Sun GD, Wu YC, Liang RH *et al.* A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 28(5): 852–867 Sept. 2013. DOI 10.1007/s11390-013-1383-8

A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges

Guo-Dao Sun[1,2] (孙国道), Ying-Cai Wu[1,*] (巫英才), *Member*, *IEEE*
Rong-Hua Liang[2] (梁荣华), *Member*, *CCF*, *IEEE*
and Shi-Xia Liu[1] (刘世霞), *Member*, *CCF*, *Senior Member*, *IEEE*

[1] *Internet Graphics Group, Microsoft Research Asia, Beijing 100080, China*

[2] *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China*

E-mail: {v-gusun, yingcai.wu}@microsoft.com; rhliang@zjut.edu.cn; Shixia.Liu@microsoft.com

Received August 6, 2013; revised August 17, 2013.

Abstract Visual analytics employs interactive visualizations to integrate users' knowledge and inference capability into numerical/algorithmic data analysis processes. It is an active research field that has applications in many sectors, such as security, finance, and business. The growing popularity of visual analytics in recent years creates the need for a broad survey that reviews and assesses the recent developments in the field. This report reviews and classifies recent work into a set of application categories including space and time, multivariate, text, graph and network, and other applications. More importantly, this report presents analytics space, inspired by design space, which relates each application category to the key steps in visual analytics, including visual mapping, model-based analysis, and user interactions. We explore and discuss the analytics space to add the current understanding and better understand research trends in the field.

Keywords visual analytics, information visualization, data analysis, user interaction

1 Introduction

Recent advances in computing and storage technologies have made it possible to create, collect, and store huge volumes of data in a variety of data formats, languages, and cultures[1]. Effective analysis of the data to derive valuable insights enables analysts to design successful strategies and make informed decisions. Various numerical/algorithmic approaches such as data mining and machine learning methods have been used to automatically analyze the data. Although these approaches have proven their usefulness in many practical applications, they still face significant challenges such as algorithm scalability, increasing data dimensions, and data heterogeneity. Furthermore, these methods may not be perfect under all analysis scenarios. Users often have to provide their knowledge to iteratively refine the methods. If complex, interesting patterns are discovered, it is usually difficult to understand and interpret the findings in an intuitive and meaningful manner[2].

To address these challenges, visual analytics has been developed in recent years through a proper combination of automated analysis with interactive visualizations. The emergence of visual analytics can be largely attributed to the strong need of homeland security of the United States to analyze complex data, such as incomplete, inconsistent, or potentially deceptive information, since the September 11, 2001 terrorist attacks[3]. The analysis requires that humans should become involved to evaluate the data to respond in a timely manner.

Thomas and Cook presented the first widely-accepted roadmap for visual analytics to meet the practical requirement in their seminal book[3]. In the book, visual analytics is defined as "The science of analytical reasoning assisted by interactive visual interfaces". Later, the VisMaster Coordinated Action community, funded by the European Union, updated the roadmap and provided a more specific definition of visual analytics: "Visual analytics combines automated analysis

---

Survey

The work is partly supported by the National Natural Science Foundation of China under Grant No. 61070114, the Program for New Century Excellent Talents in University of China under Grant No. NCET-12-1087, and the Zhejiang Provincial Qianjiang Talents of China under Grant No. 2013R10054.

*Corresponding Author

with interactive visualizations for effective understanding, reasoning and decision making on the basis of a very large and complex dataset"[2].

These pioneering researches[2-3] define the scope of the field and discuss future research challenges that the field will face. Subsequently, a large number of visual analysis techniques have been developed. The rapid technical developments in the field have greatly promoted the use of visual analysis techniques in different domains to solve real-world problems, such as network traffic analysis[4], engaging education[5-6], concepts[5], sport analysis[7], database analysis[8], and biological data analysis[9-11]. As a result, visual analytics has been gaining more and more attention from both industry and academia. With the growing popularity of visual analytics, there is an increasing need for a comprehensive survey covering the recent advances of the field.

Our motivations in conducting this survey are twofold. First, we aim to review the most recent developments of visual analysis techniques and applications and provide a concise but broad review of the field. To the best of our knowledge, the surveys of visual analytics published in the last few years mainly focus on some narrow topics of visual analytics, such as visual analysis of time-oriented data[12], spatio-temporal data[13], or network data[14]. A comprehensive survey that reviews the current research of visual analytics is still absent.

Second, this survey aims to organize, classify, and compare recent research to provide a critical assessment of the research and understand current research trends. We introduce *analytics space* to organize and classify the current visual analytics research in a novel way using *key VA (Visual Analytics) steps* and *application categories*. As visual analytics is an application-driven research field[3], we classify the papers into different application categories: *space and time*, *multivariate*, *text*, *graph and network*, and *other applications*. The key VA steps refer to the key steps in the classic visual analytics process[2] including *visual mapping, model-based analysis*, and *user interactions* which have been commonly accepted in the field. We analyze the analytics space to discuss and explore the research trends.

The contributions of this paper are as follows. First, the paper presents a comprehensive survey of recent developments of visual analytics research. Second, it provides a novel classification of the results and identifies new research trends, which can help enhance the understanding of the field.

The structure of the paper is as follows. In Section 2, we introduce recent models and theories of visual analytics. Section 3 discusses the paper classification and analyzes the research trends. In Sections 4, 5, 6, and 7, we review current researches in different application categories. Finally, Section 8 concludes the paper and outlines future challenges in this research domain.

2 Theories, Models, and Frameworks

Visual analytics focuses on analytical reasoning using interactive visualizations. Shneiderman *et al.*[15] proposed a famous information seeking mantra: "Overview first, zoom/filter, details on demand" to facilitate visual data exploration. Keim *et al.*[1] indicated that only displaying the data using a visual metaphor rarely provides any insight. They extended the mantra[15] for visual analysis to gain profound insights: "Analyze first, show the important, zoom/filter, analyze further, details on demands". Compared with the information visualization mantra[15], the visual analysis mantra[1] highlights the combination of numerical/algorithmic data analysis and interactive visual interfaces.

Keim *et al.*[2] also introduced a seminal framework to depict the visual analytics process. Fig.1 illustrates the entire visual analysis process. The process starts by transforming the data (such as filtering and sampling) for further exploration. After that, a visual or an automatic analysis method is adopted separately. When automatic analysis methods are applied, approaches such as data mining methods are used to estimate models for characterizing the data. When visual data exploration is used, users directly interact with the visual interface to analyze and explore the data.

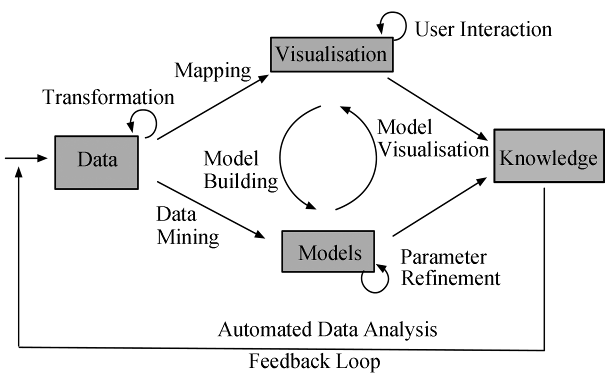

Fig.1. Visual analytics process by Keim *et al.*[2]

The combination and interaction between visual and automatic analysis methods are the key feature of visual analytics, which helps distinguish the visual analytics process from other data analysis processes. It allows for progressive refinement and evaluation of the analysis results. For instance, patterns discovered by the visual method can help refine the automatic analysis model. Thus, visual data exploration together with automatic model-based analysis can often lead to better analysis results.

Recently, researchers have introduced different means to enhance the classic information visualization process[16]. Bertini *et al.*[17] proposed overlaying the *Quality-Metrics-Driven Automation* on the classic pipeline[16]. The quality metrics can be integrated into different steps of the pipeline to automate the numerical/algorithmic data analysis and better support visual analysis and exploration. In addition, Crouser *et al.*[18] emphasized the importance of human-computer collaboration in the visual analytics process. Simoff *et al.*[19] suggested the importance of user interactions in the visual analysis process.

Some other new models and guidelines for visual analytics have also emerged in recent years, greatly boosting the advancement of the field[20-26]. Munzner *et al.*[27] divided the visual analysis design into four layers: domain problem characterization, data/operation abstraction design, encoding/interaction technique design, and algorithm design. Sedlmair *et al.*[28] introduced a methodology with nine stages (learn, winnow, cast, discover, design, implement, deploy, reflect, and write) for conducting an effective design study. Lam *et al.*[29] reviewed a large number of visualization publications and derived seven evaluation scenarios in visual analytics, thus providing a useful guidance for designing an effective evaluation procedure. An interaction model called *semantic interaction*[30] has been introduced recently. It allows users to interact with high-dimensional data in a two-dimensional (2D) view, in which the distances between data items in the view represent the similarity between the items.

A few visual analysis frameworks have also been introduced to facilitate the development of visual analytics systems. Data-Driven Documents (D3)[31] is a representation-transparent framework for rapid development of online data visualizations. It allows for direct manipulation and modification of any document elements and enables smooth animation and user interactions. WebCharts[32] is a new visualization platform that enables an application to host Javascript code. It allows for easy reuse of existing code and fast system deployment.

3 Analytics Space

In this section, we organize the papers from a novel perspective, which considers the application categories and the key steps of the visual analytics process. The key steps include *data transformation, visual mapping/layout, model-based analysis*, and *user interactions* according to the widely-accepted visual analytics model[2]. These key steps form the foundation of effective visual analytics systems. We do not consider data transformation in our classification since it is straightforward and commonly-used. We also carefully examined the sections of papers from the premier conferences of visual analytics such as IEEE InfoVis and IEEE VAST. Five categories of applications have been identified: *space and time*, *multivariate*, *text*, *graph and network*, and *other applications*. The categories not only provide a broad overview of visual analytics applications, but also differentiate recent research.

We have come up with *analytics space*, inspired by *design space*, to better understand the relationships between these key steps and different application categories. It relates each application category to specific visual analytics steps. Fig.2 illustrates the analytics space using a heatmap. Each row of the figure represents a key step of the visual analytics process and each column stands for an application category. Each cell contains one or more surveyed papers. A paper in a certain cell means that the work belongs to an application category and the techniques used can be classified into a particular key step. We also use color to visually

	Space & Time	Multivariate	Text	Graph	Others
Visual Mapping	[33-58]	[59-71]	[72-86]	[87-114]	[4-5, 7, 9-11, 25, 115-117]
Model-Based Analysis	[33, 39, 43, 45-49]	[59, 62, 64-65, 67-71, 118-119]	[72-73, 75-76, 79-80, 82, 84-86, 120]	[91, 110, 113]	[8, 25, 116]
User Interactions	[36, 39-40, 46-52, 54-55, 57]	[59-62, 64, 66-68, 70-71, 116]	[72-73, 75, 77-79, 82, 84-86, 120]	[90-91, 94, 97-102, 104-105, 107, 110-113]	[4-5, 7, 9-10, 25, 115-116]

Fig.2. Analytics space for different applications in the visual analytics process.

encode the number of papers. The darker a cell, the more papers it contains.

Fig.2 clearly indicates the imbalanced distribution of recent research in different key steps across different applications. Obviously, the second row of the figure, representing *model-based analysis*, looks lighter than the other rows. That is recent research mainly improves the visual mappings/layouts of existing algorithms and designs intuitive user interactions to solve real-world analysis problems. It is possible that the traditional information visualization research still plays an important role in the field of visual analytics. In the future, the research of visual analytics needs to be conducted towards a seamless integration of interactive visualizations and model-based analysis.

Fig.2 also reveals that the second and third columns of the figure look overall darker than the other columns. That is the research in the *text* and *multivariate* categories exhibits a more balanced structure of the visual analytics process. We speculate that text data is more complex, unstructured free text, which is difficult to analyze directly. Mining algorithms are needed to transform the unstructured data to structured information to facilitate the analysis. The multivariate data is often high dimensional. Without model-based analysis such as dimension reduction techniques, it would be almost impossible to derive any insight. In contrast, the papers in the categories of *space* & *time* and *graph* still mostly focus on visual mappings and user interactions. One possible reason is that the data used by *space* & *time* and *graph* is usually structured data. Thus, the techniques without model-based analysis suffice for the applications.

4 Space and Time

With advances in technologies, geospatial, temporal, and spatio-temporal data have been one of the most prominent and ubiquitous data types in visual analytics[121]. Finding spatial and temporal relationships and patterns in the data is needed in many analysis tasks[2]. However, the scalability and complexity of the data pose significant challenges for effective analysis, which requires both advanced computational and visualization techniques.

4.1 Analysis of Geospatial Data

Visual analytics often plays a key role in analysis of geospatial data[36]. Recent research has brought some new developments in this field[34,51,55]. Slingsby *et al.*[51] presented an interactive visual analysis system to explore and examine the results of OAC — a geodemographic classifier. The work uses OAC to classify the UK population with 41 demographic variables into a set of geographical areas that are organized in a three-level hierarchy. A set of coordinated views such as dot maps, barcharts, treemaps, and parallel coordinates plots are employed to visually analyze the OAC categories with uncertainty information. The treemaps used with spatial ordering[56] relate the node positions in the treemap to the corresponding real geographic regions.

BallotMaps[55] is an interesting interactive graphics tool based on hierarchically organized charts to facilitate analysis of spatial and non-spatial data. The tool was used to study the relationship between the number of votes received by a candidate and the position of his name on the ballot paper, and examine the associated geographical patterns. Some interesting patterns related to the 2010 local government elections in the Greater London area were discovered using the tool (see Fig.3). However, the method does not consider the voting bias patterns for different parties over time.

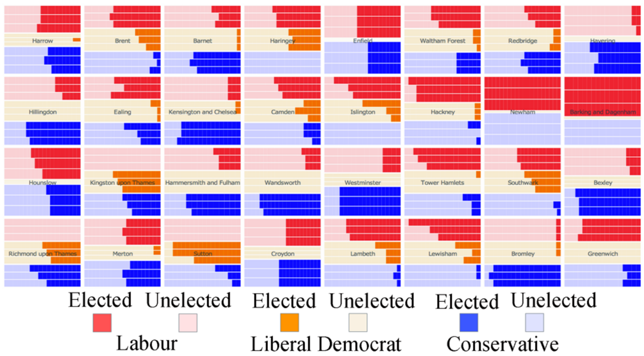

Fig.3. BallotMap of 2010 local government elections in the Greater London area[55].

4.2 Analysis of Temporal Data

Visual analytics of temporal data has attracted increasing interest in many analysis tasks and has been widely used in a variety of applications such as analysis of environmental time series[50]. This subsection reviews only recent research. For other related work, interested readers can refer to a book on visualization of time-oriented data[12].

CloudLines[44] uses a new compact visual metaphor to visualize time series in limited space. ChronoLenses[57] provides different types of lenses to explore regions of interest in time series data. Users are allowed to interact with the lenses to build analytical pipelines to facilitate exploratory analysis.

High-dimensional time series data, such as multivariate financial and economic data, is commonly found in our daily lives but is challenging for analysis. TimeSeer[39] is a useful visualization tool for exploring the high-dimensional time series data. The tool employs a set of measures, such as density, skewness, and outliers, called *scagnostics* to capture the characteristics of the data. TimeSeer displays the estimated scagnostics using a scatterplot matrix, line charts, and a set of small multiples (see Fig.4). It supports various interactions such as filtering, brushing, and drill-down.

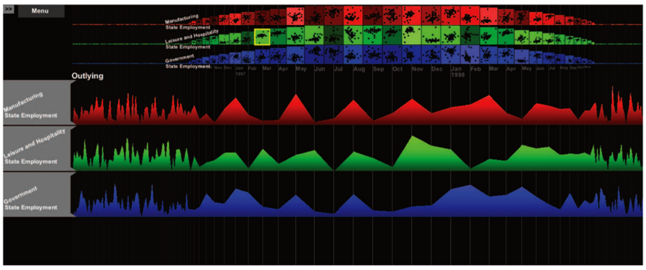

Fig.4. Visualization of a series of US Employment data using TimeSeer[39].

RankExplorer[49] is a novel visual analysis technique that combines ThemeRiver[122], color bars, and glyphs to explore ranking changes in large time series data. RankExplorer first segments the time series data into different segments. A ThemeRiver layout is used to visualize the temporal variation of each segment and the total variation of all the segments. Color bars and glyphs are embedded in the ThemeRiver layout to display inner ranking changes inside a segment and outer ranking changes between segments, respectively. The tool was used to analyze the ranking changes of temporal search queries (see Fig.5).

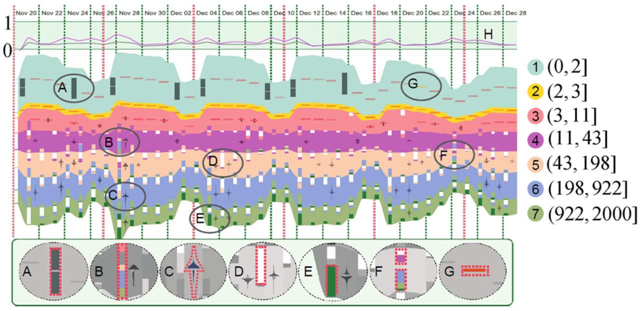

Fig.5. Visualization of the top 2000 Bing search queries using RankExplorer[49].

4.3 Analysis of Spatio-Temporal Data

Spatio-temporal visual analytics has attracted a great deal of attention. Spatio-temporal data refers to the data with both spatial and temporal information. Various methods have been used to solve real-world problems[37,43,46,58]. Nevertheless, visual analytics of spatio-temporal data remains difficult.

Trajectory visualization is a very important application of spatio-temporal visual analytics. Tominski *et al.*[52] proposed visualizing trajectories using a hybrid 2D/3D display. This display stacks 2D trajectory bands on top of a 2D map in 3D space, such that trajectories can be displayed in their spatial context, as shown in Fig.6. Density-based methods with kernel density estimation techniques[33,35,47-48] are used for visualizing a large number of trajectories on a map. Scheepens *et al.*[47] proposed using composite density maps for multivariate trajectories. Their approach uses a flexible architecture with six different operators to create, compose, and enhance density fields. Fig.7 shows a composite density map displaying multivariate trajectories of different vessel types in front of Rotterdam harbor.

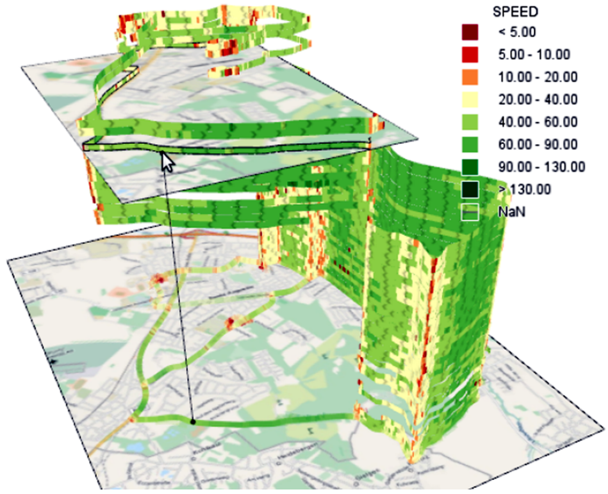

Fig.6. Stacking-based visualization of trajectories in 3D space[52].

Fig.7. Composite density maps of vessels[47].

Many applications can benefit from interactive spatio-temporal visual analytics. Maciejewski *et al.*[45]

presented a visual analytics approach to forecast hotspots, namely, unusual spatio-temporal regions. BirdVis[40] is a typical interactive spatio-temporal visualization system with coordinated views to understand bird populations.

4.4 Summary

The recent developments in spatial, temporal, and spatio-temporal visual analysis approaches indicate that this research area is growing rapidly. Nevertheless, there are still quite a few research challenges that must be addressed. One challenge is to effectively visualize realtime streaming data with a large number of time series. Additionally, effective modeling, characterization, and visualization of the uncertainty information arising from spatial-temporal data collection and transformation must also be investigated.

5 Multivariate Data

Visual analytics of multivariate data is an active research area. Numerous methods are used to explore and understand the distributions and correlations among different data dimensions[62,67,69,119,123]. These approaches can be generally classified into two broad categories: projection-based methods based on dimension reduction techniques and visual methods based on visual layouts.

5.1 Projection-Based Methods

Projection-based techniques (or dimension reduction) find "interesting" projections of high-dimensional data in low-dimensional space[123]. The techniques transform high-dimensional data to low-dimensional data while preserving some important features of the original data. Dimension reduction can help avoid the effects of "the curse of dimensionality"[124] for subsequent data analysis.

Multidimensional scaling (MDS) is widely used in this area to reduce data dimensionality. Traditional MDS uses the Euclidean distance to compute data similarity. Lee *et al.*[64] argued that the Euclidean distance cannot characterize the inter-cluster distances, thus resulting in poor data projections. They introduced a structure-based distance metric to overcome this problem in high-dimensional space to produce good projections. This method was used to explore a variety of multidimensional datasets, such as aerosol particles data and operating system data.

Heterogeneous relationships among the dimensions in high-dimensional data space are ignored in most analysis methods. Turkay *et al.*[69] proposed using representative factors to capture the grouping relationships among the data dimensions. Their method carefully chooses a set of factors including projection factors based on MDS and principal component analysis (PCA), medoid factors, and distribution model factors to represent the relationships among the data dimensions. The representative factors are integrated into the visual analytics pipeline to facilitate exploration of high-dimensional data. The method was used to analyze the data from a healthy brain aging study with 315 dimensions and successfully discovered different subsets of individuals.

Local affine multidimensional projection (LAMP)[62] is a new projection method based on an orthogonal mapping theory for handling high dimensional data. LAMP is efficient and allows users to progressively refine the results with their knowledge. The experiments provided demonstrate that LAMP outperforms other projection methods. A system developed using this method was used to correlate images and music (see Fig.8).

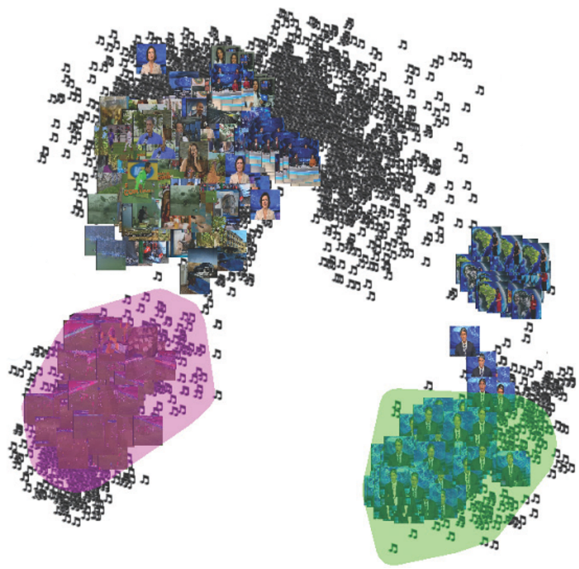

Fig.8. Visualization of image and music correlation using LAMP[62].

Paiva *et al.*[65] described an improved similarity tree technique for visual analysis of high-dimensional data. It is an alternative to traditional multidimensional projections. A platform called VisPipeline was developed to apply the technique to three image datasets and overcome the difficulty in traditional data analysis through visual feedback.

Turkay *et al.* [68] presented an interactive visual analysis approach that is performed iteratively over two spaces: the items space and the dimensions space, thus allowing for joint analysis of both items and dimensions. The approach uses PCA to map the dimensions space to items space. This technique was tested on the "Boston Neighborhood Housing Prices" dataset for understanding the relationships between different data dimensions.

5.2 Visual Methods

Visual methods leverage visualization layout algorithms such as pixel-oriented methods and parallel coordinate plots (PCPs)[25] to directly draw multivariate data for analysis.

Pixel-oriented methods visually map each multivariate data item to a pixel or block with visual attributes such as color, size, and position[63,80]. A typical recent example is DICON[59], an icon-based solution that helps compare and interpret clusters of multidimensional data. The icons representing the clusters can be embedded into various visualizations.

Traditional multivariate data visualizations such as scatterplot matrices and PCPs[125] can also be viewed as projection-based techniques, since they draw high-dimensional data in a two-dimensional space. Researchers have recently introduced flexible linked axes[60], which links a set of scatterplot matrices and/or PCPs together for analyzing high-dimensional data. This technique allows users to draw and drag axes freely, which is useful for different applications. Fig.9 shows a visualization of high-dimensional demographic data of different countries using the technique.

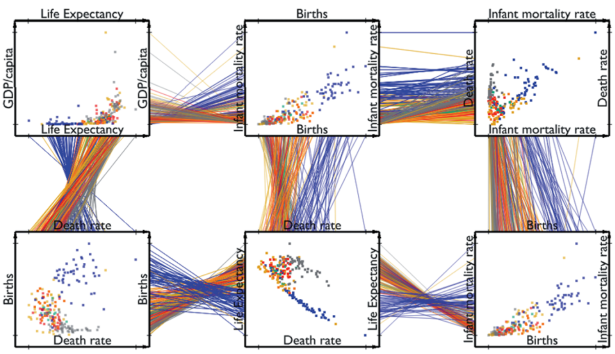

Fig.9. Flexible linked axes with scatterplots matrices and parallel coordinated plots[60].

Although PCPs are widely used in the field, they still suffer from the problems of over-plotting and clutter. Angular histograms and attribute curves were recently introduced by Geng *et al.*[61] to overcome these problems. They are able to explore the correlation in the data by investigating the density and slopes of the histogram. This work was evaluated on real-world animal tracking datasets and was compared with traditional parallel coordinates plots and histograms.

5.3 Summary

This section reviewed and discussed recent approaches to visual analysis of multivariate data. The approaches are categorized into two classes, namely, projection-based methods and visual methods. Although notable successes have been achieved, it is still difficult to understand data with a large number of dimensions due to the "curse of dimensions"[124]. Projection-based approaches based on dimension reduction can deal with data that has many dimensions, but understanding the projected data is often challenging. On the other hand, visualization approaches cannot handle data with many dimensions, but the results created by these approaches are intuitive to understand and interpret. Towards this end, Yuan *et al.*[70] made an early attempt to combine PCPs and MDS. A seamless integration of two kinds of methods is an interesting direction.

6 Text Data

Text can be found almost everywhere in billboards, newspapers, books, social media sites, and so on. With the advance of technologies, a tremendous amount of text data is being produced, collected, and stored each day. However, effective analysis of the text data is challenging for two reasons. First, the text data is often free, unstructured text corpora. The data is inherently ambiguous due to natural language ambiguity. Second, the volume of the text data is usually huge. This prevents analysts from reading the entire text corpora.

Many visual analytics techniques and applications have been developed in recent years to address these problems. They often leverage model-based analysis algorithms such as topic modeling methods[126-127] from natural language processing (NLP) to turn unstructured text into structured information, which can be used readily by subsequent interactive visualization approaches[75,79,84-85,120].

6.1 Topic-Based Methods

Topic-based methods extract topics or events from text corpora and visually explore the extracted information using different visualization techniques. It has been reported that the temporal information associated with the documents in text corpora is very important for investigative analysis of the data[74]. Recent researches, such as EventRiver[79], Visual Backchannel[77], and TextFlow[75], mostly analyze and track on the temporal evolution and diffusion of events, topics, or activities.

TextFlow[75] integrates topic mining techniques into interactive visualizations to visually analyze the evolution of topics over time (see Fig.10). It uses a few text mining algorithms to model topic evolution trends, detect critical events, and find keyword correlations. Three visual views including a topic flow view, a timeline view, and a word cloud are employed to interactively visualize the mining results and gain insights into the massive text data.

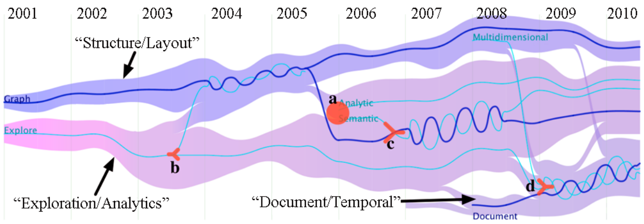

Fig.10. Visualization of topic evolution illustrating the merging and splitting patterns of topics over time by Textflow[75].

Whisper[72] is a system for visual analysis of information diffusion. It uses a visual metaphor, "sunflower", to design a hierarchical social-spatial layout for visualizing the propagation of a typical event over time on Twitter.

More recently, Xu *et al.*[85] studied the competition among topics through information diffusion on social media as well as the impact of opinion leaders on competition. They developed a system with three views: a timeline visualization with an integration of ThemeRiver and storyline visualization[128] to visualize the competition, radial visualizations of word clouds to summarize the relevant tweets, and a detailed view to list all relevant tweets. The system was used to illustrate the competition among six major topics, such as economy, election, and welfare, during the 2012 United States presidential election on Twitter. This work found that different groups of opinion leaders such as the media and grassroots played different roles in the competition (see Fig.11).

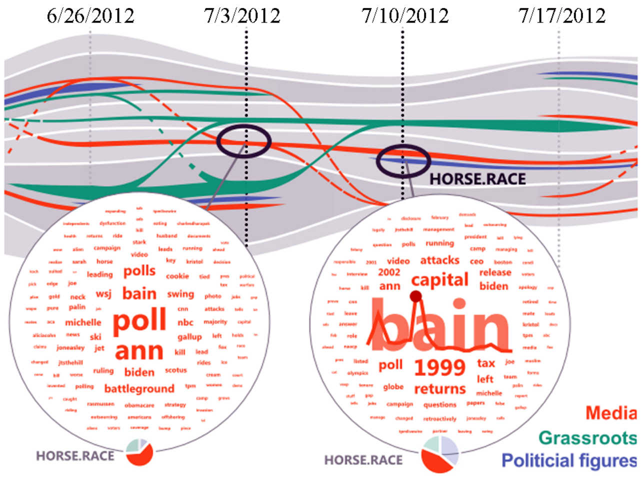

Fig.11. Visualization of topic competition and the impacts of opinion leaders on the competition on social media[85].

this issue, researchers introduced methods such as the force-based algorithm[76] (see Fig.12) and the seam-carving algorithm[84] to produce semantic-preserving word clouds. This can ensure that the keywords that co-occur frequently in the source text are placed close to one another in the word clouds.

Fig.12. Visualization of dynamic text corpora using a context-preserving word cloud visualization[76].

6.2 Feature-Based Methods

Feature-based methods use various features such as word-level features[83] and document-level features[82] to visualize text.

Word clouds are a commonly used method and have received a great deal of attention in the last few years. This method provides an intuitive visual summary of document collections by displaying the keywords in a compact layout. Keywords that appear more frequent in the source text are drawn larger. A variety of algorithms such as Wordle[83] and ManiWordle[78] have been proposed to create good word cloud layouts. However, the semantic relationships between the keywords in the original text are lost in the layouts. To handle

FacetAtlas[73] integrates a node-link diagram into a density map to visually analyze the multifaceted relations of documents. The tool was used to explore a document collection with over 1 500 articles. Interesting multifaceted relations between different diseases were discovered. DAViewer[86] was designed to help linguistics researchers study the discourse of language through a tree layout using interactive visualization.

Oelke *et al.*[80] described an interesting visual analysis application for answering "How to make your writings easier to read". Their work uses a semi-automatic method to choose proper features from 141 candidate readability features. They developed a visual analysis system called VisRA with three views, including the corpus view, the block view, and the detail view, to explore the feature values of text corpora at different levels of detail.

6.3 Summary

This section mainly introduced recent topic-based and feature-based methods for visual analysis of text data. Both kinds of methods are commonly used to solve practical problems. Although some successes have been achieved, visual analytics of text data still faces a few challenges. It is still difficult, if not impossible, for current methods to handle large amounts of text data. More efficient text mining and NLP algorithms, as well as scalable interactive visualizations, are needed to address this issue. Another challenge is to handle the natural language ambiguity and the uncertainty that arises from the text mining algorithms. Finally, text data is often accompanied by multimedia data such as images and videos, which are even more challenging for analysis. Heterogeneous text data with images and videos can be complementary. It may allow users to explore the data from different perspectives. Thus, effective analysis of the heterogeneous text data is worth further study.

7 Graph and Network

Visual analysis of graphs is an important application of visual analytics. This section covers only the research published in the last few years and classifies the work into two general categories: graph layout methods and clutter reduction methods. Interested readers can refer to a complete survey[14] for more details about the past research.

7.1 Graph Layout Methods

Graphs can be visually represented by matrix visualization[100], node-link diagrams[92], or hybrid views of node-link diagrams and matrix visualization[101].

Matrix visualization is widely used to represent networks[100]. For instance, RelEx[107] employs matrix visualization to help car engineers visually analyze information communication in in-car networks. Nevertheless, matrix visualization does not work well for sparse networks. Compressed matrices[93] explore the characteristics of a network and rearrange the matrix visualization for a compact layout. It was used to discover subnetworks in a large network. Quilts[88] is also a matrix-based method for visualizing very large layered graphs such as flow charts.

Node-link diagrams are one of the most prevalent visual representations for graphs[91-92,106,112]. In node-link diagrams, nodes are linked with directed or undirected edges to indicate the relationships of the nodes. Node-link diagrams have been successfully used to explore and understand different kinds of traditional network data such as social networks[91,114] and paper citation networks[94].

TreeNetViz[98] draws a node-link diagram in a radial layout to visualize both the hierarchical structure and network relationships in a social network. Fig.13 shows TreeNetViz that displays both the hierarchical structure (such as schools and departments) and the network relationship at different scales. Apart from traditional networks, researchers have also employed node-link diagrams to visually analyze some other interesting data such as set data[87] and interaction networks[105,109].

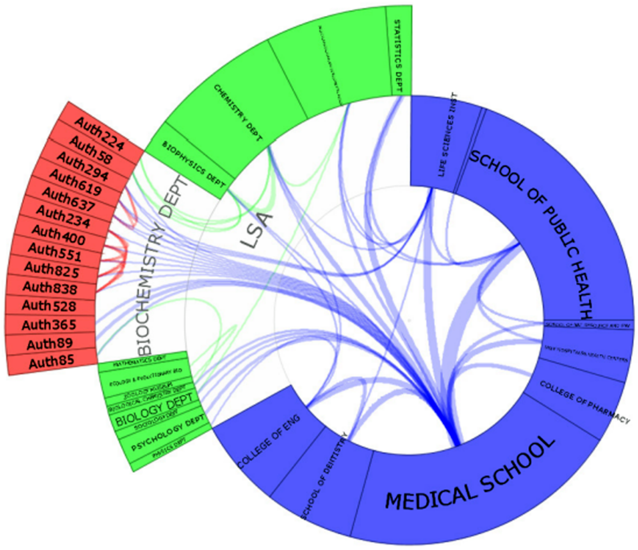

Fig.13. Visualization of a compound graph by TreeNetViz[87].

Analysis of elements of sets and their relationships is an important task. LineSets[87] is a new visual analysis technique for set data. It uses curves to link the elements across different sets to intuitively reveal the element relationships. Compared with traditional methods such as Euler diagrams, LineSets can reduce cluttered information and handle complex situations when many sets overlap. The technique was used to visualize sets of geospatial elements such as restaurants on a map to facilitate visual search tasks. It was also employed to analyze communities in social networks (see Fig.14).

StoryLine visualization has emerged recently as a new and effective means to analyze dynamic relationships such as the temporal interactions among the characters in a movie[105,109]. It is a new form of node-link diagrams. In a storyline visualization, a character is represented by a line, and the temporal interactions between characters are encoded by the convergence and divergence relationships of the corresponding lines over time.

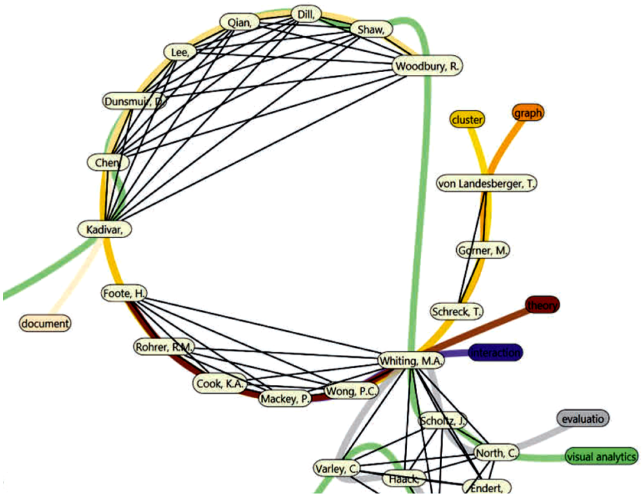

Fig.14. Visualization of a co-authorship network by Linesets[87].

Tanahashi and Ma[109] described a set of design principles, such as the principles for reducing line crossings and wiggles, for creating proper storyline layouts, and used a genetic optimization algorithm to automate the layout generation process. Although the approach is effective for creating aesthetically-appealing, compact layouts, the layout generation process is time consuming. Thus, it does not support user interactions.

Recently, StoryFlow[105] was developed to create good storyline layouts quickly for interactive visualization (see Fig.15). It uses an efficient hybrid optimization framework with an integration of discrete and continuous optimization. The efficient framework enables a set of useful real-time user interactions such as bundling and straightening. Furthermore, the approach can faithfully convey the hierarchical relationships between entities in the created layouts. StoryFlow was successfully used to study the dynamic interactions between opinion leaders on social media in the context of 2012 US presidential election.

7.2 Clutter Reduction Methods

Visual clutter is a commonly-found problem in information visualization[95]. With ever increasing sizes of networks, reducing visual clutter has become even more important for visual analysis of large networks.

Edge bundling[102-103] is an effective technique to reduce visual clutter and improve the readability of node-link diagrams by bundling related edges along an adjacent path. Hierarchical edge bundling[102] takes advantage of the hierarchy information in compound graphs to bundle edges of the graphs. The technique was used to explore and understand a software system (with hierarchically organized components) and the call graph between the components. The work was extended to bundle edges in a graph using a force-directed technique without the need of hierarchy information[103]. Selassie *et al.*[108] improved the force-directed edge bundling method to take into account the directional information, such that high-level directional edge patterns can be revealed intuitively.

Apart from forced-based methods, recent research introduces other methods for edge bundling such as the geometry-based technique[92] (see Fig.16) and the skeleton-based technique[96] (see Fig.17). The geometry-based technique uses a control mesh to attract edges to some control points on the mesh, thus generating edge bundles. In contrast, the skeleton-based method extracts the skeleton of a graph and forces its edges to be close to the skeleton. Compared with the geometry-based technique, the skeleton-based method can generate smoother bundling results while maintaining the graph structure[96], and can be easily accelerated by graphics hardware as it is an image-based method. Both methods were applied to understand US migrations network data (see Fig.16 and Fig.17). Luo *et al.*[106] introduced a new method to reduce ambiguity in edge-bundling results and enable detail-on-demand visualization. The system was evaluated using a co-authorship network.

Parallel edge splatting[90] is a new clutter reduction technique for visual analysis of large graphs. It overcomes the over-plotting problem by rearranging graph nodes on different parallel vertical axes, and connecting the nodes between the axes using directed, colored edges. This technique is capable of visualizing the evo-

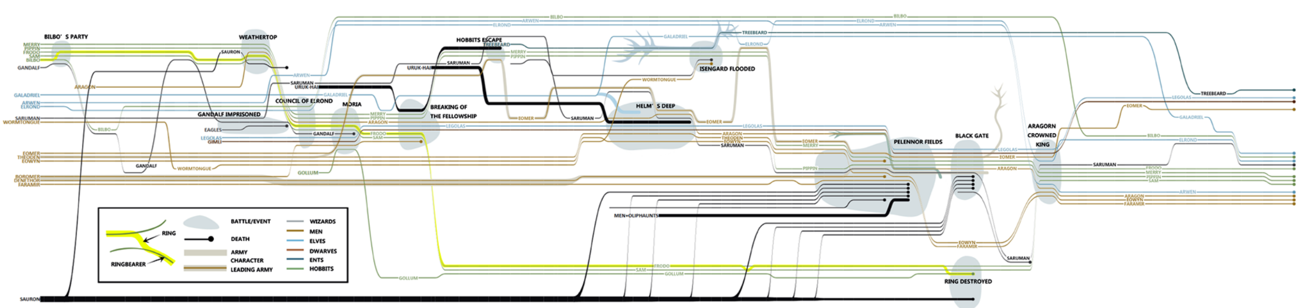

Fig.15. Storyline visualization of movie *The Lord of the Rings* by StoryFlow[105].

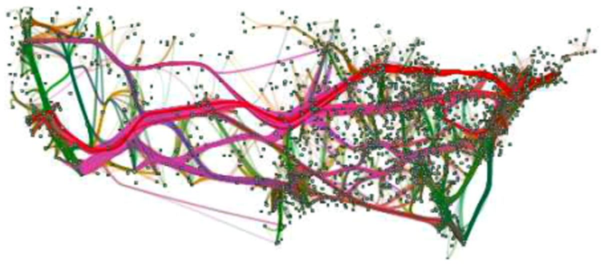

Fig.16. Geometry-based edge clustering[92].

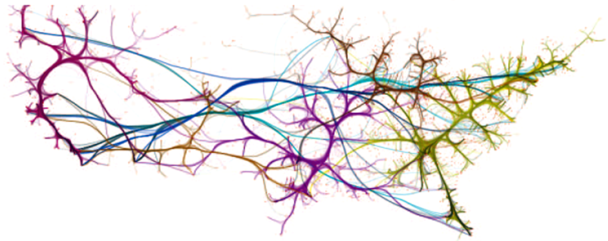

Fig.17. Skeleton-based edge bundling[96].

lution of a dynamic graph. Zinsmaier *et al.*[112] presented a fast hardware-assisted layout technique using the information of edge cumulation and node density to reduce visual clutter. It also enables interactive level-of-detail rendering of large graphs. The technique was used to visually analyze large real world graphs to detect patterns.

7.3 Summary

This section reviewed recent research of visual analysis of graphs. Although we have witnessed rapid developments in the research area, it is still very difficult to visually analyze and explore large graphs, let alone extreme-scale graphs with billions of edges. For instance, current edge-bundling techniques usually handle thousands of edges, but they may not work well for larger graphs. Clutter reduction in large graphs needs to be studied in the future.

Another possible direction is to combine model-based analysis methods, such as graph partition and frequent pattern detection methods, with interactive visualizations. Model-based analysis can help filter out a great deal of irrelevant information while preserving interesting patterns in the graphs. Interactive visualizations, on the other hand, allow analysts to work closely with the model-based analysis process to evaluate the results for sensemaking.

8 Conclusions and Future Challenges

This state-of-the-art report reviewed recent research in the field of visual analytics. It represented a comprehensive overview of many advances in visual analytics techniques and applications to gain a better understanding of the cutting-edge research in the field. In particular, the report classified the work of visual analytics research in a novel and systematic manner according to the types of the applications and the steps in the visual analytics process that the work focuses on. Additionally, through the analysis and comparison across different paper categories, this report identified the trends and recent developments in visual analytics. Furthermore, we divided the literature review into several broad application categories such as space and time analysis, text analysis, and network analysis. Next, we discuss and summarize the key challenges of the future visual analytics research.

Scalability. The explosion of data in recent years presents a significant challenge to existing techniques for visualizing big data interactively. While recent visual analytics techniques can handle small or intermediate-size data, most of them are not scalable to extreme-scale data. With the advance of parallel computing technologies, researchers have started to employ powerful computational hardware such as GPUs to accelerate the performance of visualization layout algorithms[47,112]. Nevertheless, the hardware-based parallel acceleration cannot keep pace with the data explosion rate. To overcome these issues, a variety of new visual analytics mechanisms such as bottom-up methods[91] and in-situ analysis[99] have been proposed in recent years. It is expected that scalable visual analytics techniques and methodologies will continue to attract substantial interest in the future.

Storytelling. Storytelling methods have received a great deal of attention over the past several years in visualization[105,109,129-130]. Narrative, interactive visualizations are also widely used in data-driven journalism to engage more users and reach a wider audience[81]. Typical visual analytics applications usually include a step for creating reports on the findings of the analysis. Interactive, storytelling visualizations can benefit the reports by communicating the findings more effectively for sensemaking, as narrative visualizations can convey the entire story behind the patterns found in the analysis. For example, storytelling visualizations can provide in-depth insights into why there are such patterns. Nevertheless, storytelling (or narrative) visual analytics is still in its infancy. The basic definition and the usage guidelines of storytelling techniques are heuristics and subjective. The fundamental theories for storytelling visual analytics are worth further study, and may involve multi-discipline research of human perception and cognition, human computer interaction, and visualization.

Trustworthiness. Uncertainty information may arise and spread in different steps of an analytics process[26]. Uncertainty modeling and visualization play a critical role in ensuring the reliability and trustworthiness of the analytics process. Trustworthy visual analytics with effective uncertainty modeling and visualization enables users to explicitly consider the uncertainty information, so that informed decisions can be made[26]. Many techniques have been proposed to quantitatively characterize the uncertainty information and intuitively display the information[26,51,131]. However, due to the complexity of different visual analytics applications, there are still no widely accepted techniques. In the future, research on trustworthiness will be extended.

Evaluation. It is always important to assess the effectiveness of visual analytics systems[29]. Visual analytics practitioners use various approaches such as case studies, expert review, or formal/informal user studies to evaluate the usability and effectiveness of the systems[132]. Each method has its own strengths and weaknesses. For instance, a well-designed formal user study can provide robust and valuable user feedback to identify potential problems with the systems. However, it is time-consuming to conduct a formal study and it may be difficult to provide high-level insights. A typical visual analytics system is rather complex and may involve multiple data analysis and visualization components, which poses a great challenge to evaluate the system. Effective evaluation of a visual analytics system is expected to gain more interest in the field.

Provenance. Keeping track of a visual analytics process has become prominent in the field, as the records allow analysts to be informed of where they have been and where they are now[133]. One straightforward usage of provenance information is to allow for redo/undo user interactions, or to avoid repeated analysis processes. Furthermore, the provenance information of the gained insight can facilitate the review and evaluation of the knowledge or findings. The advance of collaborative visual analytics highlights the importance of an effective mechanism for recording insight provenance, such that collaborating users can share and exchange their knowledge and insight judiciously. Nevertheless, existing simple history mechanisms, such as the Photoshop-style history mechanism, may not work well in complicated, collaborative scenarios, for instance, when users work remotely on the same problem and need to frequently exchange their findings. It is foreseeable that research into this topic will need to continue.

References

[1] Keim D A, Mansmann F, Schneidewind J, Ziegler H. Challenges in visual data analysis. In *Proc. the IEEE Conference on Information Visualization*, Oct. 2006, pp.9-16.

[2] Keim D A, Kohlhammer J, Ellis G, Mansmann F. Mastering the Information Age: Solving Problems with Visual Analytics. Florian Mansmann, 2010.

[3] Thomas J J, Cook K A. Illuminating the Path: The Research and Development Agenda for Visual Analytics. National Visualization and Analytics Ctr, 2005.

[4] Landge A G, Levine J A, Bhatele A, Isaacs K E, Gamblin T, Schulz M, Langer S H, Bremer P, Pascucci V. Visualizing network traffic to understand the performance of massively parallel simulations. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2467-2476.

[5] Block F, Horn M S, Phillips B C, Diamond J, Evans E M, Shen C. The DeepTree exhibit: Visualizing the tree of life to facilitate informal learning. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2789-2798.

[6] Ma J, Liao I, Ma K L, Frazier J. Living Liquid: Design and evaluation of an exploratory visualization tool for museum visitors. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2799-2808.

[7] Pileggi H, Stolper C D, Boyle J M, Stasko J T. Snapshot: Visualization to propel ice hockey analytics. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2819-2828.

[8] Wang X, Zhou X, Wang S. Summarizing large-scale database schema using community detection. *Journal of Computer Science and Technology*, 2012, 27(3): 515-526.

[9] Albers D, Dewey C, Gleicher M. Sequence Surveyor: Leveraging overview for scalable genomic alignment visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2392-2401.

[10] Pretorius A J, Bray M A, Carpenter A E, Ruddle R A. Visualization of parameter space for image analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2402-2411.

[11] Das K, Majumder A, Siegenthaler M, Keirstead H, Gopi M. Automated cell classification and visualization for analyzing remyelination therapy. *The Visual Computer*, 2011, 27(12): 1055-1069.

[12] Aigner W, Miksch S, Schumann H, Tominski C. Visualization of Time-Oriented Data. Springer, 2011.

[13] Andrienko G, Andrienko N, Demsar U, Dransch D, Dykes J, Fabrikant S I, Jern M, Kraak M J, Schumann H, Tominski C. Space, time and visual analytics. *International Journal of Geographical Information Science*, 2010, 24(10): 1577-1600.

[14] von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, van Wijk J J, Fekete J D, Fellner D W. Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum*, 2011, 30(6): 1719-1749.

[15] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. the IEEE Symposium on Visual Languages*, Sept. 1996, pp.336-343.

[16] Card S K, Mackinlay J D, Schneiderman B. Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, 1999.

[17] Bertini E, Tatu A, Keim D A. Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2203-2212.

[18] Crouser R J, Chang R. An affordance-based framework for human computation and human-computer collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2859-2868.

[19] Simoff S J, Böhlen M H, Mazeika A (eds.). Visual Data Mining. Springer, 2008.

[20] Dasgupta A, Kosara R. Adaptive privacy-preserving visualization using parallel coordinates. *IEEE Transactions on*

Visualization and Computer Graphics, 2011, 17(12): 2241-2248.

[21] Hullman J, Adar E, Shah P. Benefitting InfoVis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2213-2222.

[22] Kandel S, Paepcke A, Hellerstein J M, Heer J. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2917-2926.

[23] Wickham H, Hofmann H. Product plots. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2223-2230.

[24] Wu Y, Liu X, Liu S, Ma K L. ViSizer: A visualization resizing framework. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(2): 278-290.

[25] Wu Y, Wei F, Liu S *et al.* Opinionseer: Interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 1109-1118.

[26] Wu Y, Yuan G, Ma K L. Visualizing flow of uncertainty through analytical processes. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2526-2535.

[27] Munzner T. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 921-928.

[28] Sedlmair M, Meyer M, Munzner T. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2431-2440.

[29] Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(9): 1520-1536.

[30] Endert A, Fiaux P, North C. Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2879-2888.

[31] Bostock M, Ogievetsky V, Heer J. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2301-2309.

[32] Fisher D, Drucker S M, Fernandez R, Ruble S. Visualizations everywhere: A multiplatform infrastructure for linked visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 1157-1163.

[33] Adrienko N, Adrienko G. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(2): 205-219.

[34] Afzal S, Maciejewski R, Jang Y, Elmqvist N, Ebert D S. Spatial text visualization using automatic typographic maps. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2556-2564.

[35] Andrienko G, Andrienko N, Hurter C, Rinzivillo S, Wrobel S. Scalable analysis of movement data for extracting and exploring significant places. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(7): 1078-1094.

[36] Andrienko N, Andrienko G. Exploratory Analysis of Spatial and Temporal Data. Springer Berlin, 2006.

[37] Bak P, Mansmann F, Janetzko H, Keim D A. Spatiotemporal analysis of sensor logs using growth ring maps. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 913-920.

[38] Buchin K, Speckmann B, Verbeek K. Flow map layout via spiral trees. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2536-2544.

[39] Dang T N, Anand A, Wilkinson L. TimeSeer: Scagnostics for high-dimensional time series. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(3): 470-483.

[40] Ferreira N, Lins L, Fink D, Kelling S, Wood C, Freire J, Silva C. BirdVis: Visualizing and understanding bird populations. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2374-2383.

[41] Haunert J H, Sering L. Algorithms for labeling focus regions. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2583-2592.

[42] Haunert J H, Sering L. Drawing road networks with focus regions. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2555–2562.

[43] Kim S, Maciejewski R, Malik A, Jang Y, Ebert D S, Isenberg T. Bristle Maps: A multivariate abstraction technique for geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(9): 1438-1454.

[44] Krśtajic M, Bertini E, Keim D A. CloudLines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2432-2439.

[45] Maciejewski R, Hafen R, Rudolph S *et al.* Forecasting hotspots — A predictive analytics approach. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(4): 440-453.

[46] Maciejewski R, Rudolph S, Hafen R *et al.* A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(2): 205-220.

[47] Scheepens R, Willems N, van de Wetering H *et al.* Composite density maps for multivariate trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2518-2527.

[48] Scheepens R, Willems N, van de Wetering H, van Wijk J J. Interactive visualization of multivariate trajectory data with density maps. In *Proc. the IEEE Symposium on Pacific Visualization*, Mar. 2011, pp.147-154.

[49] Shi C, Cui W, Liu S, Xu P, Chen W, Qu H. RankExplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2669-2678.

[50] Sips M, Kothur P, Unger A, Hege H C, Dransch D. A visual analytics approach to multiscale exploration of environmental time series. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2899-2907.

[51] Slingsby A, Dykes J, Wood J. Exploring uncertainty in geodemographics with interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2545-2554.

[52] Tominski C, Schumann H, Andrienko G, Andrienko N. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2565-2574.

[53] Wang Y S, Chi M T. Focus+context metro maps. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2528-2535.

[54] Wongsuphasawat K, Gotz D. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2659-2668.

[55] Wood J, Badawood D, Dykes J, Slingsby A. BallotMaps: Detecting name bias in alphabetically ordered ballot papers. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2384-2391.

[56] Wood J, Dykes J. Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 2008, 14(6): 1348-1355.

[57] Zhao J, Chevalier F, Pietriga E, Balakrishnan R. Exploratory analysis of time-series with chronolenses. *IEEE Transactions*

on Visualization and Computer Graphics, 2011, 17(12): 2422-2431.

[58] Sun G, Liang R, Wu F, Qu H. A web-based visual analytics system for real estate data. *Science China Information Sciences*, 2013, 56(5): 052112(13).

[59] Cao N, Gotz D, Sun J, Qu H. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2581-2590.

[60] Claessen J H, van Wijk J J. Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2310-2316.

[61] Geng Z, Peng Z, Laramee R S, Walker R, Roberts J C. Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2572-2580.

[62] Joia P, Paulovich F V, Coimbra D *et al.* Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2563-2571.

[63] Keim D A, Hao M C, Dayal U. Hierarchical pixel bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 2002, 8(3): 255-269.

[64] Lee J H, McDonnell K T, Zelenyuk A, Imre D, Mueller K. A structure-based distance metric for high-dimensional space exploration with multi-dimensional scaling. *IEEE Transactions on Visualization and Computer Graphics*, 2013, to be appeared.

[65] Paiva J G, Florian L, Pedrini H, Telles G, Minghim R. Improved similarity trees and their application to visual data classification. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2459-2468.

[66] Steinberger M, Waldner M, Streit M *et al.* Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2249-2258.

[67] Tatu A, Albuquerque G, Eisemann M *et al.* Automated analytical methods to support visual exploration of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(5): 584-597.

[68] Turkay C, Filzmoser P, Hauser H. Brushing dimensions — A dual visual analysis model for high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2591-2599.

[69] Turkay C, Lundervold A, Lundervold A J, Hauser H. Representative factor generation for the interactive visual analysis of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2621-2630.

[70] Yuan X, Guo P, Xiao H, Zhou H, Qu H. Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 1001-1008.

[71] Lex A, Schulz H J, Streit M *et al.* VisBricks: Multiform visualization of large, inhomogeneous data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2291-2300.

[72] Cao N, Lin Y R, Sun X *et al.* Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2649-2658.

[73] Cao N, Sun J, Lin Y R *et al.* Facetatlas: Multifaceted visualization for rich text corpora. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 1172-1181.

[74] chul Kwon B, Javed W, Ghani S *et al.* Evaluating the role of time in investigative analysis of document collections. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(11): 1992-2004.

[75] Cui W, Liu S, Tan L *et al.* TextFlow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2412-2421.

[76] Cui W, Wu Y, Liu S *et al.* Context-preserving dynamic word cloud visualization. *IEEE Computer Graphics and Applications*, 2010, 30(6): 42-53.

[77] Dork M, Gruen D, Williamson C, Carpendale S. A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 1129-1138.

[78] Koh K, Lee B, Kim B, Seo J. Maniwordle: Providing flexible control over wordle. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 1190-1197.

[79] Luo D, Yang J, Krstajic M, Ribarsky W, Keim D A. EventRiver: Visually exploring text collections with temporal references. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(1): 93-105.

[80] Oelke D, Spretke D, Stoffel A, Keim D A. Visual readability analysis: How to make your writings easier to read. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(5): 662-674.

[81] Segel E, Heer J. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 1139-1148.

[82] Strobelt H, Oelke D, Rohrdantz C *et al.* Document cards: A top trumps visualization for documents. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 1145-1152.

[83] Viegas F B, Wattenberg M, Feinberg J. Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 1137-1144.

[84] Wu Y, Provan T, Wei F, Liu S, Ma K L. Semantic-preserving word clouds by seam carving. *Computer Graphics Forum*, 2011, 30(3): 741-750.

[85] Xu P, Wu Y, Wei E, Peng T Q, Liu S, Zhu J H, Qu H. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): to be appeared.

[86] Zhao J, Chevalier F, Collins C, Balakrishnan R. Facilitating discourse analysis with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2639-2648.

[87] Alper B, Riche N, Ramos G, Czerwinski M. Design study of linesets, a novel set visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2259-2267.

[88] Bae J, Watson B. Developing and evaluating quilts for the depiction of large layered graphs. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2268-2275.

[89] Brandes U, Nick B. Asymmetric relations in longitudinal social networks. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2283-2290.

[90] Burch M, Vehlow C, Beck F, Diehl S, Weiskopf D. Parallel edge splatting for scalable dynamic graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2344-2353.

[91] Crnovrsanin T, Liao I, Wu Y, Ma K L. Visual recommendations for network navigation. *Computer Graphics Forum*, 2011, 30(3): 1081-1090.

[92] Cui W, Zhou H, Qu H, Wong P C, Li X. Geometry-based edge clustering for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2008, 14(6): 1277-1284.

[93] Dinkla K, Westenberg M A, van Wijk J J. Compressed adjacency matrices: Untangling gene regulatory networks. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2457-2466.

[94] Dork M, Riche N H, Ramos G, Dumais S. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2709-2718.

[95] Ellis G, Dix A. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 2007, 13(6): 1216-1223.

[96] Ersoy O, Hurter C, Paulovich F V, Cantareiro G, Telea A. Skeleton-based edge bundling for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2364-2373.

[97] Feng K C, Wang C, Shen H W, Lee T Y. Coherent time-varying graph drawing with multifocus+context interaction. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(8): 1330-1342.

[98] Gou L, Zhang X. TreeNetViz: Revealing patterns of networks over tree structures. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2449-2458.

[99] Hadlak S, Schulz H J, Schumann H. In situ exploration of large dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2334-2343.

[100] Henry N, Fekete J D. MatrixExplorer: A dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 2006, 12(5): 677-684.

[101] Henry N, Fekete J D, McGuffin M J. Nodetrix: A hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 2007, 13(6): 1302-1309.

[102] Holten D. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 2006, 12(5): 741-748.

[103] Holten D, van Wijk J J. Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, 2009, 28(3): 983-990.

[104] Hurter C, Telea A, Ersoy O. MoleView: An attribute and structure-based semantic lens for large element-based plots. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2600-2609.

[105] Liu S, Wu Y, Wei E, Liu M, Liu Y. StoryFlow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): to be appeared.

[106] Luo S J, Liu C L, Chen B Y, Ma K L. Ambiguity-free edge-bundling for interactive graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(5): 810-821.

[107] Sedlmair M, Frank A, Munzner T, Butz A. RelEx: Visualization for actively changing overlay network specifications. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2729-2738.

[108] Selassie D, Heller B, Heer J. Divided edge bundling for directional network data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2354-2363.

[109] Tanahashi Y, Ma K L. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2679-2688.

[110] Yang J, Liu Y, Zhang X, Yuan X, Zhao Y, Barlowe S, Liu S. PIWI: Visually exploring graphs based on their community structure. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(6): 1034-1047.

[111] Yuan X, Che L, Hu Y, Zhang X. Intelligent graph layout using many users' input. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2699-2708.

[112] Zinsmaier M, Brandes U, Deussen O, Strobelt H. Interactive level-of-detail rendering of large graphs. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2486-2495.

[113] Correa C D, Crnovrsanin T, Ma K L. Visual reasoning about social networks using centrality sensitivity. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(1): 106-120.

[114] Martins R M, Andery G F, Heberle H, Paulovich F V, de Andrade Lopes A, Pedrini H, Minghim R. Multidimensional projections for visual analysis of social networks. *Journal of Computer Science and Technology*, 2012, 27(4): 791-810.

[115] Bowman B, Elmqvist N, Jankun-Kelly T J. Toward visualization for games: Theory, design space, and patterns. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(11): 1956-1968.

[116] Trimm D, Rheingans P, desJardins M. Visualizing student histories using clustering and composition. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2809-2818.

[117] Nocaj A, Brandes U. Organizing search results with a reference map. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2546-2555.

[118] Albuquerque G, Lowe T, Magnor M. Synthetic generation of high-dimensional datasets. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2317-2324.

[119] Lu Z, Liu C, Zhang Q, Zhang C, Fan D, Yang P. Visual analytics for the clustering capability of data. *Science China Information Sciences*, 2013, 56(5): 052110(14).

[120] Heimerl F, Koch S, Bosch H, Ertl T. Visual classifier training for text document retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2839-2848.

[121] Keim D A, Andrienko G, Fekete J *et al.* Visual analytics: Definition, process, and challenges. In *Information Visualization: Human-Centered Issues and Perspectives*, Kerren A, Stasko J, Fekete J *et al.* (eds.), Springer, 2008, pp.154-175.

[122] Havre S, Hetzler B, Nowell L. Themeriver: Visualizing theme changes over time. In *Proc. the IEEE Symposium on Information Visualization*, Oct. 2000, pp.115-123.

[123] Keim D A, Kriegel H P. Visualization techniques for mining large databases: A comparison. *IEEE Transactions on Knowledge and Data Engineering*, 1996, 8(6): 923-938.

[124] Bellman R E. Dynamic Programming. Courier Dover Publications, 2003.

[125] Inselberg A, Dimsdale B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proc. the 1st IEEE Symposium On Visualization*, Oct. 1990, pp.361-378.

[126] Song Y, Pan S, Liu S, Wei F, Zhou M X, Qian W. Constrained co-clustering for textual documents. In *Proc. the 24th AAAI Conference on Artificial Intelligence*, July 2010, pp.581-586.

[127] Zhang J, Song Y, Zhang C, Liu S. Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora. In *Proc. the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, July 2010, pp.1079-1088.

[128] Munroe R. Xkcd 657: Movie narrative charts. http://xkcd.com/657, Aug. 2013.

[129] Hullman J, Drucker S, Riche N H, Lee B, Fisher D, Adar E. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2013, to be appeared.

[130] Lee B, Kazi R H, Smith G. SketchStory: Telling more engaging stories with data through freeform sketching. *IEEE Transactions on Visualization and Computer Graphics*, 2013, to be appeared.

[131] Correa C D, Chan Y H, Ma K L. A framework for uncertainty-aware visual analytics. In *Proc. the IEEE Symposium on Visual Analytics Science and Technology*, Oct. 2009, pp.51-58.

[132] Tory M, Moeller T. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 2005, 25(5): 8-11.

[133] Jankun-Kelly T, Ma K L, Gertz M. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, 2007, 13(2): 357-369.

Guo-Dao Sun is a Ph.D. candidate at Zhejiang University of Technology, Hangzhou. He is a research intern at the Internet Graphics Group, Microsoft Research Asia, Beijing, from 2013 to 2014. His research interests include information visualization and visual analytics.

Ying-Cai Wu is a researcher in the Internet Graphics Group at Microsoft Research Asia. His main research interests include visual analytics of social media data, uncertainty data visualization and modeling, and visual analytics of large-scale user log data. He received the B.Eng. degree in computer science and technology from the South China University of Technology in 2004 and the Ph.D. degree in computer science from the Hong Kong University of Science and Technology in 2009. Prior to his current position, he was a postdoctoral researcher at the Visualization and Interface Design Innovation (VIDi) research group in the University of California, Davis, from June 2010 to March 2012.

Rong-Hua Liang received the B.Sc. degree in computer science from Hangzhou Dianzi University in 1996, and the Ph.D. degree in computer science from Zhejiang University in 2003. He worked as a research fellow at the University of Bedfordshire, UK, from April 2004 to July 2005 and as a visiting scholar at the University of California, Davis, US, from March 2010 to March 2011. He is currently a professor of computer science, and vice dean of the College of Information Engineering, Zhejiang University of Technology, Hangzhou. His research interests include computer vision, information visualization and visual analytics.

Shi-Xia Liu is the lead researcher in the Internet Graphics Group at Microsoft Research Asia (MSRA), Beijing. Her research interest mainly focuses on interactive, visual text analytics and interactive, visual graph analytics. She received her B.S. and M.S. degrees in computational mathematics from Harbin Institute of Technology, a Ph.D. degree in computer-aided design and computer graphics from Tsinghua University. Before she joined MSRA, she worked as a research staff member and research manager at IBM China Research Lab, where she managed the Departments of Smart Visual Analytics and User Experience.