Interactive Network Visualisation of Important Evidence Linked to Missing GASTech Employees

Jessica Tan Soo Cheng
Singapore Management University
jessica.tan.2020@mitb.smu.edu.sg

Liu Nian
Singapore Management University
nian.liu.2020@mitb.smu.edu.sg

Lee Suling
Singapore Management University
sulinglee.2020@mitb.smu.edu.sg

ABSTRACT

This study covers the information and events of the Kronos Incident of the VAST2021 Challenge related to the missing GASTech employees. In face of the avalanche of data related to the incident, it would be necessary to use the best in class methods to efficiently decipher the data and detect the links in such time-sensitive analysis. For this purpose, a R-Shiny app is developed for interactive visualisation of the communication amongst GASTech employees across email and other online mediums. The network visualisation is presented in 3 parts: (i) email communication network, (ii) network of Credit Card Spending Locations; and (iii) Microblog Network Visualisation.

Through the employment network visualisation, the study is able to gain meaningful insights into the communication and activities of GASTech employees in the days leading up to the said disappearance of GASTech employees and their subsequent rescue by the police.

1. MOTIVATION

Using interactive R-Studio's interactive network visualisation packages, this study aims to showcase the use of network visualization to organise the information from various mediums such as emails, social media tweets, micro-blog and credit card transaction data into interactive and user-friendly formats for easy deciphering of the data that can enable the user to trace the links to the missing employees. The study also attempts to provide a more software-reliant approach based on R-tools for network detection and visualisation compared to the mostly Excel and Visio studio-based processes used in the VAST2014 Challenge on the same subject.

2. LITERATURE REVIEW

The followed literature was reviewed for this study:

Tianjin University ("TJU_Cai-MC2") – Outstanding Results Presentation Supported by Visualizations*:* This top-prize winning team of 6 student-cum-faculty members spent 180 hours using D3, MySQL, Excel and Visio to chart and link the dots between events and individuals over timelines to identify key individuals and the organizational network of POK. Elaborate charts and graphs were made to map out the inner circle as well as extended network of POK, identifying their leaders and contributors, as well as explaining the occurrence of important events. In particular, the relationships and interpersonal linkages of the POK with GASTech were detected via visualisation of news and email data.

Tianjin University ("TJU-Gao-MC1") - Honorable Mention for Effective Use of Coordinated Visualizations: This 5-student team spent 150 hours using D3.js, Highcharts, Vis., Python, Excel and Visual Studio to graphically address the MC1 questions. Drawing up well-coordinated and aesthetically pleasing graphs, the team traced and identified the key members of POK, their infiltration of GASTech as employees as well as the events on the timeline leading up to the disappearance of GASTech's executives. They provided strong persuasion that POK is behind the disappearance of the GASTech executives.

University of Buenos Aires ("UBA-Avila-MC1") – Honorable Mention for Effective Timeline Visualizations: This 3-member student team spent 60 hours using Tableau, JavaScript, Excel and Rapidminer to respond to the MC1 requirements. Through Tableau's smooth graphics, the leadership team of POK were identified and their affiliates working within GASTech were extracted and linked by visualizing news and email data. The submission lacked the detailed analysis and intricate charting of the prior 2 submissions but perhaps have its simplicity and easy comprehension as reasons for the Honourable Mention.

Tianjin University("TJU-Yang-MC3") - Honorable Mention for good support for streaming and forensic analysis: The dashboard consists of a timeline, wordcloud, map and display of the tweets at the bottom allowing the user to view the events as they unfold by the timeline. If the component plots are linked and filtered, the user will be able to perceive the information effectively. However, it is not seen from the dashboard how RT, mentions and hashtags were analysed. In a social network, the "wisdom of the crowd" can be revealed in examining frequent retweets or mentions. Moreover, following influential and knowledgeable authors can lead to quicker identification of important events instead of filtering and examining the many messages to extract the required information manually.

3. DESIGN FRAMEWORK

The application is a R-Shiny app designed for interactive network visualisation of the activities of GASTech employees and local communities related to the disappearance of some GASTech employees. The network visualisation is presented in 3 parts:

(i) Email communication network of GASTech employees
(ii) Network of Credit Card Spending Locations; and
(iii) Microblog Network Visualisation.

By accessing communication in GASTech's email and microblog text messages as well as tracing the transaction activities and locations of key individuals, the user would be able to detect meaningful patterns and important insights to assist with uncovering the causes leading to the disappearance of the GASTech employees.

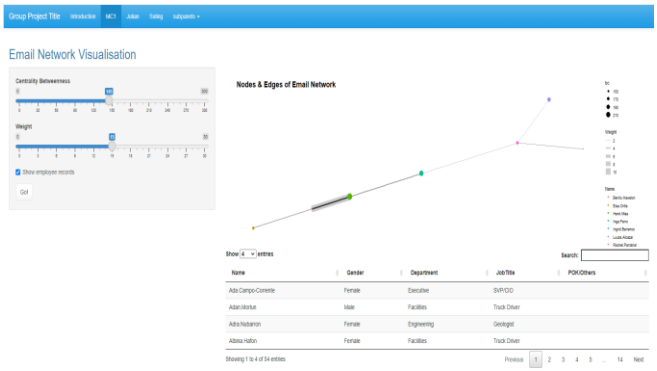
3.1 GASTech Email Network

The network visualisation of GASTech's email communication covers the 2 work-weeks from 6th to 17th January 2014 leading up to the disappearance of some of GASTech's employees over the weekend of Sunday 19 Jan 2014. For visual display of the work and social communities interacting on GASTech's email platform during the 2 work-weeks, the procedures of social network analysis were undertaken to derive the nodes and edges as well as measurements of centrality betweenness and weights of GASTech's email network.

The deployment of the email network visualisation on R-Shiny takes the form of a graphical display of the network nodes and edges connecting the employees communicating on the email platform. Users accessing this module of the R-Shiny app could select and define the centrality betweenness and weights of the network visualisation to identify the core participants in the email network from which the emails were initiated and/or generated. In addition, a data table providing key information on the employees' background is retrievable with a tick in the user panel. The user could input the names of the key email participants into the search panel of the data table to obtain their background information, such as work department, job title and POK connections (activist group suspected of involvement in the disappearance of the employees), if any. For viewing of GASTech employees linked to POK, the user could click on the upward arrow on the POK column title in the data table to retrieve the list.

Overall, this module of the R-Shiny app would be useful for detecting deviant communities and/or abnormal communication patterns in GASTech's email network to gain insights related to the disappearance of the GASTech employees.

Figure 1: Email Network Visual

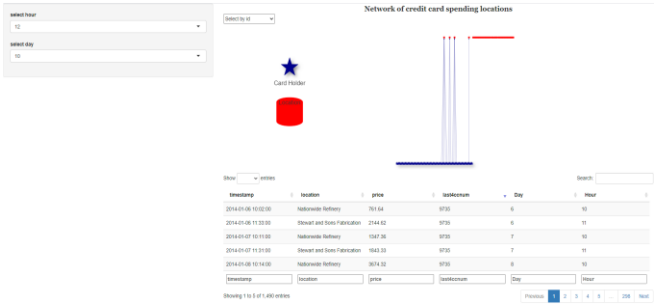


3.2 Network of Credit Card Spending Locations

The network visualisation of credit card transaction data covers the 2 calendar weeks from 6th January 2014 to the disappearance of GASTech employees on 19 January 2014. The visualisation is built on the principles of bipartite graph theory in which the set of cardholders and the set of locations of card expenditure are designated as the bipartite graph nodes with the timestamp tracked for the card spending as edges linking the cardholders with the spending locations.

The implementation of the network of card spending locations in R-Shiny takes the shape of a top-down bipartite link graph in which card locations at the top portion of the graph is connected to cardholders at the base of the graph by lines representing the timestamp for card spending incurred at the linked locations. The user is also supported with the use of a data table providing details of the card transaction. The user could also retrieve the bipartite graph for specified card ID, or by hour and day of the card transaction through selection in the user panel.

Figure 2: Bipartite graph of card spending locations



3.3 Microblog Text Message Network

The network graph of text messages is a collection of tweets and call centre emergency dispatch records from 17:00 to 21:35 on 23rd January 2014, which reveals a sequence of events that led to the capture of kidnappers and release of hostages related to the earlier disappearance of GASTech employees.

Typical of a social network, tweets messages can contain any number of mentions, RTs, hashtags, and links. These features are important as they can reveal real-time or trending information based on the "wisdom of the crowds". For example, messages with high RTs shows that the network is focused on at that moment and influential authors are those with many mentions.

Selection of data sets by these features and filtering by groups are interactive functions that were build into the application. The Shiny application also allows for the graph to be animated, thereby allow users to view changes in the network over time. Linking the network to the data table allow users to further drill down the details of the events.

Figure 3: Network graph of text messages

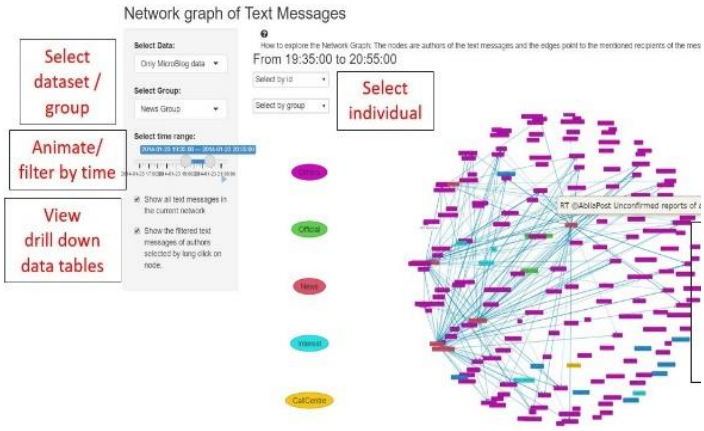
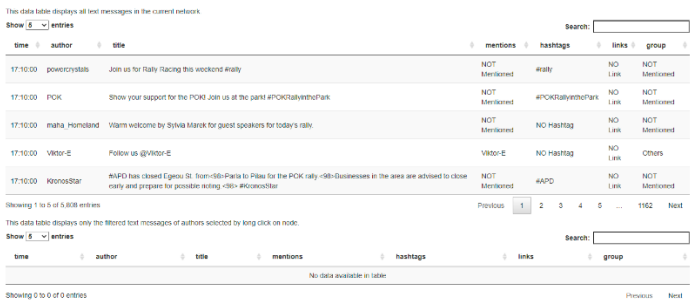


Figure 4: Data table displaying network text messages



4. DEMONSTRATION – USE CASE

The following is the demonstration of a probable use case for the R-Shiny app covering its 3 components of (i) GASTech email network, (ii) network of card spending locations and (iii) microblog text message network.

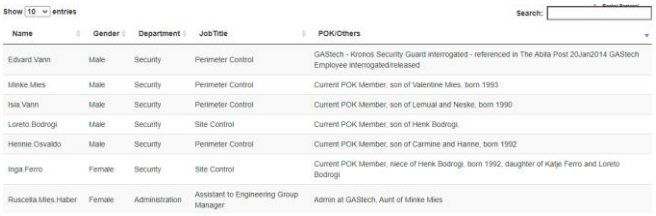
(i) GASTech email network: To obtain a quick review of unusual activities on GASTech's email network, the user could select the centrality betweenness and weight of the email interactions in the app module's the user panel. High centrality betweenness indicates that the email participant (as node) is a "bridge" in the network. Hence these nodes facilitate communication between participants from different clusters in the network. Based on emails generated, the weight denotes the number of participants a node communicates with in the network. Thus, weights can show the influence of the node in the network.

Figure 5: Use case review of unusual email activities

In the use case example at Figure 6 above, the centrality betweenness is set at 160 to detect heavy communication on the email network. It then became apparent that the most frequent communication during the 2 weeks leading up to 19 January 2014, the eventful day of GASTech employees' disappearance, occurred through Henk Mies (driver) and Inga Ferro (security staff and POK member). Incidentally, Mies is also the family name of senior POK member Valentine Mies. At the same time, there had also been frequent emails between Inga Ferro and Rachel Pantanal, who is one of the employees found to be missing on 19 January 2014.

The user could further retrieve the entire list of POK members who have infiltrated into GASTech as employees by clicking on the upward arrow in the POK column tab of the attached data table. In the data table extract at Figure 7 below, it is noted that 5 of GASTech's 11-member security team are POK members whilst another security staffer, Edvard Vann is a close relative of senior POK members Mandor and Isia Vann.

Figure 6: List of GASTech employees with links to POK



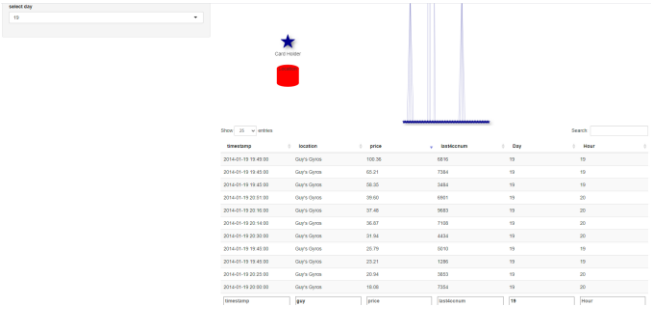
Name	Gender	Department	JobTitle	POK/Others
Edvard Vann	Male	Security	Perimeter Control	GASTech - Kronos Security Guard interrogated - referenced in The Abila Post 20Jan2014 GASTech Employee interrogated/released
Minke Mies	Male	Security	Perimeter Control	Current POK Member, son of Valentine Mies, born 1993
Isia Vann	Male	Security	Perimeter Control	Current POK Member, son of Lemuel and Neske, born 1990
Loreto Bodrogi	Male	Security	Site Control	Current POK Member, son of Henk Bodrogi.
Hennie Osvaldo	Male	Security	Perimeter Control	Current POK Member, son of Carmine and Hanne, born 1992
Inga Ferro	Female	Security	Site Control	Current POK Member, niece of Henk Bodrogi, born 1992, daughter of Katje Ferro and Loreto Bodrogi
Ruscella Mies Haber	Female	Administration	Assistant to Engineering Group Manager	Admin at GASTech, Aunt of Minke Mies

(ii) Network of card spending locations: For easy detection of unusual card spending patterns by GASTech employees in the 2 weeks leading up 19 January 2014, the user could access the card spending locations network module of the R-Shiny app by selecting the date and hour to view the bipartite graph displaying the links between card spending locations and the cardholders.

In addition, to view any unusual gathering of GASTech employees by their spending locations, the user could simply

scroll the pages of the data table as an initial step to gleam the frequented locations and subsequently input the location names input the search tab of the data table to retrieve the details which would include the timestamp and card spending for the location. For example, by inputting "Guy's Gyros" into the location column search tab, it is found that 11 GASTech employees had been gathered at the Guy's Gyros from 19:45 pm to 20:51 pm on 19 January 2014, the eventful day of GASTech employees' disappearance.

Figure 7: Card spending showing large employee gathering



(iii) Microblog Text Message Network: To visualise the fluid exchanges on social media platforms, the user could select the microblog dataset on the R-Shiny module and select the time period to display the mentions network graph and data table. Users could further drill down on the network by clicking on the nodes of interest to dynamically display the text messages in the data table below. Using the animate button, users can also see how the network evolves over time. In particular, we see a spike in network activity after 18:50 where several messages from the news sources about a fire was retweeted. There was another spike in activity in the Interest Group of nodes as they posted messages about their first-hand account of the police shoot off in real-time.

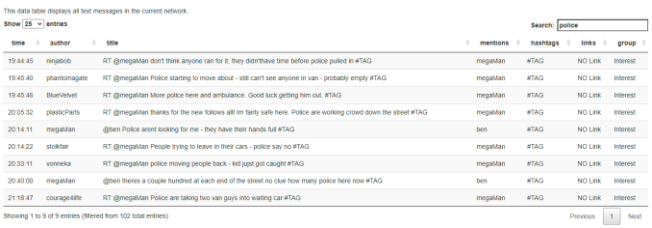
Some key events identified by analysing the network are as follows:

Figure 8: Key events and the associated Network Graph settings

Time	Event	Network Graph
17:00 – 19:08	POK Rally	Select for "Has hashtags" data set, shows #POKrally to be trending.
18:42 – end	Fire at Dancing Dolphin	Select for "Has mentions" and "News" shows that news sources have the most mentions. Drill down on specific new sources like "Homeland Illuminations" to obtain key reports of fire.
9:37 – 21:20	Police standoff and hostages released	Select for "Has mentions" and "Interest Group" shows how their network of mentions evolves over the evening. Drill down on their messages

Figure 9 illustrates the data table displayed following selection of "Has Mentions" and "Interest Group" in the user panel as well as the word "Police" inserted into the search tab of the data table.

Figure 9: Breaking news and information



From the above data table, we see "megaMan" reporting on the police shootout as an observer based on the tweet exchanges displayed.

5. KEY INSIGHTS & OBSERVATIONS

From the above methods used to decipher information and detect communities related to the disappearance and subsequent rescue of the GASTech employees, it is apparent that the enormous analytical tasks of organising and visualising the huge and varied collection of data had been substantially simplified by the use of network visualisations compared to the previous approaches employed by participants in the 2014 Vast Challenge on the same subject.

In particular, the use of network visualisation for detecting abnormal patterns and informal communities interacting on GASTech's email network grants the user almost immediate access to the identification of individuals who may have links to the disappearance of GASTech employees, such as the POK members who have joined the GASTech security team and heavily used the company's email network in the days leading up to the disappearance.

In addition, the visualisation of card spending data through bipartite graphical display had been critical in providing the evidence of irregular gathering of a large number (11) of GASTech employees at Guy's Gyros on 19th January 2014, the day which some GASTech employees were found to be missing.

Using network graph to visualise the social media exchanges of text messages not only reveals the sequence of events but also the relationships between participants in the network. For example, we found by animating the network that messages by the News group were the most mentioned and retweeted especially at the start of crisis situations like the fire and police shoot off. We see the mentions network rapid growing when the fire started with different reactions from the network, ranging from support for the Fire department to speculation to despair more homes were evacuated.

Filtering for various data set reveals that a lot of spam text messages contained links but none of them has mentions or RTs. We also saw that the mentions network was stable with the same participants mentioned and retweeting over each period of time (persistent edges in the network when animated). Interestingly the official channels of Abila Fire Department and Police Department did not have many

mentions despite their updates about the fire and shoot out. We suspect that this is due to the relatively low followers of these channels, but this cannot be confirmed as we do not have the followers information in the network.

Based on the above observations and insights gained, the analysis and visualisation of data surrounding the GASTech employees' disappearance had been a fruitful analytics exploration for the study team.

6. FUTURE WORK

The analytical methods and approaches employed in this here could be extended to other datasets such as the Twitter social media network, shopping transaction records etc. In our development of the R-Shiny app, we found that some of these areas could be improved upon, but were not executable due to limitations of our datasets, such as:

(i) Allowing for more centrality measures to reveal different aspects of influential nodes in the network as different centrality measures can reveal different aspects of the key nodes identified by high centrality in the network.
(ii) Linking transaction spending to a location map for more intuitive visualisation of clusters of location patterns can better how the movement of personnel
(iii) Including followers as receiver nodes in the network for text messages so that messages with no mentions can also be displayed in the graph network of that a more comprehensive view of the network can be created.

7. CONCLUSION

The study has showcased 3 different network visualisations and illustrated how they could be applied to uncover important insights using different datasets. In particular, network graphs have enabled the visual display of vast amounts of data in intuitive and interactive formats which substantially expediate the process of social community and network detection on online media platforms.

8. REFERENCES

[1] Entry Name: "TJU_Cai-MC2", VAST Challenge 2014 Mini Challenge 1. https://www.cs.umd.edu/hcil/varepository/VAST%20Challenge%202014/challenges/MC1%20-%20Disappearance%20at%20GASTech/entries/Tianjin%20University%20-%20Cai/

[2] Entry Name: "TJU-Gao-MC1", VAST Challenge 2014 Mini Challenge 1. https://www.cs.umd.edu/hcil/varepository/VAST%20Challenge%202014/challenges/MC1%20-%20Disappearance%20at%20GASTech/entries/Tianjin%20University%20-%20Gao/

[3] Entry Name: "UBA-Avila-MC1", VAST Challenge 2014 Mini Challenge 1. https://www.cs.umd.edu/hcil/varepository/VAST%20Challenge%202014/challenges/MC1%20-%20Disappearance%20at%20GASTech/entries/University%20of%20Buenos%20Aires%20-%20Avila/

[4] Entry Name: "TJU-Yang-MC3", VAST Challenge 2014, Mini Challenge 3. https://www.cs.umd.edu/hcil/varepository/VAST%20Challenge%202014/challenges/MC3%20-%20Real-Time,%20Streaming%20Social%20Media/entries/Tianjin%20University/