

Short Paper

Lim Jin Ru (Alethea)
Singapore Management University
jinru.lim.2020@mitb.smu.edu.sg

Mayurapriyann Arulmozhi Baskaran
Singapore Management University
mayurapriya.2020@mitb.smu.edu.sg

Yeo Ding Run (Darius)
Singapore Management University
dingrun.yeo.2020@mitb.smu.edu.sg

Associate Professor Kam Tin Seong
Singapore Management University
tskam@smu.edu.sg

ABSTRACT

In the roughly twenty years that Tethys-based GASTech has been operating a natural gas production site in the island country of Kronos, it has produced remarkable profits and developed strong relationships with the government of Kronos. However, GASTech has not been as successful in demonstrating environmental stewardship.

In January, 2014, the leaders of GASTech are celebrating their new-found fortune as a result of the initial public offering of their very successful company. In the midst of this celebration, several employees of GASTech go missing. An organization known as the Protectors of Kronos (POK) is suspected in the disappearance, but things may not be what they seem. As experts in visual analytics, our group are called in to help the law enforcement from Kronos and Tethys using interactive visual analytics and R packages where relationships that led up to the kidnapping are analyzed through the visNetwork graph, anomalies with the Plotly scatterplots and boxplots of email headers and suspicious behaviors are identified through the igraph bipartite graph of credit card transactions and interactive movement path using tmap of vehicle GPS data.

1. INTRODUCTION

With the pressing time urgency in solving crimes, it is necessary to maximize the use of analytics tools to enhance the law enforcers' abilities to pick up hints and clues in order to speed up their investigation and to close the case. The R packages for data analytics provide an excellent tool for analysts to customize various visualizations according to the business use cases. The R packages share an underlying design philosophy, grammar, and data structures and this strong advantage allows analysts to combine various functionalities of different packages in an integrated manner.

2. MOTIVATION OF THE APPLICATION

The motivation behind this application is driven by the need for an efficient end-to-end environment of data pre-processing, data visualization creation and analysis as well as to host the visualizations on a web browser for the end users' convenient use, allowing them to perform further what-if analysis, manipulating parameters to see the impact of different approaches/perspectives without a coding background prerequisite. Shiny is an R package for developing interactive web applications and its benefits include ease-of-use, diversity, shareability and elegant design. It also makes development work more efficient especially since it has high reproducibility as end users can easily replicate or make tweaks with the source code. Shiny is also open source which reduces the barrier of entry for new users who would want access to a visualization tool that allows them to customize visualizations.

Our research and development efforts for this Shiny application were motivated by the interactive and user-friendly web-enabled client-based visual analytics tool to aid in the identification of anomalies, suspicious behaviours and patterns. The Shiny web application built for the analysis is highly customisable and easy to use with user interactivity on visualizations. It aims to provide the law enforcement agencies in Kronos and Tethys with an analytic tool for discovering visual behaviours and patterns of the network relationships in emails, credit card transactions and car movement paths. Specifically, it supports the following analysis:

- To perform community visualization by creating a bipartite network analysis of the credit card transaction data to understand the relationship between the credit cards and the merchants.
- To create visualizations on the stationary points using the vehicle tracking data to uncover the formal/informal relationships among the employees and to map the employees to their corresponding credit cards.

3. REVIEW AND CRITIC ON PAST WORK

There have been some existing network visualizations using R performed with the same GASTech employees' email dataset but there are still room for improvement especially in utilizing or integrating more functionalities of R packages.

For the review on community detection network visualization (Figure 1) there was no label to indicate the identity of

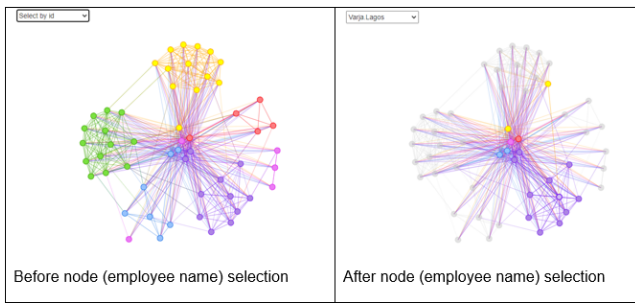


Figure 1: Network Visualization for Node Selection from Gao Yu (2020) [3]

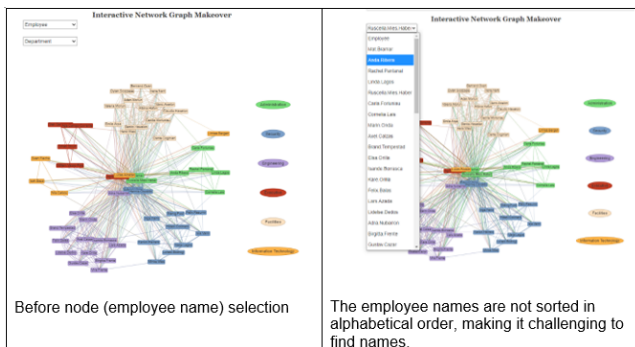


Figure 2: Network Visualization using Employee Name as Node from Christine (2019) [1]

each node (employee) and no label to explain the meaning of the colors. On the other hand, Figure 2's community detection network visualization contains labels which aid in the understanding of the identification of each node (employee) and a legend to explain the meaning of the colors. The edge link also retained their color after the node have been selected which could be confusing for users as there are too many lines that are not directly related to the nodes and nearest neighbors.

Both of the visualizations shown in Figure 1 and Figure 2 above did not provide other filter functions for the end users which would have otherwise allowed the end users more flexibility in zooming in on targeted analysis that they would like to perform.

A research paper by Seal & Wild [5] looked at the visualization of network plots using Shiny application. The application had a module on the network properties of the bipartite network graph such as modes, density, compute centrality measures such as degree, density, betweenness of the nodes and total number of interactions (Figure 3). This is useful to the users to help understand the centrality measures that will aid in their decision making and the visualization of the bipartite network graph.

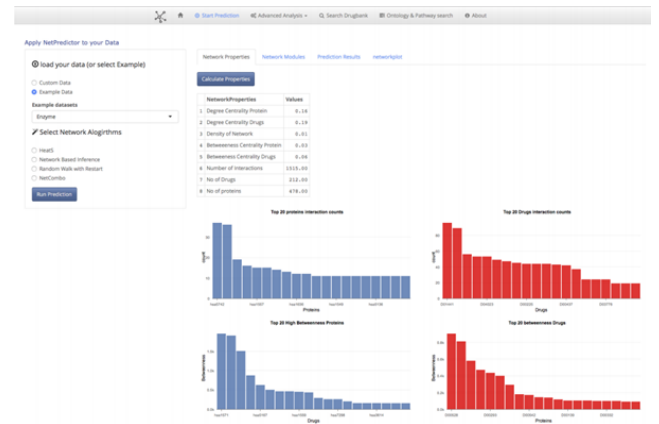
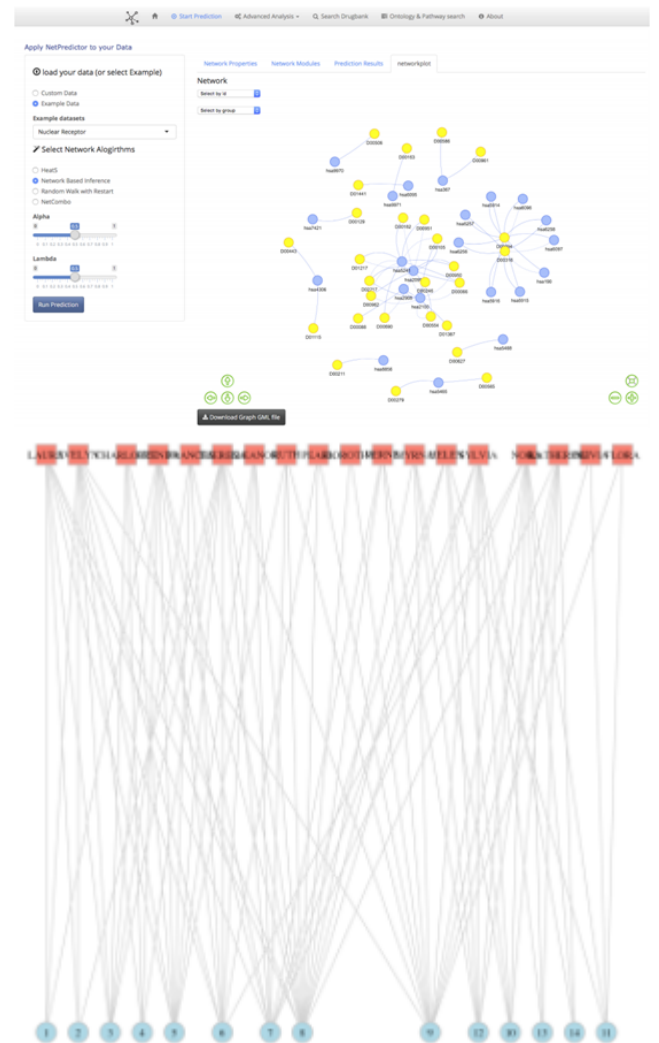


Figure 3: Network Properties from Seal and Wild (2018) [5]



The application also contains a network plot module (Figure 4) using the visNetwork package. The true interactions are denoted by the bold lines while the predicted interactions are displayed in dashed lines with drop down lists provided to select specific nodes and groups. Although this network plot module is useful for visualizing the network, it is not as intuitive as the other two network visualizations.

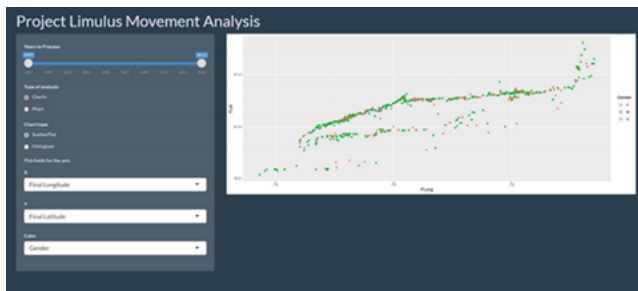


Figure 4: Movement Path Analysis' Shiny App from Youssef et al. (2019) [6]

plot is useful to show the network, the layout of the network visualization could be further improved to better show the bipartite relationship of the data using the bipartite network layout since the dataset used is a two-projected (bipartite) network of drugs and proteins. In view of this problem, we research to explore alternative visualization methods to analyze the bipartite network data in our credit card transactions dataset which contains the credit card numbers and the locations. A good visualization method will be using the bipartite-specific layout of the igraph package (Figure 5) to better visualize and represent the network that makes it easier to understand and incorporating features to draw interest.

Youssef et al [6] utilizes the leaflet package to visualize and analyse the movement of the horseshoe crabs. Shiny application is used to give the users the control to look at the chart view and map view of the crab movements. There is also an slider bar present to specify the year range and a dropdown to select the gender (Figure 6).

References

- [1] Christine 2019. IS428 Data Visualization Makeover2 Christine. (Nov. 2019).
- [2] Fenner, M. 2012. One-click science marketing. Nature Materials. 11, 4 (Mar. 2012), 261–263.
- [3] Gao, Y. 2020. In-class exercise 11: Network. (Apr. 2020).
- [4] Professinal Android 4 Application Development: 2012.
- [5] Seal, A. and Wild, D.J. 2018. Netpredictor: R and Shiny package to perform drug-target network analysis and prediction of missing links. BMC Bioinformatics. 19, 265 (Jul. 2018).
- [6] Youssef, I. et al. 2019. An R Analytics Tool to Visualize and Analyze the Movement of Horseshoe Crabs in Long Island Sound. 2019 IEEE Long Island Systems, Applications and Technology Conference (LISAT). (May 2019).