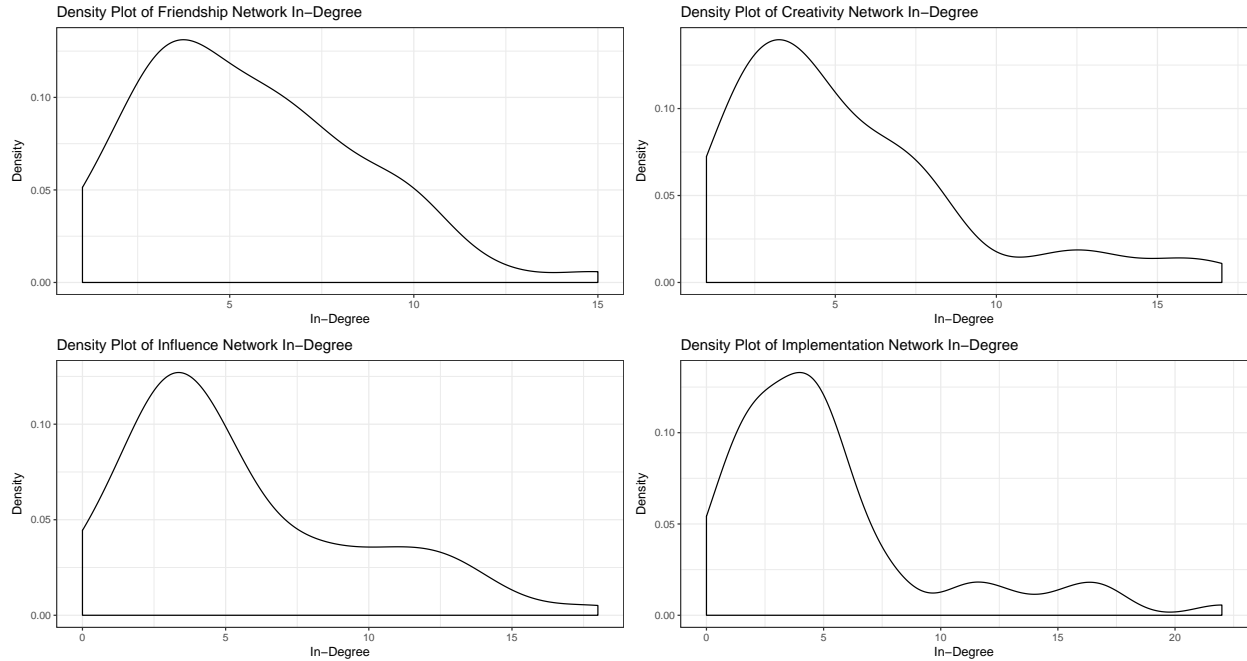# Analytics in Business Group Project

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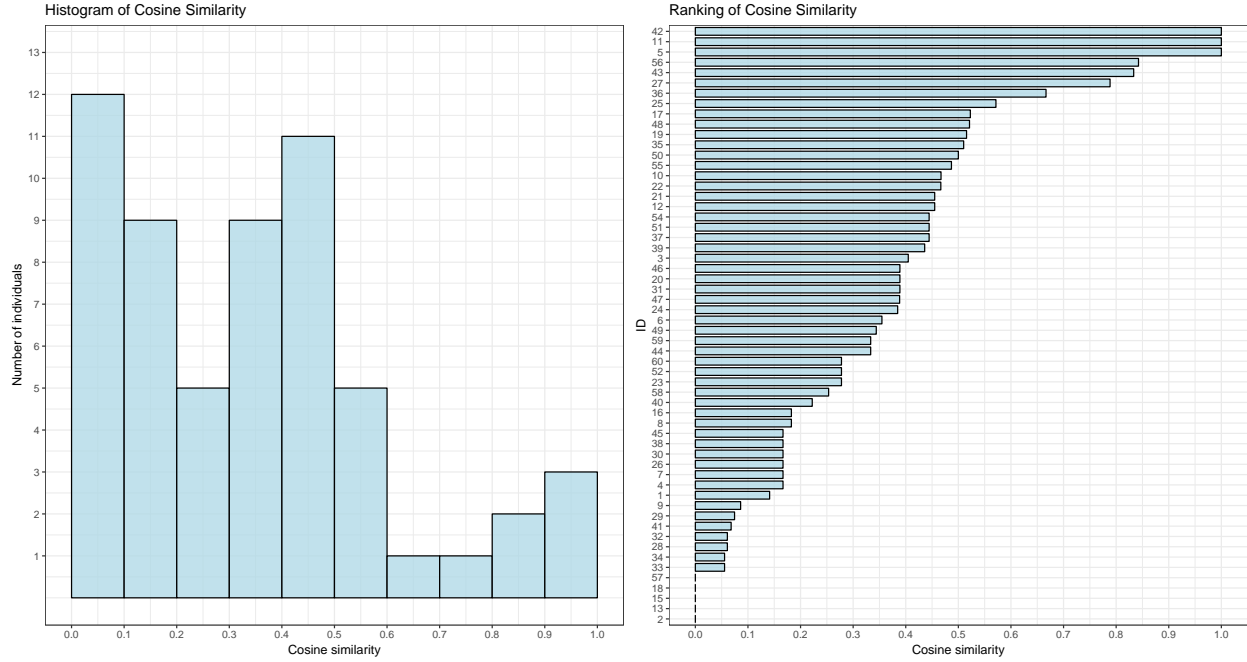
Below table shows the average similarity score and Z-score for each individual (order by similarity score).

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
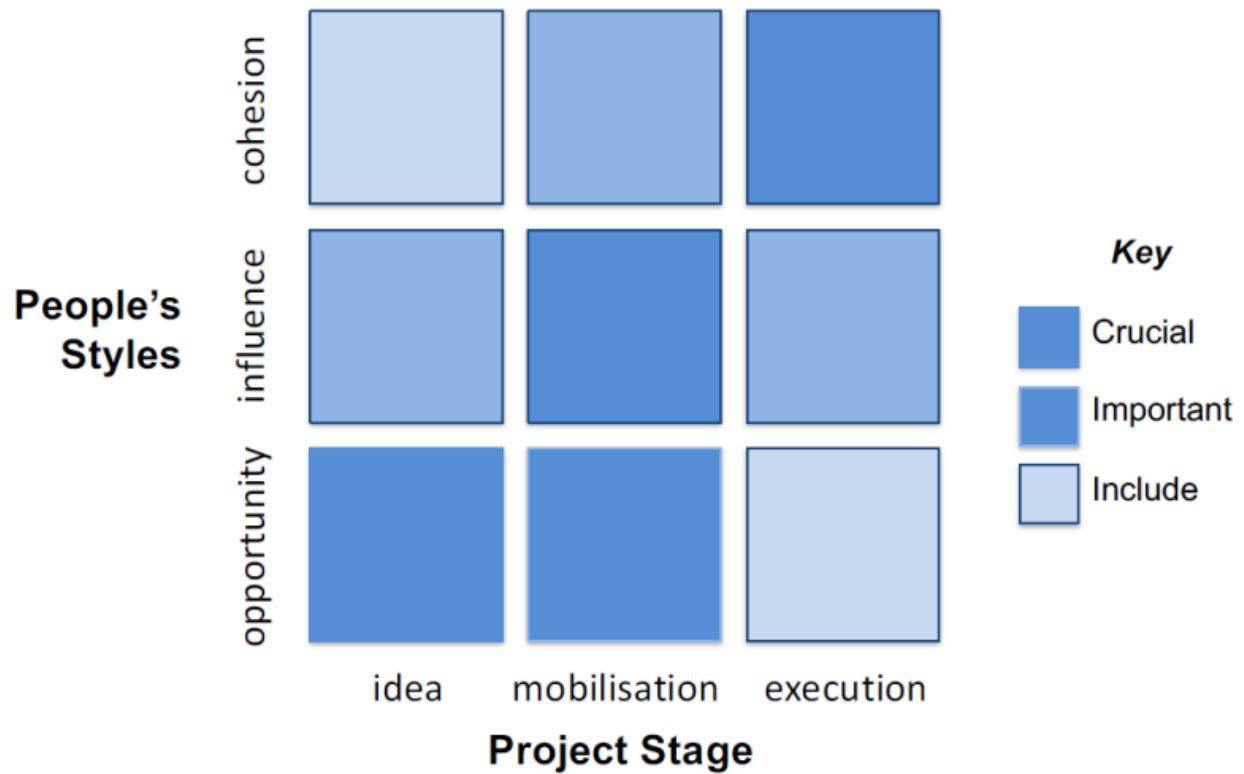
Figure 1: People-Picking for Projects

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	0.25	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 - ID Rankings

Firstly, let's look at the ranking of in-degree centrality of the three networks. They are divided by the in-degree of the friendship network to obtain the cost-benefit ratio. The below table ranks the nodes for each of the three networks (friendship in-degree of node #27 is zero, thus it is artifically set to 0.5 to avoid "zero-cost" circumstances) using this cost-benefit ratio.

Table 10: Ranking of Nodes in terms of In-Degree Centrality

	creativity	influence	implementation
1st	30	30	30
2nd	8	34	8
3rd	15	8	27
4	42	52	15
5	49	21	1
6	7	20	23
7	27	7	31
8	34	1	7
9	20	15	20
10	50	23	52
11	21	27	6
12	24	45	21
13	38	38	2
14	45	22	22
15	60	24	25
16	19	41	42
17	39	19	19
18	41	2	46
19	6	6	11
20	28	9	24
21	33	10	26
22	1	11	33
23	25	39	18
24	35	25	48
25	40	50	14
26	47	18	16
27	48	57	40
28	57	5	41
29	51	14	43
30	59	16	44
31	13	26	45
32	16	37	50
33	26	33	57
34	43	28	29
35	10	40	9
36	22	49	10
37	37	56	37
38	53	43	39
39	55	29	53
40	14	4	55
41	46	13	13
42	3	17	38
43	18	51	49
44	23	53	60
45	36	54	5
46	44	55	32
47	54	58	17
48	58	48	4
49	56	32	56
50	9	35	12
51	52	31	35
52	29	42	51
53	4	44	59
54	11	60	54
55	2	12	28
56	5	59	3
57	31	46	34
58	12	3	36

	creativity	influence	implementation
59	17	36	47
60	32	47	58

Team Metric

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

- Party Type
- Whether a candidate has been voted by team leader for the tasks
- Betweenness: High leverage (Design)
- In-Degree of Creativity Network (Design)
- Eigen Centrality: High influence (Lobbying)
- In-Degree of Influence Network (Lobbying)
- Closeness: High cohesion (Implementation)
- In-Degree of Implementation Network (Implementation)

The two tables below tabulate the weightings of the criterion used to construct the final metrics.

Table 11: Preferred Party Size for Team Selection (Weight = 0.15)

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

Table 12: 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

Design Team Ranking

The ranking of the design team based on the new creativity metric is as below:

ID	Creativity Metric	Creativity In-Degree	Friendship In-Degree	Betweenness
27	Inf	1	0	0.00000
30	14.4964337	11	1	13.47460
34	5.5766639	2	1	17.42599
8	3.3921420	3	1	106.04675
15	2.7791667	5	2	0.00000
49	2.4292618	7	3	137.64460
39	1.8863339	5	4	39.19423
45	1.7227231	4	3	96.53696
20	1.6812143	12	7	166.09037
38	1.6645686	4	3	74.48627
50	1.5879581	13	8	239.40071
7	1.4237099	17	8	80.05487
24	1.4183667	6	4	45.06039
3	1.3886613	2	3	121.06502
44	1.3861846	2	3	120.12593
1	1.3113743	3	3	16.16201
42	1.2916701	7	3	36.18082
57	1.2440341	6	6	329.88131
31	1.1796985	1	3	117.42959
23	1.1614107	2	3	34.89756
52	1.1050661	4	7	202.87577
26	1.0916006	4	5	188.35695

ID	Creativity Metric	Creativity In-Degree	Friendship In-Degree	Betweenness
60	1.0899515	4	3	186.48780
21	1.0622683	15	10	142.33039
37	1.0542794	3	4	78.58043
28	0.9879882	8	7	114.94440
35	0.9778812	5	5	35.14418
47	0.9531921	1	1	72.28512
16	0.9487869	4	5	90.88476
11	0.8821614	2	4	62.81051
41	0.8291145	5	4	137.73882
58	0.7868719	4	6	99.72016
13	0.7827005	5	6	185.58332
33	0.7644160	16	15	217.26193
48	0.7513743	7	7	122.21650
19	0.7244627	13	10	136.67637
6	0.7061301	7	6	135.45469
2	0.6655539	3	7	175.06082
22	0.6411356	6	8	186.37606
46	0.6135085	7	10	100.51236
51	0.6087607	9	10	272.35707
56	0.5978503	5	8	47.79516
25	0.5416014	8	8	86.89877
9	0.5381137	3	5	216.07405
59	0.5178148	7	8	135.43881
29	0.5169706	5	9	199.14212
43	0.4794138	4	5	100.41275
18	0.4669831	4	6	162.75583
12	0.4634222	2	7	51.38665
40	0.4610819	3	3	23.63464
55	0.4590250	3	4	92.47484
54	0.4371956	4	6	137.90757
10	0.4089746	3	4	65.90589
36	0.3459370	2	3	55.57248
17	0.3442103	3	12	119.29474
14	0.3148026	7	10	0.00000
53	0.2848214	3	4	0.00000
5	0.2444942	2	5	91.27267
4	0.2309100	1	2	52.53303
32	0.1079593	1	5	73.68362

Lobbying Team Ranking

The ranking of the lobby team based on the new influence metric is as below:

ID	Influence Metric	Influence In-Degree	Friendship In-Degree	Eigen Centrality
30	18.5579075	17	1	0.2824265
34	16.7119202	13	1	0.2823795
8	3.8838938	4	1	0.2845182
15	2.2206054	4	2	0.1458528
1	2.0923842	6	3	0.3996524
23	1.8391873	6	3	0.2879860
11	1.7547124	4	4	0.2579766
52	1.6467980	16	7	0.5745472
3	1.6305246	0	3	0.4401284
47	1.5350129	0	1	0.2401134
49	1.5282100	2	3	0.3136156
24	1.5021059	5	4	0.4450061
39	1.4579584	4	4	0.4584425
44	1.4423484	1	3	0.3164431
31	1.3795791	1	3	0.2887602
37	1.3129402	3	4	0.4095977
7	1.2323846	17	8	0.6597561
20	1.1850436	15	7	0.4694557
21	1.1769142	22	10	0.7981309
50	1.1767373	7	8	0.6093845

| ID	Influence Metric	Influence In-Degree	Friendship In-Degree	Eigen Centrality
35	1.1057094	2	5	0.5173991
4	1.0777311	1	2	0.3492070
45	1.0397453	5	3	0.3597986
22	1.0337946	10	8	0.6818688
26	1.0049177	4	5	0.3559966
38	0.9824128	4	3	0.3752081
57	0.9481101	5	6	0.4374619
16	0.9436568	4	5	0.3073649
19	0.8778689	12	10	0.7039080
48	0.8605046	3	7	0.5856534
2	0.8442780	8	7	0.3635735
28	0.8196417	5	7	0.4574079
58	0.7572261	3	6	0.3336447
13	0.7553972	3	6	0.3318702
10	0.7548661	4	4	0.4080203
56	0.7523368	5	8	0.5129516
29	0.7338182	5	9	0.6179031
25	0.7004809	7	8	0.7466455
46	0.6990546	1	10	0.8518746
6	0.6945717	6	6	0.5344624
33	0.6750000	11	15	1.0000000
12	0.6748642	2	7	0.4134802
17	0.6680650	6	12	0.8321731
41	0.5081433	5	4	0.2149901
60	0.5019801	1	3	0.2854089
59	0.4855190	2	8	0.6657073
9	0.4578365	5	5	0.2646960
40	0.4500662	2	3	0.2218188
51	0.4469554	5	10	0.6502959
18	0.4436618	5	6	0.3317099
5	0.4276252	4	5	0.2814075
36	0.3679381	0	3	0.2669875
14	0.3560979	8	10	0.3759439
43	0.3416421	3	5	0.2538447
32	0.3378422	2	5	0.2915228
42	0.3362258	1	3	0.2123069
55	0.3266288	2	4	0.2250004
54	0.3003728	3	6	0.2740721
53	0.1318182	2	4	0.1047173
27	NaN	1	0	0.1543776

Implementation Team Ranking

The ranking of the implementation team based on the new implementation metric is as below:

ID	Implementation Metric	Implementation In-Degree	Friendship In-Degree	Closeness Centrality
30	14.5409839	17	1	0.0047170
8	10.8419282	4	1	0.0046083
47	7.2830803	0	1	0.0045872
34	3.0991728	13	1	0.0048309
4	2.7837543	1	2	0.0048544
1	2.4230329	6	3	0.0038760
15	2.2697639	4	2	0.0037736
42	2.1712930	1	3	0.0046296
40	1.9911634	2	3	0.0046083
45	1.9608655	5	3	0.0043860
60	1.8227448	1	3	0.0046729
38	1.7658397	4	3	0.0042553
24	1.6257978	5	4	0.0040984
41	1.4943652	5	4	0.0040486
36	1.4866695	0	3	0.0048077
31	1.4785152	1	3	0.0046296
23	1.4505993	6	3	0.0044248
55	1.4285479	2	4	0.0046729

ID	Implementation Metric	Implementation In-Degree	Friendship In-Degree	Closeness Centrality
10	1.3826802	4	4	0.0042017
6	1.3699517	6	6	0.0043103
43	1.3063609	3	5	0.0046512
22	1.2620448	10	8	0.0041841
18	1.2410864	5	6	0.0048309
9	1.1892893	5	5	0.0044053
25	1.1186487	7	8	0.0042194
44	1.1182827	1	3	0.0045872
32	1.0983441	2	5	0.0045045
5	1.0836447	4	5	0.0043103
11	1.0342450	4	4	0.0044053
53	1.0011905	2	4	0.0002825
19	0.9697070	12	10	0.0040323
20	0.9689679	15	7	0.0044643
52	0.9648628	16	7	0.0043860
29	0.9615149	5	9	0.0042918
49	0.9527448	2	3	0.0046729
7	0.9492541	17	8	0.0042735
26	0.9315934	4	5	0.0047847
48	0.9189536	3	7	0.0042194
2	0.9140500	8	7	0.0047170
21	0.8401103	22	10	0.0042017
54	0.8232093	3	6	0.0045872
16	0.8150246	4	5	0.0045455
13	0.7946349	3	6	0.0047170
33	0.7942022	11	15	0.0042918
37	0.7718289	3	4	0.0043103
57	0.7622128	5	6	0.0047847
46	0.7487590	1	10	0.0042553
39	0.7479513	4	4	0.0040650
14	0.6953216	8	10	0.0002825
50	0.6616486	7	8	0.0044444
59	0.6510344	2	8	0.0043290
3	0.5953385	0	3	0.0046512
51	0.5560459	5	10	0.0042017
35	0.4832831	2	5	0.0040650
12	0.3867233	2	7	0.0037594
17	0.3832304	6	12	0.0041152
56	0.3828496	5	8	0.0035088
28	0.3185609	5	7	0.0037594
58	0.2374110	3	6	0.0041152
27	NaN	1	0	0.0066667

Communities

Based on the Friendship network, four communities have been identified. To ensure sufficient diversity in each team, Design and Lobby teams must not have more than 2 members from the same community, while for Implementation team the limit is 3 members from the same community.
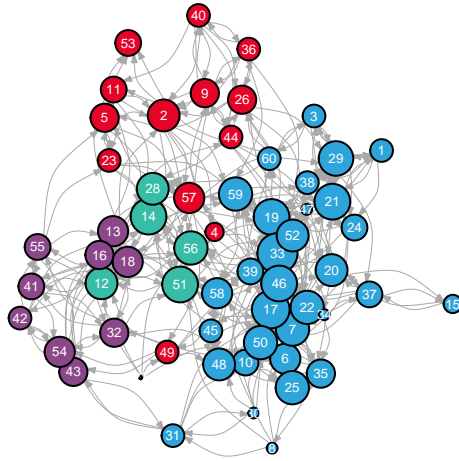
Four Communities in Friendship Network

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. Jaccard similarity is always smaller than cosine similarity.

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.

Value of Cosine and Jaccard Similarity

Ranking of Cosine Similarity

Ranking of Jaccard Similarity