Dynamic DinoFun World Communication Graph

VAST 2015 Mini Challenge #2

Ting Li* Qi Liao†

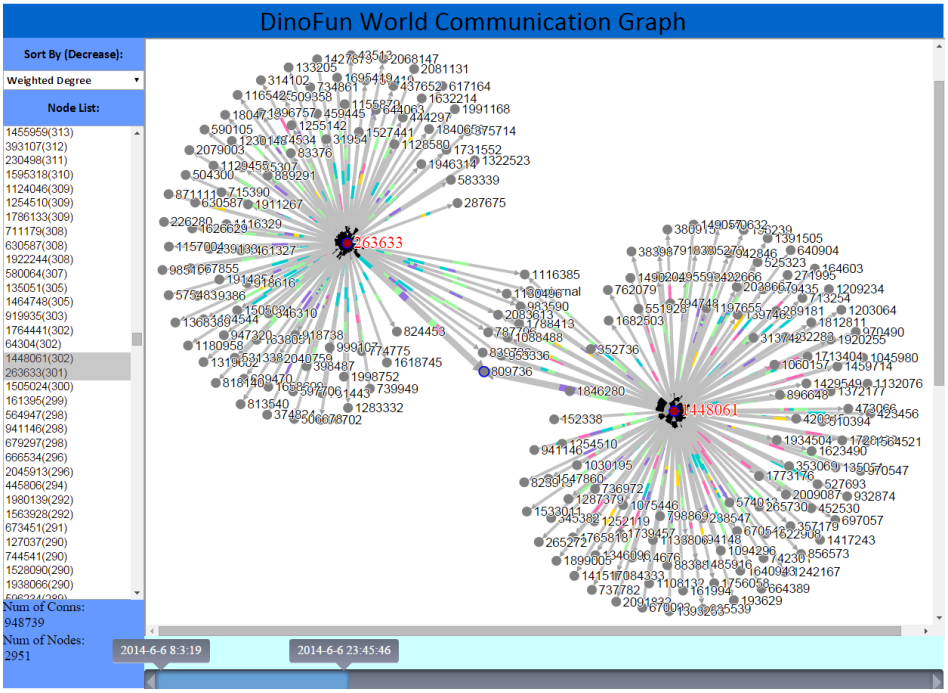

Figure 1: Overview of DWCG visualization tool. Communication IDs are sorted by degrees (left panel) and can be selected for analysis; a time bar (bottom) for narrowing down the investigation time period; additional information (e.g., number of involved IDs) are available (left bottom text); dynamic graphs show communication patterns of when and where (right panel).

Abstract

We designed and implemented DinoFun World Communication Graph (DWCG) using Data-Driven Documents (D3) technologies for solving Mini-Challenge 2(MC2) of VAST 2015. DWCG takes full advantage of communication edges by dividing them into segments based on the time and locations where communications occur. Through case study via the challenge dataset, we demonstrate how this novel design is useful to detect the spatiotemporal communication patterns in dynamic graphs.

1 Introduction

VAST Challenge 2015 MC2 includes three days' (Friday, Saturday and Sunday) communication data from DinoFun World Park, which contains 9430 IDs (park visitors or park services) and 4153329 communications. In each record, there are four attributes, i.e., time, from (the sender ID), to (the recipient ID) and location (area where communications occurred). The task is to find out interesting communication patterns and how the patterns may help to identify crime (vandalism) during one weekend in the park.

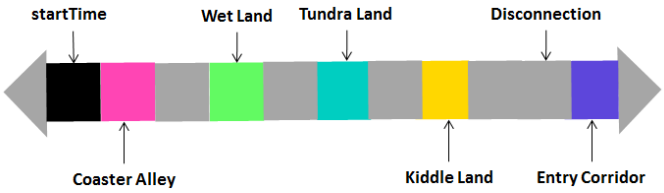

Figure 2: Each edge in the graph is divided into segments by time. Colors represent communication locations.

The challenge lies in how we show the dynamics and shift of patterns over time of a large number (millions) of communication records. For better human perception, we adopt a static view approach, i.e., using a single graph view for analyzing dynamic communication data. Interesting information will be focused on when and where those communications occur.

Graph visualizations are useful for many security problems [2]. We designed DWCG as a dynamic graph [1] problem. Nodes in DWCG are used to represent IDs and edges among IDs indicate communications. We utilize and divide the edges to show the detailed information about when and where each communication happens as discussed in the following sections.

*e-mail: li2t@cmich.edu, Department of Computer Science, Central Michigan University.

†e-mail: liao1q@cmich.edu, Department of Computer Science, Central Michigan University.

IEEE Conference on Visual Analytics Science and Technology 2015
October 25–30, Chicago, Il, USA
978-1-4673-9783-4/15/$31.00 ©2015 IEEE

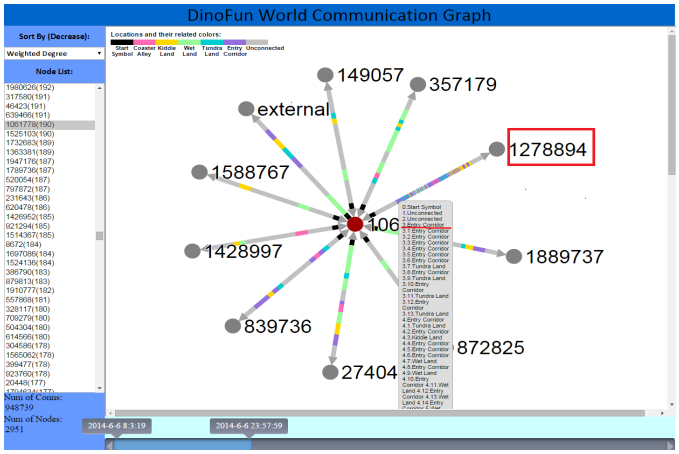

Figure 3: 1278894's communications with other IDs usually start at Entry corridor between 12pm and 9pm.

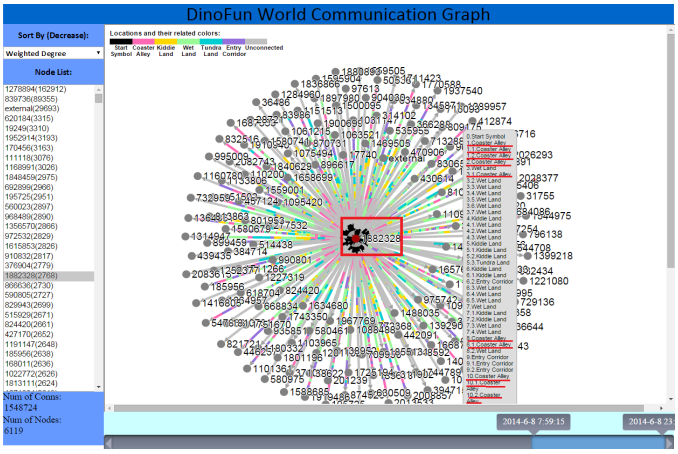

Figure 4: On Sunday, 1882328's communications with other IDs follow a pattern: a lot of its communications happen at Coaster Alley during 9 am to 12 pm and 8 pm to 10 pm.

2 DWCG DESIGN

The dynamics of communications among IDs reflect not only on communications' topology, namely connections may be created or dropped over time, but also on different locations that communications may occur. To show the dynamics of communications in the graph, we divide edges according to time. For example, in DWCG, each edge contains 11 segments. We use the first one segment to indicate the start time which is denoted with black color. If there are 10 hours of data in total, then each of the other ten segments will stand for one hour. In addition, we use different colors to denote communications locations. For example, pink, green, blue, yellow and purple colors are used to represent Coaster Alley, Wet Land, Tundra Land, Kiddle Land and Entry Corridor respectively. Gray color is used to indicate disconnection between two IDs. The dynamic graph visualization and color coding are illustrated in Figure 2.

Considering the magnitude of communications in the entire time period, we add a time bar into our visualization, so that user can choose whichever interesting time period to analyze. Once the time bar is selected, the total number of IDs and communications during the selected period will be shown in DWCG's left bottom text (see Figure 1). The visualization tool also sorts IDs in the selected period in descending order according to their weighed degrees (repeated connection will be counted) or unweighed degrees (repeated connection will not be counted). User can choose any interesting IDs in the left node list and related information about the selected IDs will be shown in the right graph. Figure 1 show the overview of the DWCG.

3 CASE STUDY

In this section, we illustrate DWCG through a few examples from the challenge dataset.

3.1 Pattern 1

When using DWCG to analyze MC2's data, we found that 1278894 is the ID which has the largest volume of communications. To find out its communications pattern, we analyzed all of its communications and learned that 1278894's communications always begin at Entry Corridor and only happen between 12pm and 9 pm everyday (see Figure3).

We learned from Figure3 that on Friday, the communication between 1278894 (highlighted with red color) and 1061778 starts at the third segment (around 12 pm) at Entry Corridor (purple color), and ends at the ninth segment (near 9 pm). Compared with other IDs, 1278894 has more frequent communications with 1061778.

From this pattern, we concluded that 1278894 is located at Entry Corridor. Since it only appears from around 12 pm to 9 pm, we assumed that it is a specific service in the park at Entry Corridor, such as a check-in system.

3.2 Pattern 2

From the data description provided by MC2, we knew that there were two shows daily in Park's Grinosaurus Stage, and the crime happened at Creighton Pavilion, both of them belong to Coaster Alley. On Sunday, there are many IDs which communicate with other IDs at Coaster Alley during specific time periods, shown in Figure4.

As we can see from Figure4, ID 1882328 (highlighted with red color) communicates busily with other IDs, such as 412874 and 1160780, and communications among them follow similar patterns. Namely, 1882328 tends to communicate a lot with other IDs at Coaster Alley (pink color) from around 9am to 12pm (1st-3rd segments). 1882328 also makes many communications at Coaster Alley near 8pm (8th segment), and it tends to communicate at Entry Corridor (purple color), Wet Land(green color), and again comes back to Coaster Alley around 10 pm. This pattern highlights the periods of 9 am to 12 pm and 8 pm to 10 pm on each day, which may be related to the show time and crime discovery time. Based on this pattern and along with other patterns, we made the assumption that the crime is discovered around 7:40pm on Sunday.

4 CONCLUSION

We developed DWCG as a dynamic graph visualization tool that is designed for VAST challenge 2015. DWCG shows advantages in revealing spatiotemporal communication patterns for each ID in terms of when and where the communication occurs. The separation of edges via colors helps us identify interesting patterns that lead to assumptions about when the crime was discovered in the park. In the future, we will try to add more efficient interactions into the visualization and apply them into other applications for further evaluation.

REFERENCES

[1] D. Eppstein, Z. Galil, and G. F. Italiano. *Algorithms and Theoretical Computing Handbook*, chapter chapter 8: Dynamic graph algorithms. CRC Press.

[2] H. Shiravi, A. Shiravi, A. Ghorbani, et al. A survey of visualization systems for network security. *IEEE Transactions on Visualization and Computer Graphics*, 18(8):1313–1329, 2012.