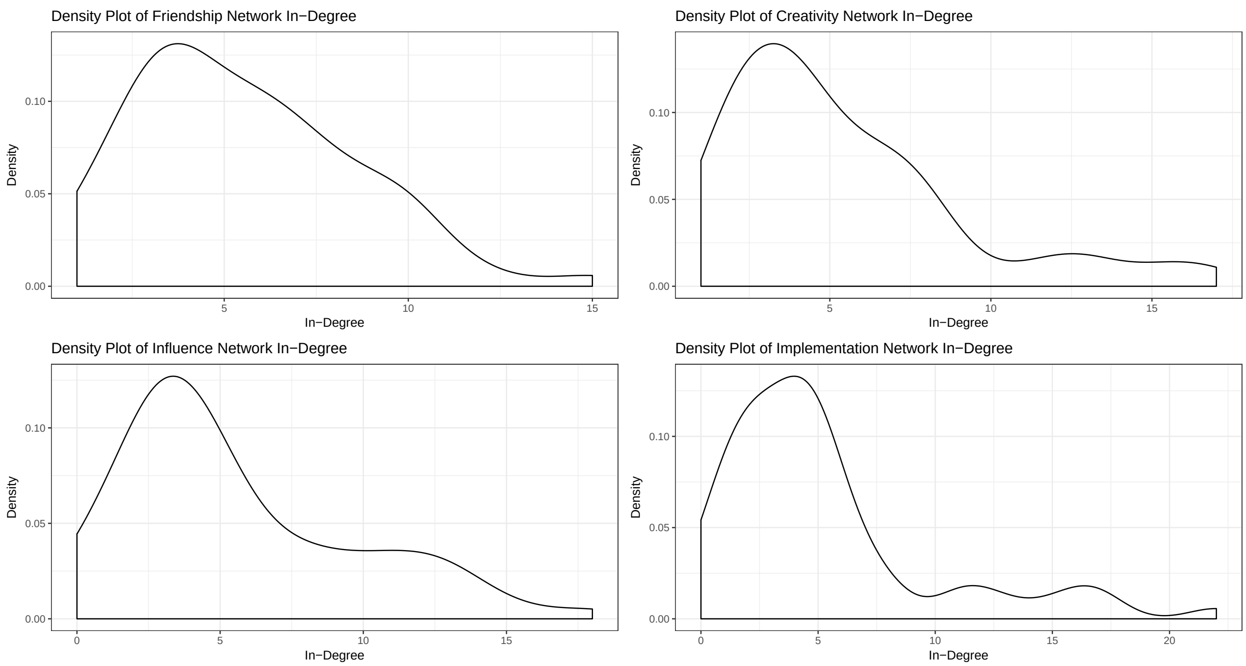
Analytics in Business Group Project

Group 1

13 Dec 2016

Question 1: Regressions

The density plots of the in-degrees for the four networks are plotted below.



The in-degree is a count data. Judging from the above density plots, we could use the discrete probability distribution (e.g. Poisson, Negative Binomial) for the regression model. An overdispersion test has been performed and there is evidence of overdispersion (i.e. mean is not equal to variance) in the data, especially for implementation network (P-Value < 0.01). Therefore, negative binomial regression will be a better fit.

The below three tables shows the regression results of the three networks:

Table 1: Regression of Creativity Network In-Degree

	Creativity Network In-Degree	
	Poisson	*negative binomial*
	(1)	(2)
friendInDegree	0.123***	0.125***
	(0.017)	(0.022)
Constant	0.938***	0.927***
	(0.125)	(0.154)
N	60	60
Log Likelihood	−144.529	−142.057
θ		10.575* (5.561)
Akaike Inf. Crit.	293.058	288.114
Notes:	***Significant at the 1 percent level.	
	**Significant at the 5 percent level.	
	*Significant at the 10 percent level.	

Table 2: Regression of Influence Network In-Degree

	Influence Network In-Degree	
	Poisson	*negative binomial*
	(1)	(2)
friendInDegree	0.141***	0.143***
	(0.016)	(0.024)
Constant	0.836***	0.824***
	(0.125)	(0.168)
N	60	60
Log Likelihood	−152.968	−148.562
θ		6.903** (3.157)
Akaike Inf. Crit.	309.936	301.124
Notes:	***Significant at the 1 percent level.	
	**Significant at the 5 percent level.	
	*Significant at the 10 percent level.	

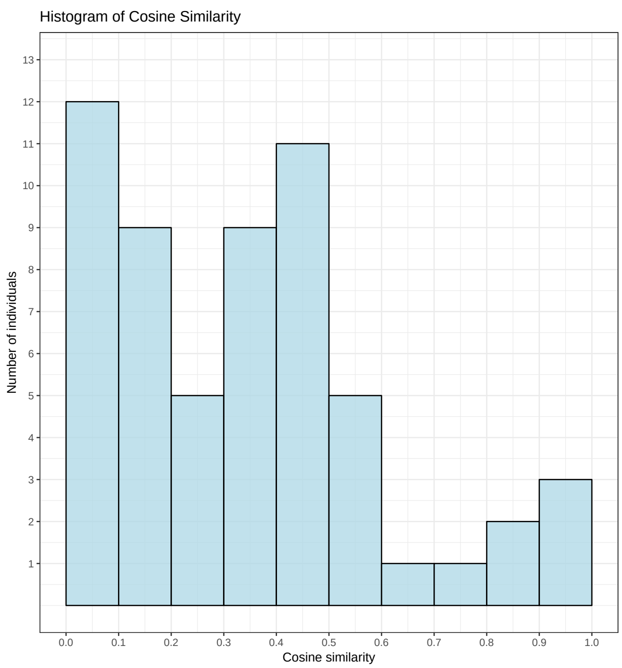
Table 3: Regression of Implementation Network In-Degree

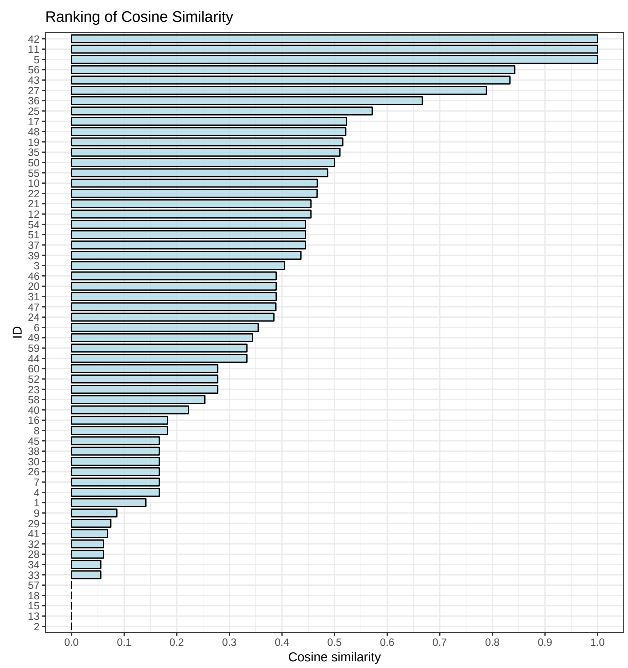
	Implementation Network In-Degree	
	Poisson	*negative binomial*
	(1)	(2)
friendInDegree	0.099***	0.094***
	(0.017)	(0.032)
Constant	1.069***	1.097***
	(0.125)	(0.213)
N	60	60
Log Likelihood	−185.067	−157.439
θ		2.589*** (0.722)
Akaike Inf. Crit.	374.134	318.879
Notes:	***Significant at the 1 percent level.	
	**Significant at the 5 percent level.	
	*Significant at the 10 percent level.	

As seen from the tables, the in-degree of friendship network is very statistically significant. Hence higher popularity is expected to increase one's chances in being picked for the other three networks.

Question 2: Cosine Similarity

Three different values were calculated for the similarities between Friendship-Creativity, Friendship-Influence, and Friendship-Implementation picks. The final score averages the three individual scores. The below graphs display the distribution of the average scores and the ranking of each individuals based on the scores (lower similarity score indicates higher flexibility).





Below table shows the average similarity score and Z-score for each individual (order by similarity score).

Table 4: Average Cosine Similarity Score Ranking

ID	Average Score	Z-score
5	1	2.5
11	1	2.5
42	1	2.5
56	0.842	1.89
43	0.833	1.85
27	0.788	1.68
36	0.667	1.21
25	0.571	0.848
17	0.523	0.661
48	0.521	0.655
19	0.516	0.633
35	0.51	0.612
50	0.5	0.573
55	0.487	0.522
10	0.467	0.446
22	0.467	0.445
12	0.455	0.4
21	0.455	0.4
37	0.444	0.359
51	0.444	0.359
54	0.444	0.359
39	0.436	0.327
3	0.405	0.207
20	0.389	0.146
46	0.389	0.146
31	0.389	0.146
47	0.388	0.144
24	0.385	0.129
6	0.355	0.0142
49	0.344	-0.0272
44	0.333	-0.068
59	0.333	-0.068
23	0.278	-0.282
52	0.278	-0.282
60	0.278	-0.282
58	0.253	-0.376
40	0.222	-0.495
8	0.183	-0.648
16	0.183	-0.648
4	0.167	-0.709
7	0.167	-0.709
26	0.167	-0.709
30	0.167	-0.709
38	0.167	-0.709
45	0.167	-0.709
1	0.141	-0.807
9	0.0861	-1.02
29	0.0745	-1.06
41	0.068	-1.09
28	0.0609	-1.12
32	0.0609	-1.12
33	0.0556	-1.14
34	0.0556	-1.14
2	0	-1.35
13	0	-1.35
15	0	-1.35
18	0	-1.35
57	0	-1.35
14	NA	NA
53	NA	NA

Question 3: Leaders

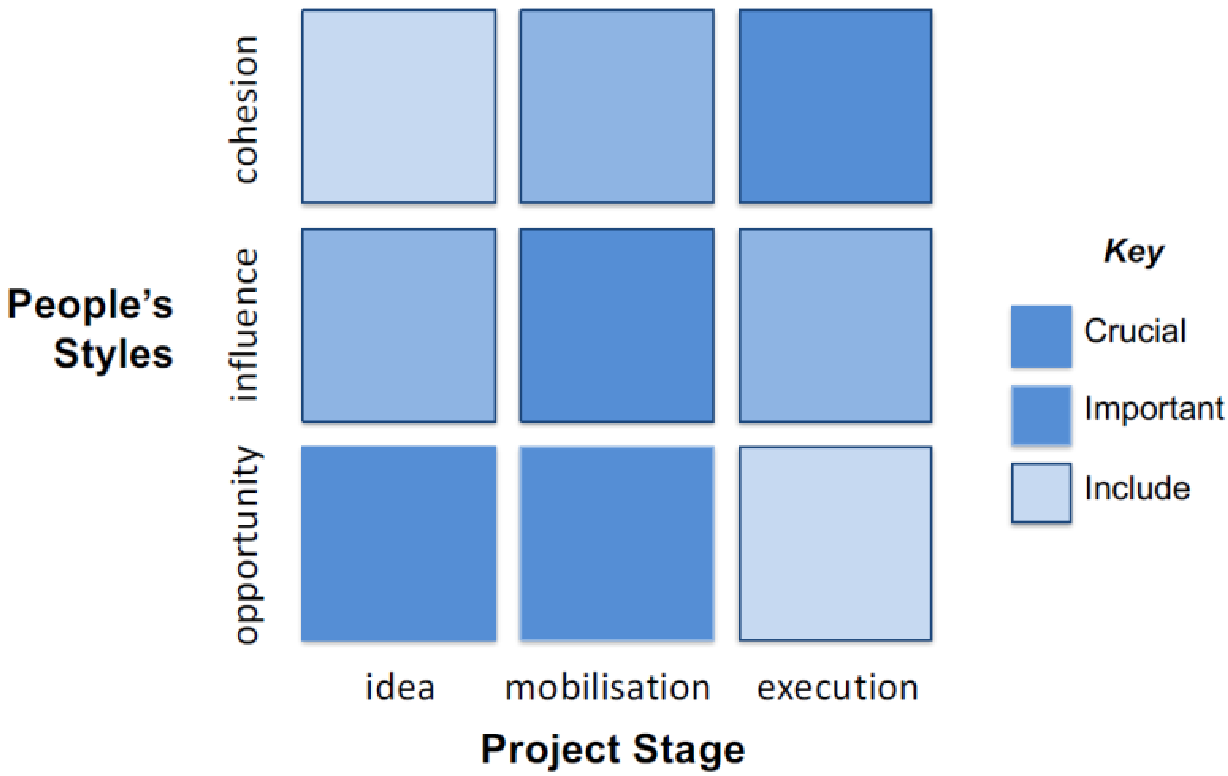


Figure 1: People-Picking for Projects

With reference to the above framework, the leader for each task must have the guestlist listed in below table:

Table 5: Guest List Criteria for Each Leder

Tasks	Guest List
Design	Expansive
Lobbying	Expansive, Exclusive
Implementation	Cohesive

To select the best leader, the in-degree centrality scores from each category are weighted according their respective importance in each task. The detailed weightings is tabulated in the below table.

Table 6: Weighting Each Score for Leader Selection

Tasks	Flexibility	Friends	Advice	Creative	Influence	Impl
Design	0.25	0.05	0.30	0.40	0	0
Lobbying	0.25	0.05	0.30	0	0.40	0
Implementation	0.2	0.1	0.35	0	0	0.35

Lobbying Leader

The following table lists the top 3 leader candidates for lobbying. We select person 33 to be the leader.

Table 7: Top 3 Lobbying Leader Candidates

ID	Flexibility	Friends	Advice	Influence	Final Rank
33	6.5	1	4.5	8	1
34	6.5	57.5	3	6	2
30	16.5	57.5	12.5	2.5	3

Design Leader

The following table lists the top 3 leader candidates for design. Since person 33 has already been selected for lobbying, we select the next best candidate - person 30.

Table 8: Top 3 Design Leader Candidates

ID	Flexibility	Friends	Advice	Creative	Final Rank
33	6.5	1	4.5	2	1
30	16.5	57.5	12.5	7	2
57	3	23.5	15	19	3

Implementation Leader

The following table lists the top 3 leader candidates for implementation and person 19 is chosen as the leader.

Table 9: Top 3 Implementation Leader Candidates

ID	Flexibility	Friends	Advice	Implementation	Final Rank
19	48	5	1	4.5	1
14	59	5	9	11.5	2
18	3	23.5	36.5	17.5	3

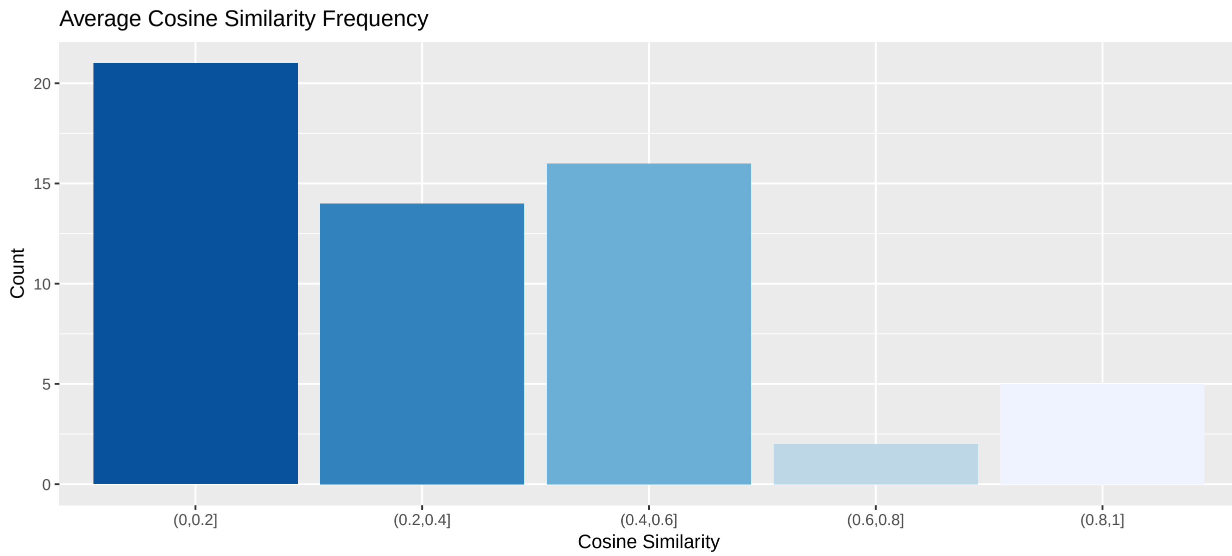
Question 4 - WIP

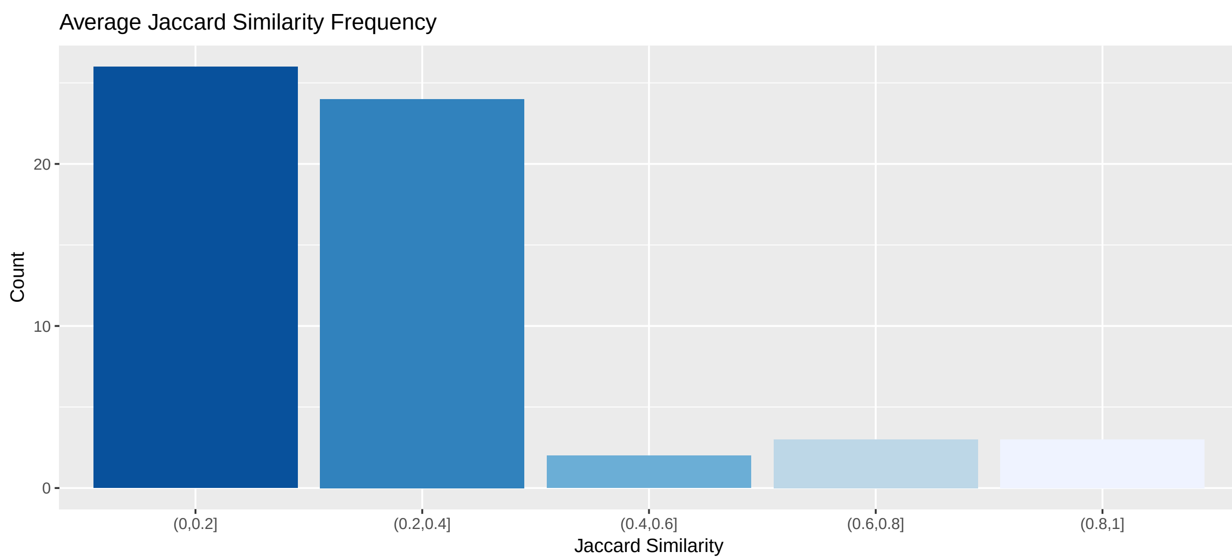
Question 5 - WIP

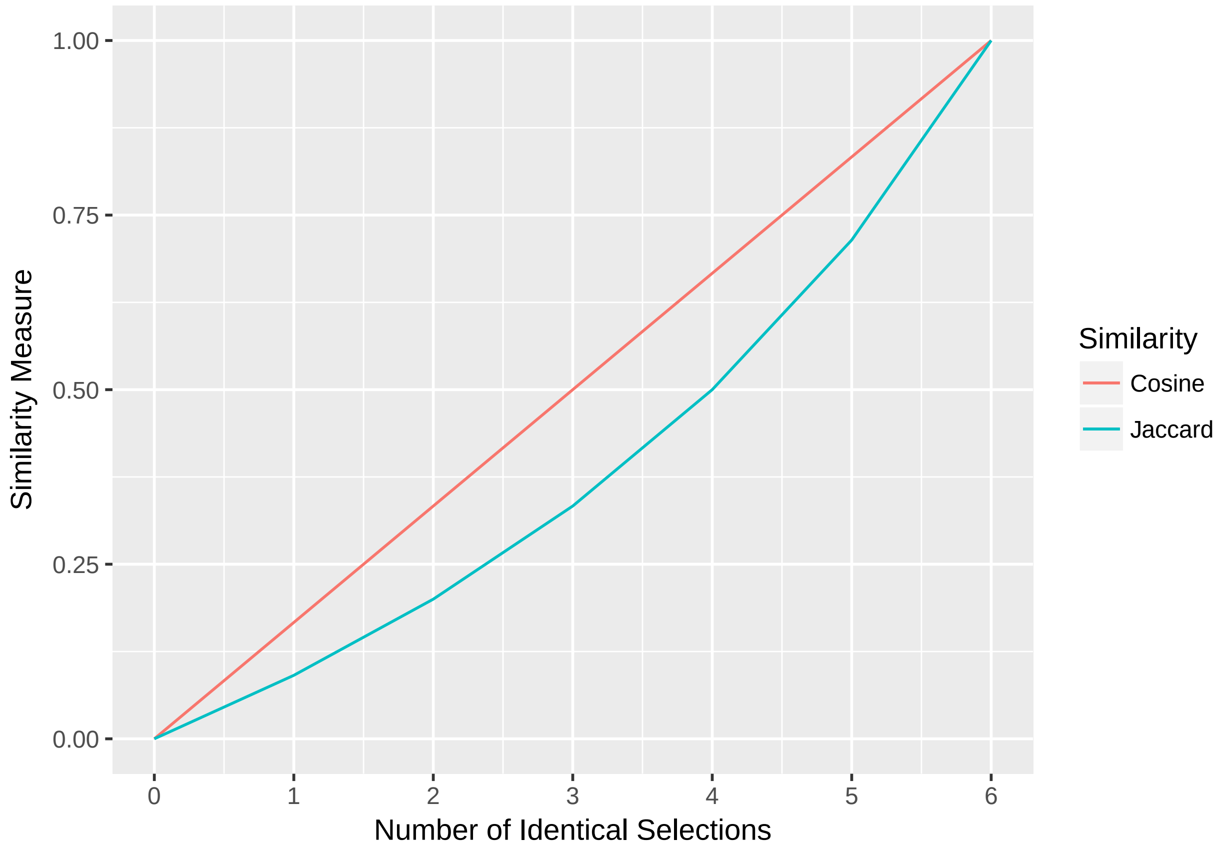
In the context of the assignment, we are comparing a two 60-element binary vectors. Each vector should have six 1s. In this case, cosine similarity grows linearly with the number of identical selection. In contrast, Jaccard similarity coefficient grows slower with small number of identical selections. As the number of identical selections gets larger, Jaccard Similarity increases at a faster rate. From the plot, we can observe that Jaccard similarity is always smaller than cosine similarity.

Paccard similarity coefficient may be better for the purpose of leader selection. A person may know the strengths and weaknesses of his/her close friends better and a small number of identical selections may indicate thoughtfulness in selecting team members for different tasks. On the contrary, a person, who selected completely different team for each tasks, might have chosen his/her members in completely random fashion (without giving deep thoughts).

Therefore, a better similarity measure should penalise small number of identical selections lesser than large number of identical selections (since the likelihood of "inflexibility" goes up as number of identical selection increases). Paccard similarity coefficient has this property.


Average Cosine Similarity Frequency
Count
20
15
10
5
0
(0,0.2]
(0.2,0.4]
(0.4,0.6]
(0.6,0.8]
(0.8,1]
Cosine Similarity


Average Jaccard Similarity Frequency
Count
20
10
0
(0,0.2]
(0.2,0.4]
(0.4,0.6]
(0.6,0.8]
(0.8,1]
Jaccard Similarity


1.00
0.75
0.50
0.25
0.00
Similarity Measure
0
1
2
3
4
5
6
Number of Identical Selections
Similarity
Cosine
Jaccard