Analytics Practicum Midterm – Team 17

Aayush Garg, Prekshaa Uppin, Akshita Dhandhania

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**

Through this project, we aim to provide a way of comprehending these factors through **deep data analysis and patterns observed in communication interactions of employees**

our project

our client

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**

OBJECTIVES

Perform **Feature Engineering** to create **a new 'Trust Score' algorithm**. A trust score is an aggregate weightage that shows the strength of communication tie between two employees in a social network.

01

02

Develop a dashboard that displays various **metrics** that would quantify the collaboration between employees and identify the most influential employees

Research and validate the potential of **a Hybrid Centrality** calculated from email communication data as a measure of influence score.

03

Email Data

Columns	Column Explanation
Date	Date of the E-mail
Remote IP	If the email exchange is external, then this column shows the external person's email
Remote	The trust sphere employee who is receiving the email
Remote Domain	Always trustsphere
Local	E-mail address of the person sending the email
Local Domain	Domain of the person who is sending the email
Originator	Inbound, Outbound or Internal
Direction	Always trustsphere in this case
Domain Group	Email Header (Subject Line)
Subject	Type of message: email/im/voice/sms
Inbound Count	Number of emails received
Outbound count	Number of emails sent
Size	Size of the message
Msgid	Encoded Message ID

Email Data

**Data from
11/26/2016 8:00 am
to
02/01/2017 00:00 am**

Total Number of
Columns: 14

Total Number of
Rows: 121154

Filter out inbound and
outbound emails &
filter data to only keep
email data

-> 45855 rows of data
remaining

Originator
internal
inbound
internal
internal
internal
internal
inbound
outbound
outbound

Subject
em
em
im
em
em

Email Data

Remove System Email occurrences from dataset:

Local	Remote
• heartbeat@trustsphere.com (1658)	• crm.report@trustsphere.com (6197)
• trustvault.selfservice@trustsphere.com (95)	• customer.care@trustsphere.com (5)
• tv.reports@trustsphere.com (1394)	• sg.boardroom@trustsphere.com (25)
• And more	• And more

After removing system emails, **29797 rows** left

Email Data

Remove unnecessary columns such as:

Remote IP
Remote Domain
Local Domain
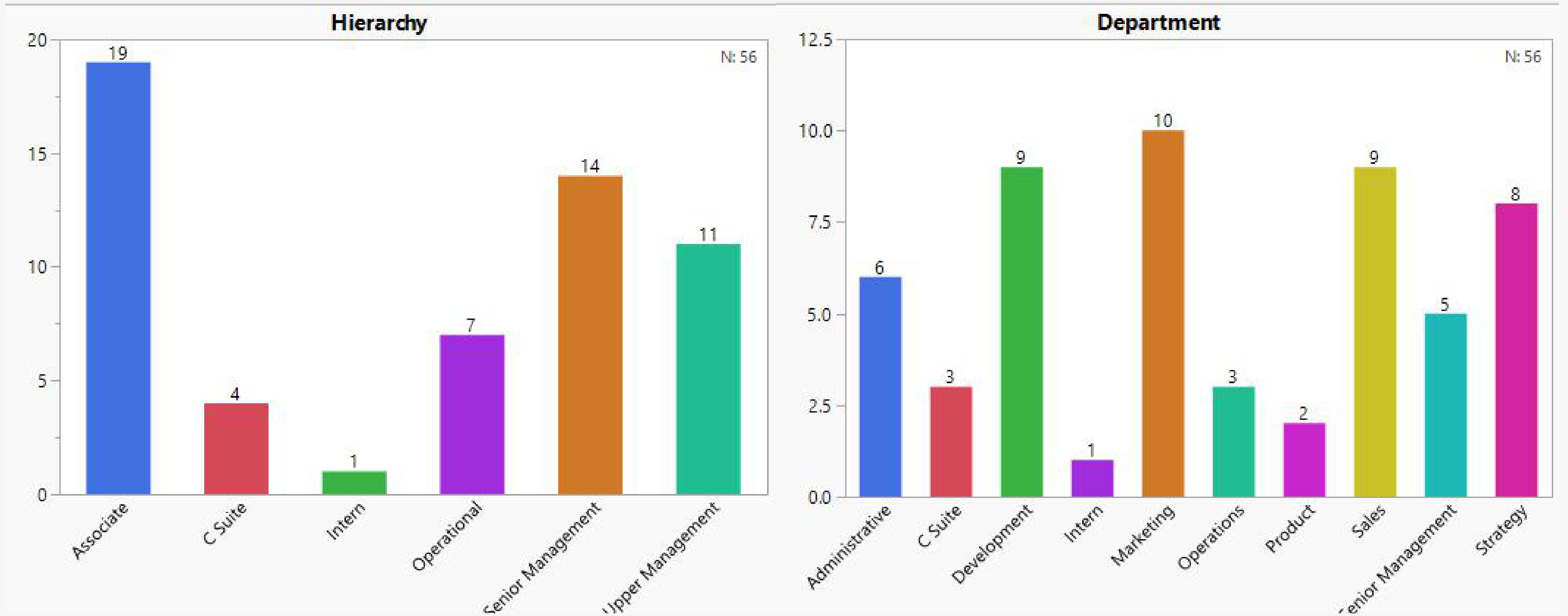Direction
Inbound count
Outbound count
Subject

No missing data

Staff Data
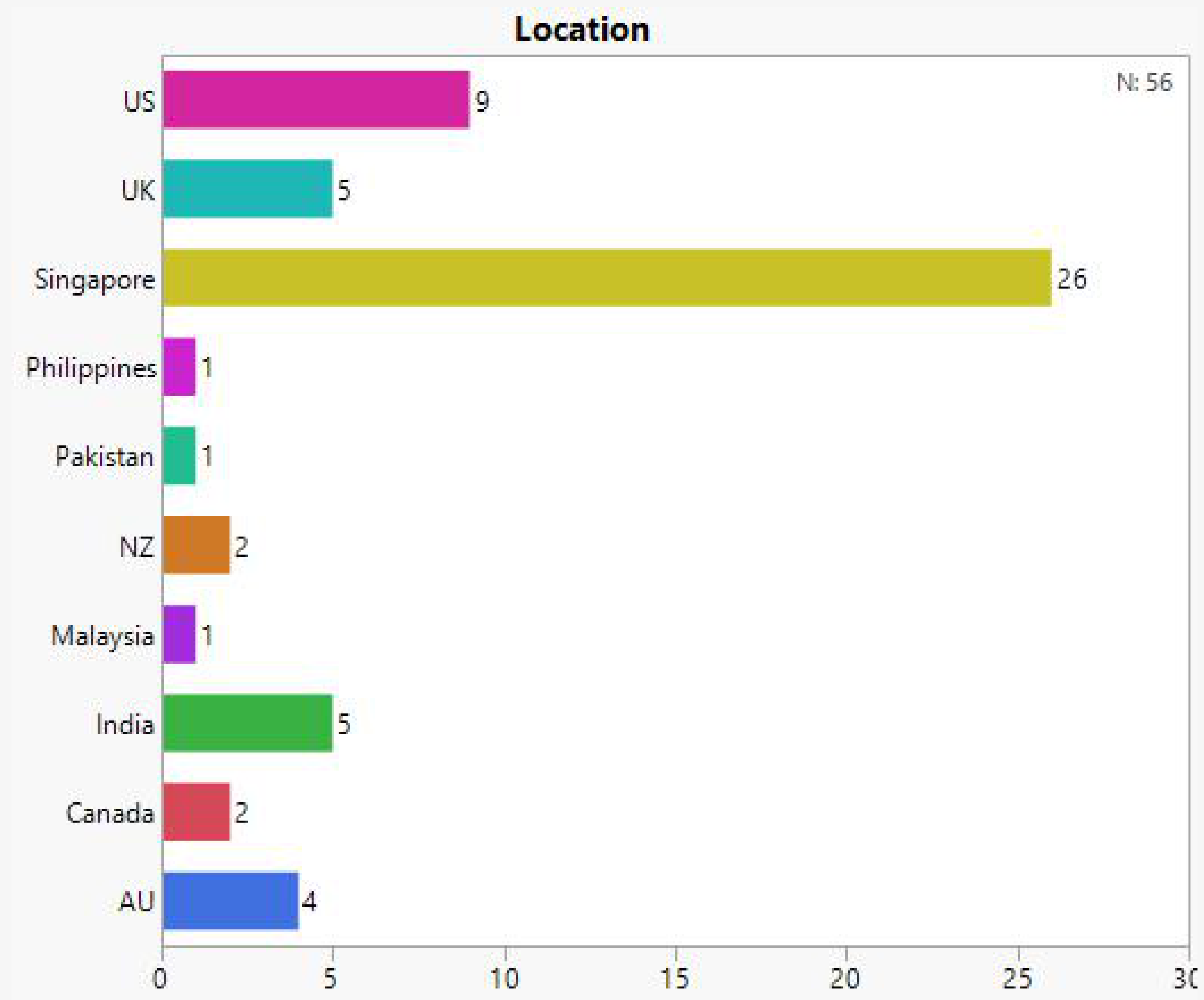
Number of Rows: 56

Columns	Column Explanation
Name	Name of the employee
Heirarchy	Designation of the employee
Department	Department of the employee
Location	The location where the employee is based

Staff Data

Hierarchy

N: 56

Associate	19
C Suite	4
Intern	1
Operational	7
Senior Management	14
Upper Management	11

Department

N: 56

Administrative	6
C Suite	3
Development	9
Intern	1
Marketing	10
Operations	3
Product	2
Sales	9
Senior Management	5
Strategy	8

Staff Data

Data Exploration: Email Data vs Staff Data

From Email Data (Last 10 weeks)		Staff List			
id		**Name**	**Hierarchy**	**Department**	**Location**
adesh.goel@trustsphere.com		Adesh Goel	C Suite	C Suite	Singapore
alistair.weatherill@trustsphere.com		Ajay Rana	Associate	Sales	India
amanda@trustsphere.com		Alistair Weatherill	Upper Management	Operations	UK
		Ananya Deshpande	Associate	Strategy	Singapore
annabel.koh@trustsphere.com		Annabel Koh	Associate	Strategy	Singapore
antony.ebelle@trustsphere.com					
aravind.mp@trustsphere.com					
		Anoshia Naseer	Associate	Strategy	Pakistan
arun.sundar@trustsphere.com		Arun Sundar	C Suite	Senior Management	Singapore
bersileus.sacamay@trustsphere.com					
brian.lebahn@trustsphere.com		Brian Lebahn	Upper Management	Senior Management	US
bryan.acedo@trustsphere.com		Bryan Acedo	Senior Management	Development	Philippines
dawn.radecki@trustsphere.com		Dawn Radecki	Upper Management	Marketing	US
deokant.payasi@trustsphere.com					
dev.menon@trustsphere.com		Dev Menon	Upper Management	Sales	Singapore
		Elizabeth Botes	Upper Management	Marketing	US
esther.tan@trustsphere.com		Esther Tan	Operational	Administrative	Singapore
gabrielle.tremblay@trustsphere.com		Gabrielle Tremblay	Intern	Intern	Singapore
gladys.opone@trustsphere.com		Gladys Opone	Senior Management	Development	US
grace.siew@trustsphere.com		Grace Siew	Operational	Administrative	Singapore
graham.wells@trustsphere.com		Graham Wells	Upper Management	Development	UK
greg.newman@trustsphere.com		Greg Newman	Upper Management	Product	Singapore
hana.owens@trustsphere.com		Hana Owens	Associate	Strategy	Singapore
hansel.than@trustsphere.com		Hansel Than	Associate	Development	Malaysia
irene.lim@trustsphere.com		Irene Lim	Senior Management	Sales	Singapore
		Jeff Martin	Associate	Sales	US
johnny.helmberger@trustsphere.com		Johnny Helmberger	Associate	Sales	US
kathleen.tan@trustsphere.com		Kathleen Tan	Operational	Administrative	Singapore
kit.perez@trustsphere.com		Kit Perez	Senior Management	Development	Singapore
kunal.thukral@trustsphere.com		Kunal Thukral	Senior Management	Strategy	Singapore
lemuel.caldito@trustsphere.com		Lemuel Caldito	Associate	Development	Canada
manish.goel@trustsphere.com		Manish Goel	C Suite	C Suite	Singapore
manmeet.nayyar@trustsphere.com		Manmeet Nayyar	Associate	Sales	India
mark.padginton@trustsphere.com		Mark Padginton	Upper Management	Strategy	AU
mark.whittenburg@trustsphere.com					
michael.knight@trustsphere.com		Michael Knight	Associate	Senior Management	US
oscar.veronese@trustsphere.com		Oscar Veronese	Upper Management	Marketing	Singapore
paolo.aquino@trustsphere.com		Paolo Aquino	Senior Management	Development	Canada
penny.phonenaing@trustsphere.com		Penny Phonenaing	Associate	Marketing	Singapore
priya.bagga@trustsphere.com		Priya Bagga	Upper Management	Marketing	Singapore
radha.chopra@trustsphere.com		Radha Chopra	Senior Management	Marketing	AU
regina.foo@trustsphere.com		Regina Foo	Associate	Marketing	Singapore

Highlighted in Pink:

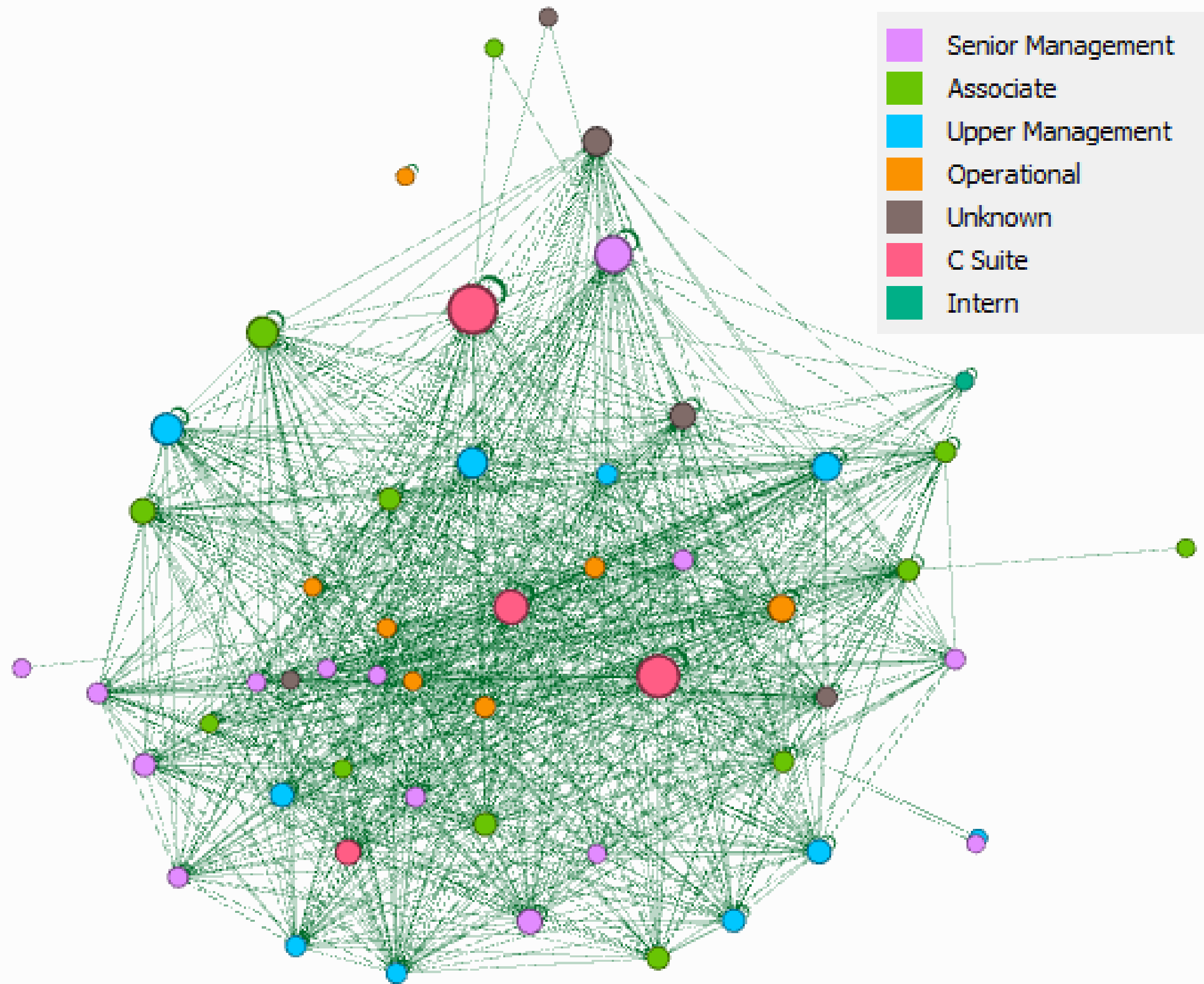• Employees in Staff List but no internal emails in the last 10 weeks

• 8 Employees

Highlighted in Yellow:

• Employees with Email Records in the last 10 weeks but not part of the staff list

• 5 Employees

Data Exploration: Network

Legend:
- Senior Management (purple)
- Associate (green)
- Upper Management (blue)
- Operational (orange)
- Unknown (brown)
- C Suite (pink)
- Intern (teal)

Node: Each employee
Node Color: Hierarchy
Node Size: Eigenvector Centrality

• No weights for edges – purely based on quantity

• Many Senior Management and Upper Management Employees seem to have a low centrality score

• Possibly a biased solution

• Need for feature engineering to add weight that removes the bias

METHODOLOGY

01
Understand the Scope of the project

02
Explore and Clean Data

03
Perform Feature Engineering on Email Data

05
Design suitable metrics that can be calculated from the weighted network

04
Using the features, create an aggregate score (new Trust Score) that will be used as weight for the communication network

Create and send out a Survey to create test data that would validate the hybrid centrality metric (influential score)

06
Test multiple hybrid centrality equations against survey results and finalize the algorithm
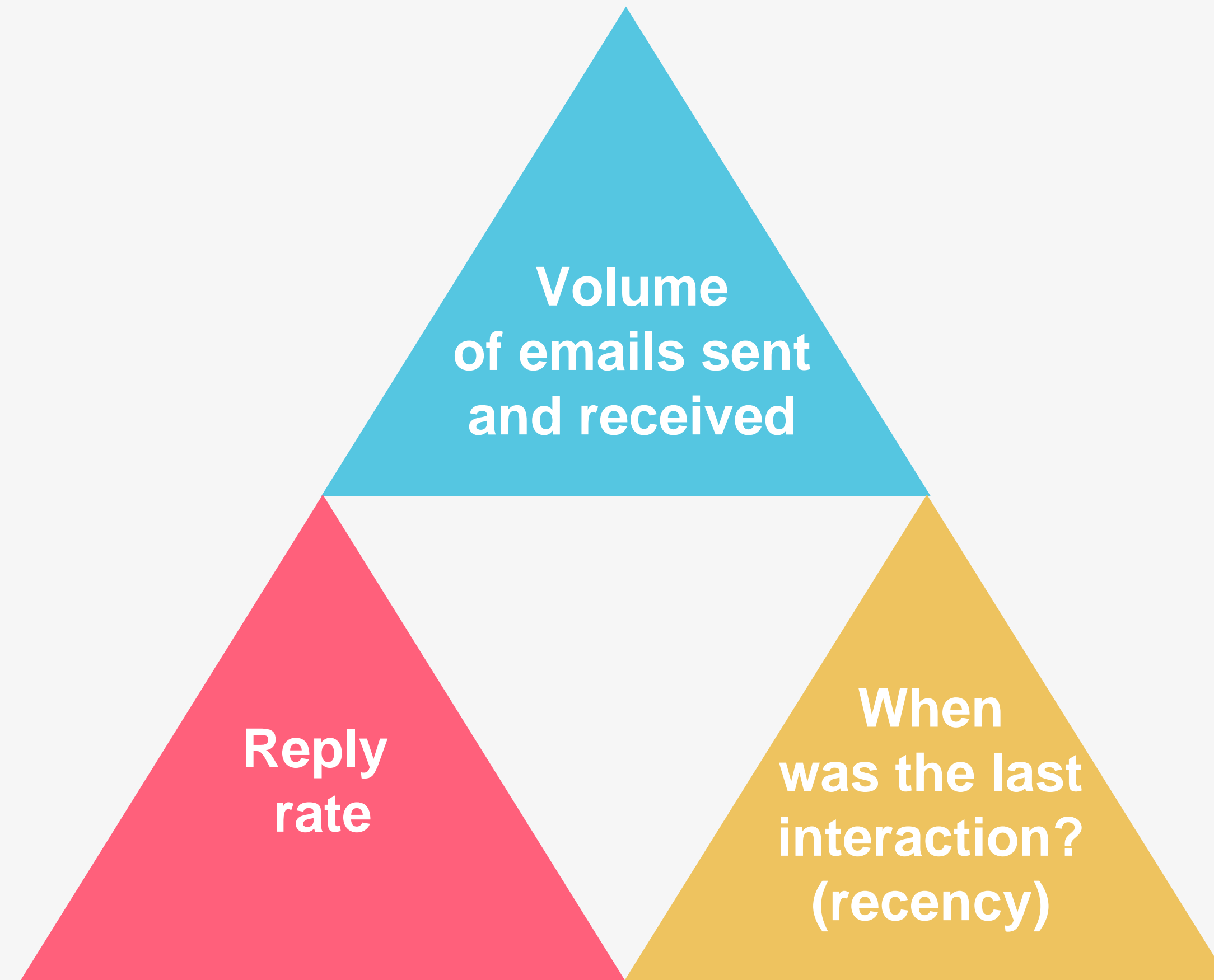
07
Develop a dashboard to display all metrics using R

08
Deliver to client

AS-IS TRUST SCORE

Based on high level features such as

Volume of emails sent and received

Reply rate

When was the last interaction? (recency)

FEATURE ENGINEERING

SUBJECT LINE WEIGHTAGE	**EMAIL EXCHANGE RATIO**	**AVERAGE EMAIL EXCHANGE SIZE**	**EMAIL CHAIN RATIO**	**RATE OF EXCHANGE**
quality	frequency	quantity	variety	regularity

FEATURE 1 – SUBJECT LINE WEIGHTAGE

SETTINGS ON SAS EM

Data → Text Parsing → Text Filter
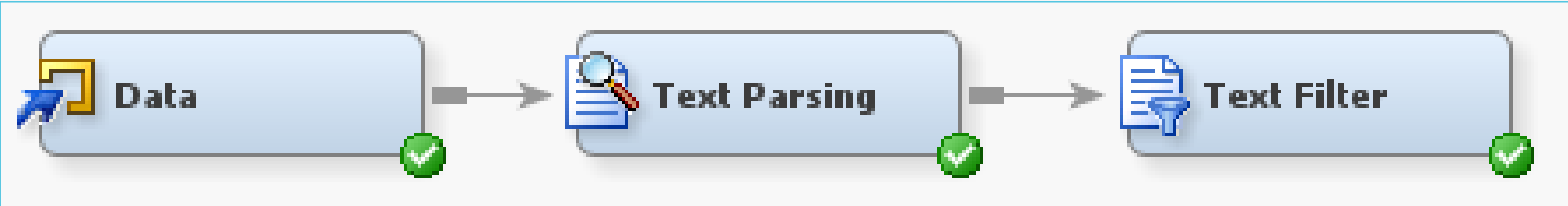
TEXT PARSING

Ignore Types of Entities

Choices
- Address
- Company
- Currency
- Date
- Internet
- Location
- Measure
- Organization
- Percent
- Person
- Phone
- Product
- Prop_misc
- Ssn
- Time

OK | Clear | Cancel

Ignore Types of Attributes

Choices
- Abbr
- Alpha
- Mixed
- Num
- Punct

OK | Clear | Cancel

Ignore Parts of Speech

Choices
- Aux
- Conj
- Det
- Interj
- Noun
- Num
- Part
- Pref

OK | Clear | Cancel

Stop List-EMWS1.TextParsing_stopList

Term
'd
'll
'm
're
's
've
18th
7am
872-240-3212
Automatic reply
a
aboard
about
above
according
accordingly
across
actually
adesh

TEXT FILTER

Weightings	
Frequency Weighting	Log
Term Weight	Inverse Document Frequency
Term Filters	
Minimum Number of Documents	30
Maximum Number of Terms	.
Import Synonyms	...

Create Equivalent Terms

Select term to represent entire set

Term	Role
sugar	Prop
sugarcrm	Prop

FEATURE 1 – SUBJECT LINE WEIGHTAGE

JUSTIFICATION — To find important emails based on the subject line

RESULTS

Term	1/weight
meeting	0.200965
sugar	0.200361
update	0.187758
poc	0.177022
discussion	0.169808
opportunity	0.168067
sales	0.163212

Date	Target	Source	Originator	Domain group	Size	Msgid	Subject Weightage
1/31/2017 23:58	arun.sundar@trustsphere.com	hana.owens@trustsphere.com	internal	Call with HCLI	21358	<AM4PR0201MB1	0
1/31/2017 23:52	steve.allam@trustsphere.com	tom.butler@trustsphere.com	internal	RE: SOW_DBS_hk(2)-1 (003).docx	872031	<VI1PR0202MB29	0.138419167
1/31/2017 23:52	dev.menon@trustsphere.com	tom.butler@trustsphere.com	internal	RE: SOW_DBS_hk(2)-1 (003).docx	872031	<VI1PR0202MB29	0.138419167
1/31/2017 23:52	annabel.koh@trustsphere.com	tom.butler@trustsphere.com	internal	RE: SOW_DBS_hk(2)-1 (003).docx	872031	<VI1PR0202MB29	0.138419167
1/31/2017 23:52	arun.sundar@trustsphere.com	tom.butler@trustsphere.com	internal	RE: SOW_DBS_hk(2)-1 (003).docx	872031	<VI1PR0202MB29	0.138419167
1/31/2017 23:44	shaun.keating@trustsphere.com	warren.tait@trustsphere.com	internal	Re: Resignation	50644	<336252EC-0E96-4	0
1/31/2017 23:42	warren.tait@trustsphere.com	shaun.keating@trustsphere.com	internal	Re: Resignation	42933	<C2C21FCF-6298-4	0
1/31/2017 23:41	shaun.keating@trustsphere.com	warren.tait@trustsphere.com	internal	Resignation	19411	<DA248A80-1B76-	0
1/31/2017 23:41	mark.padginton@trustsphere.com	warren.tait@trustsphere.com	internal	Resignation	19411	<DA248A80-1B76-	0
1/31/2017 23:41	brian.lebahn@trustsphere.com	manish.goel@trustsphere.com	internal	Re: meet with Graco next week	220240	<ed14fb28-4444-c	0.137540144
1/31/2017 23:40	radha.chopra@trustsphere.com	mark.padginton@trustsphere.com	internal	FW: Sales Ready Lead > +40	70289	<A47EE504-69F4-4	0.142603932
1/31/2017 23:40	shaun.keating@trustsphere.com	mark.padginton@trustsphere.com	internal	FW: Sales Ready Lead > +40	70289	<A47EE504-69F4-4	0.142603932
1/31/2017 23:40	priya.bagga@trustsphere.com	mark.padginton@trustsphere.com	internal	FW: Sales Ready Lead > +40	70289	<A47EE504-69F4-4	0.142603932
1/31/2017 23:38	manish.goel@trustsphere.com	brian.lebahn@trustsphere.com	internal	RE: meet with Graco next week	156697	<HE1PR0202MB26	0.137540144
1/31/2017 23:36	shaun.keating@trustsphere.com	warren.tait@trustsphere.com	internal	FW: Sales Ready Lead > +40	23870	<E2BA850F-40BA-	0.142603932
1/31/2017 23:36	mark.padginton@trustsphere.com	warren.tait@trustsphere.com	internal	FW: Sales Ready Lead > +40	23870	<E2BA850F-40BA-	0.142603932
1/31/2017 23:35	shaun.keating@trustsphere.com	manish.goel@trustsphere.com	internal	Re: IM screen	184126	<2427ee97-1109-5	0
1/31/2017 23:33	dawn.radecki@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Sales Ready Lead > +40	120900	<985545a8-223d-7	0.142603932
1/31/2017 23:32	arun.sundar@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Manish - approval pls	93906	<b68ba518-9fce-a	0
1/31/2017 23:32	adesh.goel@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Manish - approval pls	93906	<b68ba518-9fce-a	0
1/31/2017 23:29	manish.goel@trustsphere.com	arun.sundar@trustsphere.com	internal	Re: Manish - approval pls	92621	<228053DF-F3C1-4	0
1/31/2017 23:29	adesh.goel@trustsphere.com	arun.sundar@trustsphere.com	internal	Re: Manish - approval pls	92621	<228053DF-F3C1-4	0
1/31/2017 23:29	arun.sundar@trustsphere.com	manish.goel@trustsphere.com	internal	Re: YC list	54081	<292dc4be-8416-3	0.125580811
1/31/2017 23:28	arun.sundar@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Manish - approval pls	110991	<70e83838-8ecf-e	0
1/31/2017 23:28	arun.sundar@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Manish - approval pls	110343	<70e83838-8ecf-e	0
1/31/2017 23:28	adesh.goel@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Manish - approval pls	110343	<70e83838-8ecf-e	0
1/31/2017 23:27	arun.sundar@trustsphere.com	manish.goel@trustsphere.com	internal	Re: Recommended Partner in UK?	46480	<d00af210-05e9-1	0.132450331
1/31/2017 23:20	shaun.keating@trustsphere.com	tom.butler@trustsphere.com	internal	RE: New release candidate for Sugar v1.2 - (manif	148279	<VI1PR0202MB29	0.166980592

FEATURE 2 – Email Exchange Ratio

FORMULA

$$\frac{N_{ab}}{N_a + N_b - N_{ab}}$$

N_{ab}: Number of emails exchanged between A and B
N_a: Number of emails sent by A
N_b:Number of emails sent by B

SQL STATEMENT

```
select t1.employee1, t1.employee2, (t1.total/(IFNULL(t2.total,0) + IFNULL(t3.total,0) - t1.total)) as
EmailExchangedRatio from

(select least(`Remote`,`Local`) as employee1, greatest(`Remote`,`Local`) as employee2, count(*) as total from
midterm group by least(`Remote`,`Local`), greatest(`Remote`,`Local`)) as t1

left join

(select `Local` as employee, count(*) as total from midterm group by `Local`) as t2

on t2.employee = t1.employee1

left join

(select `Local` as employee, count(*) as total from midterm group by `Local`) as t3

on t3.employee = t1.employee2
```

JUSTIFICATION

Number of emails exchanged between two employees shows how much they interact, collaborate and share information

RESULTS

employee1	employee2	EmailExchangedRatio
adesh.goel@trustsphere.com	adesh.goel@trustsphere.com	0.0021
adesh.goel@trustsphere.com	alistair.weatherill@trustsphere.com	0.0710
adesh.goel@trustsphere.com	annabel.koh@trustsphere.com	0.0076
adesh.goel@trustsphere.com	antony.ebelle@trustsphere.com	0.0277
adesh.goel@trustsphere.com	aravind.mp@trustsphere.com	0.0042
adesh.goel@trustsphere.com	arun.sundar@trustsphere.com	0.1065
adesh.goel@trustsphere.com	bersileus.sacamay@trustsphere.com	0.0014
adesh.goel@trustsphere.com	brian.lebahn@trustsphere.com	0.0369
adesh.goel@trustsphere.com	bryan.acedo@trustsphere.com	0.0082
adesh.goel@trustsphere.com	dawn.radecki@trustsphere.com	0.0230
adesh.goel@trustsphere.com	dev.menon@trustsphere.com	0.0255
adesh.goel@trustsphere.com	gabrielle.tremblay@trustsphere.com	0.0013

FEATURE 3 – Average Email Exchange Size

FORMULA

Average (Size of all emails exchanged between A and B)

SQL STATEMENT

```
select least(`Remote`,`Local`) as employee1, greatest(`Remote`,`Local`) as employee2, avg(size)
from midterm group by least(`Remote`,`Local`), greatest(`Remote`,`Local`)
```

RESULTS

JUSTIFICATION

Larger email size shows larger amount of information exchange

employee1	employee2	avg(size)
adesh.goel@trustsphere.com	adesh.goel@trustsphere.com	55786.3333
adesh.goel@trustsphere.com	alistair.weatherill@trustsphere.com	188445.8462
adesh.goel@trustsphere.com	annabel.koh@trustsphere.com	1191951.0714
adesh.goel@trustsphere.com	antony.ebelle@trustsphere.com	258309.8182
adesh.goel@trustsphere.com	aravind.mp@trustsphere.com	767172.2000
adesh.goel@trustsphere.com	arun.sundar@trustsphere.com	253348.0875
adesh.goel@trustsphere.com	bersileus.sacamay@trustsphere.com	44036.0000
adesh.goel@trustsphere.com	brian.lebahn@trustsphere.com	149990.2553
adesh.goel@trustsphere.com	bryan.acedo@trustsphere.com	23118.1111
adesh.goel@trustsphere.com	dawn.radecki@trustsphere.com	482241.5641
adesh.goel@trustsphere.com	dev.menon@trustsphere.com	38610.7544
adesh.goel@trustsphere.com	gabrielle.tremblay@trustsphere.com	44036.0000
adesh.goel@trustsphere.com	gladys.opone@trustsphere.com	32002.6667

Feature 4 – Email Chain Ratio

FORMULA

$$1 - \frac{N_u}{N_{ab}}$$

N_u: Number of emails exchanged between A and B with unique subject lines

N_{ab}: Number of emails exchanged between A and B

JUSTIFICATION

Number of emails with unique subject lines shows number of different conversations taking place between employees

RESULTS

employee1	employee2	ratio
adesh.goel@trustsphere.com	adesh.goel@trustsphere.com	0.0000
adesh.goel@trustsphere.com	alistair.weatherill@trustsphere.com	0.3205
adesh.goel@trustsphere.com	annabel.koh@trustsphere.com	0.3571
adesh.goel@trustsphere.com	antony.ebelle@trustsphere.com	0.3409
adesh.goel@trustsphere.com	aravind.mp@trustsphere.com	0.0000
adesh.goel@trustsphere.com	arun.sundar@trustsphere.com	0.1778
adesh.goel@trustsphere.com	bersileus.sacamay@trustsphere.com	0.0000
adesh.goel@trustsphere.com	brian.lebahn@trustsphere.com	0.1915
adesh.goel@trustsphere.com	bryan.acedo@trustsphere.com	0.0000
adesh.goel@trustsphere.com	dawn.radecki@trustsphere.com	0.3077
adesh.goel@trustsphere.com	dev.menon@trustsphere.com	0.7719
adesh.goel@trustsphere.com	gabrielle.tremblay@trustsphere.com	0.0000
adesh.goel@trustsphere.com	gladys.opone@trustsphere.com	0.0000
adesh.goel@trustsphere.com	grace.siew@trustsphere.com	0.1176
adesh.goel@trustsphere.com	graham.wells@trustsphere.com	0.3000
adesh.goel@trustsphere.com	greg.newman@trustsphere.com	0.1471
adesh.goel@trustsphere.com	hana.owens@trustsphere.com	0.4118

SQL STATEMENT

```
SELECT t1.employee1, t1.employee2, (1-(t1.uniqueEmails/t1.total)) as ratio from

(SELECT least(`Remote`,`Local`) as employee1, greatest(`Remote`,`Local`) as employee2, count(*) as total,
count( distinct `Domain group`) as uniqueEmails FROM midterm group by least(`Remote`,`Local`),
greatest(`Remote`,`Local`)) as t1
```

Feature 5 – Rate of exchange of emails

FORMULA

$$\frac{N_{ab}}{c}$$

N_{ab}: Number of emails exchanged between A and B
c: number of weeks

JUSTIFICATION

Rate of exchange of emails shows how regularly employees interact with one another

RESULTS

employee1	employee2	total
adesh.goel@trustsphere.com	adesh.goel@trustsphere.com	0.3000
adesh.goel@trustsphere.com	alistair.weatherill@trustsphere.com	7.8000
adesh.goel@trustsphere.com	annabel.koh@trustsphere.com	1.4000
adesh.goel@trustsphere.com	antony.ebelle@trustsphere.com	4.4000
adesh.goel@trustsphere.com	aravind.mp@trustsphere.com	0.5000
adesh.goel@trustsphere.com	arun.sundar@trustsphere.com	34.3000
adesh.goel@trustsphere.com	bersileus.sacamay@trustsphere.com	0.1000
adesh.goel@trustsphere.com	brian.lebahn@trustsphere.com	4.7000
adesh.goel@trustsphere.com	bryan.acedo@trustsphere.com	0.9000
adesh.goel@trustsphere.com	dawn.radecki@trustsphere.com	3.9000
adesh.goel@trustsphere.com	dev.menon@trustsphere.com	5.7000
adesh.goel@trustsphere.com	gabrielle.tremblay@trustsphere.com	0.1000
adesh.goel@trustsphere.com	gladys.opone@trustsphere.com	0.3000
adesh.goel@trustsphere.com	grace.siew@trustsphere.com	1.7000
adesh.goel@trustsphere.com	graham.wells@trustsphere.com	1.0000
adesh.goel@trustsphere.com	greg.newman@trustsphere.com	3.4000

SQL STATEMENT

```
(select least(`Remote`,`Local`) as employee1, greatest(`Remote`,`Local`) as employee2, (count(*)/10) as total
from midterm group by least(`Remote`,`Local`), greatest(`Remote`,`Local`))
```

NEXT STEPS

01
Understand the Scope of the project

02
Explore and Clean Data

03
Perform Feature Engineering on Email Data

05
Design suitable metrics that can be calculated from the weighted network

04
Using the features, create an aggregate score (new Trust Score) that will be used as weight for the communication network

Create and send out a Survey to create test data that would validate the hybrid centrality metric (influential score)
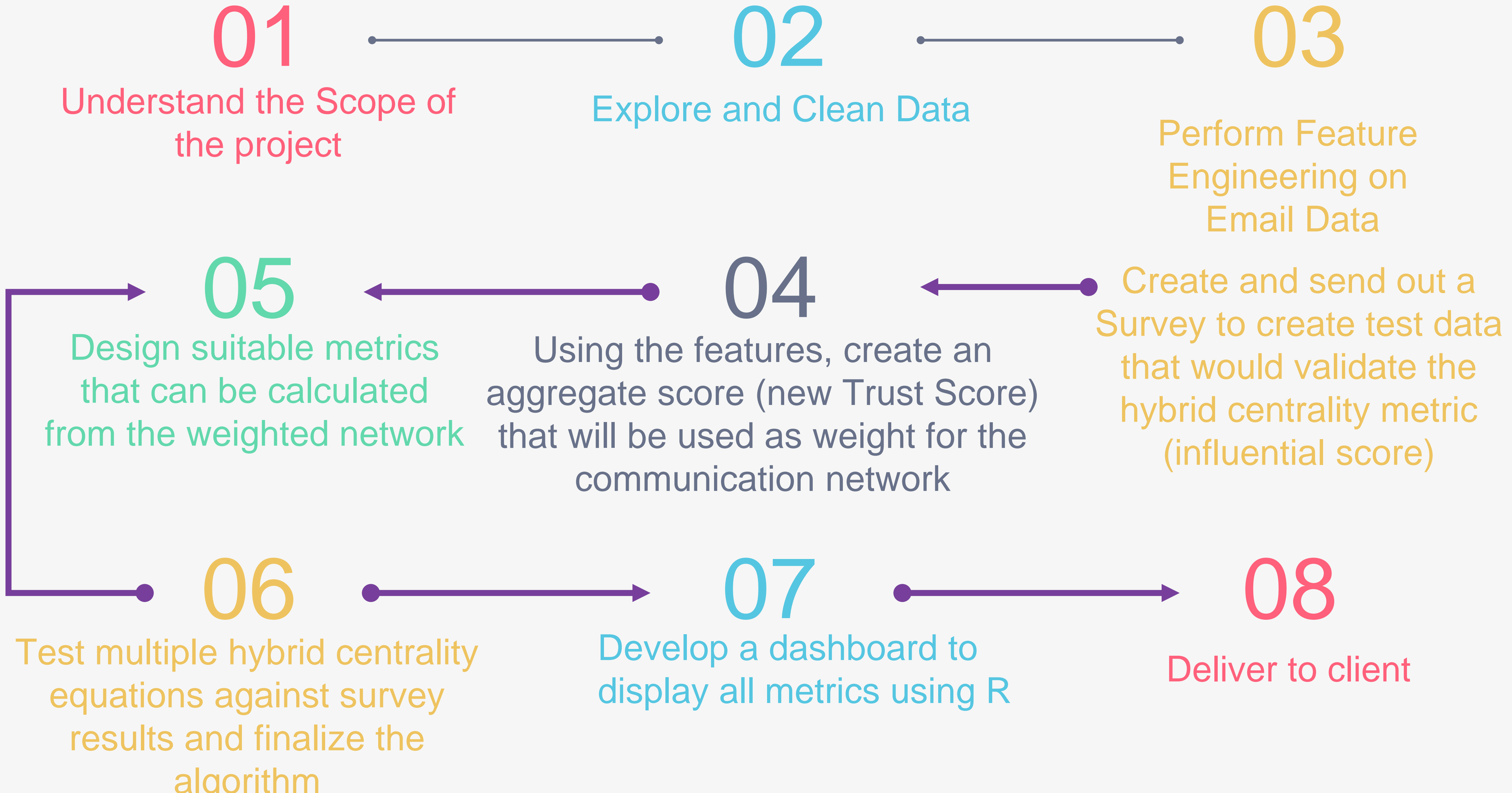
06
Test multiple hybrid centrality equations against survey results and finalize the algorithm

07
Develop a dashboard to display all metrics using R

08
Deliver to client

SURVEY

OBJECTIVE

Our aim is to use the survey to validate if an email exchange network is a good tool to calculate influence score.

We define Influence Score as the extent to which an individual sways information flow in the workplace.

6 TYPES OF INFORMATION FLOW (INFLUENCE)

SOCIAL
How many times do you interact with the following colleagues regarding business topics, within a month?

DECISION MAKING
How many times do you consult the following colleagues if you have a work-related decision to make within a week?

INFORMATION SHARING
How many times do you receive job related information from the following colleagues within a month?

SUPPORT
How many times do you discuss your career prospects and progression with the following colleagues in a year?

PROBLEM SOLVING
How many times do you seek help from the following colleagues for business/technical related problems within a week?

IDEA GENERATION
How many times do you discuss, share or brainstorm novel ideas with the following colleagues, in a quarter?

WEEK 14

Deliver to client

08

FINISH

THANK YOU