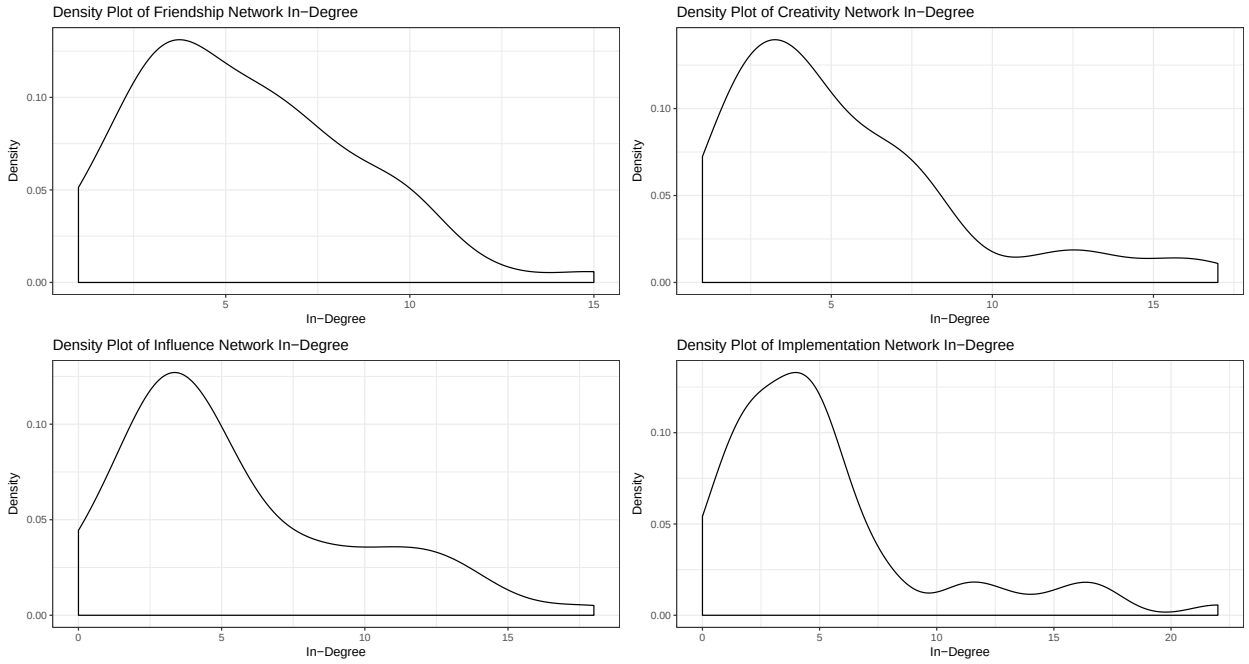
Analytics in Business Group Project

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. As seen from the tables, the in-degree of friendship network is very statistically significant. Hence higher popularity is expected to increase the votes one receives in the other 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

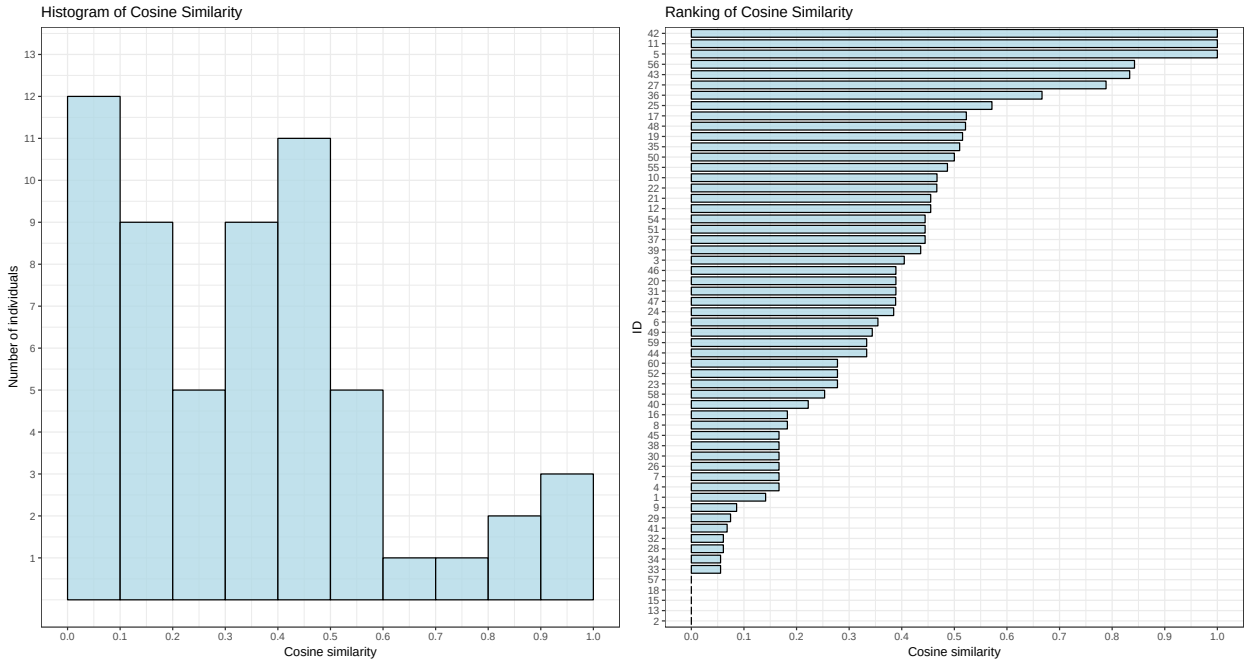
	Influence Network In-Degree	
	Poisson	*negative binomial*
	(1)	(2)
friendInDegree	0.141***	0.143***
	(0.016)	(0.024)
Constant	0.836***	0.824***
	(0.125)	(0.168)
N	60	60
Log Likelihood	−152.968	−148.562
θ		6.903** (3.157)
Akaike Inf. Crit.	309.936	301.124
Notes:	***Significant at the 1 percent level.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

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

ID	Average Score	Z-score
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

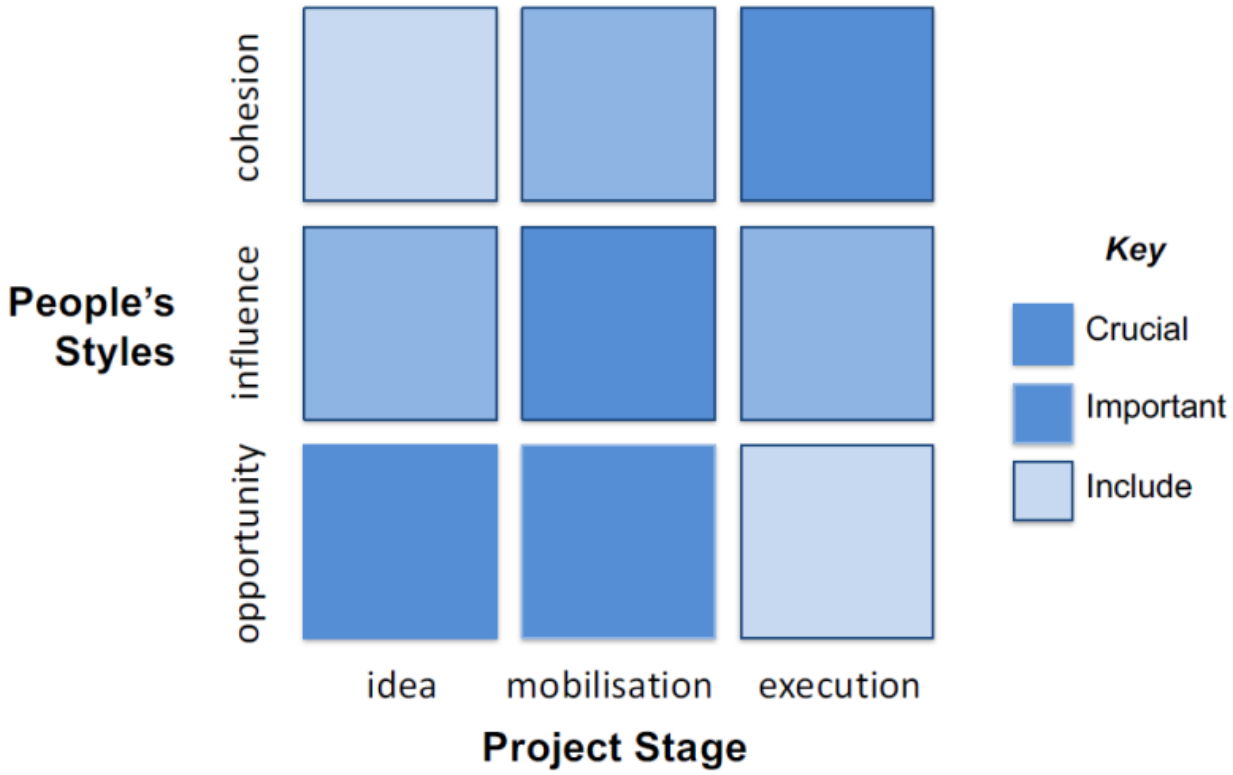


Figure 1: People-Picking for Projects

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	0.275	0.05	0.275	0.40	0	0
Lobbying	0.275	0.05	0.275	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 - ID Rankings

Team Metric

The following measures have been used to construct the final metric for each tasks:

- Party Type
- Whether a candidate has been picked by team leader for the task
- Betweenness: Indicates information flow (Design)
- In-Degree of Creativity Network: Perceived Creativity (Design)
- Eigenvector Centrality: Indicates influence (Lobbying)
- In-Degree of Influence Network: Perceived Influence (Lobbying)
- Closeness: Indicates cohesiveness (Implementation)
- In-Degree of Implementation Network: Perceived Execution Skills (Implementation)

The two tables below tabulate the weightings of the criteria used to construct the final metrics. The final metric is then divided by the in-degree of the friendship network.

Note that the friendship in-degree of node #27 is zero. We do not rank this node and set its final metrics to zero.

Table 10: Preferred Party Size for Team Selection (Weight = 0.15)

Tasks	Tight	Exclusive	Expansive
Design	L	M	H
Lobbying	L	H	M
Implementation	H	M	L

Table 11: Metric Weighting for Team Selection

Tasks	Pick	Betweenness	Creativity	Eigen	Infl	Closeness	Impl
Design	0.15	0.15	0.55	0	0	0	0
Lobbying	0.15	0	0	0.35	0.35	0	0
Implementation	0.15	0	0	0	0	0.15	0.55

Team Member Ranking

The ranking of the team members for the respective task is tabulated below:

Table 12: Ranking of Individuals for the Three Tasks

Rank	Design	Design Metric	Lobby	Lobby Metric	Impl	Impl Metric
1	30	7.4981554	30	8.4739177	30	7.5211986
2	34	2.8844813	34	7.5190966	8	5.6078939
3	15	2.1562500	8	2.0089106	47	3.7671105
4	49	2.0941912	1	1.4287794	4	2.1598093
5	8	1.7545562	52	1.4211853	1	2.0888214
6	39	1.7074574	11	1.3070673	42	1.8718043
7	20	1.6149595	23	1.2105063	15	1.7610237
8	50	1.5400456	15	1.1603835	40	1.7165202
9	45	1.4851061	24	1.0784148	45	1.6904013
10	38	1.4349729	7	1.0545756	34	1.6030204
11	7	1.3807531	21	1.0441225	60	1.5713318
12	24	1.2838664	39	1.0384537	38	1.5222756
13	3	1.1971218	3	1.0306247	24	1.4716273
14	44	1.1949867	22	1.0026025	41	1.3526582
15	57	1.1796875	50	1.0006073	6	1.2990922
16	1	1.1304951	20	0.9776281	55	1.2930821
17	42	1.1135087	49	0.9424224	36	1.2816116
18	52	1.0615167	37	0.9071873	31	1.2745821
19	21	1.0439534	45	0.8963322	10	1.2515640
20	31	1.0169815	44	0.8684038	23	1.2505167
21	26	1.0163178	19	0.8627332	22	1.2239659

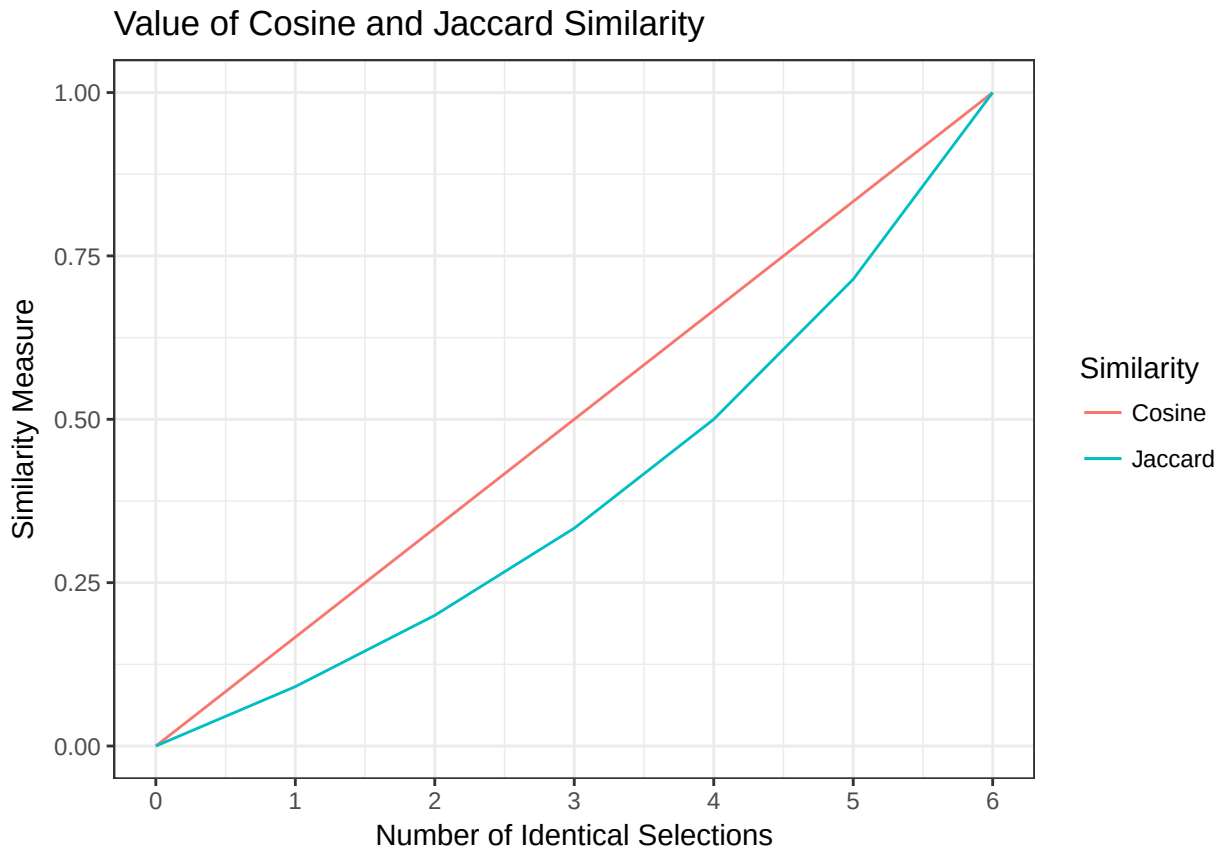
Rank	Design	Design Metric	Lobby	Lobby Metric	Impl	Impl Metric
22	23	1.0012162	38	0.8469075	43	1.2162671
23	37	0.9543046	4	0.8361707	18	1.1768923
24	28	0.9490527	48	0.8265931	9	1.1072694
25	60	0.9396134	31	0.8142924	25	1.0848964
26	35	0.9104411	35	0.8044536	32	1.0225963
27	16	0.8833533	47	0.7939722	5	1.0089106
28	11	0.7985081	29	0.7169488	44	0.9640368
29	33	0.7644160	13	0.7163250	19	0.9529879
30	41	0.7504916	57	0.7115699	29	0.9394111
31	58	0.7461716	26	0.7106130	11	0.9361700
32	13	0.7422160	10	0.6832839	20	0.9307819
33	48	0.7217635	25	0.6793457	52	0.9268387
34	19	0.7119720	6	0.6586456	7	0.9206128
35	6	0.6696062	16	0.6535770	53	0.9062500
36	2	0.6393251	2	0.6502917	48	0.8827387
37	22	0.6217910	28	0.6266263	2	0.8780284
38	46	0.6029308	33	0.6000000	26	0.8673456
39	51	0.5982648	56	0.5890120	21	0.8256256
40	56	0.5798117	46	0.5745019	49	0.8213318
41	25	0.5252600	17	0.5685558	33	0.7942022
42	29	0.5050862	58	0.5305592	54	0.7806295
43	59	0.5021911	12	0.4875543	16	0.7588160
44	9	0.5010024	59	0.4708697	13	0.7535331
45	47	0.4930304	41	0.4599573	46	0.7358493
46	43	0.4463507	51	0.4392493	57	0.7227880
47	12	0.4451592	60	0.4327415	37	0.6986382
48	18	0.4428288	9	0.4262616	14	0.6833333
49	55	0.4154968	18	0.4207138	39	0.6770249
50	54	0.4145820	5	0.3981338	50	0.6416851
51	40	0.3974844	40	0.3879881	59	0.6313911
52	10	0.3701925	14	0.3499583	51	0.5464589
53	17	0.3412430	43	0.3180805	3	0.5132229
54	14	0.3093750	36	0.3171880	35	0.4499532
55	36	0.2982215	32	0.3145427	17	0.3799267
56	53	0.2578125	55	0.2956554	12	0.3714830
57	5	0.2276325	42	0.2898498	56	0.3712981
58	4	0.1791543	54	0.2848362	28	0.3060068
59	32	0.1005138	53	0.1193182	58	0.2251311
60	27	0.0000000	27	0.0000000	27	0.0000000

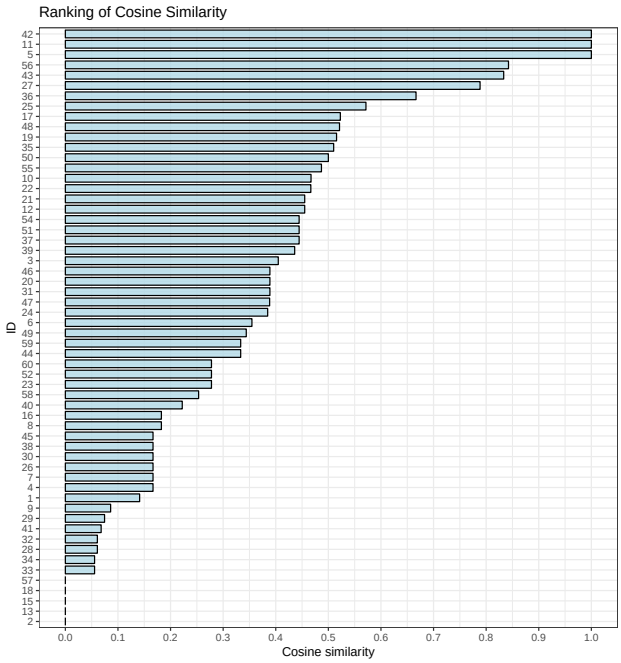
Question 5 - Jaccard Similarity Coefficient

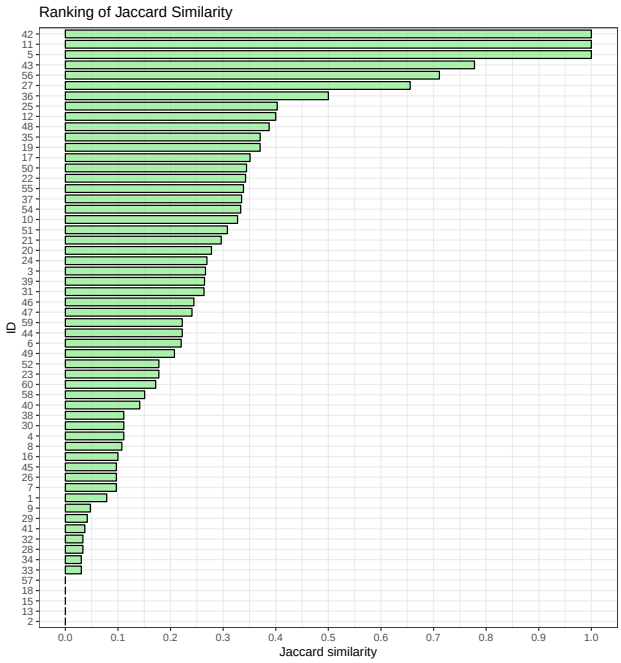
In Question 2, we compares two 60-element binary vectors. Each vector should have six 1s. From the below plot, we can see that cosine similarity grows linearly with the number of identical selection. In contrast, Jaccard similarity coefficient grows slower initially. As the number of identical selections gets larger, Jaccard Similarity increases at a faster rate.

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).

Compared to cosine similarity, Jaccard Similarity Coefficient punishes 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). Hence Jaccard Similarity may be better suited for computing the average similarity score.




Ranking of Cosine Similarity
ID
Cosine similarity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Ranking of Jaccard Similarity
ID
Jaccard similarity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0