Computer Assignment # 1

Question 1:

• Identify the top 20 most trusted individuals in the network.

|  |  |
| --- | --- |
| Rank | User ID |
| 1 | 18 |
| 2 | 143 |
| 3 | 737 |
| 4 | 790 |
| 5 | 136 |
| 6 | 1179 |
| 7 | 1719 |
| 8 | 118 |
| 9 | 4416 |
| 10 | 780 |
| 11 | 27 |
| 12 | 128 |
| 13 | 34 |
| 14 | 1516 |
| 15 | 40 |
| 16 | 1 |
| 17 | 791 |
| 18 | 28 |
| 19 | 1619 |
| 20 | 1621 |

Question 2:

• How many (and which) members should the company contact in order to achieve its “50% coverage” goal? Assume that any individual who is provided this incentive will definitely review the product.

• Formulate the problem as an optimization model.

Let be a binary variable of if a person is selected. =1 if person i is selected; =0 otherwise.

be a binary variable of if person p is informed. =1 if person i is informed; =0 otherwise.

N be the total number of people in the network.

Mi be the number of people trusted by person i.

Subject to

• Provide a pseudo-code and explanation for the method you used (it should be a reasonable method, but need not necessarily guarantee finding the “optimal” solution)

Question 3:

• How many days does it take to inform 40% of the members?

Two days.

• How long (days) does it take for the process to terminate?

Eight days.

• At the end of the process (termination), what fraction of members are “informed”?

74.99% of members are “informed” after eight days.

• Discuss how we can use a Maximum Flow model to address the above question.

We will use the most trusted individual as the source node and create a dummy sink node. There are arcs from people are trusted to people that trust them with zero cost and unlimited capacity. Except for the source node, every node has an arc pointing to the dummy sink node with cost c= -1 and capacity u=1. There is another arc starting from the dummy sink node to the source node with zero cost and unlimited capacity. Then the minimum cost flow solution maximizes the flow from source to sink.

Question 4:

• What are the top 5 nodes with the highest local clustering coefficients?

• What is the (average) clustering coefficient of the whole network?

Question 5

Other questions and additional data needed

1. Data on the cost to motivate different persons. Motivate a celebrity is usually more expensive than motivate a normal person. With the cost to motivate different persons, we can answer the question of what is the least amount of money to achieve 50% of coverage instead of the least number of people.
2. Data of level of the trust. There are different levels of trust, strongly trust or somewhat trust. If the level of trust can be collected, it can help to further define the level of coverage.
3. Data on the time to pass the trust information. The time to obtain information from people you trust can be different depending on how often you check their status etc. Knowing this data can further answer the coverage in a certain time frame.

Some other applications:

1. It can be used in the online advertisement. Find the targeted customers and focus on them.
2. It can be used in the warehouse/hospital/cell tower location problem. Suppose we have the data of the coverage of each possible location, we can use the similar methods to determine the location.
3. It can be used in the political campaign when the candidates want to find the best person to endorse for them.