Group Assignment 2

Team 7
Network Analytics
Imperial Business School

March 11, 2020

1 Exercie 1

1.1 Leaders and Opinion-makers

We have been provided 2 datasets. The first transcribes sent emails and the second received ones. It appears those 2 datasets are not actually symmetric, so we keep the maximum of the 2 values to have a merged matrix of our network.

It is represented as a directed graph of 81 nodes, one for each student, and 2570 weighted directed edges which stand for the mails addressed. The weight of the edge is given by the number of mails sent from one to student to the other.

This graph of this email interaction is sparse since $2570 \ll 19\,440$, the maximum possible number of edges. A visualization of this graph shows that the cohort is mainly divided into 2 clusters with only a few students that link these two clusters.

The following matrix displays all the students who are outstanding. For each of them, it shows $\frac{c-avg}{std}$ where c is a centrality measure.

Node	Clustering	Indegree	Outdegree	Betweenness	Eigenvector
15	4.9669	-2.6047	-2.713	-1.3136	-2.455
64	1.8776	-1.7842	-1.6227	-1.2592	-1.546
42	0.1221	-1.9893	-1.7439	-1.1284	-2.0145
5	-0.5524	-1.9893	0.679	-1.0314	-1.916
74	-0.0724	-1.5791	-0.7747	-1.0618	-1.4765
62	1.7139	-0.8611	0.0733	-0.9925	-0.3606
58	2.3823	-0.5534	-0.0479	-1.1424	-0.1082
55	-1.691	-0.1431	0.4367	0.7851	-0.4128
43	-1.6596	0.2672	-0.2901	0.8587	-0.0068
3	-1.622	0.8826	0.8001	1.6018	0.7461
13	-0.9737	-0.7585	1.6481	0.5086	-0.9748
52	-1.7425	0.2672	1.6481	1.4803	0.0387
41	-0.8412	1.6006	-1.2593	-0.0613	1.5255
72	-1.0332	1.6006	0.3156	1.313	0.8062
46	-0.3315	1.7031	0.9213	1.0079	1.7551
61	-0.7174	1.7031	1.6481	2.3185	1.8895
21	-0.7651	2.4211	1.527	3.0648	2.4537
31	-1.713	1.9083	1.7693	3.6551	1.375

Since the objective for our client is to identify who are the leaders and opinion-makers, we will focus more on the green colored values in the above table.

Student 41, 72 and 46 seem to be students who received a lot of mails, but their outdegree measure is relatively low - hinting that these 3 students are receiving a lot of "information" but not sending out "information." Furthermore, their eigenvector centrality measure are low, hinting that these 3 students might be part of the influential friend circle, but they are not really voicing their opinions.

Low clustering centrality (55, 43, 3, 52, 31) could mean here that these students are the link between the 2 isolated clusters of the cohort, especially for student 3 and student 52 who have a high betweenness centrality score. Therefore, because of their high betweenness centrality score, student 3 and student 52 could possibly be event makers, people who spread information. Student 31 is "special", since this student also has high indegree and eigenvector centralities, meaning he is quite popular/sociable.

Lastly, 61 and 21 have high centralities for indegree, outdegree, betweenness and eigenvector. Consequently, they are opinion-leaders.

2 Exercice 3

2.1 Linear Program: Primal form

Decision Variables:
f: total flow from 1 to 5
x_{12}: flow from node 1 to node 2
x_{13}: flow from node 1 to node 3
x_{14}: flow from node 1 to node 4
x_{24}: flow from node 2 to node 4
x_{25}: flow from node 2 to node 5
x_{35}: flow from node 3 to node 5
x_{43}: flow from node 4 to node 3
x_{45}: flow from node 4 to node 5

Objective Function:
maximize flow: f

Constraints:
Node 1: -f $= -x_{12} - x_{14} - x_{13}$
Node 2: $x_{12} = x_{24} + x_{25}$
Node 3: $x_{13} + x_{43} = x_{35}$
Node 4: $x_{14} + x_{24} = x_{43} + x_{45}$
Node 5: $x_{25} + x_{45} + x_{35} =$ f
capacity 1: $x_{12} \leq 5$
capacity 2: $x_{13} \leq 2$
capacity 3: $x_{14} \leq 1$
capacity 4: $x_{24} \leq 3$
capacity 5: $x_{25} \leq 1$
capacity 6: $x_{35} \leq 7$
capacity 7: $x_{43} \leq 1$
capacity 8: $x_{45} \leq 4$

x_{12},x_{13} ....., $x_{45} \in N$

2.2 Linear Program: Dual form

To define the dual problem, we could use the matrix method.

Dual Decision Variables:
x_1: dual variable issued from node 1
x_2: dual variable issued from node 2
x_3: dual variable issued from node 3
x_4: dual variable issued from node 4
x_5: dual variable issued from node 5
x_{12} : dual variable issued from capacity 1
x_{13} : dual variable issued from capacity 2
x_{14} : dual variable issued from capacity 3
x_{24} : dual variable issued from capacity 4
x_{25} : dual variable issued from capacity 5
x_{35} : dual variable issued from capacity 6
x_{43} : dual variable issued from capacity 7
x_{45} : dual variable issued from capacity 8

Objective Function:
minimize capacity : $5x_{12} + 2x_{13} + 1x_{14} + 3x_{24} + 1x_{25} + 7x_{35} + 1x_{43} + 4x_{45}$

Constraints:
Constraint 1: $x_1 - x_5 >= 1$
Constraint 2: $-x_1 + x_2 + x_{12} >= 0$
Constraint 3: $-x_1 + x_3 + x_{13} >= 0$
Constraint 4: $-x_1 + x_4 + x_{14} >= 0$
Constraint 5: $-x_2 + x_4 + x_{24} >= 0$
Constraint 6: $-x_2 + x_5 + x_{25} >= 0$
Constraint 7: $-x_4 + x_3 + x_{43} >= 0$
Constraint 8: $-x_4 + x_5 + x_{45} >= 0$

x_{12},x_{13} ....., $x_{45} \in N$
x_1,x_2 ....., $x_5 \in Z$

2.3 Max-flow/Min-cut theorem

The Max-flow/Min-cut theorem states that the maximum amount of flow passing from the source to the sink in a network is less-than/equal to the total weight of the edges of the minimum cut.

Indeed, for each edge the flow is below or equal to its capacity. Consequently, for each cut, the flow is always below the sum of the capacities of the edges.

This intuitively makes sense because the minimum cut set will contain the edges with the least amount of capacities, so the amount of flow that can be passed through the network is always limited by these smallest edges. So even if the network has edges with higher capacities, these capacities will not be used to their fullest.

2.4 Minimum Capacity Cut

Into the dual problem, each dual variable, issued from edge constraints, would be a dummy variable to tell whether this edges will be into the cut or not and each variable issued from node constraints would be a dummy variable to tell whether the node is in S or not.

Here, we have the minimum cut equal to 7 with nodes 1 and 2 in S and edges (1,3), (1, 4), (2, 4), (2, 5) in the cut.