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.

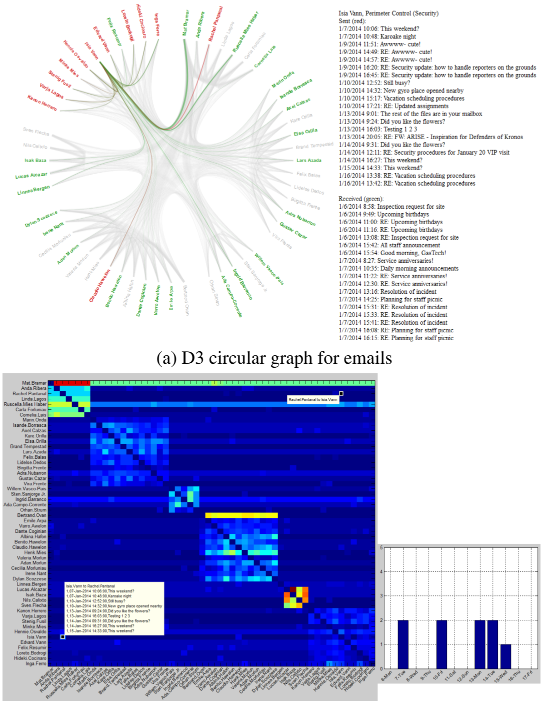
Index Terms: H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Theory and methods

1 Mini Challenge 1

The goals of MC1[1] 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.



Isia Vann, Perimeter Control (Security)
Sent (red):
1/7/2014 10:06: This weekend?
1/7/2014 10:48: Karoake night
1/9/2014 11:51: Awwww- cute!
1/9/2014 14:49: RE: Awwww- cute!
1/9/2014 14:57: RE: Awwww- cute!
1/9/2014 16:20: RE: Security update: how to handle reporters on the grounds
1/9/2014 16:45: RE: Security update: how to handle reporters on the grounds
1/10/2014 12:52: Still busy?
1/10/2014 14:32: New gyro place opened nearby
1/10/2014 15:17: Vacation scheduling procedures
1/10/2014 17:21: RE: Updated assignments
1/13/2014 9:01: The rest of the files are in your mailbox
1/13/2014 9:24: Did you like the flowers?
1/13/2014 16:03: Testing 1 2 3
1/13/2014 20:05: RE: FW: ARISE - Inspiration for Defenders of Kronos
1/14/2014 9:31: Did you like the flowers?
1/14/2014 12:11: RE: Security procedures for January 20 VIP visit
1/14/2014 16:27: This weekend?
1/15/2014 14:33: This weekend?
1/16/2014 13:38: RE: Vacation scheduling procedures
1/16/2014 13:42: RE: Vacation scheduling procedures

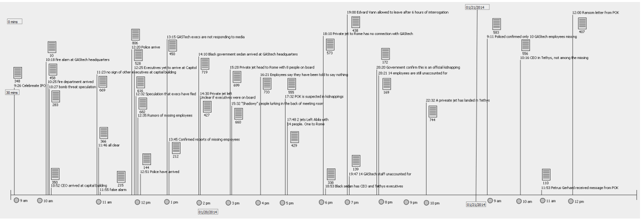
Received (green):
1/6/2014 8:58: Inspection request for site
1/6/2014 9:49: Upcoming birthdays
1/6/2014 11:00: RE: Upcoming birthdays
1/6/2014 11:16: RE: Upcoming birthdays
1/6/2014 13:08: RE: Inspection request for site
1/6/2014 15:42: All staff announcement
1/6/2014 15:54: Good morning, GasTech!
1/7/2014 8:27: Service anniversaries!
1/7/2014 10:35: Daily morning announcements
1/7/2014 11:22: RE: Service anniversaries!
1/7/2014 12:30: RE: Service anniversaries!
1/7/2014 13:16: Resolution of incident
1/7/2014 14:25: Planning for staff picnic
1/7/2014 15:31: RE: Resolution of incident
1/7/2014 15:33: RE: Resolution of incident
1/7/2014 15:41: RE: Resolution of incident
1/7/2014 16:08: RE: Planning for staff picnic
1/7/2014 16:15: RE: Planning for staff picnic

(a) D3 circular graph for emails

Isia.Vann to Rachel.Pantanal
1,07-Jan-2014 10:06:00,This weekend?
1,07-Jan-2014 10:48:00,Karoake night
1,10-Jan-2014 12:52:00,Still busy?
1,10-Jan-2014 14:32:00,New gyro place opened nearby
1,13-Jan-2014 09:24:00,Did you like the flowers?
1,13-Jan-2014 16:03:00,Testing 1 2 3
1,14-Jan-2014 09:31:00,Did you like the flowers?
1,14-Jan-2014 16:27:00,This weekend?
1,15-Jan-2014 14:33:00,This weekend?

(b) Matlab matrix view for emails

Figure 1: Email analysis



9:26 Celebreate IPO
10:18 fire alarm at GAStech headquarters
10:25 fire department arrived
10:27 bomb threat speculation
10:52 CEO arrived at capital building
11:23 no sign of other executives at captial building
11:46 all clear
11:55 false alarm
12:20 Police arrive
12:25 Executives yet to arrive at Capitol
12:32 Speculation that execs have fled
12:35 Rumors of missing employees
12:51 Police have arrived
13:15 GASTech execs are not responding to media
13:45 Confirmed reports of missing employees
14:10 Black government sedan arrived at GAStech headquarters
14:30 Private jet left
Unclear if executives were on board
15:20 Private jet head to Rome with 8 people on board
15:32 "Shadowy" people lurking in the back of meeting room
16:21 Employees say they have been told to say nothing
17:32 POK is suspected in kidnappings
17:48 2 jets Left Abila with
14 people. One to Rome
18:10 Private jet to Rome has no connection with GAStech
18:53 Black sedan has CEO and Tethys executives
19:00 Edvard Vann allowed to leave after 6 hours of interrogation
19:47 14 GAStech staff unaccounted for
20:20 Government confirm this is an official kidnapping
20:21 14 employees are still unaccounted for
22:32 A priveate jet has landed in Tethys
9:11 Policed confirmed only 10 GAStech employees missing
10:16 CEO in Tethys, not among the missing
11:53 Petrus Gerhard received message from POK
12:00 Ransom letter from POK

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.

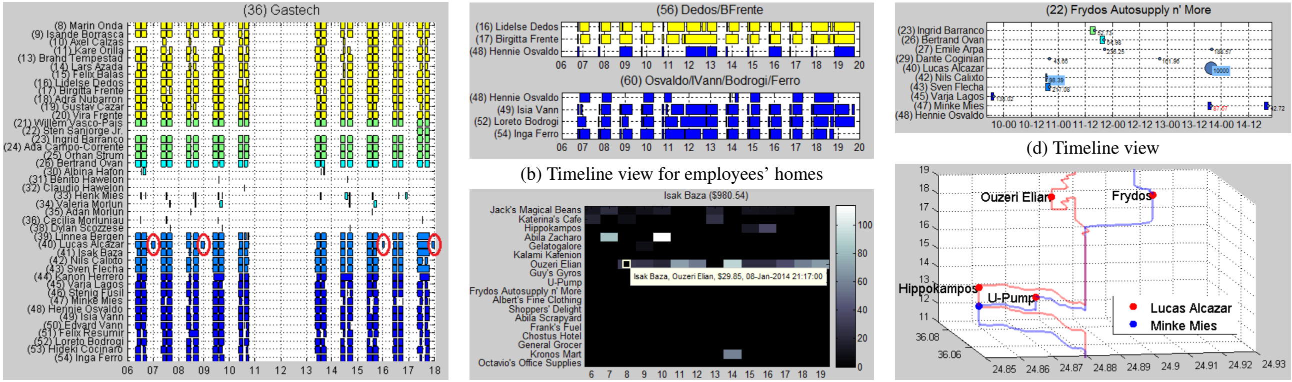
2 Mini Challenge 2

The task of Mini Challenge 2[2] was to find unusual patterns in employees' daily lives from their credit/debit/loyalty card records and

*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

[1] Video: http://www.cc.gatech.edu/gvu/ii/challenge/GT-14-MC1.mp4

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



(a) Timeline view for GAStech

(b) Timeline view for employees' homes

(c) Matrix view for spending data

(d) Timeline view

(e) Space-time cube view

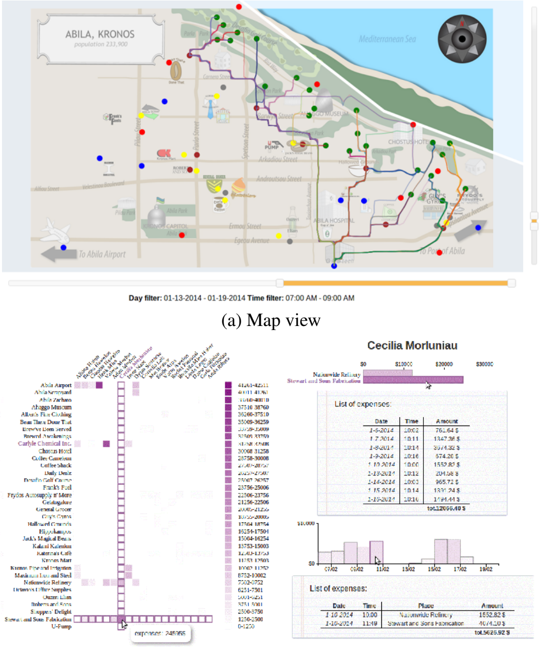
Figure 3: Analytics-based tool

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' houses. 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 first[3] 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 tool[4] 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 date (see Fig. 4(b)).



Day filter: 01-13-2014 - 01-19-2014 Time filter: 07:00 AM - 09:00 AM

(a) Map view

(b) Spending view

Figure 4: Web-based tool

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/