VISUALISATION DASHBOARD-
Visualising Internal Emails of GAStech

LIM Hui Ting

Jonathan Eduard Chua LIM

Abstract - **Email has changed the way coworkers communicate in the workplace. With millions of emails exchanges occurring per day within an organisation, the company's server has accumulated a sizeable email archive over the years. Running analysis on these emails would allow company to gain valuable insights about both internal and external communication of the company. However, companies tend to focus more on external communication to its various stakeholders and has little emphasis on the analysis of internal emails conversations. With the various benefits that analysing internal emails can bring, we seek to explore the connectivity and frequently discussed topics among employees of an organisation by using GAStech email exchanges as our case study. A visualisation storyboard is created using force directed graph and concept map. Users will be able to have a better overview of the connectivity and frequently discussed topics among the coworkers as the storyboard will equip them with an application that can highlight the data and a tooltip at the side to provide additional information.**

Index Terms – Graph Visualisation, GASTech, Email Analysis, D3.JS

INTRODUCTION

Email has revolutionised both internal and external communication in the workplace since the late 1990s. Every day, millions of emails exchanges occurred between the companies and its various stakeholders, and among the coworkers. Important information is transmitted and disseminated more efficiently thus allowing a continuous flow of information among the relevant parties. Efficient communication via email also allows employees and executives to be able to provide timely response to any issue that may arise. With the onset of mobile and cloud revolutions, coworkers are no longer limited to using desktop to access their email accounts. They will be able to access the email server wherever they are.

These email communications left a virtual paper trail that allows a more efficient retrieval of important information as compared to printed documents. As the company's server has accumulated a sizeable email archive over the years, running analysis on these emails would allow company to gain valuable insights about both internal and external communication of the company.

MOTIVATION

Companies usually neglect internal emails analysis as they tend to focus more on external communication, such as email marketing, in order to send better and more targeted emails to their customers. Insights into how employees interact within the organisation are equally valuable as it allows the company to understand its internal communications as whole.

Furthermore, emails are often used for investigations to uncover foul play of various types. For example, two founders of a Syracuse development company are being suspected of engaging in bribery based on their email exchanges. This means that analysing internal emails will be able to help the company to identify any potential red flags with regards to employees' misconduct or performance.

With the various benefits that analysing internal emails can bring, we seek to analyse the connectivity and frequently discussed topics among employees of an organisation by using GAStech email exchanges as our case study.

The objectives that we would like to address through the development of a dashboard are as follows:
1. To understand the interactivity among GAStech employees
2. To visualise keywords used in the email headers that are of interest to the organisation

RELATED WORKS

As inspirations to our final product, our group had considered a handful of visualisations. One of these visualisation is word cloud. A word cloud illustrates how a word or a phrase is coded by attributes such as colour, font size or weight can imply its importance or the frequency of appearance (Halvey, & Keane, 2007). This is relevant to us

because our dataset consists of texts in the form of email subject headers. Though the words and phrases available in an email subject is lesser than the actual body of the email, there are still insights that can be obtained from email subject headers. The ability of being able to visualise key words and phrases in our final visualisation will enable the end user to gather insights and discover the relationship and communication within GAStech Kronos.

Figure 1 - Word Cloud Visualisation from Wordle.com

Figure 1 is an example of a word cloud visualisation generated by Wordle, a website specialising in word cloud, from a Donation Usability report. This report outlines how a Non-Governmental Organisations (NGOs) and charities are able to attract donors and volunteers into their website, in an effort to entice them to participate. The contents of the reports were obtained from surveys from 60 charities. Key words are seen, such as "users" "donation" "homepage" which are prominent word that can be seen, and these words can be beneficial to charities and NGOs into thinking about how to keep the community engaged. The words are also in different colours, which allows the users the ease of reading the words in a limited space.

On the other hand, there are redundant words present in this visualisation. Words such as "one", "two" and "page" does not give meaning to the end user, and it could possibly be part of a phrase. Based on this visualisation, word cloud has a limit of the number of words, and it is not suitable to visualise phrases, which could carry more meaning than individuals words alone. Also, there are a lot of under-utilised space in this visualisation, which resulted from the different shapes that the letters of each word occupy. Some words such as "like" are forced to be placed within a bigger word "donation", which can be misleading to the end user.

Moving on, another visualisation that we drew inspiration from is a visualisation of a network of Digital Humanities professionals, researchers, engineers and enthusiasts who are in Twitter. This graph visualisation shows who is

following who, within this network. This is related to the dataset we are using because it measures the interaction of people in a digital sphere, which could be different from the physical interaction face to face, for example.
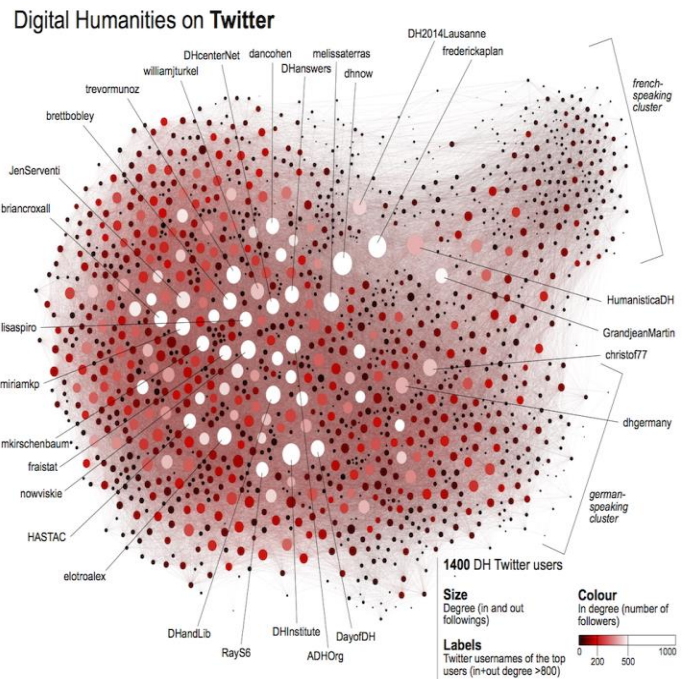

Figure 2 – Twitter Network of Digital Humanities individuals

Looking at Figure 2, a network of about 1400 Twitter users who are related in the Digital Humanities field are present in this network. This visualization was created using Gephi, a program that specializes in Graph Visualisation. There is good use of extra space, to identify the key people within this network, especially those who have attained 800 in-degree and out-dergee follows. This mean that a person has followed and are being followed by minimally 800 people and more.

Another useful visualisation tool that is currently being used is the colour of the nodes. For this visualisation, the colours exhibit weight, in such as the darker the shade of red you are, the more people following you on Twitter. The size of the circle demarcates the degree, or the extent of someone being followed and that same person following other people.

The advantages of this is that there are many different metrics to measure different characteristics, so as to give the user different options of measures to focus on.

However, it may be hard to quantify the extent of the degree using a circle. It is difficult to accurately measure the actual degree of a particular person, let alone compare the circle's size with another person.

DATA TRANSFORMATION

The data was obtained in the form of CSV format from VAST challenge 2014. It consists of two CSV files namely email exchanges, specifically email headers, among various employees and employee records which has the particulars of the various employees. These two data files are subsequently joined using the email addresses as the common attribute. As the data structure required for graph visualisation and text visualisation is different, a more indepth discussion on the individual data preparation steps are highlighted in the next segment.

I. Graph Visualisation

For the Force Directed Graph to be visualised, it will need key characteristics from the data. The essential information needed are "source", "target", "value", "groupsource" and "grouptarget". The "source" and "target" columns are similar to the sent and received personnel in the email context. The "source" is the sender of the email, whilst the "target" is the receiver of the email.

"Value" is a calculated value, in which Microsoft Excel's Countifs function was used to count the occurrence of a particular two people, Person A, sending email to Person B.

"Groupsource" and "Grouptarget" pertain to the department that these employees belong to. "Groupsource" is the department that the email sender belongs to, whilst the "grouptarget" is the department that the email receiver belongs to. To obtain this data, the department of both the email sender and receiver was obtained from the combining of the two datasets, and thus their columns are renamed "groupsource" and "grouptarget". This information will be important towards identifying which department the employee belongs to, as part of the analysis.

II. Text Visualisation

Email headers consist of "RE:" or "FW:" to indicate if they are new email, forward email or reply email. As our main objective is to analysis the keywords used in the email headers and not the type of email, "RE:" and "FW:" are cleaned out from the email headers column.

The cleaned email headers are then tokenised using the Text Explorer function in JMP. Initially, we planned to conduct a latent class analysis on these resulting terms to identify clusters of employees who tend to discuss business related content in the email and those who don't. However, the results of the clustering are unable to identify these clusters clearly. Thus, we manually search through the terms and phrases and identify words and terms that are business and non-business related. We have only identified some of the words that are of interest to us to prevent visual cluttering of the graph.

Business Related		Non-business Related	
Audit	Politics	Babysitting	Golf
Cased-Hole	Profit	Birthdays	Guys Night Out
Conference	Report	Casino	Hackathon
Data	Schedule	Catering	Lottery
Downhole	Staff Picnic	Coffee	Lunch
GAStech	Strange Network Activity	Concert	Plants
GIS	Training	Coupon Club	Sick
IPO	VIP Visit	Cute	Spam
Meeting	Virus Detected	Don't Text and Drive	Traffic Advisory
Openhole	Wellhead	Flowers	Vacation
Policy		Funny	

Table 1: Key Words from Text Analysis

The resulting variables used for the text visualisation are name of employee, keywords, department and emails sent.

CHALLENGES

We have faced several challenges during the design and implementation of the application due to the team's unfamiliarity with coding. One of the key challenges that our group face is the integration of the graphs. Initially, we thought of using D3 to integrate the graphs. However, we find it difficult to do that due to the limited functions available and the need to combine two datasets of different sources. Since the two visualisations are relating to different aspects namely graph visualisation and text visualisation, we decided to create a storyboard instead of a single dashboard using javascripts to combine the two visualisations.

FINAL APPLICATION DESIGN

Having resolved the challenges faced during the implementation and design process, we are able to develop an application that consist of the layout and interactivity that we have in mind. The follow sections discuss the choice for the visualisations used in the storyboard.

I. Force Directed Graph

Force Directed Graphs are known to be an aesthetically nicer method to build a visualisation of a particular network, and also using the information present within the graph to create the graph itself, rather than using domain specific knowledge (Kobourov, 2012). This results into a graph that is easier for the end user to grasp, without having any prior knowledge about networks and graph visualisation. Within the visualisation, there are features that make it more interactive for the user.
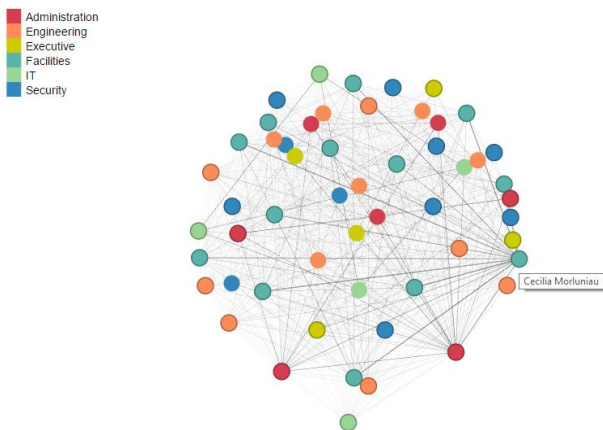

Figure 3: Final Application's Force Directed Graph

As seen on Figure 3, one key feature is the interactivity of the map by clicking the nodes to find out how many emails has a particular person sent within a given time frame. The number of people whom a person has sent to can reflect his/her role in the company, or that the person helps diffuse important information. In these situations, those who control information in the company's context are potentially influential and this person may be able to have power or exert power, based on the information known.

Also, it is crucial to know who are those employees who sent the highest number of emails or who interact often with the different departments. These employees who interact with other departments in a regular basis is in a position into forming stronger interdepartmental teamwork, which is beneficial for the company in the long term. (Ellinger, Keller & Ellinger, 2000). With the use of a legend and node

colour coding, the user is able to see the different departmets and how they interact with each other.
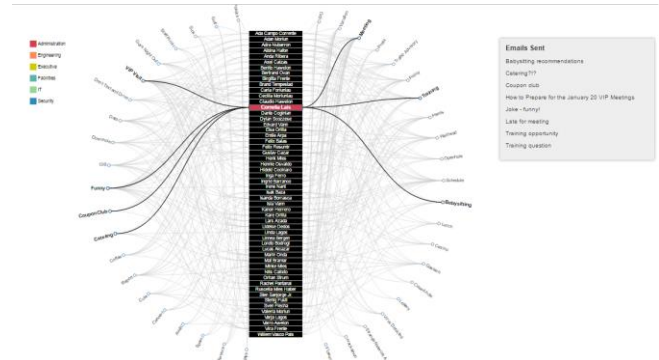
II. Concept Map

Figure 4: Final Application's Concept Map

Concept maps are graphical tools for organizing and representing knowledge (Cañas & Novak,2009). It is similar to a chord diagram where inter-relationships between the two variables, namely keywords of interest and employees are displayed. This provides an overview of the keywords used in the email headers by individuals and the topic of the conversations that they had among the other coworkers. The outer node represents the keywords of interest and the inner node represents the employees.

When the users hover the names of the employees, the keywords of interests related to this employee will be highlighted. The colour of the rectangle represent the department that the employee is in. A tooltip at the side will also appear to display the actual email headers send by the employee. This provides additional information regarding the keywords of interests in order to determine if the employee indeed frequently use emails to discuss non-work related topics.

Users can also hover the keywords on the outernodes. The employees who use this keyword in their email headers will also be highlighted. This helps the users to identify employees who have most likely engaged in similar conversations and also shows the interactivity among the various departments and within individual department.

FUTURE WORK

There are a few areas of improvements that we would like to suggest for our current visualisations in order to increase the usability of the application.

Our visualisation provides greater insight into the interactivity among the employees and various types of conversations that they had based on the email headers. However, a comprehensive analysis on the email content are

not covered in our visualisation which paves way for further research in this area. The application can be further improved by allowing users to visualise both the keywords of the email headers and email content. This would allow the users to have a more comprehensive overview of the interaction at various organisational levels.

Increasing the ability of the application to handle a larger dataset instead of the current sample size of two weeks would be valuable for further analysis. This would further enhance the adaptability of the application to large organisations. In order to prevent visual cluttering in graphs for large dataset, a drag and drop visualisation can be considered. This would also further improve on the usability of the application as users can changed the visualisation according to their preferences.

CONCLUSION

With the ease of storing digital information nowadays, there will be the wide availability of different types of digital information to study to find out insights from their communication and interactions There are many potential applications that our analysis can provide for different types of companies who may want to investigate a suspected fraud case discreetly or to examine the different dynamics that exists within the workplace. With the present technology, there will be a widespread reach of tools, such as D3.JS, that can be used to do particular tasks and enable companies and stakeholders to fully understand internal communication and interactions within the company.

ACKNOWLEDGEMENT

REFERENCES

[1] Halvey, M., & Keane, M. T. (2007.). An Assessment of Tag Presentation Techniques. Retrieved November, from http://www2007.org/htmlposters/poster988/

[2] Ellinger, A. E., Keller, S. B. and Ellinger, A. D. (2000), Developing Interdepartmental Integration: An Evaluation of Three Strategic Approaches for Performance Improvement. Performance Improvement Quarterly, 13: 41–59. doi:10.1111/j.1937-8327.2000.tb00175.x

[3] Cañas, A. J., Novak, J.D. (2009), The Theory Underlying Concept Maps and How to Construct and Use Them; Institute for Human and Machine Cognition (IHMC), from http://cmap.ihmc.us/docs/theory-of-concept-maps#[1]