CSCI 4190 Introduction to Social Networks
Project Report
Task set 3: Simulate cascading behaviors in networks

Name: Chow Man Chun StudentID: 1155108980

1. Abstract

This social network analysis (SNA) project involves a simulation of cascading behaviors in networks by using the model from Slashdot's dataset. The experiment will use this data to simulate a network how people change or maintain their decision when they are affected by the information from neighbors. Stanford Network Analysis Platform (SNAP) and several programming tools are used to simulate the network.

2. Objective

In this project, the main objective is to observe how many payoff of decision make a complete cascade. If the payoff of each node is different, how the result change. Also, we will observe how many adopting key nodes to cause a complete cascade. If adopting key nodes are concentrated or discrete, how the result change. At the last, we want to find out the major clusters of the model.

3. Methodology

3.1 Data Source

The dataset comes from Slashdot. Slashdot is a technology-related news website. The user can submit news to them. The editor will wdit them and post it on the homepage as a blog form. Slashdot allows user to tag other users as friends or foes. The dataset is available in the SNAP Datasets of Stanford Large Network Dataset Collection (http://snap.stanford.edu/data) free of charge. The file contains Slashdot Zoo social network from February,2009.

3.2 Tools

We used SNAP for Python which is developed by the Standford University to do the social network analysis. We run and compile the simulation code on the MacOS Catalina. In addition, we also deployed some third party packages to facilitate the analysis and make up for the lack of SNAP.

- Matplotlib (https://matplotlib.org/)
  - A plotting library for the Python and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications.

3.3 Experiment Design

Since the cascading behaviors in networks need an undirected graph. We change the graph from directed to undirected.

In this project, we assume that each node represents a human' choice, and each edge in the network represents a contact between two nodes. In each contact with people who have different decision, there would be "payoff a" and "payoff b". They mean that the benefit of people will get when they choose decision A or decision B. In order to simplify the progress of simulating cascading behavior in networks, we directly use threshold to replace payoffs. We define 1 wave of cascading behavior as 1 time step.

In task1, we would like to choose 35000 nodes as initial adopting key nodes randomly. We run the program 8 times to let initial adopting key nodes spread out their decision, and then we will change the threshold to see what is the difference of changing threshold and payoff.

In task2, we would like to choose 0.75 as the threshold. We run the program 10 times to let initial adopting key nodes spread out their decision, and then we will change the number of initial adopting key nodes to see the difference.

In task3, we would like to find the clustering coefficient of each degree and the average clustering coefficient to seek which group will be the obstacle of causing a complete cascade.

In this study, we would mainly focus on the following factors that are suspected to have impact on the cascading behavior, which are:

1. the threshold
2. the number of initial adopting key nodes
3. the position of initial adopting key nodes
4. the clustering coefficient

When we measure the effect of these factors, we will maintain the other factor as a constant in the model. For example, when we are doing the experiment of task1, we will not change the number of initial adopting key nodes.

4 Data Statistics

Table 1 shows the network statistics of Slashdot dataset generated by the function "PrintInfo" of SNAP.

Type:	Directed	Self Edges:	78303
Nodes:	82168	BiDir Edges:	1086763
Edges:	582533	Closed triangles:	1102922
Zero Deg Nodes:	0	Open triangles:	72659933
Zero InDeg Nodes:	0	Frac. of closed triads:	0.014952
Zero OutDeg Nodes:	0	Connected component size:	1
NonZero In-Out Deg Nodes:	82168	Strong conn. comp. size:	1
Unique directed edges:	1086763	Approx. full diameter:	11
Unique undirected edges:	582533	90% effective diameter:	4.77592

Table 1. Network statistics generated by "PrintInfo" in SNAP

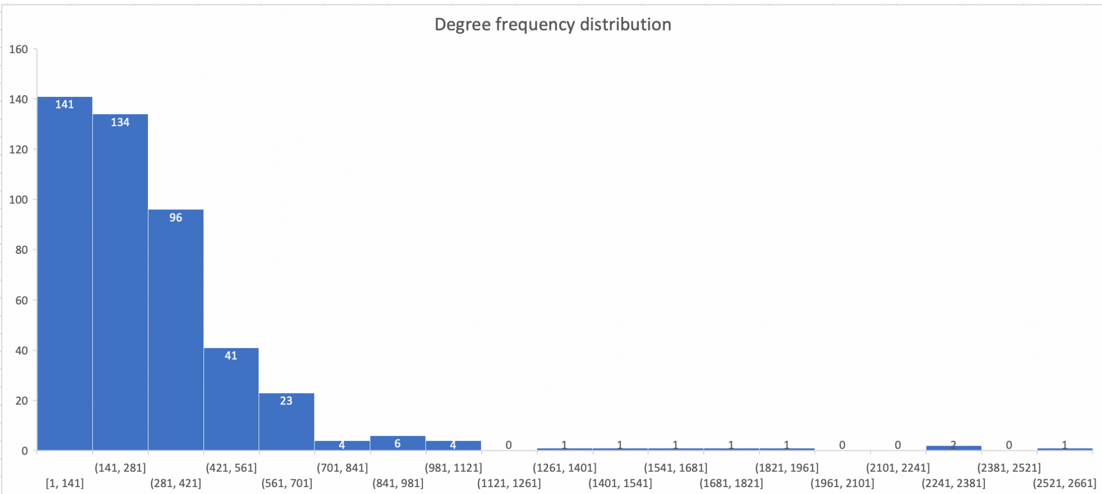


Image 1. Degree frequency distribution

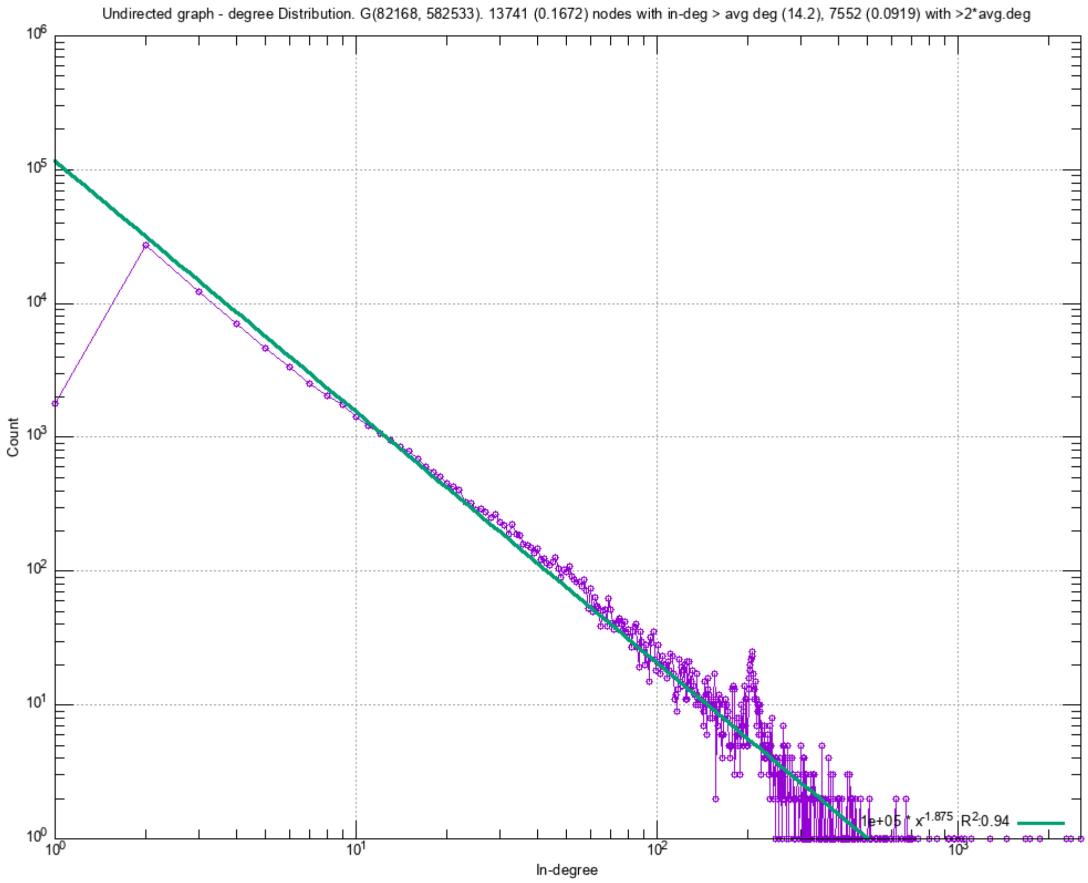


Image 2. The number of nodes with different level of degrees

The minimum degree	1
The maximum degree	2553
Average degree of each node	13.226
The maximum number of nodes of a degree	Degree 2 – 27371 nodes
The number of nodes with minimum degree	1803
The number of nodes with maximum degree	1

Table 2. Degree frequency distribution and summary

According to above image and tables, the network has no zero-degree nodes. It is a huge connected component and all nodes have at least one edge with other nodes. The diameter is large. It means that each network node is alienate from each other relatively. It is worth to pay attention that there are 12 nodes which have more than 1000 degree, which implies that it is easy to have a complete cascade if they are chosen to be an initial adopting key node.

5 Result

5.1 Effect of threshold

In this part, we are interested in how the threshold would affect the number of nodes that change their decision. We set the following conditions for all runs:

- 35000 initial adopting key nodes
- 0<=threshold<=1
- The decision of all nodes except initial adopting key nodes is "B". The only other decision is "A"

The data below shows the effect of different threshold on the situation with 35000 initial adopting key nodes.

Threshold/times	1	2	3	4	5	6	7	8	Total
0.04	4640	1862	325	36	6	2	0	0	6871
0.26	4815	1758	245	37	7	0	0	0	6862
0.37	4730	1860	331	26	1	0	0	0	6948
0.41	4821	1763	269	25	1	0	0	0	6879
0.47	4765	1727	299	34	1	0	0	0	6826
0.58	4827	1841	286	39	4	0	0	0	6997
0.68	4782	1812	294	34	4	0	0	0	6926
0.69	4645	1807	272	35	6	0	0	0	6765
0.84	4785	1650	292	38	6	0	0	0	6771
0.90	4853	1831	261	29	0	0	0	0	6974

Table 3. The number of affected nodes of each wave with different threshold

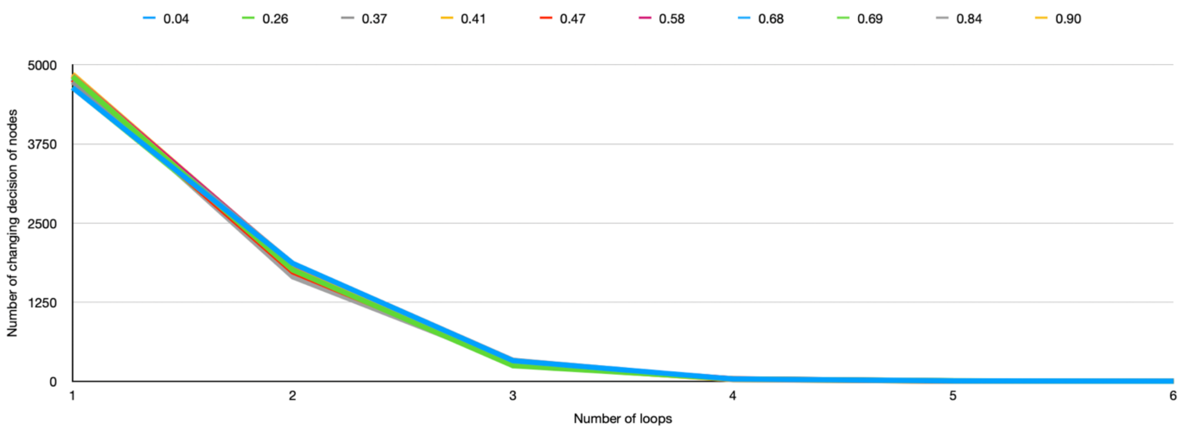


Image 3. The effect of threshold with each wave

5.2 Effect of number of initial adopting key nodes

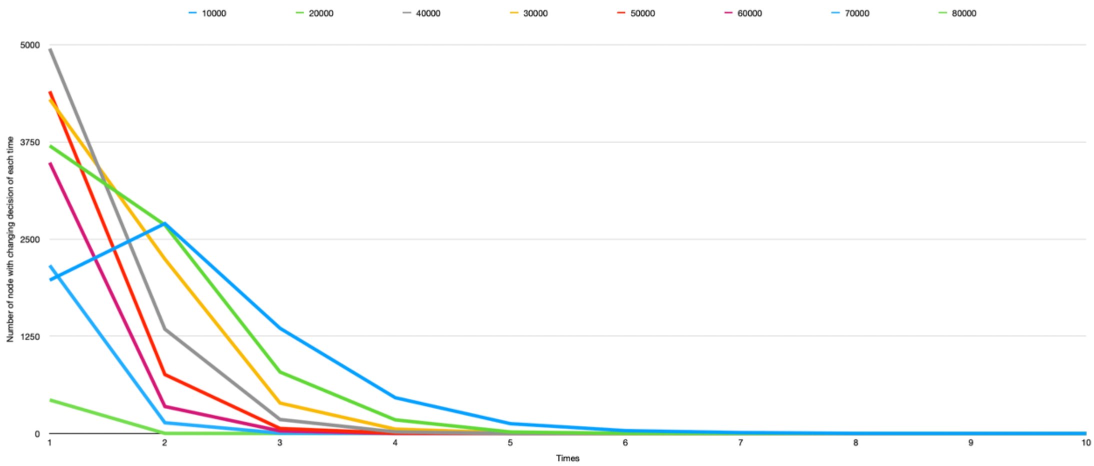
In this part, we are interested in how the number of initial adopting key nodes would affect the number of nodes that change their decision. We set the following conditions for all runs:

- Threshold = 0.75
- The decision of all nodes except initial adopting key nodes is "B". The only other decision is "A"

The data below shows the effect of different number of initial adopting key nodes on the situation with threshold 0.75.

initial adopting nodes/times	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
10000	1972	2700	1353	460	126	36	11	1	0	0	6659
20000	3701	2682	790	175	19	1	0	0	0	0	7368
30000	4298	2243	391	56	8	1	1	0	0	0	6998
40000	4950	1342	179	20	0	0	0	0	0	0	6491
50000	4401	758	64	5	0	0	0	0	0	0	5228
60000	3484	347	32	0	0	0	0	0	0	0	3863
70000	2162	138	4	0	0	0	0	0	0	0	2304
80000	431	1	0	0	0	0	0	0	0	0	432

Table 4. The number of affected nodes of each wave with different number of initial adopting key nodes



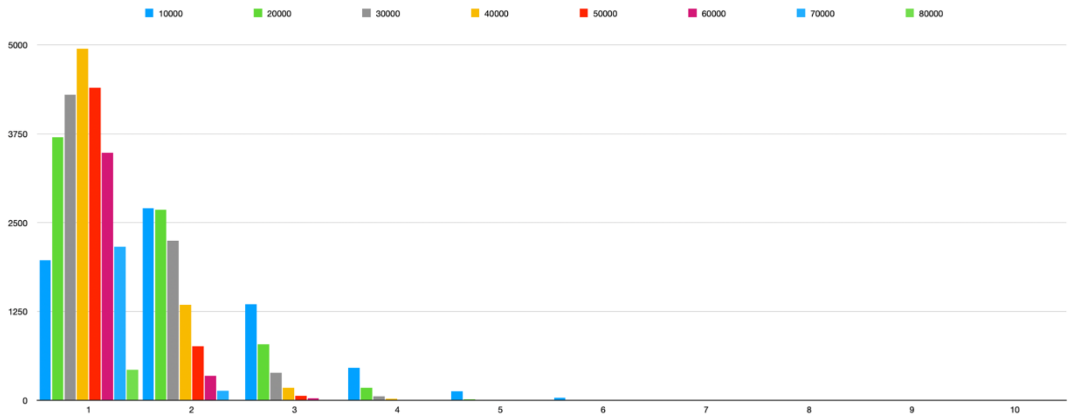


Image 4/5. The effect of affected nodes with each wave

5.3 Effect of position of initial adopting key nodes

In this part, we are interested in how initial adopting key nodes places would affect the number of nodes that change their decision. We set the following conditions for all runs:

- 35000 initial adopting key nodes
- The total degree of all initial adopting key nodes is ~93000
- Threshold = 0.75
- The decision of all nodes except initial adopting key nodes is “B”. The only other decision is “A”

The data below shows the effect of different position of initial adopting key nodes on the situation with threshold 0.75. In order to increase the accuracy of experiment, we have test 3 times in each case. The first case with “A_” means that all initial adopting key nodes place on nodes with degree between 1 to 3. “B_” means that all initial adopting key nodes place on nodes with degree between 3511 to 2553.

Test case/times	1	2	3	4	5	6	7	8	Total
A1	4770	1897	320	48	13	0	0	0	7048
A2	4542	2043	391	54	8	2	0	0	7040
A3	4573	1999	357	42	5	0	0	0	6976
B1	4876	1718	302	41	5	1	0	0	6943
B2	4790	1819	277	44	7	1	1	0	6939
B3	4779	1777	290	27	0	0	0	0	6873

Table 5. The number of affected nodes of each wave with different position of initial adopting key nodes

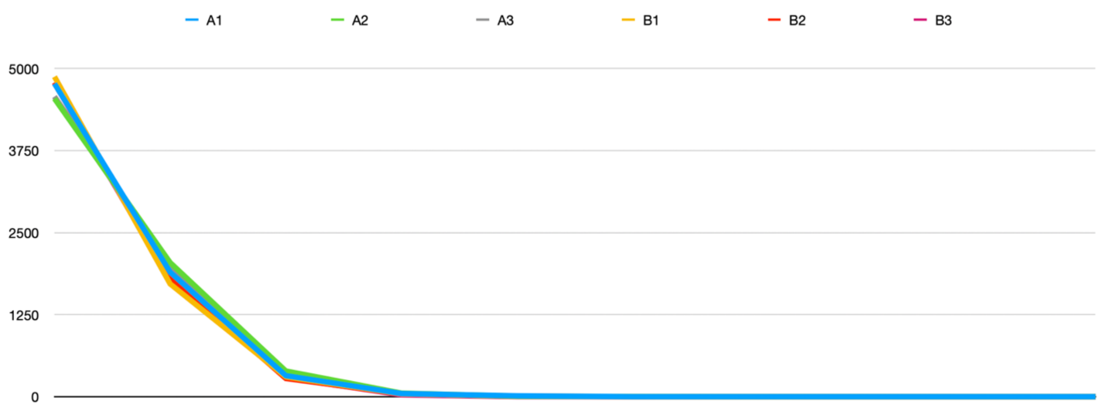


Image 6. The effect of affected nodes with each wave

5.4 The clustering coefficient

In this part, we are interested in finding the cluster according to the clustering coefficient.

The data below shows the distribution of clustering coefficient of the graph. The average of clustering coefficient is 1.023. Since the range of node of degree is large, the file "analysis_4.py" will create a list of printing the clustering coefficient of each degree.

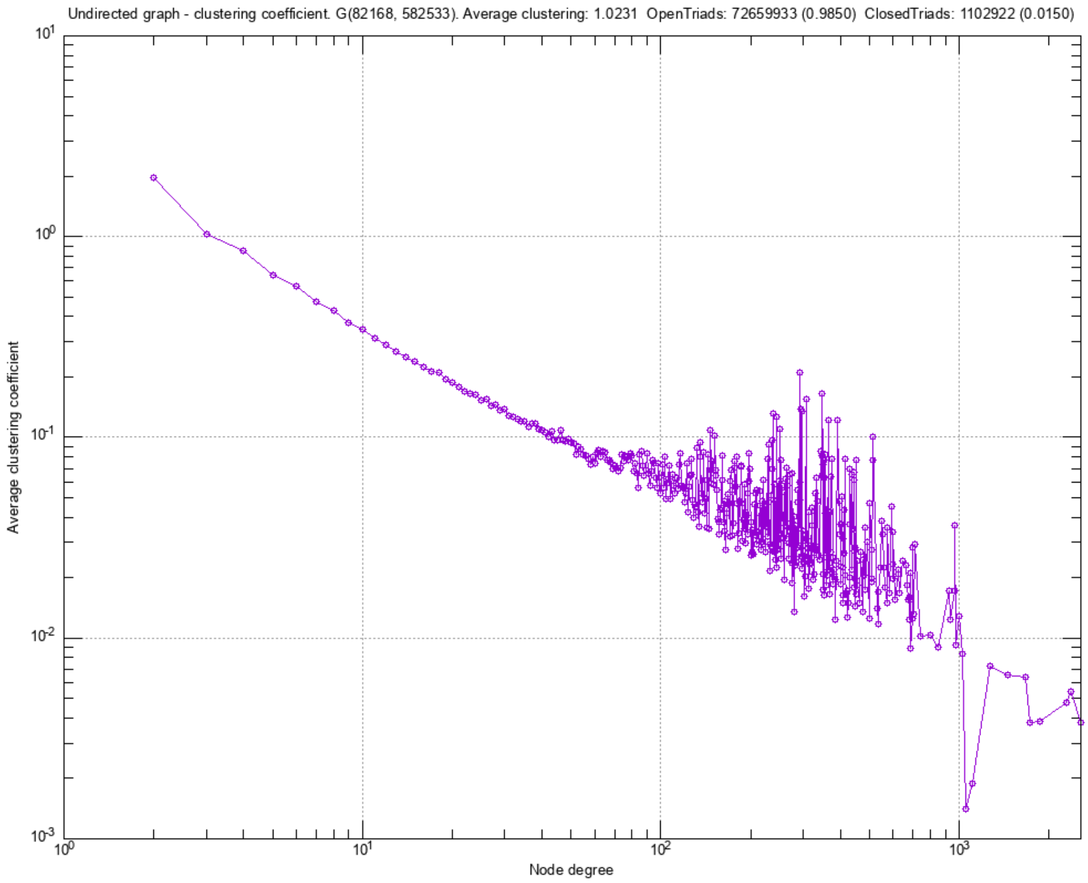


Image 7. The distribution of clustering coefficient

6 Discussion

6.1 Factors affecting cascading behaviors in networks

6.1.1 Threshold

According to our simulation results, we can see that the effect of changing threshold is low. We have done the experiment from the low threshold 0.04 to the higher 0.9. In the first time of each threshold, the initial adopting key nodes can affect 4700 - 4800 nodes. The difference between each threshold is not large. As we can see the line chart of the number affected node increase of each threshold with each wave, they have a similar trend to affect nodes. It proves that the threshold is not the main factor to affect cascading behaviors in networks. Although the lower threshold 0.04 has affected node in the 6^{th} time and the higher threshold 0.9 has stopped affecting from 5^{th} time, their total number of affected nodes is similar. As a result, the number of thresholds is not the main factor.

6.1.2 Number of initial adopting key nodes

According to our simulation results, we can see that the effect of changing number of initial adopting key nodes is high. We have done the experiment from the low number

10000 to the higher 80000. In the first time of each number, there is a large difference between 10000 nodes and 80000 nodes. There is a descending order of node increasing. The reason of this result may be related to the number of available nodes. When the number of initial adopting key nodes is low, the number of available nodes is large; When the number of initial adopting key nodes is large, the number of available nodes is low.

In the other hand, although the total number of affected nodes between 10000 nodes to 40000 nodes is similar, the spreading times of 10000 nodes is more than 40000 nodes. 10000 initial adopting nodes require 8 times to spread out their decision, but 40000 initial adopting nodes only require 4 times. It shows that the number of initial adopting nodes and the spreading times they used has the inverse proportion. When the number of initial adopting nodes is small, they need more times to spread out. When the number is large, they need less times to spread out. As we can see the image of the effect of affected nodes with each wave, each of test case has clear line in the chart. It proves that the number of initial adopting key nodes is the main factor of affecting cascading behaviors in networks.

6.1.3 Position of initial adopting key nodes

According to our simulation result, we can see that the effect of changing position of initial adopting key nodes is low. We have done the experiment of placing nodes on only 1<=degree<=3 or 351<=degree<= 2553. We do 3 times of each test case for accuracy. The difference between each test case is not large. As we can see the line chart of the number affected node increase of each test case, they have a similar trend to affect nodes. It shows that position of initial adopting key nodes is also not the main factor. Wherever we choose the initial adopting key node, if the sum degree is similar, the effect is also similar.

6.1.4 Clustering coefficient

According to the information of graph, the average clustering coefficient is 1.023. It shows that the network exists a neighborhood that every neighbor of node is also connected to other nodes. There is no major block cluster in the network. However, when we look at the list of the clustering coefficient of each degree. We would find that only degree 2 and 3 has the clustering coefficient which is more than 1. It shows that the node of degree 2 and 3 is the only major group in the network. It is easy to have complete cascade in this subgroup, but not easy in other degree, especially the node of degree larger than 600.

7 Conclusion

To conclude, after the simulation of cascading behaviors in networks. We find that the number of initial adopting key nodes is the main factor of affecting cascading behaviors in networks. The impact of other factors such as position of initial adopting key nodes and the threshold is low. From the clustering view, the nodes of high degree become an obstacle of the network to obtain a complete cascade.

8 Reference

SNAP: Stanford Network Analysis Project. (n.d.). Retrieved from http://snap.stanford.edu

9 Remarks

- 'analysis_1.py" is the code for doing experiment 5.1
- 'analysis_2.py", "analysis_2_1.py" are the code for doing experiment 5.2
- 'analysis_3.py" is the code for doing experiment 5.3
- 'analysis_4.py" is the code for doing experiment 5.4