the topics and themes of paper is often a time-consuming process.
- **D2. Providing program chairs with multiple potential session arrangements.** Given a set of papers, there are several possible session setups.
- **D3. Helping program chairs conduct an in-depth analysis of the topic relevance of individual papers.** Papers may be related in different ways. The system should enable program chairs to figure out how papers may be related easily and show the distance among papers from the perspective of topics.
- **D4. Supporting the management of sessions through interactive visualization.** To create and modify sessions, program chairs need to manipulate objects associated with sessions, papers, and topics.

3.2 System Pipeline

Based on the design goals we proposed above, we designed a web-based visual analytics system to help conference organizers construct conference sessions. Our system is a user-centered system that integrates automatic suggestions with interactive visualization. Figure 2 illustrates the pipeline. In our method, paper metadata are first processed by natural language processing. Metadata of the papers include the key attributes of papers, such as paper titles, abstracts, keywords, authors, and affiliations. Among them, keywords are no required but we can improve the automatic clustering result with them. Firstly, the title, abstract and keywords of a paper are merged as the original text material. To reduce the impacts of common words, such as "interface", "user" in the HCI domain, users can define a set of stop words. After stop-word removal and lemmatization, text data are sent to the topic modeling module, implemented with LDA [3].

In the on-line system, users should first specify the number of sessions and the numbers of papers in these sessions, which may be different for different conferences. Based on these parameters, processed paper data will be sent to a topic modeling module to

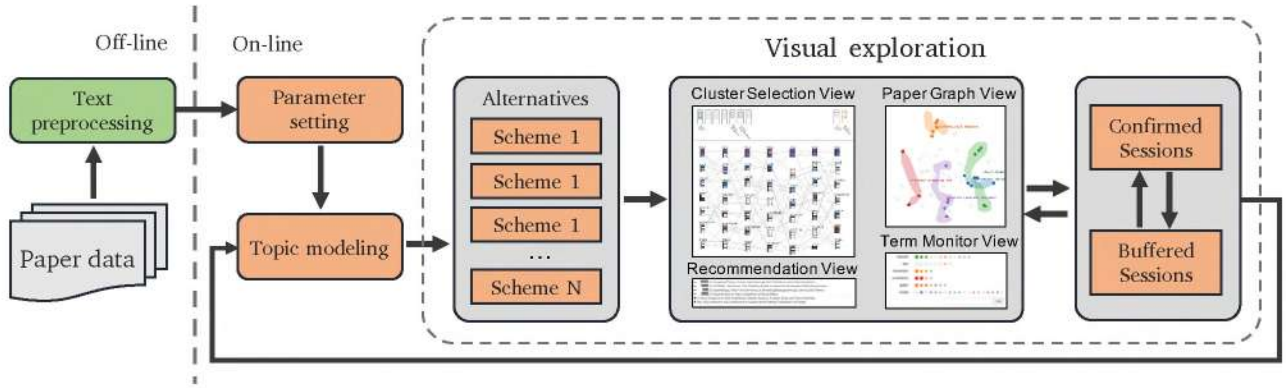

Figure 2: Visual analytics system pipeline. It enables users to iteratively select alternatives session partitions from topic modeling, construct sessions and solve conflicts with multiple coordinated visualization interfaces.

generate topics of each paper and produce several multiple partition schemes (D2). With the information on paper topics, users can analyze papers and construct sessions (D1). With visualized and pre-computed alternative schemes, users start their visual exploration of sessions, paper topics, and the relationships among papers, sessions, and topics (D3). Our system provides multiple linked views to help organize conference sessions (D4).

4 Interactive Visual Analytics System

In this section, we will describe the various views in the system and the details of the system implementation.

4.1 System Views

As illustrated in Figure 1, the user interface of our system consists of six coordinated views and all these views are interactive and linked with each other.

4.1.1 Cluster Selection View

Cluster Selection View provides users with the situational awareness of the paper selection process. It is divided into two regions: candidate region and session region (Figure 1(a)).

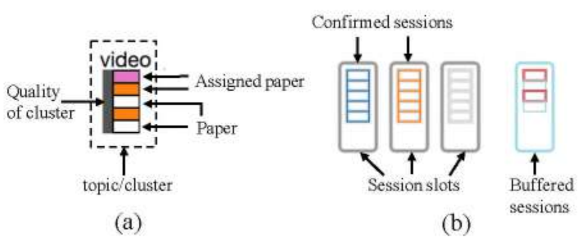

Figure 4: Visual encodings of the Cluster Selection View. (a) A paper session in the candidate region. (b) Confirmed and buffered sessions in the session region.

In the candidate region, each column represents a scheme of partition from topic modeling. In each column, there are several blocks representing topics in topic models (Figure 4(a)). On the left of each block is a bar with different width and color from deep to light gray, indicating the quality of the corresponding topic. Cells in each block represent papers. Cells of the same paper in adjacent columns are linked with lines so that users can see where individual papers are in different partition schemes. The session region has two sections, confirmed section and buffered section (Figure 4(b)). In the confirmed section, several slots for sessions are lined up according to a user's configuration. When the user tries to add a session, the operation is allowed only if there is an available slot that meets the size constraint, and there is no conflict with other confirmed sessions. Otherwise, the session is added to the buffer section. The conflicted papers that appear in multiple sessions are highlighted in red for warning. The user can add, remove or move papers in the buffer.

4.1.2 Paper Graph View

Our system visualizes papers with a node-link diagram style to represent an overview of papers (Figure 1(b)). Each dot represents a paper, the color of which encodes its current session. The view uses t-distributed stochastic neighborhood embedding (t-SNE) as the layout algorithm, which can generate a 2D layout along with a given distance matrix [15]. In this view, users can examine the relationship among papers based on the topic modeling outputs. In addition, to visualize the distribution of paper in each session, we provide a bubble set [8] visualization. The contour of each bubble with a unique color indicates the region of the selected session. We expect the paper within the same session bubble should be distributed closely.

4.1.3 Term Monitor View

To gain the idea of selected paper's topic distribution, we provide the Term Monitor View (Figure 1(c)). For each keyword selected by the users, we extract the paper containing it and visualize it as a circle. The y-axis is the keyword axis, the order of which indicates the selecting order by default. Users can re-order them by multiple criteria, including paper count, fitness, etc. The x-axis is laid out for each paper, and the size of the paper circle encodes the occurring time of the keywords. The color of the circle encodes the currently selected topics the paper belongs to. If it belongs to multiple topics, the circle becomes a pie chart with multiple colors. The interaction supported in the view is to enhance the understanding of the current division of the papers.

4.1.4 Recommendation View

Recommendation View helps program chairs conduct an in-depth analysis of the session and access potential relevant papers, as shown in Figure 1(d). The view lists papers in the currently selected session, where each paper has a bar to indicate the similarity between the paper with other papers in this session. Here, the similarity is measured by TF-IDF and cosine similarity. We provide buttons of "+" and "-" to allow users to directly edit the session.

4.1.5 Session List View & Paper List View

Session List View (Figure 1(e)) displays detailed information of currently confirmed sessions. With this view, the user can better

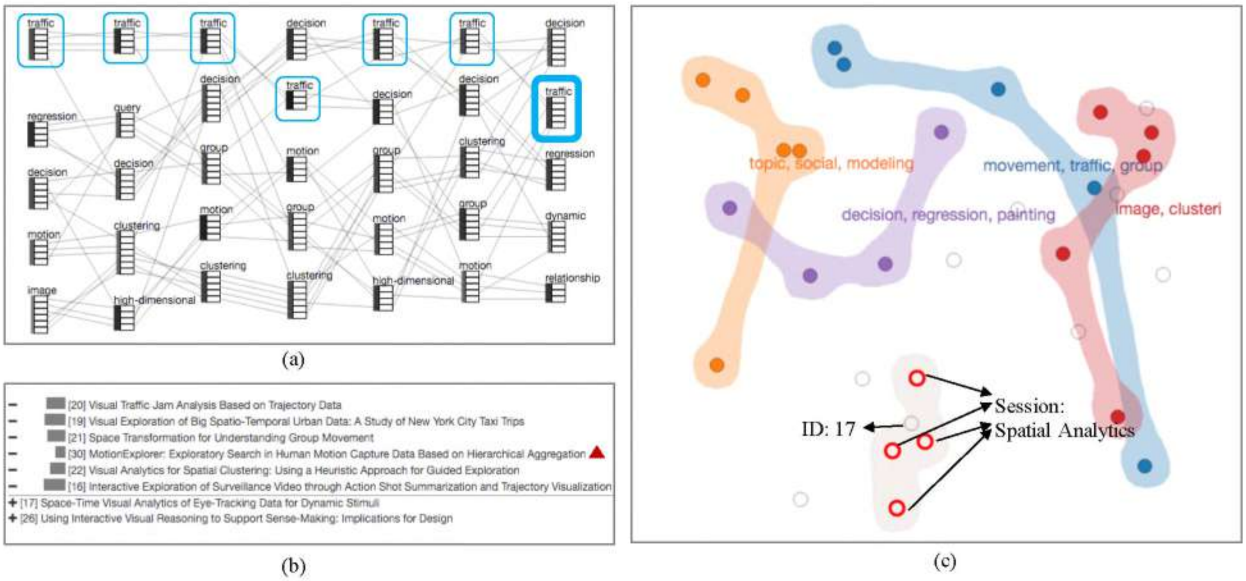

Figure 3: Case study interaction phases. (a) Selecting highly reliable topics in the Cluster Selection View. (b) Recommendations for suggested related papers in the selected section. (c) Exploring the Paper Graph View to selected related papers.

understand the primary topics of sessions. During exploration and classification, accessing original information of documents is important to understand why these documents should be grouped as a cluster and to make decisions. The Paper List View (Figure 1(f)) lists the original information of all papers. Each paper is shown with its title, abstract, keywords, and authors. In this view, users can also type terms for searching.

4.2 Multiple-linked View Analysis

Multiple views are all coordinated and provide a set of user interaction tools for effective exploration through paper metadata and management of session grouping. When the mouse hovers on a paper, the corresponding visual components in all views will be highlighted. The similar highlight reaction triggers when users hover on a candidate group, buffer sessions or confirmed sessions. Users can click a paper to add it to a selection set or click it again to remove it from the current selection set. In the Paper Graph View, users can also drag an area to select multiple papers at one operation. A "create" button is provided in the Paper Graph View to allow user to create a new session with current selected papers. To monitor keywords, users can add a new word to the list of monitoring not only by typing terms in the Term Monitor View but also by selecting the keywords in Session List View.

4.3 System Implementation

Our system was developed with a client-server architecture. The client is built with HTML and Javascript. The server-side is implemented with NodeJS for web services and MongoDB for data storage and management. For data processing tools on the server-side, we implemented a topic-modeling module with JGibbLDA [17], and used Standford CoreNLP [16] for natural language processing.

5 Results

In this section, we present an exploration case to show how our system can help a user iteratively construct conference sessions based on the accepted papers. We also compare the result produced by our system with the sessions that the conference actually had.

A typical exploration process can be described as follows. After session settings are given, the system provides session slots that exactly match the specified configuration. Users then explore paper topics and pre-computed possible paper clusters via coordinate views. In this case, we use the data from the IEEE Visual Analytics Science and Technology Conference (VAST) 2013, which contains 32 papers. During exploration, the user identifies a session with a topic on "traffic" (Figure 3(a)) and the paper 30 is much not similar to other papers in this session in Recommendation View (Figure 3(b)). By investigating the content, the user decides to move the paper to the "temporal analytics" session. In the Paper Graph View, Paper 17 placed near other papers is in this session (Figure 3(c)), and its fitness to "temporal analytics" is also indicated in the Term Monitor View (Figure 1(c)). After these observations, the user moves Paper 17 into the session and finally forms a satisfactory session.

We compare the actual sessions of the VAST 2013 with the result created by following the exploration process illustrate above, which can be seen in Figure 5. VAST 2013 has 7 sessions but contained one session of 3 papers, one session of 4 papers, and 5 sessions of 5 papers. From the aspect of size balancing, our case study created 4 sessions of 5 papers and 3 sessions of 4 papers, which is more balanced than the partition in practice. Specifically, we find 3 sessions exactly match in the two schedules. These sessions are "Text and Social Media", "High Dimensional Data" and "Sensemaking and Collaboration". All other sessions match at least 3 papers.

In our case study, Paper 16 of the session "Spatial Analytics" focuses on human trajectory visualization, which is also related to space and movement analytics. Paper 18 is moved to the session "Modeling and Decision Making" and is renamed as "Decision Making and Provence". Indeed, "Provence" and "Decision making" are closely related. Concentrated on the session "Images and Video", it is divided into the sessions "Temporal Analytics" and "Clustering". Therefore, image-related papers and Paper 31 make the new session "Clustering". Compared to the session "Temporal Analytics" in the two partitions, the extra paper, Paper 17, is reasonable to be classified into this session because it is also related to temporal visual analytics. As for Paper 31, it can be classified as either temporal analytics or clustering. In our case study, it is moved to the session "Clustering" due to size constraints. The strategy that splits the session "Images and Video" and creates a new session "Clustering" makes the session more coherent.

VAST 2013 Sessions		Results of the Our System	
Session: Modeling and Decision Making	0 1 2	0 1 2 18	Session: Decision Making and Provence
Session: Text and Social Media	3 4 5 6 7	3 4 5 6 7	Session: Text and Social Media
Session: High Dimensional Data	8 9 10 11 12	8 9 10 11 12	Session: High Dimensional Data
Session: Images and Video	13 14 15 16 17	13 14 15 31	Session: Clustering
Session: Space and Movement	18 19 20 21 22	16 19 20 21 22	Session: Spatial Analytics
Session: Spatial Analytics	23 24 25 26	23 24 25 26	Session: Sensemaking and Collaboration
Session: Temporal Analytics	27 28 29 30 31	17 27 28 29 30	Session: Temporal Analytics

Figure 5: The sessions of VAST 2013 and results of the case study. The number in the circle indicates the paper id which can be seen in the supplemental paper. Green indicates papers are in the session of VAST 2013 but not in our case results, while orange indicates the opposite.

6 Conclusion and Future Work

In this work, we present a system to generate coherent sessions from conference papers with multiple constraints. Users are highly involved in the process and tools are provided to support decision-making. Combining with data processing algorithms such as topic modeling and interactive visualization, our system supports an iterative exploration phase, which can narrow down the problem scale, solve the conflicts, and split the sessions with coherent papers inside. Although this work focuses on helping people construct conference paper sessions, our approach could be applied to many other scenarios, such as creating schedules in large activities, arranging classes for university students, etc.

We envision to extend our work in several ways. Firstly, we could import more clustering methods into the system for model comparison. With multiple models, users can have more choices and evaluate the situation more accurately. Secondly, we could provide more hints for users to evaluate the effectiveness of each interaction step in decision-making, and provide recommendations of the following series of decisions, which help to avoid conflicts in a pre-warning manner, and further make it easy for conference committee members to conduct paper partitions. Finally, we could use the proposed results of multiple users as inputs, such as from multiple committee members, to further enhance the partition results. A collaborative system that enables multiple users to be involved in the problem can really help decision-making.

Acknowledgments

We thank the anonymous reviewers for their valuable comments. This work is supported by National Key Research and Development Program of China (2016QY02D0304) and NSFC No. 61872013.

References

[1] E. C. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher. Serendip: Topic model-driven visual exploration of text corpora. In *Proc. VAST'14*, pp. 173–182, 2014. doi: 10.1109/VAST.2014.7042493

[2] A. P. Bhardwaj, J. Kim, S. Dow, D. R. Karger, S. Madden, R. Miller, and H. Zhang. Attendee-sourcing: Exploring the design space of community-informed conference scheduling. In *Proceedings of the Seconf AAAI Conference on Human Computation and Crowdsourcing, HCOMP 2014, November 2-4, 2014, Pittsburgh, Pennsylvania, USA*, 2014.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[4] K. Chen and L. Liu. ivibrate: Interactive visualization-based framework for clustering large datasets. *ACM Trans. Inf. Syst.*, 24(2):245–294, 2006. doi: 10.1145/1148024

[5] L. B. Chilton, J. Kim, P. André, F. Cordeiro, J. A. Landay, D. S. Weld, S. P. Dow, R. C. Miller, and H. Zhang. Frenzy: collaborative data organization for creating conference sessions. In *Proc. CHI'14*, pp. 1255–1264, 2014. doi: 10.1145/2556288.2557375

[6] J. Choo, C. Lee, C. K. Reddy, and H. Park. UTOPIAN: user-driven topic modeling based on interactive nonnegative matrix factorization. *IEEE Trans. Vis. Comput. Graph.*, 19(12):1992–2001, 2013. doi: 10.1109/TVCG.2013.212

[7] J. Chuang, C. D. Manning, and J. Heer. Termite: visualization techniques for assessing textual topic models. In *Proc. AVI'12*, pp. 74–77, 2012. doi: 10.1145/2254556.2254572

[8] C. Collins, G. Penn, and M. S. T. Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1009–1016, 2009. doi: 10.1109/TVCG.2009.122

[9] W. Cui, S. Liu, Z. Wu, and H. Wei. How hierarchical topics evolve in large text corpora. *IEEE Trans. Vis. Comput. Graph.*, 20(12):2281–2290, 2014. doi: 10.1109/TVCG.2014.2346433

[10] M. Desjardins, J. MacGlashan, and J. Ferraioli. Interactive visual clustering. In *Proc. IUI'07*, pp. 361–364. ACM, 2007.

[11] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2002–2011, 2013. doi: 10.1109/TVCG.2013.162

[12] M. J. Gardner, J. Lutes, J. Lund, J. Hansen, D. Walker, E. Ringger, and K. Seppi. The topic browser: An interactive tool for browsing topic models. In *NIPS Workshop on Challenges of Data Visualization*, vol. 2, 2010.

[13] J. Kim, H. Zhang, P. André, L. B. Chilton, W. E. Mackay, M. Beaudouin-Lafon, R. C. Miller, and S. P. Dow. Cobi: a community-informed conference scheduling tool. In *Proc. UIST'13*, pp. 173–182, 2013. doi: 10.1145/2501988.2502034

[14] H. Lee, J. Kihm, J. Choo, J. T. Stasko, and H. Park. ivisclustering: An interactive visual document clustering via topic modeling. *Comput. Graph. Forum*, 31(3):1155–1164, 2012. doi: 10.1111/j.1467-8659.2012.03108.x

[15] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[16] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *Proc. ACL'14*, pp. 55–60, 2014.

[17] X.-H. Phan and C.-T. Nguyen. Jgibblda: A java implementation of latent dirichlet allocation (lda) using gibbs sampling for parameter estimation and inference, 2006.