Visual Analytics for Detecting Communication Patterns

VAST 2015 Mini-Challenge 2: Honorable Mention for Compelling Narrative Debrief

William Hatton*
United States Air Force Academy

Jieqiong Zhao
Mahesh Babu Gorantla
Junghoon Chae
Benjamin Ahlbrand †

Hanye Xu
Siqaio Chen
Guizhen Wang
Jiawei Zhang ‡

Abish Malik
Sungahn Ko
David S. Ebert §
Purdue University

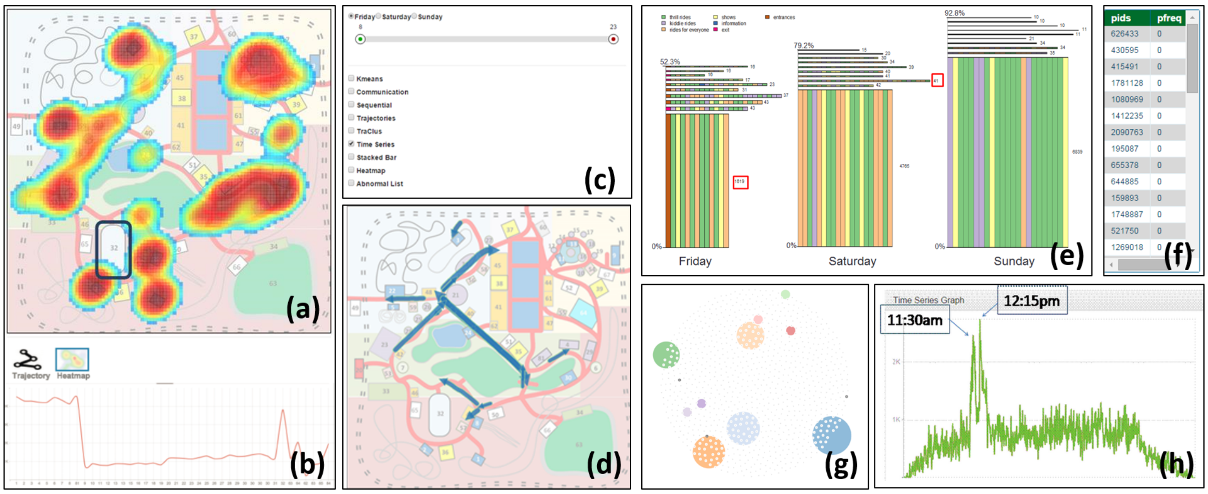

Figure 1: Our visual analytics system. Our system comprises of a map view (a), check-in frequency view (b), control panel (c), trajectory view (d), sequence clustering view (e), list view (f), clustering node-link view (g), and communication frequency graph (h).

1 Introduction

In the modern age, mobile devices provide a convenient way to send and receive messages. The meta-data associated with such messages (e.g., location, time, sender, receiver) can reveal communication patterns that provide insights into human activities and patterns. In this paper, we use the visitor communication data provided as part of the 2015 VAST Challenge to characterize the communication patterns of visitors in an amusement park over a three-day weekend. The challenge data features movement tracking (Mini-Challenge 1 (MC1)) and communication information (Mini-Challenge 2 (MC2)) of all visitors at the park. The theme of the weekend event for which the data was collected was around a star soccer player named Scott Jones in honor of his accomplishments. As part of this event, Scott appeared for two stage shows on each day of the weekend, along with a display of memorabilia related to his career at the park. Every visitor in the park was assigned a mobile device that was used to send and receive messages from others in the park. The objective of MC2 was to characterize the communication patterns of the visitors and park employees, compare different communication patterns over the three days, and discover anomalies or unusual behavior patterns that relate to a crime that occurred during the weekend. We utilized both movement data provided in MC1 and communication data provided in MC2 to answer the questions asked in MC2.

*e-mail: C16william.hatton@usafa.edu
†e-mail: zhao413|mgorantl|jchae|bahlbran@purdue.edu
‡e-mail: xu193|chen1722|wang1908|zhan1486@purdue.edu
§e-mail: amalik|ko|ebertd@purdue.edu

IEEE Conference on Visual Analytics Science and Technology 2015
October 25–30, Chicago, Il, USA

2 Visual Analytics Environment

In order to characterize the challenge data, we created a visual analytics system called ParkAnalyzer (Figure 1) that is comprised of several linked views to facilitate the exploration of the challenge data. The data are first preprocessed for individual days of the weekend and run through the movement clustering and community detection methods described in Section 3. The clustering results are then shown to users as node-link diagrams (Figure 1 (g)) and sequence clustering diagrams (Figure 1 (e)). Users can interactively select cluster subsets in order to further examine their characteristics in the other linked views. The system also utilizes density estimated heatmaps (Figure 1 (a)) to quickly identify check-in hotspots. Figure 1 (d) shows the trajectory view where we cluster trajectories into sets of similar sub-trajectories to discover common patterns [2]. Visitors that pass the filters applied by the users are shown in the list view (Figure 1 (f)). The communication frequency time series graph (Figure 1 (h)) shows the number of people who send or receive messages over time. Similarly, the check-in frequency time series view (Figure 1 (b)) shows the number of check-ins at the selected park attractions over time. The system also provides an interactive time slider widget (Figure 1 (c)) that allows users to temporally scroll through the data while dynamically updating the other linked windows. After several iterations of exploring the datasets, analysts can find the visitors with suspicious behaviors that may be related to the crime.

3 Communication Pattern Abstraction

In order to more effectively characterize the data, we apply several clustering techniques to group people that exhibit similar communication patterns. We also utilize the movement tracking data (MC1) along with communication data (MC2) to further analyze the patterns. We provide details on the techniques we apply to analyze these datasets in this section.

3.1 Clustering based on communication patters

We assume that people who communicate frequently among each other usually tend to belong to the same group (e.g., family members, friends, school field trips). In order to detect these groups, we utilize the community detection algorithm [1]. This method utilizes the network structure obtained from the communication data to discover sub-groups that are comprised of individuals that communicate frequently. Additionally, we also investigate whether people in different groups communicate among each another.

3.2 Spatiotemporal patterns in communication data

The spatial and temporal trends in the communication data can provide insights into visitor behaviors, and can also assist in the detection of anomalies. This method enabled us to discover anomalous communication spikes at certain times during the three day weekend. Furthermore, it becomes important to note that such analyses must be performed at appropriate data scales to achieve viable results. A data scale that is too coarse can overgeneralize the data, while a data scale that is too fine may limit the analysis due to data sparseness issues [4]. Accordingly, our system provides users with the ability to perform analysis at several pre-defined spatiotemporal resolutions to facilitate analysis.

3.3 Grouping people who move together

In order to discover the individuals that travel together, we apply a clustering technique to group the visitors who visit the attractions in the same order. We first utilize the longest common sub-sequence (LCS) [3] to measure the similarity of the sequences of two visitors. Next, we apply a density based clustering algorithm, to group visitors who check-in at the same attractions in a day into clusters.

4 Communication Data Analysis

In this section, we present our results from different aspects: communication patterns, identified communities, and crime situation. For analysis, we mainly utilize a communication frequency graph, clustering node-link view, and map view.

4.1 Communication Pattern Identification

Our analysis begins with a visualization of communication frequencies where we find that two individuals stand out for their large volumes of communication. We describe their communication patterns in detail. The person sends out a large group message at 12:00 PM each day, and another message to around the same amount of people every 5 minutes until the next hour (1:00 PM), and then waits an hour to begin again. The person continues the 5 minute interval messages for an hour followed by an hour long break through 9:00 PM (the last message is sent at 8:55 PM). On Friday and Saturday, the amount of messages that this person sends at 2:55 PM and 4:00 PM dips, likely due to the second performance of the day by Scott Jones. At this point, we can hypothesize that this person is a park employee simply sending out information to all the park participants. Since the amount of messages the person sends seem to fluctuate with the amount of people and the events of the park (Scott's second performance in particular), we can theorize that this person sends messages to park visitors that are not currently checked-in with information about which attractions are open or have small lines and/or other general park information.

The other person depicts a different communication pattern. This person sends between 5 and 20 messages every minute of each day. The only exception to this occurs at 12:00 on Sunday, when there is a huge spike of up to 1400 messages that slowly decline back to his/her normal over the next 45 minutes. Therefore, we hypothesize this person is also a park employee who deals with safety and security issues due to the constant activity throughout the weekend. Also, the spike in communication on Sunday would likely occur from the crime involving Scott Jones, which this person would be responsible to mitigate for.

4.2 Community Pattern Analysis

We found several patterns from the communications and movement tracking data. Small groups generally tend to communicate with other small groups. By examining the individuals that are clustered together using the communication clustering technique (Section 3.1), we find that many of them also travel together. In fact, there are several clusters of 7 or 8 people that contain smaller groups of 2-3 people who also travel together, but are communicating with everyone in the cluster. We hypothesize that these may be small groups that break off from bigger groups while in the park, or just people who make new friends and communicate with each other throughout the day.

4.3 Crime Analysis

The information about the total messages sent in the park shows how actively people are communicating over the three day weekend. To study this across the three days in an attempt to find unusual patterns that may possibly be related to the crime, we used time series graphs for total messages from park visitors on each day of the weekend. We found similar patterns throughout each day of the weekend, except between 11:30 AM and 12:15 PM on Sunday (Figure 1 (h)). On this day, it appears that there is a spike in communications in several different areas of the park. At first, the spike occurs in Wet Land, which reaches an unusual number of messages for that area. We know that the pavilion entrance is located in Wet Land. Since the crime occurred at the pavilion, we conclude that the crime was discovered at around 11:30 AM and the next half hour flurry of communications from that region was the result of the discovery. Also, with the heat map feature in our tool (Figure 1 (a)), the movement data shows that there are no check-ins to the pavilion after 12:00 PM. At 12:00 PM, we see a large spike in communication from the Entry Corridor. Since the two park employees sending messages (Section 4.1) send all their messages from the Entry Corridor, we can determine they likely sent messages notifying visitors the pavilion was closed at that time. Thus, since the pavilion was closed for the police to conduct their investigation (as stated in the news article), we further confirm that the crime occurred right before 12:00 PM and was discovered at 11:30 AM, the beginning of the communication spike.

5 Conclusion

Our visual analytics system that we developed as part of the 2015 VAST Challenge provided us with insights to characterize the communication patterns of visitors.

References

[1] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[2] J. Chae, Y. Cui, Y. Jang, G. Wang, A. Malik, and D. S. Ebert. Trajectory-based Visual Analytics for Anomalous Human Movement Analysis using Social Media. In *EuroVis Workshop on Visual Analytics (EuroVA)*, 2015.

[3] C. H. Elzinga. Sequence analysis: Metric representations of categorical time series. *Sociological methods and research*, 2006.

[4] A. Malik, R. Maciejewski, S. Towers, S. McCullough, and D. Ebert. Proactive spatiotemporal resource allocation and predictive visual analytics for community policing and law enforcement. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1863–1872, Dec 2014.