

AWARD: Efficient Toolkit Integration

Solving the Cell Phone Calls Challenge with the Prajna Project

Edward Swing*

Vision Systems & Technology, Inc.

ABSTRACT

The Prajna Project is a Java toolkit designed to provide various capabilities for visualization, knowledge representation, geographic displays, semantic reasoning, and data fusion. Rather than attempt to recreate the significant capabilities provided in other tools, Prajna instead provides software bridges to incorporate other toolkits where appropriate.

This challenge required the development of a custom application for visual analysis. By applying the utilities within the Prajna project, I developed a robust and diverse set of capabilities to solve the analytical challenge.

KEYWORDS: Information Visualization, Software Toolkit, Knowledge Representation, Toolkit Integration

INDEX TERMS: D.2.11 [Software Engineering]: Software Architectures - Domain-specific architectures; [Computer Graphics]: Methodology and Techniques - Interaction Techniques.

1 INTRODUCTION

This challenge involved a set of records of cell phone calls in Isla del Sueño over a ten-day period. Each cell phone call included only an ID of the caller and receiver, the cell phone tower that the call originated from, the date and duration of the call. This information was provided in a Comma-Separated Value (CSV) file. Supplemental information included general information about five high-level members of the Catalano-Vidro social network in Isla del Sueño. The geographic locations of the cell phone towers were not provided. Instead, I received only a map with estimated locations.

The challenge included questions about the social structure of the network. I was asked to identify the network at the end of the scenario, hopefully identifying key personalities. Furthermore, I needed to understand and characterize the changes to the social network over that ten-day period.

VSTI develops software for diverse customers who face numerous analytical challenges. The solutions developed for this challenge should apply to other analytical challenges that our customers must face.

2 DEVELOPING THE SOLUTION

2.1 Analysis of the Problem

This challenge clearly required a network display for social analysis. The temporal element implied that individuals would be

joining or leaving the network over time. Therefore, it would be necessary to incorporate a temporal component to the social network display.

I realized that the references to the tower locations might also provide clues. Therefore, I visually inspected the map provided, and estimated coordinates for the cell phone towers. These were stored in a separate data file.

2.2 Building with Prajna

The Prajna Project includes significant capabilities in a variety of technology areas. It also provides various software bridges to enable application developers to use specialized toolkits. Therefore, part of the design of the solution included selecting the appropriate elements of Prajna to include.

The first task for any analytical challenge is parsing the data. Prajna provides a data accessor for CSV records. This data accessor interprets the data according to a data description file, which identifies fields and their data types within each row of a CSV file. This utility significantly simplified the data ingestion process.

Initially, I applied Prajna's network visualization capability. During the development of this application, I developed several experimental graph arrangement techniques. Many social network displays use a Force-Directed display, and for that reason, I included a Prefuse Force-Directed display using the Prefuse bridge provided by Prajna.

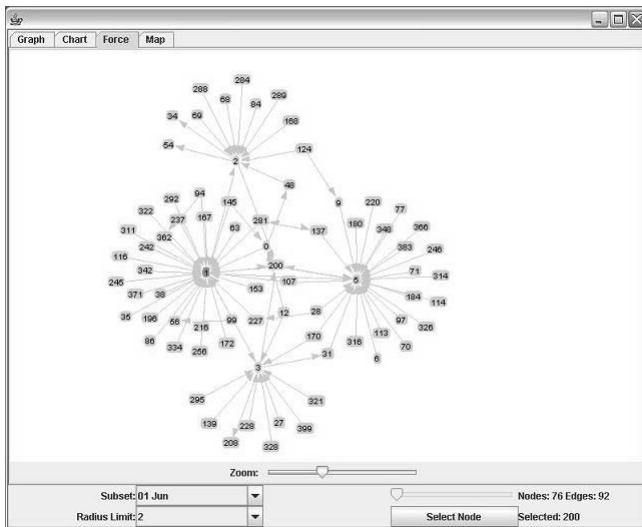
One initial concern with using a Force-Directed Display was that such displays – particularly Prefuse – are challenged when attempting to display disjoint graphs. However, after reviewing the data, I realized that the entire network of 400 nodes was in fact connected. This was a surprising feature of the data - a typical collection of phone calls would not necessarily form a connected graph.

In addition to both the Force-Directed and experimental graph displays, I realized that analysis of the caller's location, as determined by the cell phone tower, might provide useful information. I added the Prajna geographic display capabilities to the application, enhancing them to use arbitrary images for a background image. By interpolating the location of the caller among the cell phone towers accessed on any particular day, I was able to provide a geographic component to the network.

Finally, after reviewing the data, I realized that the temporal aspect of the data would also be important. I applied the Prajna bridge to JFreeChart to provide a temporal statistics display.. JFreeChart is an open source project that provides a Java toolkit for displaying graphical charts. I included charts which could illustrate calling frequency for the network.

To enhance the temporal capabilities further, I augmented the Prajna graph display with a time slider. A custom edge renderer highlighted the links between personalities over time, allowing me to finely trace the calling activity of the personalities in the network.

Email: deswing@vsticorp.com



Radius 2 Force Directed Layout on June 1

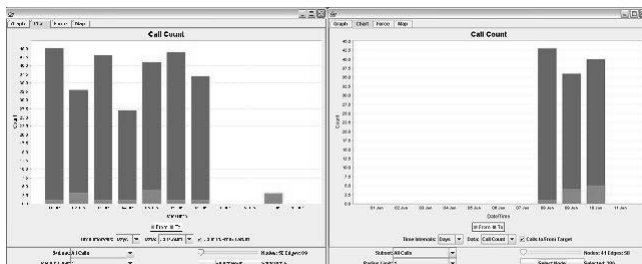
3 PERFORMING THE ANALYSIS

Once I completed the development of the application, I applied the capabilities of the tools to the challenge itself.

In order to understand the social network, I first displayed the entire network in the graph visualizations. This overwhelmed the Force-Directed Layout, and the other arrangements were unable to separate the nodes enough for any accurate analysis. Once I started filtering the display by selecting a central node and limiting the graph radius, I began to discern some patterns.

I knew Ferdinando Catalano was node 200, so he was the logical starting point. Identifying his neighbors, I quickly determined the identities of David Vidro and Esteban Catalano on June 1. No information distinguishing Jorge and Juan Catalano was available, so the two individuals were functionally interchangeable. I identified the two nodes which corresponded to them as well.

When I viewed the Force-Directed display for the entire 10-day period, I noticed two nodes (5, 306) apparently trying to occupy the same space, one of which was Esteban. The two nodes had almost identical neighbors, but did not actually communicate with each other. This suggested the two nodes might both be Esteban. I checked the pattern of both nodes over time, and realized that one ID was used through June 7th, while the other one was used almost exclusively after June 7th. This corroborated the theory that the two nodes represented the same individual.



Call Patterns of Esteban Catalano: ID 5 (left), 306 (right)

I noticed the nodes initially identified with the other principal personalities also disappeared on the same date. After analyzing the calling patterns of the neighbors, I could identify each of the new IDs that corresponded to the old IDs in the network. I also noticed during this analysis that Ferdinando Catalano's activity

level dropped somewhat after June 9th, and he did not contact the other principal personalities at all on June 10th. I do not have an explanation for this at this time.

I wanted to determine why this change took place. Viewing the call frequency on June 7th, I noticed a significant increase of call activity. I traced the calls within the network, and noticed that many of the subordinate nodes were calling the principal personalities, as if asking for information or instructions.

The final step was geographic analysis. I mapped the locations of the personalities over the ten-day period. While relatively stable until June 7th, all of the personalities moved after June 7th, sometimes significantly.

4 RESULTS

The table below lists the actual IDs of the primary personalities. The various clues I discovered during the analysis lead me to believe that some external stimulus affected the Catalano-Vidro network suddenly. I suspect that the network may have been reacting to a government crackdown, but further analysis with additional data might provide more information.

Name	Old ID	New ID
David Vidro	1	309
Estaban Catalano	5	306
Ferdinando Catalano	200	300
Jorge Vidro ?	2	397
Juan Vidro ?	3	360

Table of Primary Personalities

5 CONCLUSION

The tools developed for this challenge provided me with most of the answers I sought. Furthermore, I was able to apply the principles and design of Prajna to this challenge, demonstrating its utility. The Prajna project attempts to provide a robust toolset, leaving the development of sophisticated visualization tools for other toolkits. In this fashion, Prajna may adopt the best visualization techniques by providing a software bridge to innovative toolkits.

VSTI has already applied the techniques developed for this project to other development efforts within VSTI. The ability to trace cell phone calls, develop a social network, and represent the network over a geographic area offers a significant capability, which VSTI is actively applying to other analytical problems.

By providing an innovative architecture, which extends with software bridges to a variety of toolkits, Prajna avoids competing with the rapid pace of development across the spectrum of technology. Instead, Prajna offers developers the utilities to integrate new technology for knowledge representation in an intelligently designed architecture.

REFERENCES

- [1] E. Gamma, et al (the Gang-Of-Four), "Design Patterns: Elements of Reusable Object-Oriented Software", Addison-Wesley Professional, January 1995.
- [2] D. Gotz, M. Zhou and V. Aggarwal, "Interactive Visual Synthesis of Analytic Knowledge", Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, 2006.
- [3] J. Heer and M. Agrawala, "Software Design Patterns for Information Visualization", IEEE Transactions on Visualization and Computer Graphics (Proceedings), 2006.
- [4] JFreeChart Project (website): <http://www.jfree.org/jfreechart/>
- [5] The Prefuse visualization toolkit (website): <http://www.prefuse.org/>