

Exploring Anomalies in GAStech

VAST 2014 Mini Challenge 1 and 2

Jaegul Choo*, Yi Han*, Mengdie Hu*, Hannah Kim*, James Nugent*, Francesco Poggi† Haesun Park* John Stasko*

*Georgia Institute of Technology, †University of Bologna

ABSTRACT

We present our process and analysis for VAST 2014 Mini Challenge 1 and 2, which integrate an off-the-shelf tool, Jigsaw, rapid web-based visualization prototyping using D3, and analytics-based visualizations using Matlab.

Index Terms: H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Theory and methods

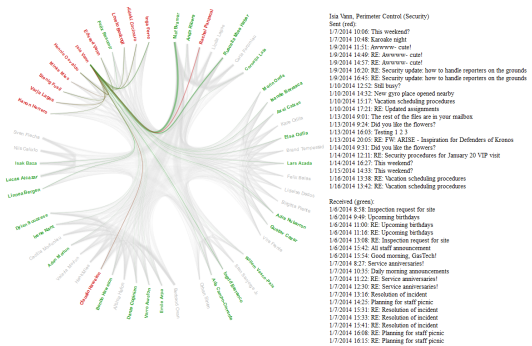
1 MINI CHALLENGE 1

The goals of MC1¹ were to understand the structure of the Protectors of Kronos (POK) organization, the connections of POK to the GASTech company, the series of events that occurred around the time of the challenge’s focus incident, and its potential causes. The provided data included semi-structured text documents (news articles, resumes, etc.), structured tabular text documents (email and employee records), an organization chart, and a map.

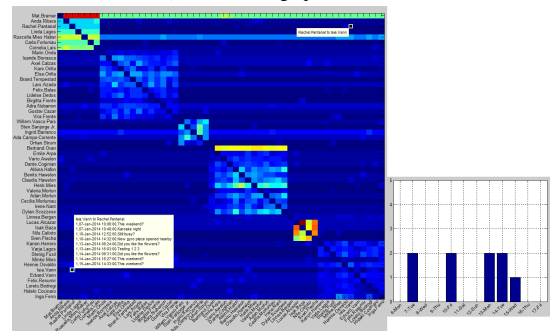
To examine the news articles and email messages, we used Jigsaw, a visual analytics system for exploring large text document collections [1]. We used Jigsaw’s List View to identify a list of news articles around the day of the incident to answer question 2. Subsequently, we deliberately inspected the identified documents in the Document View to find relevant information.

We believed that the email messages between employees would be best visualized with a graph to show the connections. This information was important for answering question 1d about connections between POK and GASTech. We used the Graph View in Jigsaw to identify the different employees involved in a specific email thread. However, the Graph view could not show an overview of the distribution of direct email exchanges between employees. Thus, we used two other visualization tools to explore the overall email connections between any given pair of employees. The first one was a circular graph visualization built using D3 toolkit. As shown in Fig. 1(a), the employees are grouped by their departments around the circumference of the graph. The amount of email between each pair of employees is encoded by the width of the connected line. Mousing over any employee, such as Isia Vann in Fig. 1(a), highlights all emails to and from this person in the circular graph and shows additional information on the side. We used this visualization to find potentially interesting connections between employees sharing last names with known POK members and other GASTech employees. The second tool was built with Matlab and employed an adjacency matrix visualization including interactive features to inspect the communication between any two employees. For example, in Fig. 1(b), Isia Vann, who we suspected to be connected with POK, seemed to be in a relationship with Rachel Pantanal, implying she may also be connected to POK.

* e-mail: jaegul.choo@cc.gatech.edu, yihan@gatech.edu,
mengdie.hu@gatech.edu, hannahkim@gatech.edu, jnugent6@gatech.edu,
fpoggi@cs.unibo.it, hpark@cc.gatech.edu, stasko@cc.gatech.edu



(a) D3 circular graph for emails



(b) Matlab matrix view for emails

Figure 1: Email analysis

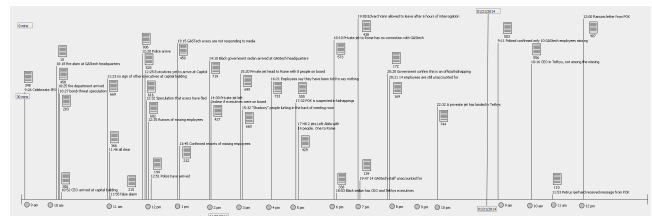


Figure 2: Tablet view for organizing events around the incident

None of these visualizations on its own was sufficient to answer any question in Challenge 1. Instead, we used information from combinations of the tools to gain a better understanding of the data. We did this by using Jigsaw's Tablet window and MS Powerpoint to visually organize our findings from the various visualizations together. The Tablet, shown in Fig. 2, provides a workspace for manually creating visual representations with lines, text boxes, and connections to documents. We created a timeline view in the Tablet for question 2 with notes and links to related documents. The view provides a flexible platform for visually organizing and presenting findings. For information learned outside of Jigsaw, we used MS Powerpoint to gather our findings.

2 MINI CHALLENGE 2

The task of Mini Challenge 2² was to find unusual patterns in employees' daily lives from their credit/debit/loyalty card records and

²Video: <http://www.cc.gatech.edu/gvu/ii/challenge/GT-14-MC2.mp4>

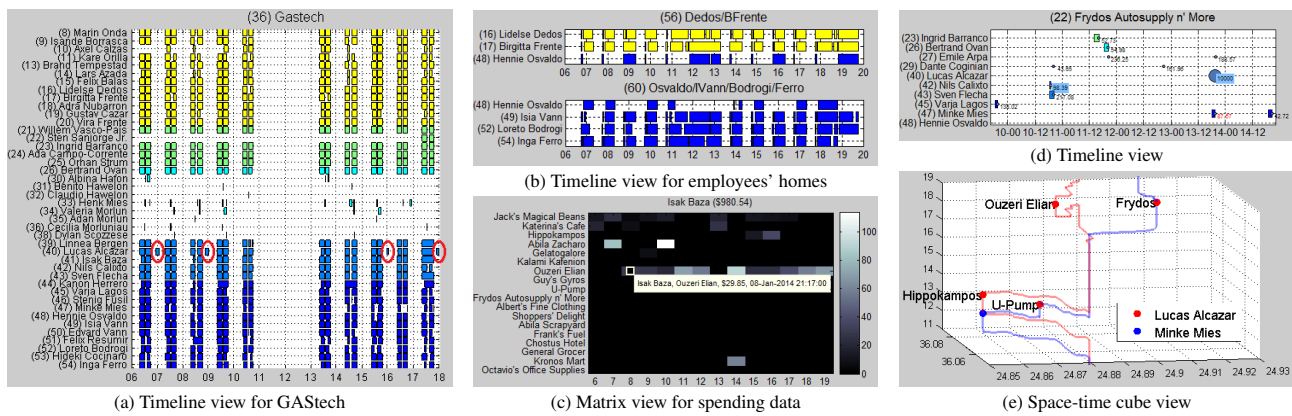


Figure 3: Analytics-based tool

the GPS tracking records of their cars. Given incomplete, inaccurate data, we first identified the exact store locations and the complete car-employee assignments. We then generated each employee's trajectory combined with his/her spending information.

Analytics-based tools. The identified information allowed us to perform analysis based on both particular locations and particular persons. For every location, we created a timeline view visualizing who stayed at the location over time, along with spending records (in the case of shops). Fig. 3(a) shows a timeline view for 'GASTech'. In this view, each row corresponds to a particular employee's timeline, and the time duration of his/her stay is represented as a rectangle along the horizontal timeline, color-coded by his/her department. This figure clearly shows regular business hours for most employees as well as Lucas Alcazar's anomalous pattern of often coming back to the office at night (red circles). Fig. 3(b) shows the timeline views of two places for employees' homes. One can see that Hennie Osvaldo sometimes stays at someone else's place at night. For particular employees, we provided the space-time cube view (Fig. 3(e)) showing their GPS trajectories in 3D where the x-, y-, and z-axes represent longitude, latitude, and time, respectively. Additionally, the matrix view shows the spending amount as a shop-by-day matrix for a particular employee and an employee-by-day one for a particular shop. For example, Fig. 3(c) indicates Isak Baza's regular visits to 'Ouzeri Elian'.

We provided flexible interactions among all the three views. For example, from the timeline view of 'Frydos Autosupply n' More' (Fig. 3(d)), where circles represent spending records with the radius proportional to the amount, we found an unusual spending of \$10,000 by Lucas Alcazar. Upon selecting it, the space-time cube that involves those employees visiting this place at the same time pops up (Fig. 3(e)). This interaction revealed that when Lucas' credit card was used at this place, he was not there while Minke Mies was. Furthermore, we can also see another transaction occurring in this manner at 'U-Pump', implying that Minke might have stolen Lucas' credit card.

Web-based tools. For further analysis, we developed two web-based visualization tools. The first³ provides a zoomable map of Abila, two sliders to filter days and hours, and a combo box to select vehicles. Circles represent shops, whose color indicates shop types (e.g., brown for cafe), and lines represent the vehicle trajectories. We used this tool both to analyze recurring employee's routines, e.g., Fig. 4(a), showing the routes of the employee commuting to GASTech in the morning during the weekdays) and to further inspect detailed unusual car movements or spending patterns. The second web tool⁴ is composed of an interactive heatmap that provides a spending summary and two coordinated bar graphs with details on each expense organized/filtered by employee, shop, and

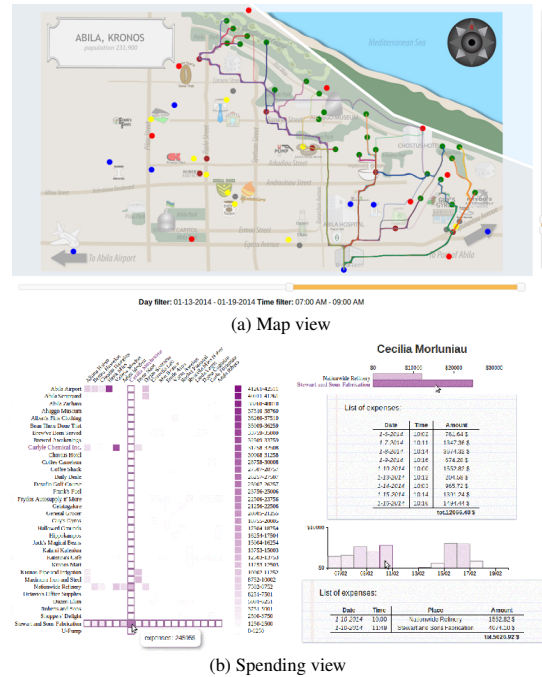


Figure 4: Web-based tool

date (see Fig. 4(b)).

3 FUTURE WORK

We plan to improve the analytics-based tools so that the basic interactions such as brushing-and-linking and details-on-demand types can be easily supported via native command-line interfaces of statistical analytics tools [2].

ACKNOWLEDGMENTS

This work was supported in part by NSF grants CCF-0808863 and IIS-0915788, the DARPA XDATA grant FA8750-12-2-0309, and the DHS VACCINE Center of Excellence.

REFERENCES

- [1] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. T. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Trans. on Visualization and Computer Graphics*, 19(10):1646–1663, 2013.
- [2] C. Lee, J. Choo, D. H. P. Chau, and H. Park. Augmenting matlab with semantic objects for an interactive visual environment. In *Proc. the SIGKDD Workshop on Interactive Data Exploration and Analytics*, pages 63–70, 2013.

³<http://eelst.cs.unibo.it/vast/map/>

⁴<http://eelst.cs.unibo.it/vast/heatmap/>