



SMU

SINGAPORE MANAGEMENT  
UNIVERSITY

ANLY482 – ANALYTICS PRACTICUM

PROJECT PROPOSAL

TEAM APA

Aayush Garg  
Prekshaa P. Uppin  
Akshita Dhandhanian

Using social network analysis theories to gather insights on  
effectiveness of collaboration across an organization

## INTRODUCTION

Human Resource Analytics is the idea of using data in the organizational context to understand different factors about employees such as their degree of collaboration and influence. Collaboration, being a crucial part of managing an organization is a valuable determinant in understanding how decisions are made and how relations are built. Furthermore, influence can provide a blueprint of the hubs of information flow and effective change in the organization. Through this project, we aim to provide a way of comprehending these factors through deep data analysis and patterns observed in communication interactions (email and instant messaging) of employees.

**TrustSphere** is a market leader in Relationship Analytics, delivering solutions through Sales Analytics, Risk Analytics and People Analytics. Their goal is to help clients find the value of their associated networks for improving key business challenges such as sales force effectiveness, enterprise-wide collaboration, participation and contribution statistics and corporate governance.

## MOTIVATION AND PROJECT OVERVIEW

People Analytics has been rated as the second-biggest overall capability gap in organizations by the Deloitte university press in 2015<sup>1</sup>. Through people analytics, companies are able to find better hires, improve retention, and find more suitable leaders. This has a direct impact on direction of the organization and hence its growth. Our team has a great opportunity to delve into Social Network Analysis, a fast-growing research field in Analytics through this project.

In this project, our focus is to develop various metrics that would quantify the collaboration between employees, identify the most influential employees and give managers a high-level view of these statistics to maintain a collaborative and efficient workplace. Currently at the company, these metrics are computed based on various sets of data that are primarily collected via pulse surveys. The survey data collection process is slow and makes it difficult for managers to view real-time insights. As an alternative, our team would be computing these metrics based on only email communication data. Since the data is always present in the IT system, an automated data pipeline can be created to compute the metrics and view them on a custom dashboard. We would also be involved in feature engineering to create an unbiased email network before the calculation of metrics.

A primary metric that our team would explore and test for value is a hybrid centrality to calculate an influential score. We are exploring a new equation that combines various

---

<sup>1</sup> <https://dupress.deloitte.com/dup-us-en/focus/human-capital-trends/2015/people-and-hr-analytics-human-capital-trends-2015.html>

SNA centralities to create a quantitative score of influence. To test the accuracy of the centrality, we are creating a survey that will be used a separate tool to calculate a more credible influential score. This result will then be used as test data for validity of our hybrid centrality equation. Our aim is to focus on the verify the potential value of the hybrid centrality methodology rather than delivering a very accurate equation.

Further, the company currently calculates its metrics via creating a social network of its employees in which the ties are weighted by an in-built 'trust score'. This trust score depends on only a few high-level factors such as reply rate of emails. Our team is going to perform feature engineering on raw email data and create our own 'trust score' to weight the ties. Our aim is to introduce more feature to reduce the bias during the calculation of metrics from the social network.

## OBJECTIVE

1. Perform Feature Engineering to create a new 'Trust Score' algorithm. A trust score is an aggregate weightage shows the strength of communication tie between two employees in a social network. (By Midterm)
2. Develop a dashboard that displays various metrics that would quantify the collaboration between employees, identify the most influential employees and give managers a high-level view of these statistics to maintain a collaborative and efficient workplace.
3. Research and validate the potential of a Hybrid Centrality (potentially a combination of betweenness and degree) calculated from email communication data as a measure of influence score.

## DATA

We are provided with an excel sheet containing a huge set of email exchange log via the TrustSphere domain. The data provided is clean (Screenshot of the data is shown below).

We will be collecting more data through a survey sent out to all employees of TrustSphere.

Date	Originator address	Recipient address	Direction	Type	Size	Msg ID	Email Subject
08/22/2016 09:09	vmikhail@sugarcrm.com	shaun.keating@trustsphere.cc	inbound	em	79207	<D3E08D9B;3F133%vmikhail@Re: Follow up Re: our meeting on the	
08/22/2016 09:03	annabel.koh@trustsphere.com	winminhtwe@yomastrategic.c	outbound	em	52346	<DB6PR0201MB208563F6FA2C Re: Relationship Analytics for Improv	
08/22/2016 09:02	annabel.koh@trustsphere.com	whmark@cwtlimited.com	outbound	em	69831	<DB6PR0201MB208547C8CC5; Re: Relationship Analytics for Improv	
08/22/2016 09:01	humanresources@news.humanr	annabel.koh@trustsphere.con	inbound	em	15063	<Kilaua687413-292960-4602; Your August 2016 digital edition of H	
08/22/2016 08:58	annabel.koh@trustsphere.com	dev.menon@trustsphere.com	internal	em	12250	<DB6PR0201MB208558CBE1B7 Accepted: Meeting with TrustSphere	
08/22/2016 08:58	annabel.koh@trustsphere.com	platimer@kpmg.com.sg	outbound	em	18531	<DB6PR0201MB2085CE95FC2C Re: Manulife	
08/22/2016 08:58	annabel.koh@trustsphere.com	eharayda@kpmg.com.sg	outbound	em	18531	<DB6PR0201MB2085CE95FC2C Re: Manulife	
08/22/2016 08:53	dev.menon@trustsphere.com	steve.allam@trustsphere.com	internal	em	14496	<owlAx00000000000000000000 An Opportunity Stage has been chan	
08/22/2016 08:53	dev.menon@trustsphere.com	arun.sundar@trustsphere.com	internal	em	13097	<owlAx00000000000000000000 An Opportunity Stage has been chan	
08/22/2016 08:53	dev.menon@trustsphere.com	adesh.goel@trustsphere.com	internal	em	14506	<owlAx00000000000000000000 An Opportunity Stage has been chan	
08/22/2016 08:53	dev.menon@trustsphere.com	alistair.weatherill@trustspher	internal	em	22982	<owlAx00000000000000000000 An Opportunity Stage has been chan	
08/22/2016 08:53	dev.menon@trustsphere.com	manish.goel@trustsphere.con	internal	em	13267	<owlAx00000000000000000000 An Opportunity Stage has been chan	

The data consists of the following attributes:

1. **Date:** date the email was sent/received
2. **Originator address:** email address of the sender
3. **Recipient address:** email address of the recipient
4. **Direction:**
  - a. 'Inbound' – email received by an employee of TrustSphere from an external sender
  - b. 'Outbound' – email sent by an employee of TrustSphere to external recipient
  - c. 'Internal' – email exchanged within TrustSphere employees
5. **Type:**
  - a. 'em' – message sent via email
  - b. 'im' – message sent via instant messaging
6. **Size:** number of characters in the message
7. **Msg ID:** unique ID given to every emailing chain
8. **Email Subject:** subject of the email

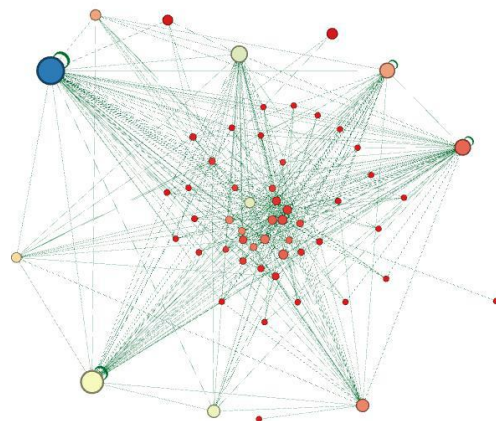
In addition, we are conducting a survey to test our hybrid centrality for influential score. It is designed in such a way that a multiple Social Networks can be created for the same employees pertaining to various influence-based factors such as perceived leadership, outside-work interaction, etc. You could find our survey via the following link:

[http://smusg.asia.qualtrics.com/SE/?SID=SV\\_6eVxySZKg8NAW2N](http://smusg.asia.qualtrics.com/SE/?SID=SV_6eVxySZKg8NAW2N)

## METHODOLOGY

The following is our methodology:

1. Understand the Scope of the project
2. Explore network and find basic trends to identify focus areas using Gephi



- a. The graph above shows a social network where each node is an Employee and every tie is an Email.
- b. There is currently no weightage to the ties at this stage

- c. Node color is set to a blue-red scale where Blue is the highest eigenvector centrality and red is the lowest eigenvector centrality. Since Eigenvector centrality is the basic measure for influence, the blue node must be the most influential employee.
  - d. The client was consulted about the blue node. We found out from the client that this person is not considered very influential in the company. Clearly, a **strong bias** was created for the number of emails sent without the weightage leading to the inaccurate result.
  - e. Hence, our team decided to spend a large portion of our time till midterm to work on feature engineering to create a relevant and accurate weight for the ties.
3. Create and send out Survey to calculate a credible influential score.
  4. Explore and perform Feature Engineering. Some features we are planning to include:
    - a. Subject Line Weightage: Based on text mining
    - b. IsReply?, isForward?, isCc?: Based on sql analysis
    - c. Reply Rate
    - d. Number of emails exchanged per week
    - e. Manager-Employee Hierarchy ranking
    - f. Features based on date patterns
  5. Weight the features to create an aggregate score (new Trust Score) for every tie.
  6. Design suitable metrics that can be calculated from the weighted network. Some examples:
    - a. Influential Score
    - b. Collaboration-Silo Index
  7. Test multiple hybrid centrality equations against survey results and finalize the algorithm.
  8. Develop a dashboard.
  9. Deliver to Client.

## SPONSOR MEETINGS AND PRIVACY

To ensure the project moves in the right direction, we will be having weekly meetings with Ridwan Ismeer, Product Manager at TrustSphere.

As we are required to access private email log of TrustSphere for the project, the team will be signing a Non-Disclosure Agreement.

## WORK PLAN

TASK		W0	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	STATUS
<b>Pre-Proposal</b>																
Gather Project Requirement	All															Completed
Create Wiki Page	Aayush, Prekshaa															Completed
Proposal Report	All															Completed
<b>Data Preparation and Exploration</b>																
Prepare Data for SNA	Akshita															
Research on SNA topics	All															
Get Familiar with SNA Softwares	All															
Exploratory Data Analysis	All															
Create Survey Questions	All															
Discuss Survey Results with Client	All															
Update Requirements and Iterate	All															
work on Hybrid Centrality	All															
Dashboard	All															
<b>Midterm</b>																
Update Wiki	Aayush, Prekshaa															
Create Midterm Presentation	Akshita															
Requirement Updates with Client	All															
<b>Post-Midterm</b>																
Revision of Scope - Larger Dataset	All															
Finalize Dashboard	All															
<b>Final Presentation</b>																
Abstract	Prekshaa															
Conference Paper	All															
Poster	Akshita															
Update Wiki	Aayush															

## REFERENCES

- ✓ <https://arxiv.org/pdf/1112.2459.pdf>
- ✓ <https://www.trustsphere.com/about-us/>
- ✓ <https://dupress.deloitte.com/dup-us-en/focus/human-capital-trends/2015/people-and-hr-analytics-human-capital-trends-2015.html>
- ✓ Karen Stephenson, 2002, Quantum Theory of Trust