

ANLY482 – ANALYTICS PRACTICUM

PROJECT PROPOSAL

TEAM APA

Aayush Garg

Prekshaa P. Uppin

Akshita Dhandhania

Using social network analysis theories to gather insights on

effectiveness of collaboration across an organization

# Introduction

During this whole decade, we have witnessed almost every industry and department hopping on to the big data wagon. In this new data-driven world, business leaders are now able to make decisions about their employees based on deep data analysis rather than the traditional methods of personal relationships, decision making based on experience, and risk avoidance. Our team will be assisting TrustSphere with such deep data analysis to help business leaders to keep better track of effectiveness of collaboration through email within the company.

**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[[1]](#footnote-1). 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. We wish to understand the basic concept and real-life application of SNA by the end of the project as well as improve our capabilities to derive deeper insights from a set of social-based data.

In this project, we will keep a focus on four main categories for the analysis, answering several questions and developing several metrics under these categories. Given below are the four categories, with some questions we are aim to answer. Over the course of the project, we will include answering more relevant questions in the project scope. Thus, the business leaders are interested in keeping a check on this area of the firm and ask various questions such as:

1. Network Strengths
	1. Find the number of relationships internally and externally (distilled by strength and date – later dates indicate there has been recent communication) of all employees. This insight will provide an understanding of which employees, departments and locations are best at building and nurturing a large number of relationships internally and externally.
2. Influence
	1. Identify top 10 employees who influence information flow within the organization. This insight can help identify agents for change as well as pinpoint employees that are overly relied on in the organization’s structure.
	2. Find the social networks of junior employees with colleagues in managerial positions. This insight will give an idea of which employee is potential turned to for advice or trusted issues by managers.
3. Collaboration
	1. Interaction within and between departments/employees, geographies. It will also highlight individual employees that collaborate well within the organization.
	2. Quantify the value an employee has on the internal and external network. It can be used to preempt the impact a departing employee will have on the network (the number of relationships that will be lost as a result of the employee leaving).
	3. Quantify a manager’s effectiveness at building relationships with his or her team and the whole organization. This behavior then can be benchmarked against ideal performers in the organization.
	4. Identify potential leaders within the organization. Research[[2]](#footnote-2) has shown that employees that build strong relationships with an organization’s various departments often possess leadership potential.
	5. Find employees who like to work in silos.
4. Email Analysis
	1. Find the average number of emails sent and received by an employee on a daily or weekly basis. How connected are these employees? Calculate a departmental average to understand collaboration between departments as well.

Answers for such questions usually rely more on qualitative observations as not many standard metrics can be directly used to track these details. Due to its general derivation, the results are often considered unreliable and hence non-actionable. The main motivation of our project is to assist TrustSphere to derive reliable and actionable insights with our data-driven approach.

# Objective

The primary objective of our project is to use social network analysis theories to create a hybrid centrality scoring method, along with other range of metrics to assess networks, influence, collaboration and email activity of employees. Additionally, we will build a comprehensive dashboard to provide TrustSphere with a clear and structured platform to view the generated metrics.

# 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.


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:**
	1. ‘Inbound’ – email received by an employee of TrustSphere from an external sender
	2. ‘Outbound’ – email sent by an employee of TrustSphere to external recipient
	3. ‘Internal’ – email exchanged within TrustSphere employees
5. **Type:**
	1. ‘em’ – message sent via email
	2. ‘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

# METHODOLOGY

We plan to take a social network analysis (SNA) approach to analyze the data since the goal is to analyze different attributes (preference to work in silos, importance, popularity) of the

actors (employees) and the relationship between them (collaboration via email). This view of the data makes it an ideal social networks data when each email address would represent a node and every email would be the relationship between both the nodes.

There are various measures of SNA proposed over the years that help determine the role and importance of a node in the network. The following are a few examples:

* Degree centrality
* Closeness centrality
* Betweenness centrality
* Eigenvector centrality

Our first step would be to explore the network using these centralities, identifying hubs, brokers and groups as well as delving into other SNA concepts discussed by various academic researchers. We would be using softwares such as Gephi, and modelling packages from R such as weighted network packages for these analysis.

Our goal is to come up with **our own hybrid centrality score** **along with a range of other metrics** that would quantify an overall importance of each node in the network. Using the insights from step one, we will be creating multiple **surveys (source of additional data)** for the employees at TrustSphere to find metrics that determine Influence, Collaboration, Network strength and email collaboration. During the process, we will be referencing to the work of Karen Stephenson and Rob Cross, both of whom specialize in the field of organizational social networks.

In the end, along with the hybrid centrality score algorithm, we will be delivering a comprehensive dynamic dashboard visualizing the most relevant measures that we identify during our project.

# SCOPE OF WORK

1. Create a hybrid centrality score as an overall comprehensive measure of the network
2. Assess Collaboration using a range of metrics
	1. Within departments
	2. Within different geographical regions
	3. Within projects
3. Identify Silos
4. Assess Influence, Network Strength and Email collaboration
5. Develop a dynamic dashboard to visualize relevant measures
6. The Scope is fluid and will become more specific as the project progresses

# 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



# 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
1. <https://dupress.deloitte.com/dup-us-en/focus/human-capital-trends/2015/people-and-hr-analytics-human-capital-trends-2015.html> [↑](#footnote-ref-1)
2. Karen Stephenson, 2002, Quantum Theory of Trust [↑](#footnote-ref-2)