Using Employee Social Networks to Better Understand Influence, Collaboration, and Network Strength

Ridwan Mohamed Ismeer, Sadhvi Ilango, Yashraj Abhiraj Jalota

**Abstract**

Relationships and networks created and maintained by employees, are a rich source of data, which can be used to support better HR decision-making. Until now, this data has been invisible to stakeholders who could benefit most. Using this data, we look to generate a range of insights, including measured strengths of existing relationships and an employee’s organizational influence that will revolutionize numerous processes and decisions. These insights make employee relationships and their networks visible and measurable, providing support to evidence-based decisions on the organization’s most valuable asset, the people. Through careful Social Network Analysis, we have provided organizations with exact metrics to identify organizational influencers, gatekeepers, highly connected individuals, immersion progress for new hires, and employee activity levels.

Table of Contents

[1. Introduction](#h.4jz554hveghw)

[1.1 Motivation: The Importance of People Analytics](#h.pvh7zpxfnzz5)

[1.2 Our Approach](#h.p9dyre8attjn)

[2. Literature Review](#h.s2yfimmh1q)

[3. Methodology - Iteration 1](#h.dcef9micc06w)

[3.1 Initial Data Cube and Data cleaning](#h.a6i57io4zhay)

[3.2 Additional Data](#h.pcvwl7q7juds)

[3.3 Social Network Analysis](#h.i9mzwluylgrl)

[4. Methodology - Iteration 2](#h.ln9ft2jrb3cd)

[4.1 Additional Data](#h.d0658qp1txo2)

[The next step was to calculate the following metrics, namely:](#h.kjh1tv6kzc6v)

[Relationship score: Based on an internally calculated metric by our client, the relationship score is weighted based on:](#h.w5f84hguwhb5)

[4.2 Designing Visualizations](#h.nnjp586cc8ty)

[4.2.1 The Chord Diagram](#h.wukt31kjec4c)

[4.2.2 Data-Linked Bar Chart and Chord Diagram](#h.bs6s4x4p6day)

[4.2.3 Social Network Graph](#h.6emhabn843ra)

[4.2.4 Bullet Graph](#h.qmqi61o0gxm6)

[4.2.5 Employee Influencer Rankings](#h.lvvruuyr908e)

[5. Discussion of Results](#h.p2413sadzzxg)

[5.1 Monitoring collaboration](#h.9o4cgee36hfk)

[5.2 Identification of Influencers within the Organization](#h.xn2jr6gb3ygz)

[5.3 Gender Distribution of Relationships](#h.s462d9kq4xxe)

[5.4 Identifying Gatekeepers and Highly Connected Employees](#h.jv5ovym23qit)

[5.5 Employee Immersion, Active Employees and other Statistics](#h.xx9ypbuni6al)

[6. Conclusion](#h.kplgv8o15qrl)

[6.1 Actionable Recommendations](#h.m1dbxog8eg39)

[6.1.1 Alteryx Software](#h.kfqprl66bja1)

[6.1.2 D3.js](#h.xooijmvdkvu1)

[6.1.3 The Process](#h.nnlntxas3gkc)

[6.1.4 Sigma.js](#h.oxbu1qs1xlyr)

[6.1.5 Client Management](#h.6c81in4cokv7)

[6.2 Limitations](#h.x6sybf2chysd)

# 1. Introduction

TrustSphere is the widely recognized market leader in Relationship Analytics. TrustSphere enables forward thinking organizations to unlock the inherent value of their own networks using their proprietary next generation technology. TrustSphere’s unique selling point lies in its ability to calculate relationship strengths between two people based simply on their communication flows. The solutions provide real-time intelligence and insights which help clients across the globe improve salesforce effectiveness, enterprise-wide sales insights and corporate governance. Their offerings currently are limited to Sales and Risk.

Currently, TrustSphere is developing new technology that delivers insights into an organization's’ most valuable asset, its people. To set the ball rolling, TrustSphere engaged our team, YSR, to help deliver preliminary insights on its own employees. With our insights and the respective methodologies used to deduce these insights, TrustSphere can move forward in the construction of their own product offering.

The complexity of this project lay in two areas:

1. The insights we were looking to deliver had to be matched with contextual social science in order for them to be relevant and significant.
2. We had severe limitations in the availability of data: we were provided with only communication metadata.

## 1.1 Motivation: The Importance of People Analytics

Understanding and managing the core of an organization - its employees - has become imperative to business decision-makers across the world. In a global survey conducted by PwC, 34% of CEO’s are “extremely concerned” about the management of talent. Workforce diversity has stalled[[1]](#footnote-1), high turnover trends are continuing, labor productivity has reduced since last year and high potential promotion numbers have dropped. Our client company’s HR and Talent department have forever relied on gut-feelings and intuitions to make important people-related decisions. The lack of evidence-based decision making within these departments are proving to be detrimental to their effectiveness. Our objectives are to utilize the huge volumes of data stored at the fingertips of their HR department to provide a precise mathematical layer to support decision-making. We also intend to delve deeper into other possible data-driven insights companies could make in relation to their people.

In short, we look to combine powerful analytics with careful social science to deliver never-before explored insights.

## 1.2 Our Approach

Our understanding of employee structures and the data provided helped define our approach to this project. Relationships and networks created and maintained by employees, are a rich source of data, which can be used to support better HR decision-making. Until now, this data has been invisible to stakeholders who could benefit most.

By performing SNA and other exploratory data analysis, we seek to enable TrustSphere’s Human Resource department’s decision-making via insights generated from employee communications, relationships and networks.

# 2. Literature Review

As the realm of People Analytics is a relatively new space, there are few organizations delving into understanding relationships with the use of data generated insights. Additionally, the firms are limited in terms of the three V’s (Volume, Variety and Velocity) of the data and the insights it offers. A report published by PWC on the *Trends in People Analytics* states that advancement in People Analytics is on the rise with the emergence of a refined data governance approach, the building of comparison data into analytical tools and also the use of predictive data to take action.[[2]](#footnote-2) What are lacking are integrated and comprehensive data collection methods in organizations to provide for a robust analysis. There are few organizations that are acting upon their findings to close the gaps by addressing root causes, deciding to deal with employees on an individual level, integrating the results to existing technologies and updating the data frequently. In the pursuit of creating flatter organizational structures to promote efficiency and flexibility, there has been an increase in informal employee networks rather than formal structures. Hence, the authors concede that there is a greater need for employee social network analysis. It reveals the level of engagement of key employees that can affect the morale and productivity of the environment, either positively or negatively.

An article published by McKinsey urges organizations to tap on the power of influencers to better harness their positive traits and increase the odds of success for the organization.[[3]](#footnote-3) An informal influencer can be broadly defined as someone who is respected and approached for advice or information. The use of influencers in a work environment could prove to be useful for implementing changes within the organization, right from planning and direction, to execution.

Further research revealed that until recently, it had been hard to test whether the performance of organizations could depend on information flow between employees.[[4]](#footnote-4) Sources such as e-mail logs, web-based calendars and online work collaborations make it possible to track information flows within firms and assess performance of teams, departments and the entire organization as a whole. However, the question remained as to what could be done with the limited data that we were able to obtain from our client. Duncan Watts and his team had a more comprehensive data set, with employee surveys to supplement their e-mail data and organization charts. It is important to note that e-mail data allows for more real time and timely feedback, compared to employee surveys. While our data was lacking in terms of employee feedback, we set to define our research parameters based on what we had access to, and with the takeaway that “computational social science works well only when powerful computation is matched with careful social science”. There are a number of key centrality measures that can be gathered from Social Network Analysis data to help us better understand interactions.[[5]](#footnote-5) The following are a few of them:

*Degree* - The number of connections an individual has.

*Closeness* - A measure of how close an individual is to others in the network.

*Eigen* *Vector* centrality - A measure of how connected an individual is to those who are highly connected.

*Betweenness* - A measure of the degree to which an individual is in a key position (gatekeeper), connecting the network.

# 3. Methodology - Iteration 1

# 3.1 Initial Data Cube and Data cleaning

Initial dataset given to us was pulled from the TrustSphere Outlook server. It contained 890,000 email exchanges and 13 variables, namely:

1. **Date:** Timestamp of when it was sent
2. **IP** **Address:** IP Address used
3. **Local:** Server used for external communications
4. **Remote:** Receiver email address
5. **Local** **Domain:** Sender email address
6. **Remote** **Domain:** Nature - Internal or External
7. **Originator:** Too many missing values to make any sense of the column
8. **Direction:** Too many missing values to make any sense of the column
9. **Domain Group:** Too many missing values to make any sense of the column
10. **Inbound Count:** Too many missing values to make any sense of the column
11. **Outbound Count:** Too many missing values to make any sense of the column
12. **Size:** Size of the email, but irrelevant to analysis
13. **Msg ID:** Too many missing values to make any sense of the column

The data included both internal and external communications denoted by variable “Remote Domain”. We chose to move forward with only internal interactions because getting context for external data would be extremely difficult and was removed from our scope.

We then went on and imputed variables 6 to 13 because they had too many missing values. We also deleted IP Address and Local because they would not add any value to the analysis done.

## 3.2 Additional Data

We now requested our client to give us a Staff-List to populate Sender/Receiver positions and Department and finally ended up with the following:

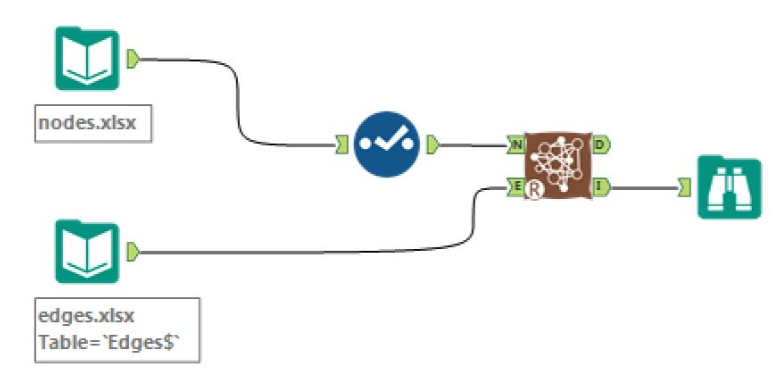
1. **Date:** Timestamp of when it was sent
2. **Remote (Renamed to Receiver):** Receiver email address
3. **Local** **Domain (Renamed to Sender):** Sender email address
4. **Sender FName:** first name extracted from email address
5. **Receiver FName:** first name extracted from email address
6. **Sender Position**
7. **Receiver Position**
8. **Sender Group**
9. **Receiver Group**

## 3.3 Social Network Analysis

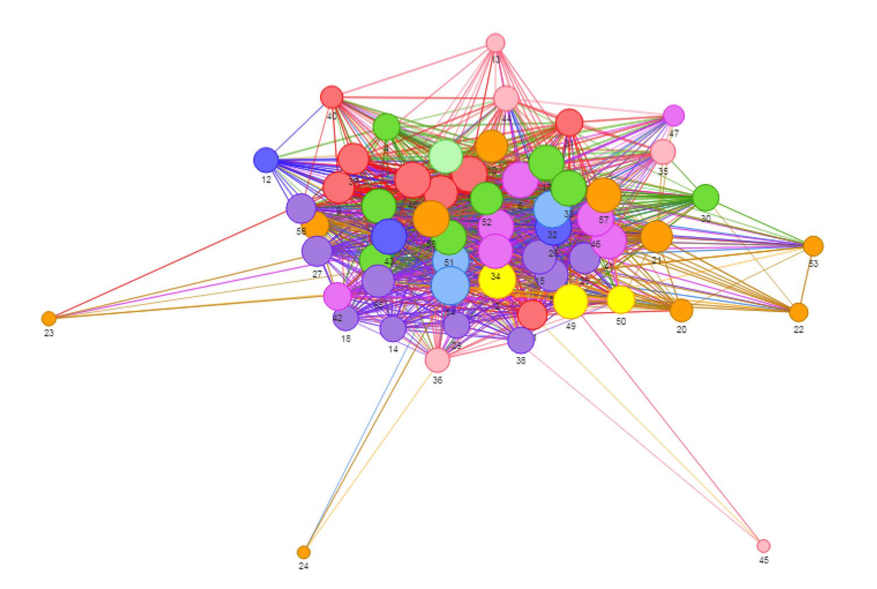
To analyze social networks, we considered using the following tools:

* R (library: igraph):​ Used to create routines for simple graphs and network analysis. It can handle large graphs very well and provides functions for generating random and regular graphs, graph visualization, centrality methods and much more. We considered using this because our analysis requirements fit well into what igraph has to offer.
* Alteryx​ ​(Predictive Analysis using R): The network analysis tool provides a way to visually interact with all kinds of data. It requires the above-discussed Nodes and Edges tables to do the analysis using a workflow UI. Alteryx also gives you access to key network data such as ​betweenness, degree ​and ​closeness.​ This data can further be used on software such as ​JMP Pro​ and ​Tableau​ to run regressions to predict how employees behave or will behave based on their network statistics.
* Gephi: We used Gephi towards the end of our project to visualize our social networks. This tool too requires nodes and edges tables similar to Alteryx.

Use Of Alteryx:



We used the Alteryx Network Analysis tool to generate our preliminary Social Network Graph and to output key network statistics measures like Betweenness, Degree Centrality and Closeness Centrality. Alteryx helped us extract key network statistics that were used in further analysis. The data required was in the form of two separate nodes and edges lists.



We found that the graph produced was insightful but the real value of the tool to us were the centrality measures.

We then worked on these network statistics to compute what we called our influencer score:

The following formula was used to scale Betweenness and Degree:

(new max - new min)

**Scaled Value**  = --------------------------------------------------

(old max - old min) x (value - old max) + (new max)

These scaled values (on scales of 0 - 100) were then summed up to calculate our Influencer Score on a total scale of 200.

# 

# 

# 4. Methodology - Iteration 2

Iteration 1 was limited to static analysis of communication flows within the organization. Even though this was in line with our client’s requirements, we realized that bar graphs and scatter plots were just not enough and that we could do more. We took on our supervisor’s feedback head on and went on to create an interactive dashboard from scratch using d3.js and other supporting libraries.

## 4.1 Additional Data

## The next step was to calculate the following metrics, namely: **Relationship score:** Based on an internally calculated metric by our client, the relationship score is weighted based on:

1. Volume of email exchange
2. Age of relationship
3. Response reciprocity
4. Frequency of exchange
5. Recency of last communication

**Influencer Score:** Calculated as a sum of scaled betweenness and degree of each employee in the network on a scale of 0 - 200.

## 4.2 Designing Visualizations

At this point we delved into the wonderful world of D3.js. We carefully considered which charts, graphs and diagrams would help us best represent the employee network.

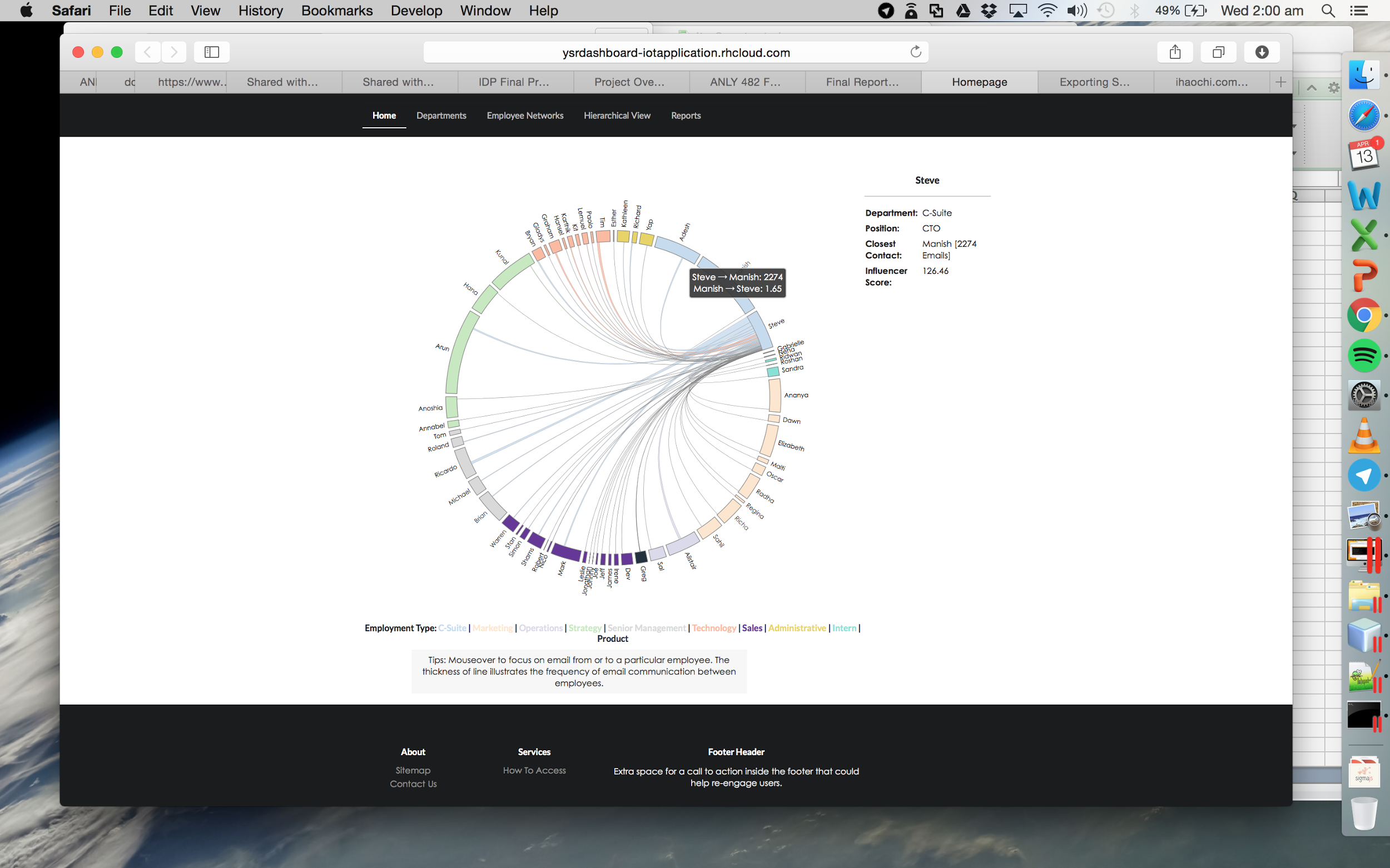
### 4.2.1 The Chord Diagram

**A] Rationale:**

We chose to use a Chord Diagram to visualize our employee email data. A chord diagram is a graphical method of displaying the inter-relationships between data in a [matrix](https://en.wikipedia.org/wiki/Matrix_(mathematics)). The data is arranged radially around a circle with the relationships between the points typically drawn as arcs connecting the data together.[[6]](#footnote-6) A chord diagram is the most aesthetically pleasing way of visualizing directed relations among entities in a space-constrained area.

**B] Data, Process & Interactivity:**

The data required for our chord was a square matrix denoting from and to number of emails along with an edges list giving information about each connection like labels and total number of emails. Ancillary information (on clicking a name) was displayed on a card on the right highlighting things like closest contact, influencer score and position in the company.

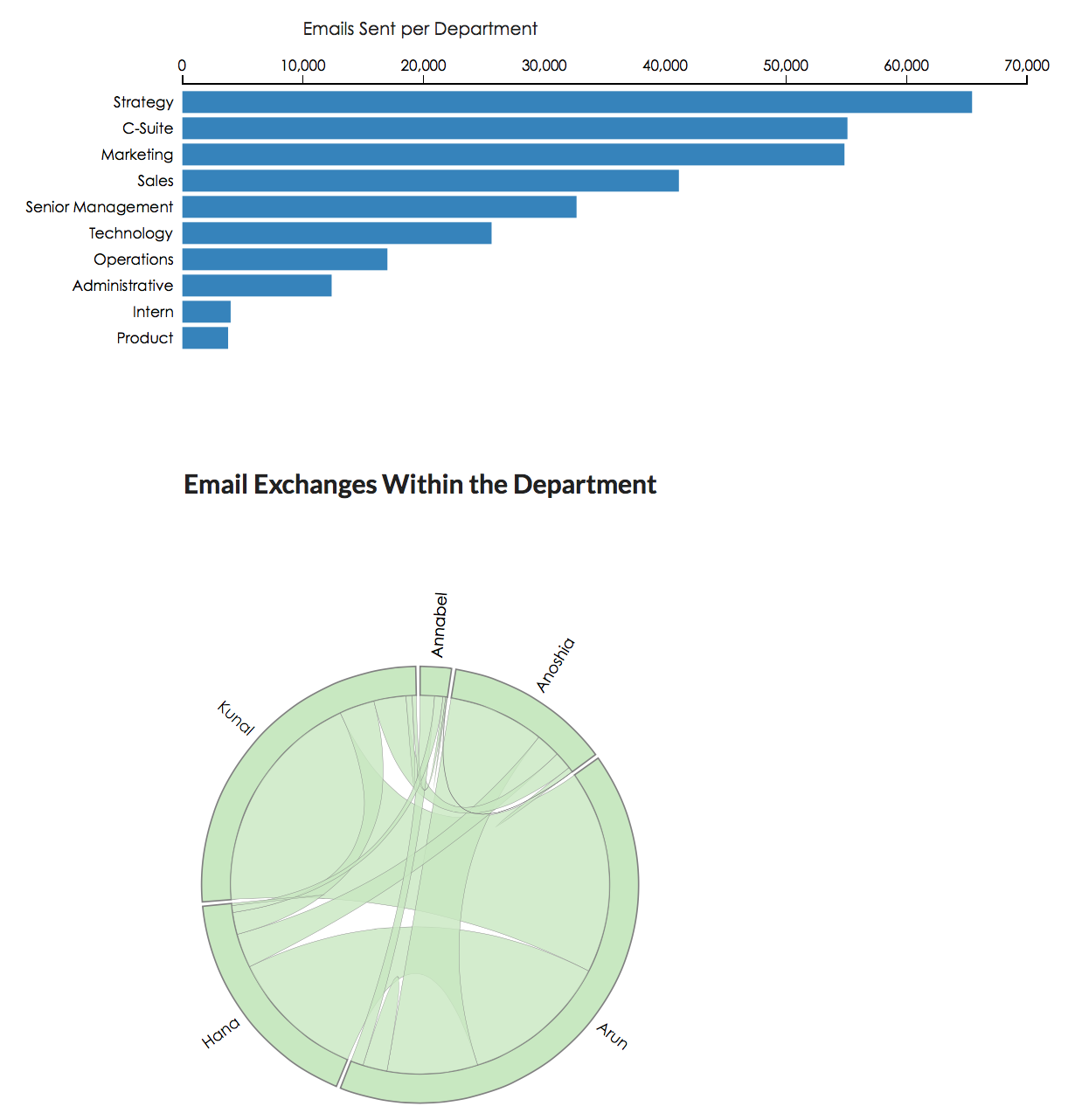


The employees are coloured by department, which the legend at the bottom helps identify. A hover over an employee's name shows the number of emails sent over the time period of the data. Hovering over the arcs will give you the number of emails sent and received from the employee that arc represents.

### 4.2.2 Data-Linked Bar Chart and Chord Diagram

**A] Rationale:**

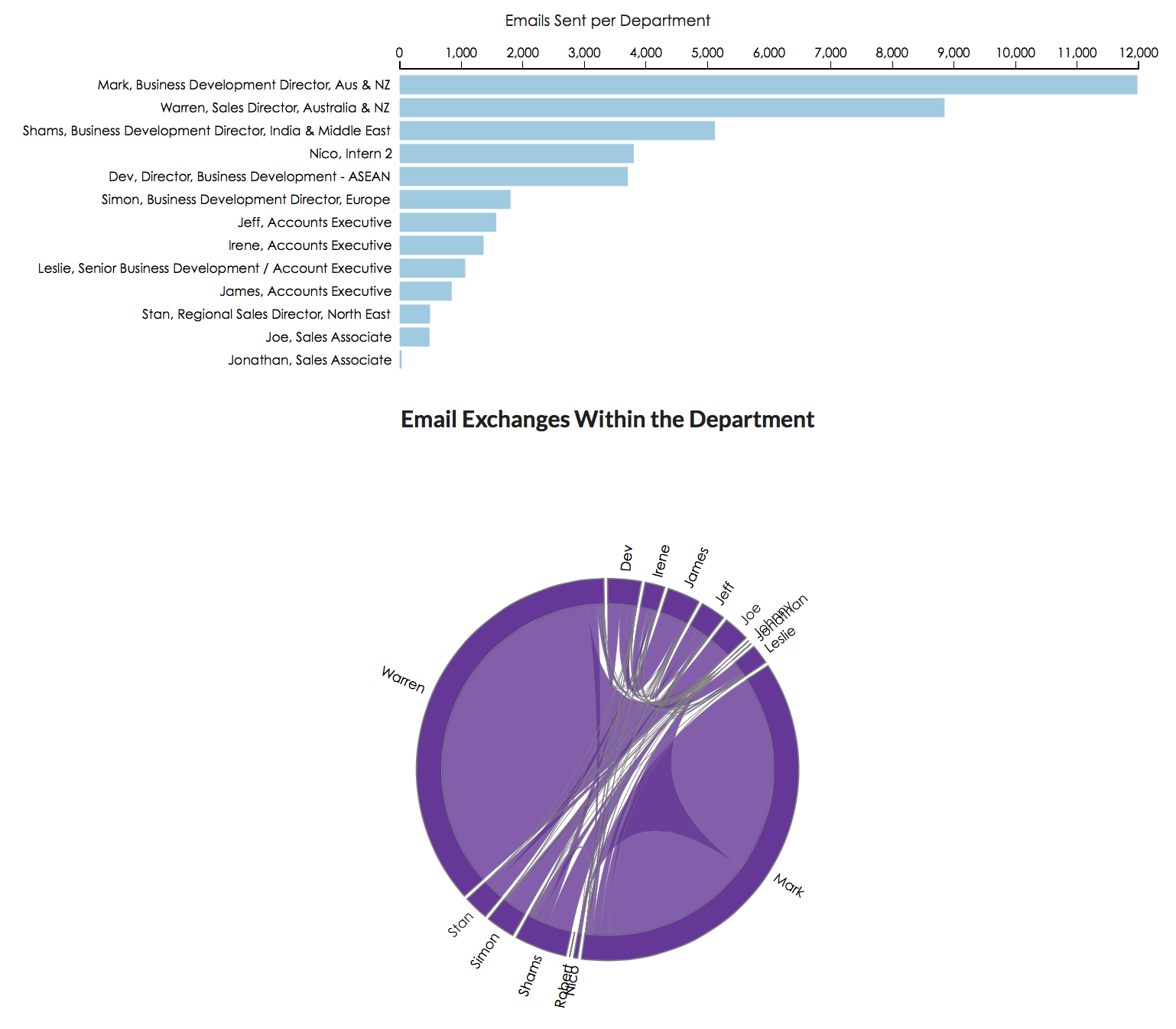
This visualization was chosen to perform a deeper analysis of within department interactions. The bar chart shows the different departments in the organization and the breakdown of number of emails sent by each department.



**B] Data, Process & Interactivity:**

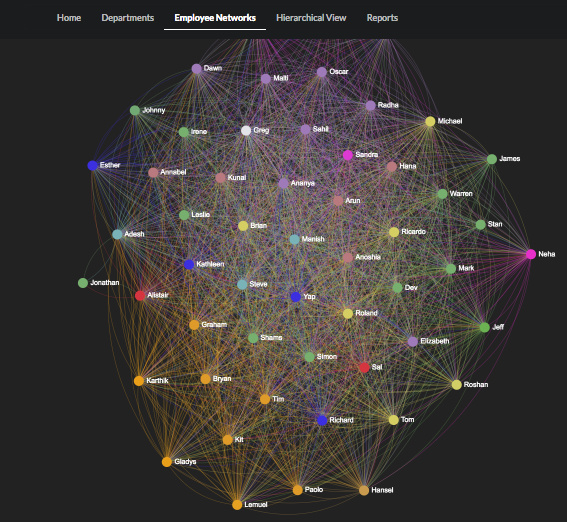
Once a bar is clicked, a drill down is activated to show the employees in that department and the number of emails sent by each person. The chord also changes accordingly to help the user understand exactly the interaction in each department.

The data required for this visualization was in the form of a key value file where departments served as keys and members as values.



### 4.2.3 Social Network Graph

A Social network using Force Atlas 2 layout was chosen to visualize betweenness and closeness amongst our employees. The data required for this was a nodes and edges list as input to Gephi, a social network visualization tool. The resulting graph was exported as a .gexf file and fed as input to SigmaJS, a javascript library dedicated to graph drawing on the web. The resulting graph showed employees with departments as colors and distance as the closeness centrality.



### 4.2.4 Bullet Graph

This graph was made in order to interactively visualize employee relationship scores across different *Levels* in the organization. We created levels for departments and coded them as follows:

**A:** C - Suite

**B:** Directors

**C:** Managers

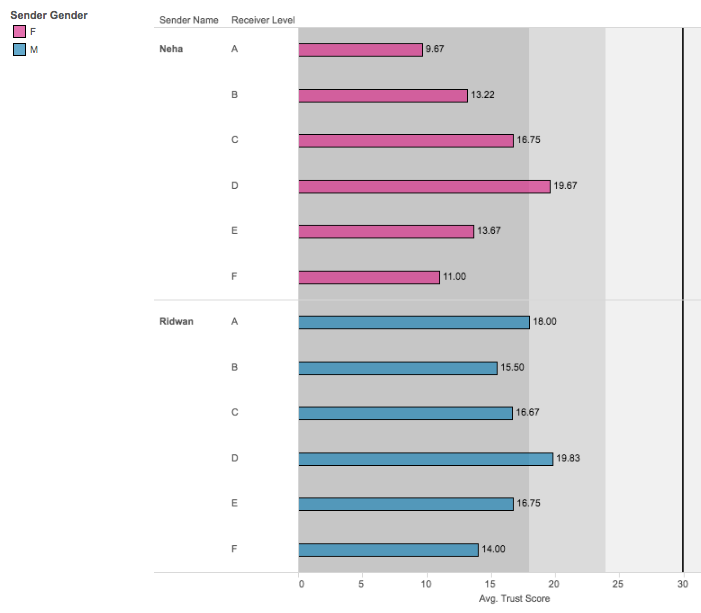
**D:** Associates

**E:** Interns

**F:** Admin

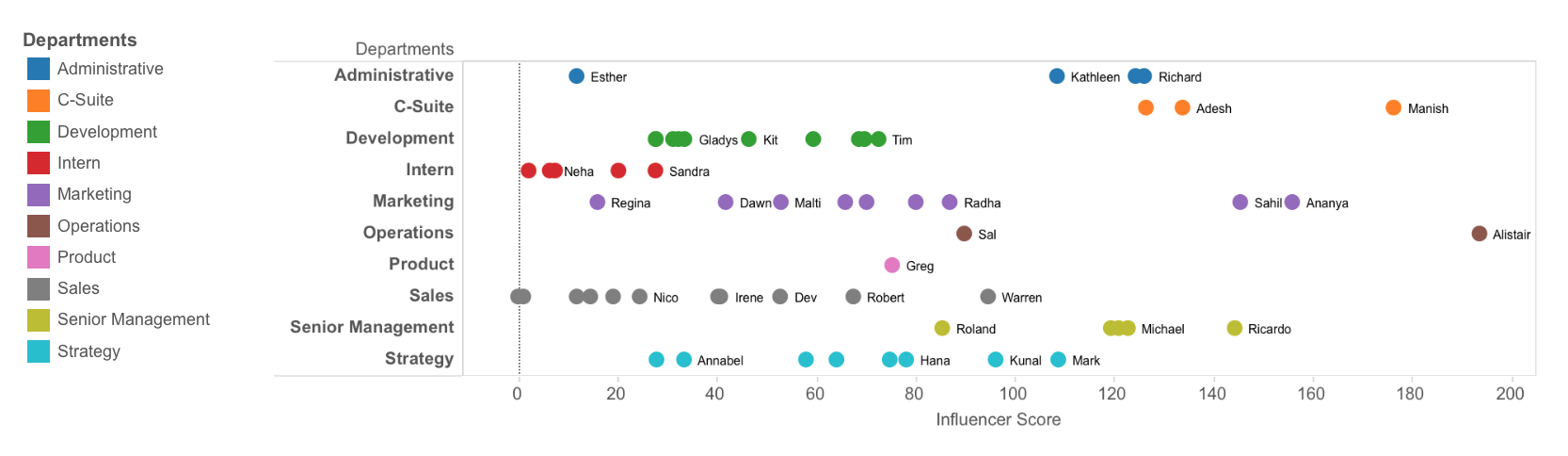
We then mapped each employee to their respective receivers broken down by level and gender. Through this, you get an overview of the employee's performance in relation to the others ticked in the filter, to the constant value 30, a good relationship score set by the client and 60% and 80% marks of 30.

We then went on to notice that males generally have higher average relationship scores than females and did an analysis of variance to confirm significance. Results of which will be discussed in the next section.



### 4.2.5 Employee Influencer Rankings

The final visualization was a graph comparing the influencer scores for the entire organization. Employees were mapped on a chart from 0 to 200 and broken down by department to distinguish employees with higher influencer scores and our so called *gate keepers and highly connected individuals*.



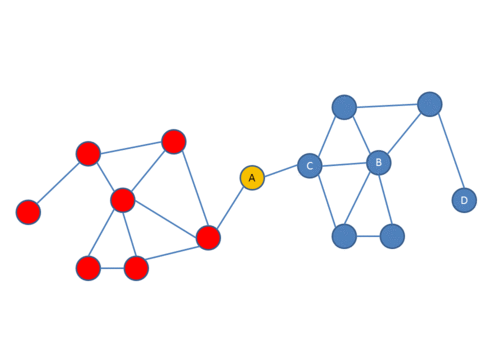
Discussions of the results from our analysis will now follow this section.

# 5. Discussion of Results

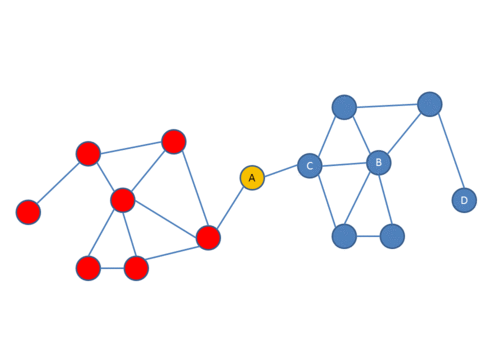
## 5.1 Monitoring collaboration

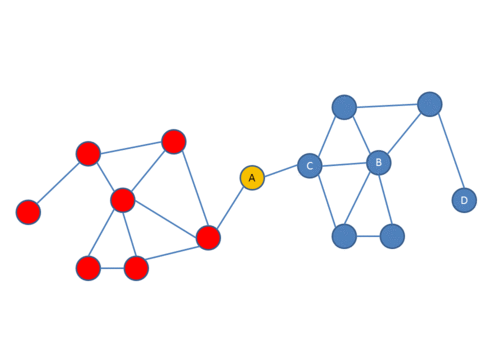
Measuring and monitoring intra and inter team collaboration activities with People Analytics allows managers to monitor and enhance collaboration and thus strengthen relationships. A preliminary social network graph was drawn with Gephi that can be viewed on our dashboard. However, our preliminary graph through Gephi did not allow us to map out exact communities. Our analysis was limited to a subjective description of possible communities.

To detail exact collaboration silos, or communities, within the organization, we implemented the Givarn-Newman algorithm on UCI-net and Netgraph to identify these exact communities. The Givarn-Newman algorithm, in essence, identifies individuals with high betweenness centralities. For example, in a simplified social network as follows, the algorithm identifies node A as one with the highest betweenness centrality.



Once the high betweenness node, in this case A, is identified it then proceeds to remove the node to display communities, or collaboration silos. We were able to set the number of communities that we desired to identify. The exact steps of the Givarn-Newman algorithm is as follows:

1. The betweenness of all existing edges in the network is calculated first.
2. The edge with the highest betweenness is removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until no edges remain.



In the application of the Givarn-Newman to our client’s social network graph, we obtained the following 4 possible silos of collaboration. Silo 1: Tom, Karthik, Lemuel, Tim, Irene, Johnny, Jeff, and Gabrielle. Silo 2: Esther, Regina, and Kath. Silo 3: Ridwan, Neha, Ananya, Sahil, Dawn, Radha, Greg. Silo 4: The biggest silo, comprising of the rest of the organization.

By using the Givarn-Newman algorithm, we could obtain an accurate representation of true collaboration silos within the organization. The silo-colored network diagram can be found in the appendix 1. Apart from Silo 3, all other silos comprised of members from spread out backgrounds, including departments, geographies, hierarchical levels and genders. Silo 3, however, contained 7 members of which 6 were from the marketing department. There was also a strong concentration of employees from only Singapore with only one member of the silo being out of Singapore.

## 5.2 Identification of Influencers within the Organization

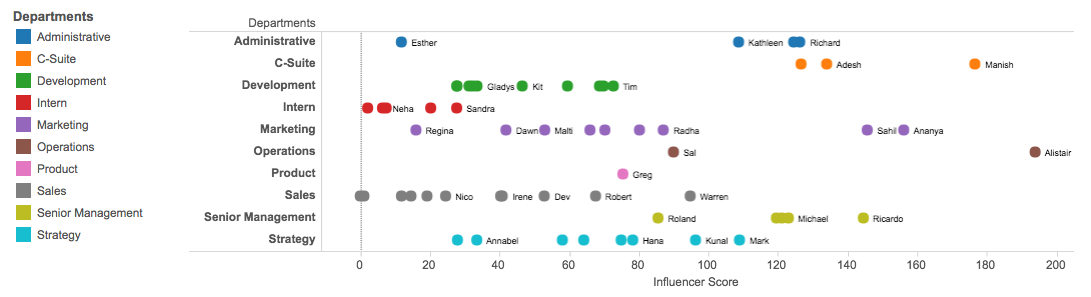
Employee resistance is the most common reason executives cite for the failure of big organizational-change efforts. Winning over skeptical employees and convincing them of the need to change just isn’t possible through mass e-mails, PowerPoint presentations, or impassioned CEO mandates. Rather, companies need to develop strong change leaders employees know and respect—in other words, people with informal influence. But there’s one problem: finding them.

Through Social Network Analysis, we understood various centrality measures of each employee. However, we understood the Eigenvalue centrality measure as encompassing all features necessary to be an influencer. The Eigenvalue is described as the following:

This measure weights the contacts according to their centralities, taking into account the whole pattern of the network and computing the weighted sum of both direct and indirect connections of every length. Therefore, having the graph G(E, V ), the adjacency matrix A, λ as the largest eigenvalue of A, and n as the number of vertices, the eigenvector centrality xi of node i can be expressed as follows (Bonacich, 2007): λxi = Xn j=1 aijxj i = 1, ..., n

Upon implementing SNA through UCI-Net and Netdraw, we were able to determine the firm’s most influential members by ranking them on their Eigenvalues. Our results showed that Manish (the CEO of the company) ranked as the most influential in the company, with Ananya and Sahil (two marketing associates on the fourth level of the organization hierarchy) ranked as second and third most influential. The complete list of employees ranked by influence can be found in the appendix 2.

In order to visually represent these results, we used Tableau’s charting tool with a further layer of context into understanding influence: the departmental designation of each employee.



## 5.3 Gender Distribution of Relationships

Research in the social sciences shows that one of the biggest impediments to females climbing up the corporate ladder is that women struggle to break into male-dominated “**old boy**” **networks.** An understanding of relationship distribution between both genders at our client company will help them develop systems and structures that help to diminish any possible gender disparities within the organization. By classifying each employee into their respective hierarchical level, we were able to find the average relationship scores between genders and these different levels.

To understand the differences we proposed the following hypotheses:

H0= There is no difference between the average relationship score between males and females at different hierarchical levels or the organization

H1= There is a difference between the average relationship score between males and females at different hierarchical levels or the organization

We conducted Analyses of Variance at 95% confidence between the average relationship score values to identify any significant differences between the two genders. Upon testing the differences for significance, we were able to come to the conclusion that males and females interacted different only with relationships at the second level of the hierarchy, i.e. directors of the company. **At 95% confidence, we were able to conclude that males had a statistically significant higher average relationship with Directors of the company.** The complete tests of significance can be found in the appendix.

## 5.4 Identifying Gatekeepers and Highly Connected Employees

Gatekeepers within an organization hold key power in deciding whether a given message will be distributed by a mass medium. Gatekeepers are key employees to identify in change management projects. They are also key players in the development of an innovative and creative workforce culture. Another key metric organizational members will look to understand is the degree of connectivity of employees in an organization. Research has proved that employees with smaller number of relationships in the organization are ones that possess information that is least likely to be known to others in the firm.

Through Social Network Analysis, we understood various centrality measures of each employee. We understood Betweenness centrality, as described below, to be a sufficiently valid metric to identify gatekeepers in an organization.

**Betweenness centrality** is an indicator of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. A node with high betweenness centrality has a large influence on the transfer of items through the network, under the assumption that item transfer follows the shortest paths.

We also understood the Degree centrality to provide the best metric to assess the overall number of relationships an employee possesses. The exact definition of Degree centrality is as follows.

**Degree centrality** is defined as the number of links incident upon a node (i.e., the number of ties that a node has).

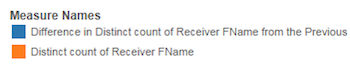
By plotting each employee on a scale of **betweenness and degree**, managers can easily identify gatekeepers and highly connected employees in the firm.

## 5.5 Employee Immersion, Active Employees and other Statistics

Through preliminary Exploratory Data Analysis, were able to calculate other important Human Resource metrics.

1. Employee Immersion

As per statistics, [22 percent of turnover](http://humancapitalist.com/these-employee-turnover-stats-should-scare-you-to-death/) takes place within the initial 45 days and 16 percent in the course of the first week. The key reason for the turnover is bad onboarding and shifting expectations pertaining to the position. As such, it is important to understand new-employee immersion figures over the first few weeks of his or her tenure. Ideally, a new employee should display a steady growth in number of internal relationships over the first few months of work. By plotting the absolute and growth values in number of new relationships for employees, managers can be on high-alert for employees that exhibit signs of disengagement.





For e.g., Annabel, a new employee displays stagnancy in the number of internal relationships she has over her first few months of work.

2. Employee Activity

The number of emails sent and received by an employee is an indication of his or her level of activity at that given point of time. Understanding employee activity levels will allow managers better workforce planning and measures of employee performance. An example of an employee activity graph can be seen in appendix 6.

# 

# 6. Conclusion

# 6.1 Actionable Recommendations

## 6.1.1 Alteryx Software

Alteryx has a simple drag and drop functionality that facilitates easy analyses of social networks. One can input the data file and and manipulate it using the functions in the toolbar above. The input data is in the form of node and edges tables. The Social Network Analysis tool abstractsthe underlying R libraries used to perform the network analysis by giving you a simple workflow. The resulting output is of two types: the D Output and the I Output. The D Output gives you a data table with Centrality Measures that can be used for further analyses. The I Output gives you an interactive network graph with basic features such as summary stats and histograms. Overall, Alteryx is a simple and easy tool to use, especially to extract centrality scores. However, it has less room for customization, compared to D3.js and Sigma.js.

### 6.1.2 D3.js

D3.js is a JavaScript library used to manipulate documents based on data. D3 helps you bring data to life using HTML, SVG and CSS. Our group thought that it would be useful to leverage on the interactive visualization examples D3 has to offer given the nature of our project. Our project wasn’t one that required exhaustive statistical modeling, or one that was based on heavy bivariate analysis. It was one that needed a dashboard to visualize key metrics in an organization in relation to employee communication behavior.

### 6.1.3 The Process

**Step 1:**

*Look at your business query and decide which visualization would best fit your needs.*

This can be done by sifting through the colossal amounts of visualizations available on D3’s homepage. Just take a look at the example and base your codes on it.

**Step 2:**

*Understand what type, structure and format of data your chosen visualization requires.*

This can be done through a simple search online or by looking at online examples of your chosen visualization. Once you find out use a tool like excel JMP Pro, or the wide array of online tools available, to format your data. Even better would be to do it through javascript or python.

**Step 3:**

*Always look for help online.*

Help is always available, may it be in the form of D3’s official documentation or stackoverflow. Don’t be afraid to ask people online. Sometimes, if you are lucky, they even write code for you.

### 6.1.4 Sigma.js

Sigma.js is a library dedicated to graph drawing. It is used to make network graphs that can easily be published on web pages. Sigma.js can be used for cases of varying complexities. You require JSON data with a nodes and edges array. These values are passed into the parser, which is then used by the Sigma.js infrastructure to create the visualization. Specific to our project, we used Sigma.js to visualize our Gephi graph on the web. We passed the .gexf, file (generated from Gephi) to our Sigma.js code, made our customizations and achieved the interactive network graph.

### 6.1.5 Client Management

There are several points to keep in mind while working for a client. In order to avoid conflict and manage relationships, one must keep in mind the following:

1. Clients might have a limited view of the fast growing field of analytics. Do not limit yourself to *only* what the client wants. He just might not know what else you are capable of.

2. There are times when the client might see you as an easy work horse. Prepare an initial list of objectives and divert from this list only if you have the resources to do so.

3. Keep your eyes on the final outcome and don’t get too engrossed dealing with petty details. Don’t waste your time treating the symptoms while ignoring the disease.

4. There might be times when the client might expect work to be completed in unrealistic time spans. Be straightforward - if it is impossible, educate them to the reality.

5. Don’t be afraid to say “no” to your client if what is asked of you will have too much of a burden on you.

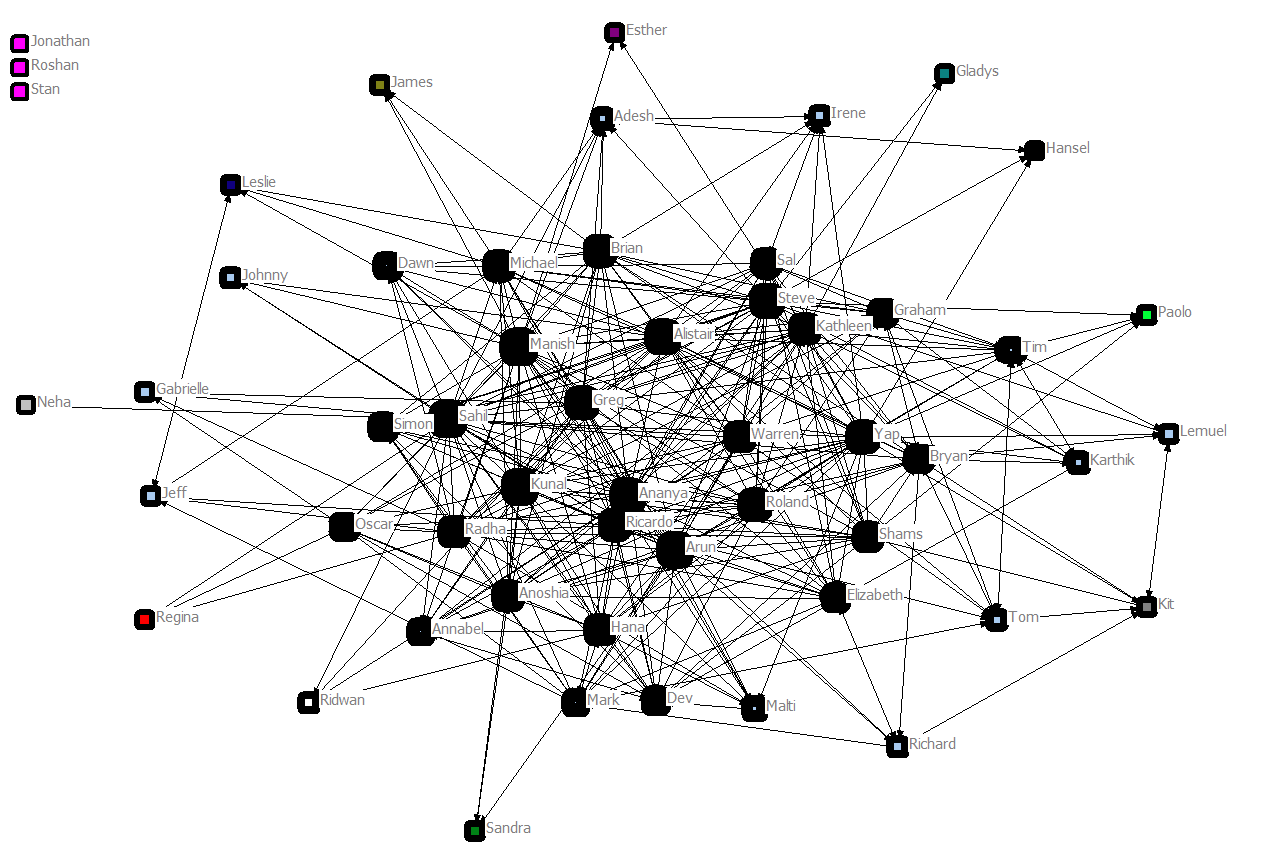
## 

## 6.2 Limitations

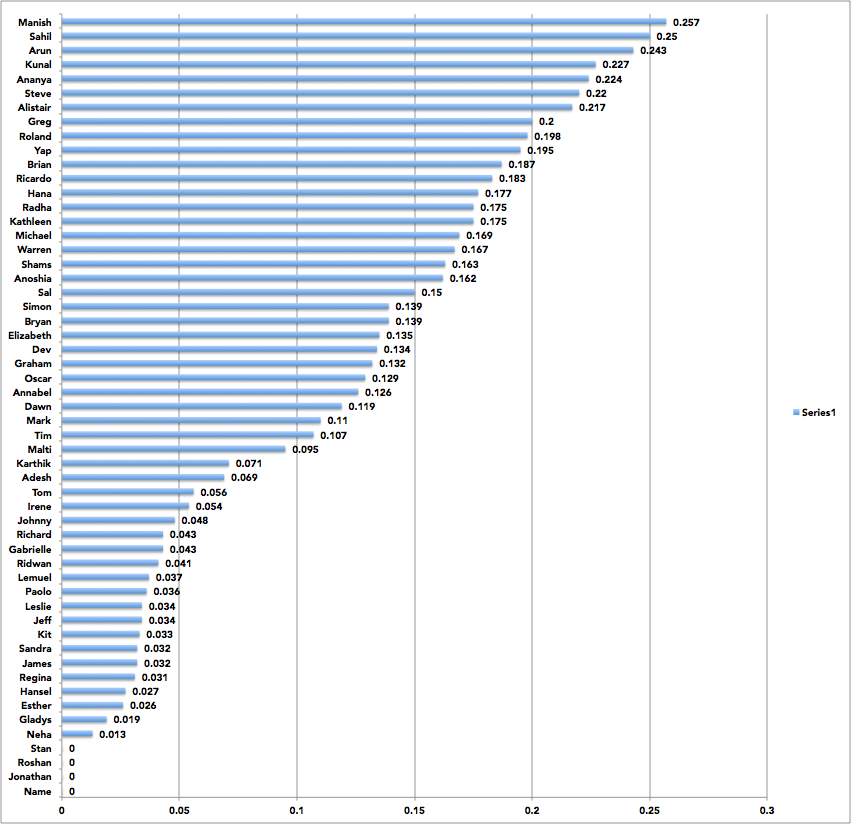
The lack of e-mail subject data is limiting to our analysis. The subjects of the e-mails could give more context with regard to the type of communication (work/personal). Limiting to *only* email data is shortsighted because e-mail data is insufficient for rich social network analysis. We should proceed to consider adding in informal employee interactions through applications such as Skype for Business and Microsoft Yammer.

Another area of improvement could be the increase of the width of data. Going beyond communication data will provide a detailed contextual analysis. We could further contextualize the insights we provide with additional background information of the employees. To solidify our claims, we could test the results by administering surveys and conducting interviews.

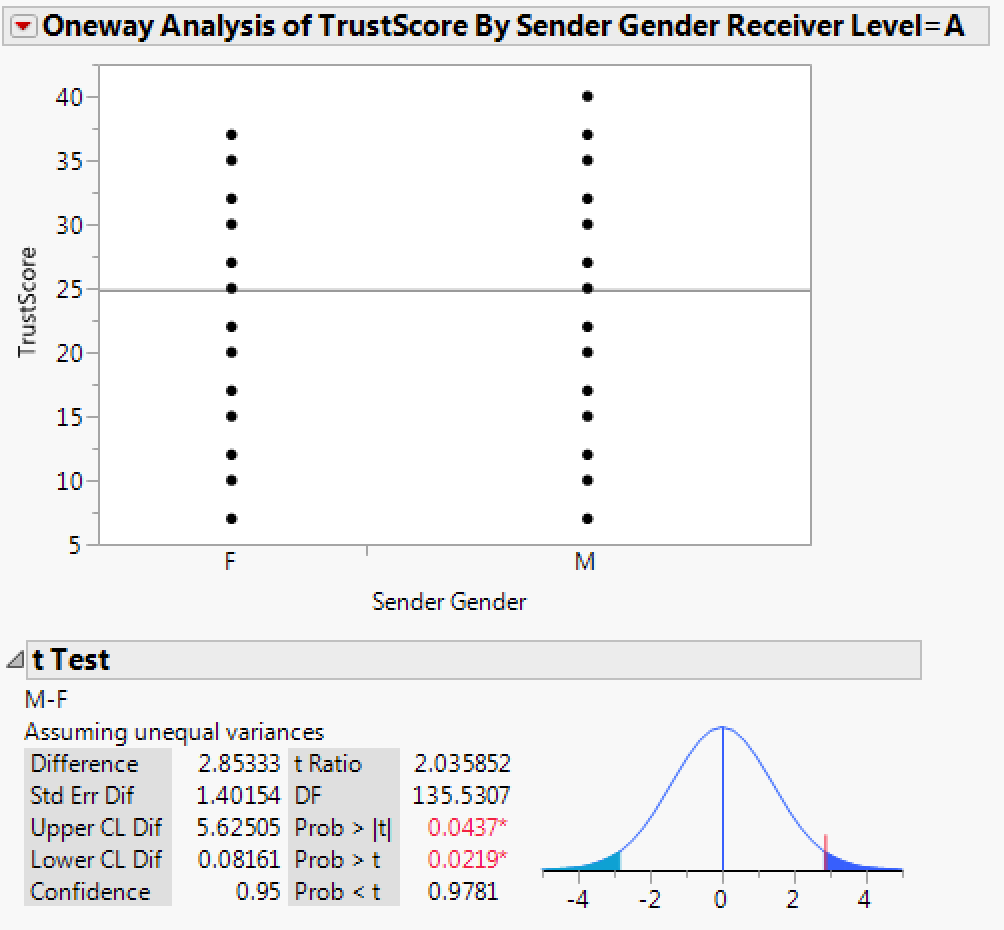
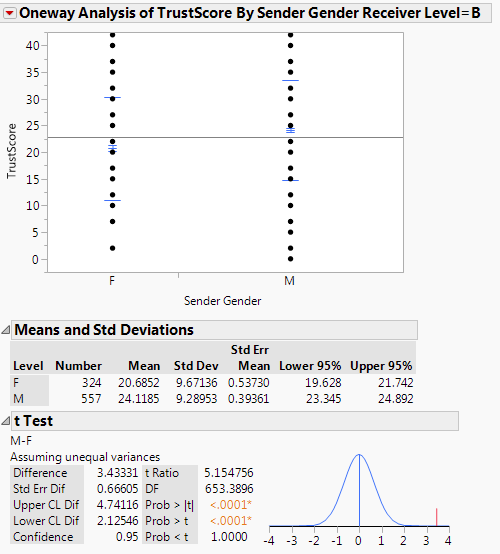
Appendix



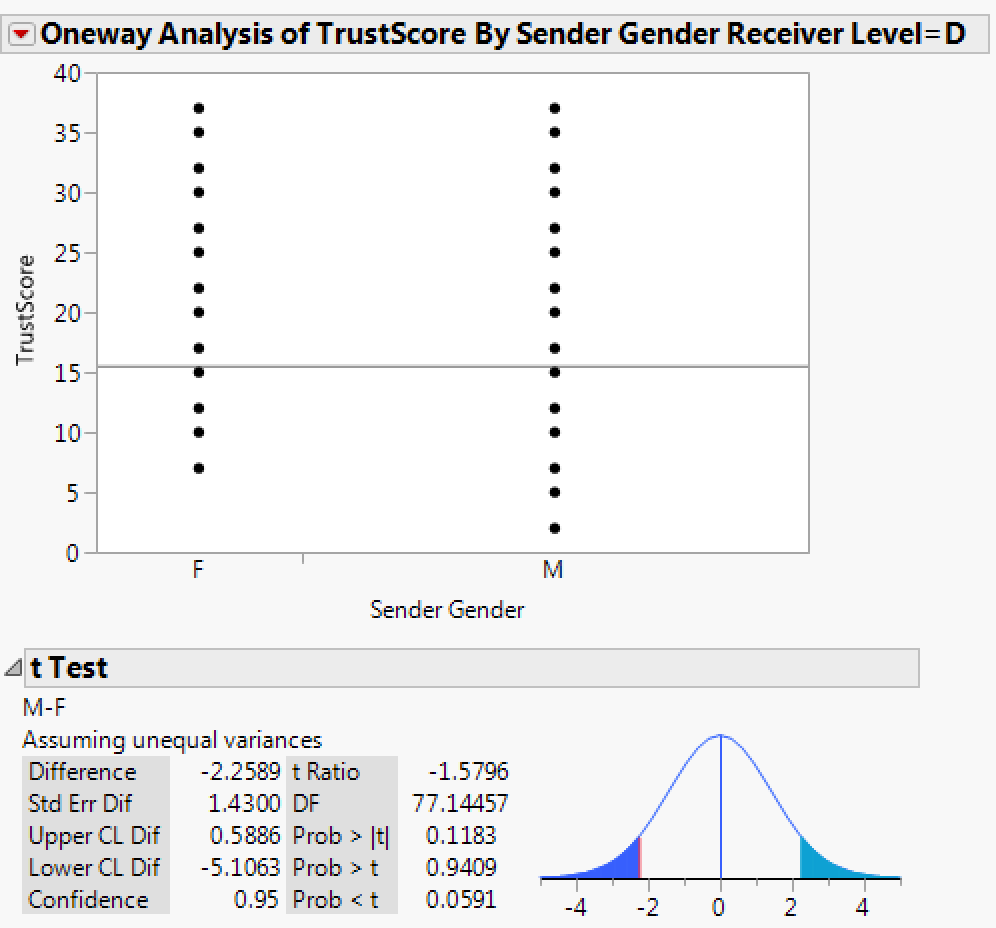
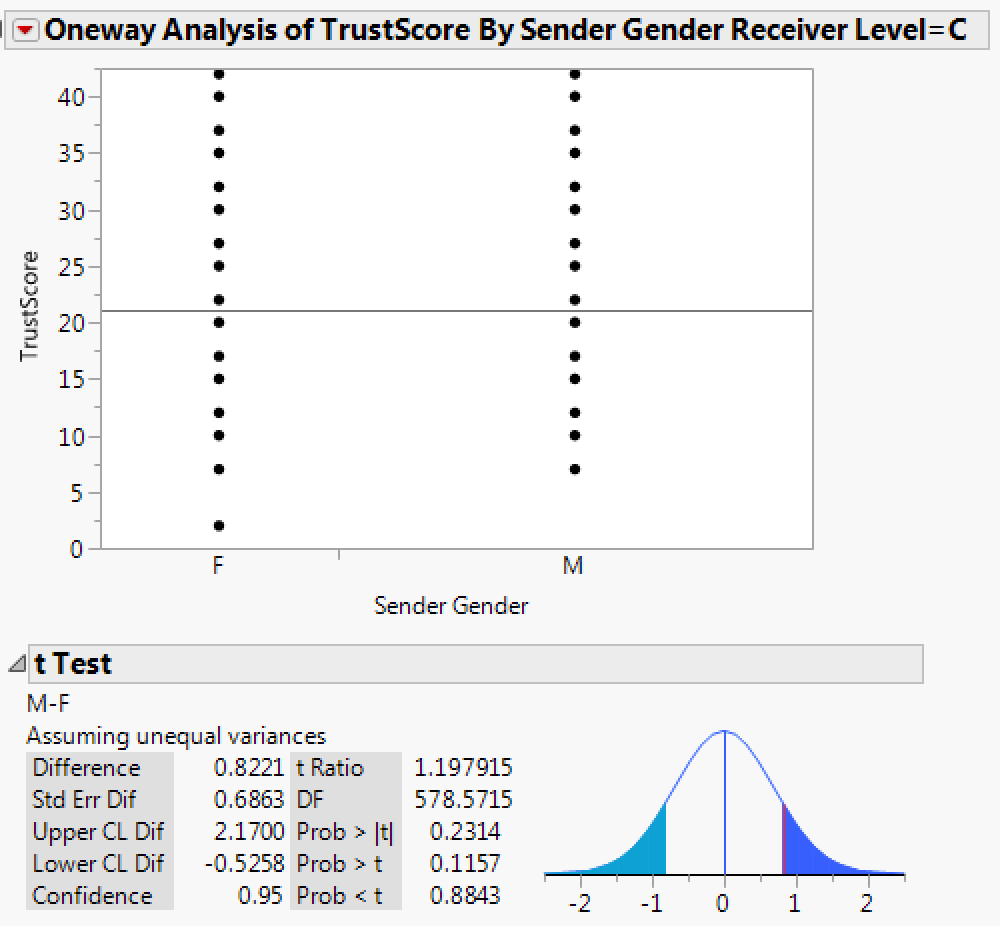
Appendix 1: Community Silos across the organization through the Givarn-Newman allocation.

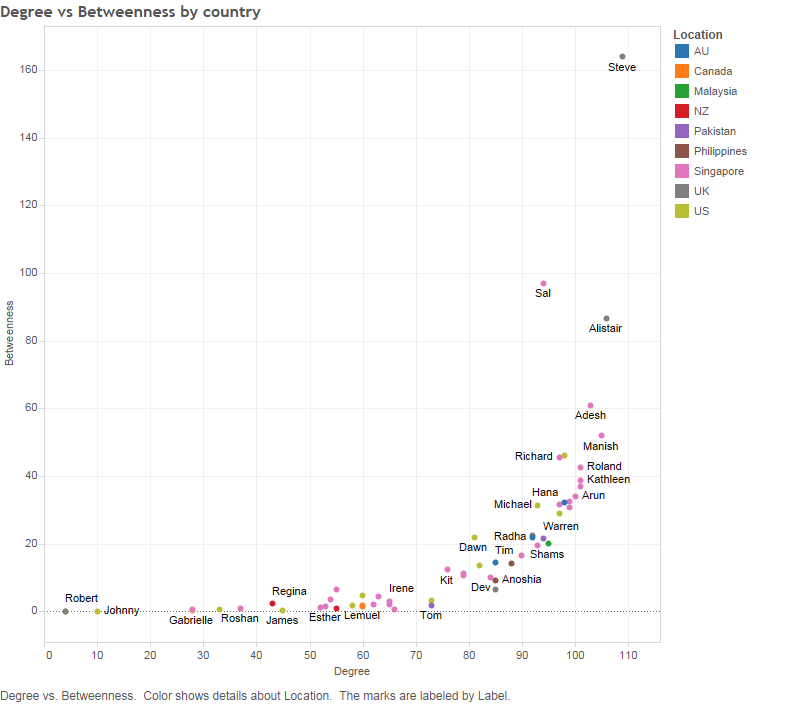


Appendix 2: Influencer Rankings of Employees

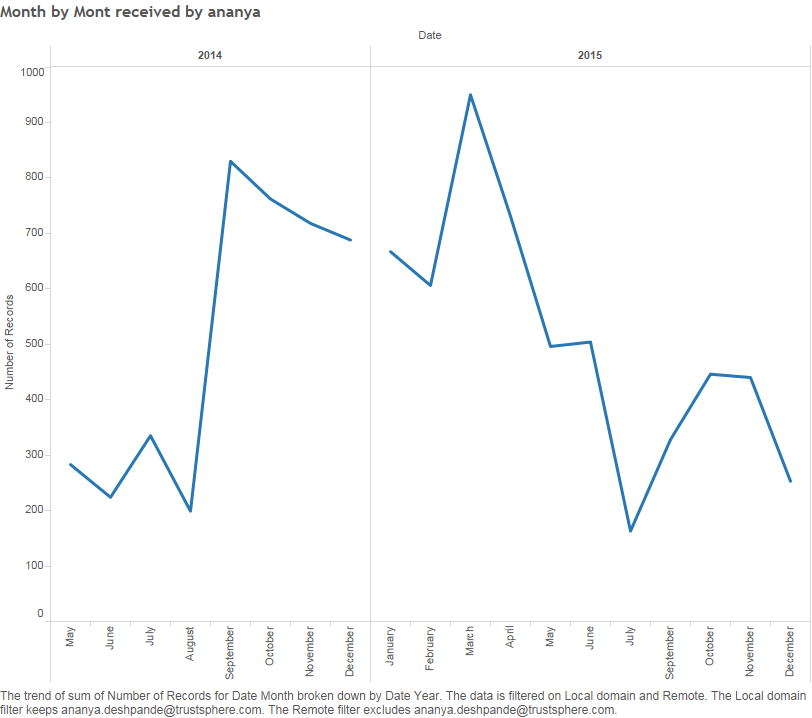


Appendix 3: Difference tests of gender relationships



Appendix 3: Difference tests of gender relationships 

Appendix 4: Gatekeepers and highly connected individuals



Appendix 5: Monthly employee activity for a single employee, Ananya Deshpande.

1. PricewaterhouseCoopers,. (2015). *Trends in People Analytics - With excerpts from the 2015 PwC Saratoga Benchmarks*. PWC. Retrieved from <http://www.pwc.com/us/en/hr-management/publications/assets/pwc-trends-in-the-workforce-2015.pdf> [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. 2Duan, L., Sheeren, E., & M. Weiss, L. (2014). *Tapping the power of hidden influencers*. *McKinsey & Company*. Retrieved 17 April 2016, from http://www.mckinsey.com/business-functions/organization/our-insights/tapping-the-power-of-hidden-influencers [↑](#footnote-ref-3)
4. Watts, D. (2016). *The Organizational Spectroscope*. *Medium*. Retrieved 17 April 2016, from https://medium.com/@duncanjwatts/the-organizational-spectroscope-7f9f239a897c#.uaq901mms [↑](#footnote-ref-4)
5. Mongeau, S. (2013). *Adopting Analytics Culture: What can be learned from social network analysis (SNA)?*. *Analyticbridge.com*. Retrieved 17 April 2016, from http://www.analyticbridge.com/profiles/blogs/adopting-analytics-culture-what-can-be-learned-from-social [↑](#footnote-ref-5)
6. https://en.wikipedia.org/wiki/Chord\_diagram [↑](#footnote-ref-6)