Multi-view Display Coordinated Visualization Design for Crime Solving Analysis

Vast Challenge 2014: Honorable Mention for Effective Use of Coordinated Visualizations

Ruimin Gao[1], Han Tao[2], Hui Chen[3], Wentao Wang[4], Jiawan Zhang[5]
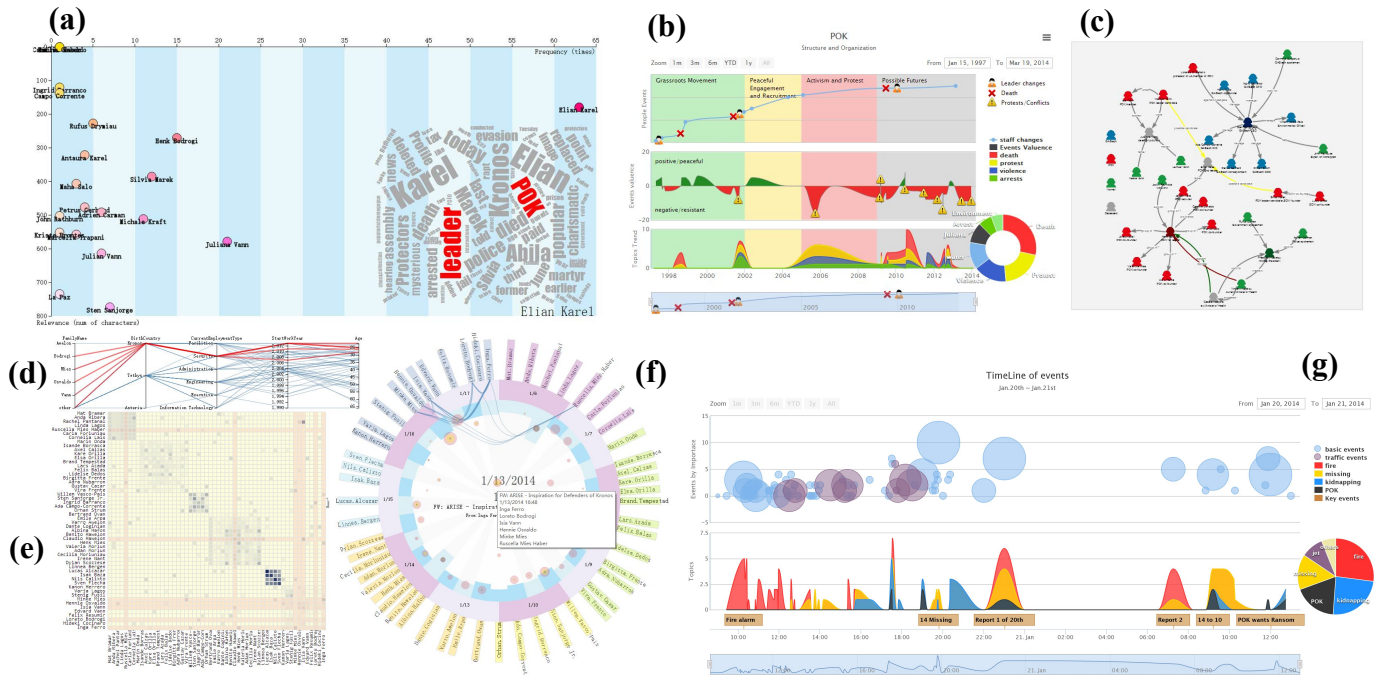
Visual Analytics Lab, Tianjin University

Figure 1: The display coordinated design for Crime Analysis. (a): coordinate graph view for finding objectives; (b): multi-lane timeline with events, group developing trend and topic changing trend; (c):interactive social network linked with timeline(b); (d):parallel coordinate graph of people's information; (e):heat-map of contacting intimacy; (f):chord graph of email relations; (g):timeline of events and topic keywords;

ABSTRACT

As terrors grow wild around the world, visualization plays an important role in helping analyze security situation and solve terrorism and emergency cases. In this paper, we present a system of multiple interactive and coordinated views to help visualize and analyze several kinds of large-scale data and complex datasets. This design assists to explore social network involved and analyze natural language used in large-scale datasets, then to make an integrated analysis and presentation of all kinds of objects involved in cases in multiple perspectives and hierarchies. We implemented datasets from VAST Challenge 2014 mini challenge 1 to verify the usability of our design.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.1.2 [User/Machine Systems]: Visual Analytics;

[1] e-mail: veslyn@tju.edu.cn
[2] e-mail: th8678@tju.edu.cn
[3] e-mail: leading2006@aliyun.com
[4] e-mail: wangwentao1990@hotmail.com
[5] e-mail: jwzhang@tju.edu.cn

1 INTRODUCTION

To solve crime cases, national and international organizations have collected and accumulated a tremendous amount of data from all kinds of sources. The most important part in crime solving is extracting and making use of effective information from multi-scale and various datasets to help analyze current cases. For this purpose, we design a system to process and integrate all kinds of data and give an effective presentation for analysis with high automation, interaction and coordination, applying to large-scale datasets like correspondence data, social network and unformatted data mainly based on natural language.

2 MULTIPLE DATASETS PROCESSING AND INTEGRATION

Unpredictability of emergency has brought difficulty in collecting effective information. Large datasets processing and integration is the precondition of analyzing security situation and solving cases.

2.1 Social Network Analysis

In our design, we extracted and integrated information from resumes, news articles and email data to analyze figures cluster relations. We estimated suspicious degree of all relevant people based on condition of events and personal profile, shown in Figure 1-(d). In consideration of topic participant number, suspicious degree of topic participants and relevance of cluster,

we present modified chord graph to represent email network relationships with each topic as a unit, as is shown in Figure 1-(f). Then we use heat-map to show the weighted email relations among all members in a cluster, shown in Figure 1-(e).

With these three graphs coordinated as one view for suspects cluster analysis, we can expand relevant suspects cluster according to a relevant topic, personal profiles and email network. As for the overall social network graph shown in Figure 1-(c), it's presented for cooperative use with timeline in later integration.

2.2 Events Analysis

We summarize text with certain strategy [1] and eliminate obvious futile information, meanwhile extract keywords from text to filter topic trend. Then we sort these worked data and extract time label in each text. Each event is initialized with relevant summarized text and a time label, and some of them are assigned with emotion trend. In the end, most text data are transformed into event elements after processing.

Noe we are able to integrate all elements into a timeline of events and keywords. Taking coordinated social network graph into consideration, we can analyze the overall changing trend of an organization or a crime case like kidnapping.

2.3 Features of Datasets integration

Data Mining One of the most important work in text visualization is data mining. We mine useful information from large-scale and messy raw data based on objective laws, then process data with efficiency and possible accuracy. For example, we take frequency of one news topic keyword in certain time as the parameter for calculating the importance degree of that specific topic keyword, as is shown in Figure 1-(g), topic lane. Frequency of events are presented as radius of event bubbles in timeline graph. And frequency of keywords appearing in each event stands for height of that event bubble.

Multi-Lane Single event timeline is not enough to present conditions of the case. So we introduce a design of multi-lanes to show changing trend of multiple properties and elements of that case as time goes by, as is shown in Figure 1-(b).

3 COORDINATED VISUALIZATION DESIGN

Single data visualization presents limited information for analysis. So we introduce cooperative visualization to expand effective information presentation for analysis in our design.

3.1 Timeline and Social Network

Timeline mainly concentrates on events and their changing trend. However, events are operated by people, personnel changes of a group and organization are on account of events. People and events are always involved in social changes. So a comprehensive overview of both people and events relevant to the focal case can help investigators know about current circumstance and make cooperative analysis.

We combine design of timeline with social network and form them into a linkage integer. As time goes by, each event shows its own information and leads to changes in the relevant social network in Figure 1-(c). Investigators can see the dynamic changes of both events, topic trend and people involved.

3.2 Social Network Centering on Group

When analyzing social network, we may find some suspects. Looking for social cluster relevant to those suspects can be helpful, for this part of social network involved with suspects could be the key investigation subject. With certain information as conditions

to filter suspects, we consider personal properties, social activities they participate in and frequency between certain people, and integrate these information for analysis. In the end, we can find the specific social cluster relevant to suspects with much efficiency and possible accuracy. Then we mark these suspicious social clusters and highlight effective information dynamically.

4 COOPERATIVE ANALYSIS OF MULTIPLE VIEWS

Individual visualization graph presents certain information for analysis that fit for its own features. With various objectives to analyze, we design four specialized views for cooperative analysis.

4.1 Coordinate Graph view

This view in Figure 1-(a) is designed to find possible objectives. We pick up all the people's names from articles that contain certain keyword of objectives, then calculate the name's frequency and relevance. With two properties as coordinates, each point stands for an objective to investigate with word cloud relevant to this objective inside.

4.2 Multi-lane timeline & social network linkage view

Multi-lane view presents events and topic changes over time. Meanwhile, it is linked with a social network view. They are both interactive and coordinated. All points are sorted and summarized from articles with time and specialized information as properties.

As is shown in Figure 1-(b), different lanes contains different information to present. Checking the events may lead to changes in the extended social network. This linkage design shows evolution dynamically. The changing trend of organization movements and topic changes are also shown.

4.3 Cluster analysis view

This view contains people's personal information, email network with each topic as a unit and intimacy of contacting among all crew, displayed as parallel coordinate graph, chord graph and heat-map. They are linked with each other and work as a linkage integer for cluster analysis. This view can help analyze a cluster of people that meet certain conditions in specific groups. Email relations and intimacy of contacting show relevant information of current chosen cluster dynamically. Shown in Figure 1-(d), (e), (f).

4.4 Event timeline view

This view is also multi-lane, showing events and topic keywords changing trend. Topic changing trend over time shows the overall evolution of topics. As is shown in Figure 1-(g).

5 CONCLUSION

Our design help to find certain patterns to solve the kidnapping case of VAST Challenge 2014 mini challenge 1. And the major functions are fully implemented. As this system still has problems on interacting efficiency and automation, we could continue our work to refine the overall design.

REFERENCES

[1] Xiao Xu, *Pyteaser* (Python Algorithm for text summarization, based on *Textteaser* by Jolo Balbin, 2012) , Github, 2013

[2] Peter von Stackelberg, *Timeline of Major Events & Trends*, datavisualization.ch, 2010

[3] Dongning Luo, et al, *EventRiver: Visually Exploring Text Collections with Temporal References,* IEEE TVCG, Vol.18, No.1, 2012