EE 232E

Graphs and Network Flows

Project 1 Report

Pingyuan Yue: 504737715

Ke Xu: 604761427

Yuanyi Ding: 404773978

5/14/2017

Table of Contents

1. Setup the graph and curve it

We just utilized the edge list file to setup the graph using function read.graph(), the graph is connected and the diameter is 8. Figure 1.1 shows the distribution of degree of this graph.

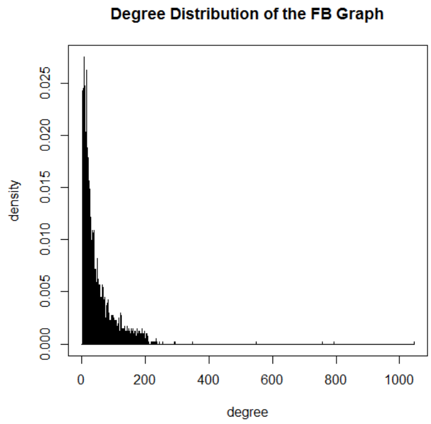

Figure1.1 Degree Distribution of generated graph

As for the fitting curve, we use the stat_smooth() function to generate the statistics model. After studying the shape of the curve, we determine to try four models, which are y ~ log(x), y ~ I(1/x*a) + b*x, y ~ I(1/x*a)+b and y ~ I(exp(1)^(a + b * x) respectively. And as we can see from figure 1.2, the curve of y ~ I(exp(1)^(a + b * x) fits best.

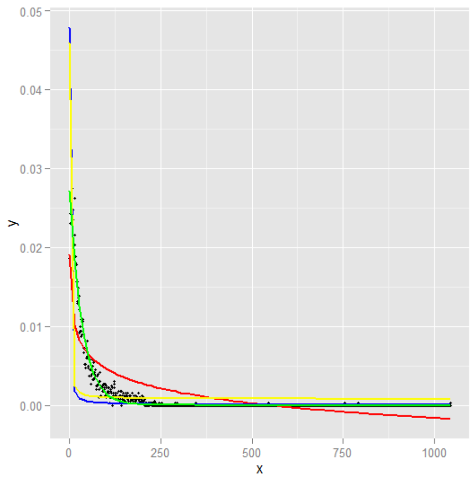

Figure1.2 Fitting curves with different models

The conclusion of the best model:

Formula: y ~ I(exp(1)^(a + b * x)) where a=-3.594 and b=-0.029

Residual standard error: 0.0006342 on 1043 degrees of freedom

Number of iterations to convergence: 15

Achieved convergence tolerance: 7.799e-07

And the MSE (total mean squared error) is 4.016458e-07 and average degree is 43.69101.

2. Generate the personal network of node 1

We generate the personal network which is the subgraph of the whole graph with node 1 and its neighbors, we can see from figure 2.1 that all the nodes (except for node 1) share a mutual friend node 1. The total edges of this personal network are 2866 and the total nodes are 348.

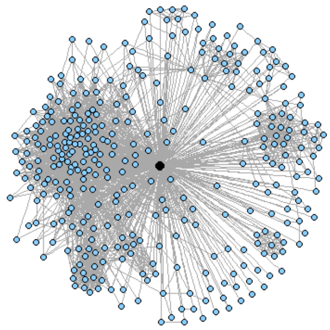

figure 2.1 personal network of node 1

3. Analyze the community structure of core nodes

We just implemented a for loop and totally found 40 such core nodes in the graph and the average degree of such nodes are 279.375. We selected node 1 to analyze, which is a core node, in details.

Figure 3.1 shows the community structure of personal network 1 using fast-greedy algorithm; figure 3.2 shows the community structure using edge-betweenness algorithm and figure 3.3 shows the community structure using infomap algorithms respectively.

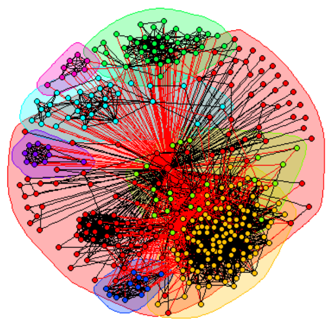

Figure 3.1 Community Structure using fast-greedy algorithm

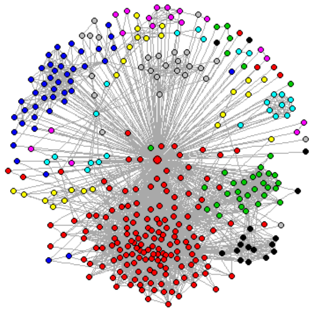

Figure 3.2 Community Structure using edge-betweenness algorithm

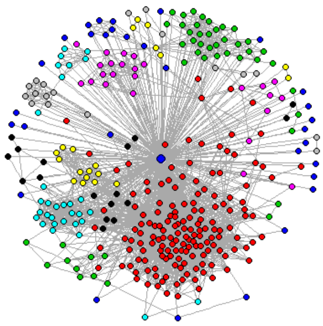

Figure 3.3 Community Structure using infomap algorithm

We see that, though determined using different algorithms, the communities in different graph have some apparent overlap, which mean these community structures have distinguish features. And it can be seen that edge-betweenness algorithm tends to break the graph into more partitions than other two algorithms.

And the modularity of 3 algorithms are 0.4131014, 0.3533022 and 0.3891185 respectively.

The following figure shows community structure of the 3 algorithms.

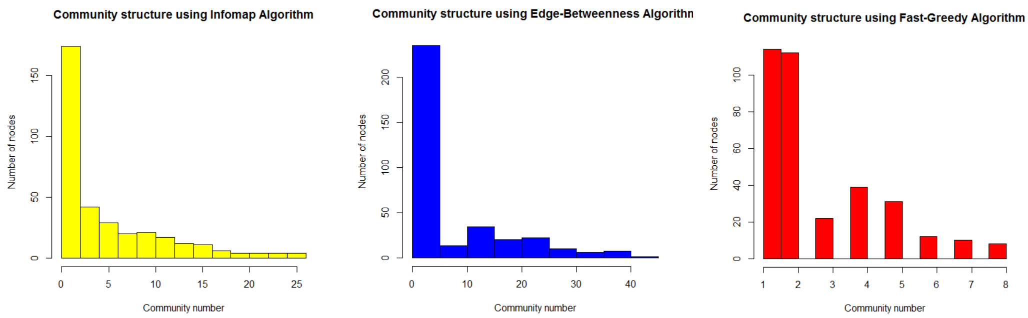

Figure 3.4 community structure of the 3 algorithms

4. Analyze the subgraph without core nodes

Firstly we generated a graph without the core node 1 and then checked the community structure again as done above using three different algorithms. The results are shown in figure 4.1-4.3.

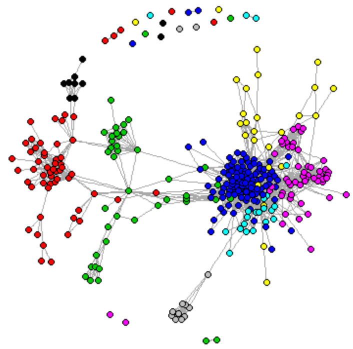

Figure 4.1 Community Structure using fast-greedy algorithm

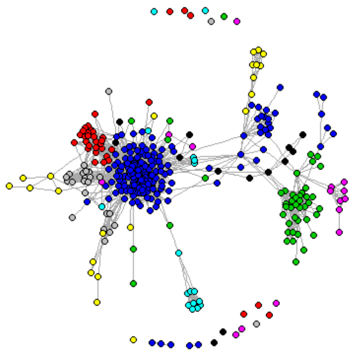

Figure 4.2 Community Structure using edge-betweenness algorithm

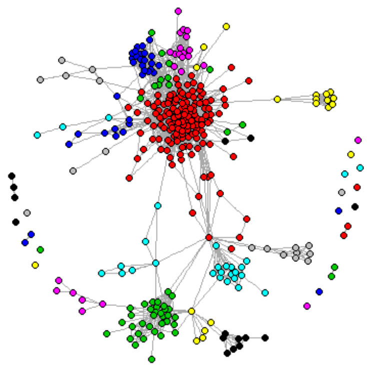

Figure 4.3 Community Structure using infomap algorithm

Looking further into those community structures, we can find out that though they are structured without the core node, the partitions are actually similar. And this can be testified by examine the modularity of structures of problem 3 and problem 4.

And the modularity of 3 algorithms are 0.4418533, 0.4161461 and 0.4180077 respectively.

The difference of modularity is about 10% between two parts.

5. Embeddedness and Dispersion

In this problem, we are required to calculate embeddedness and dispersion about the nodes in the network. Embeddedness is the number of mutual friends, which means the larger the embeddedness is, the more mutual friends you have while dispersion is the sum of distance among all mutual friends, that is to say, the larger the dispersion is, the more likely mutual friends don't know well about each other. The calculation and analysis are shown below.

It is found in problem 3 that there are 40 core nodes in the network. The distribution of embeddedness and dispersion over all core nodes are plotted below:

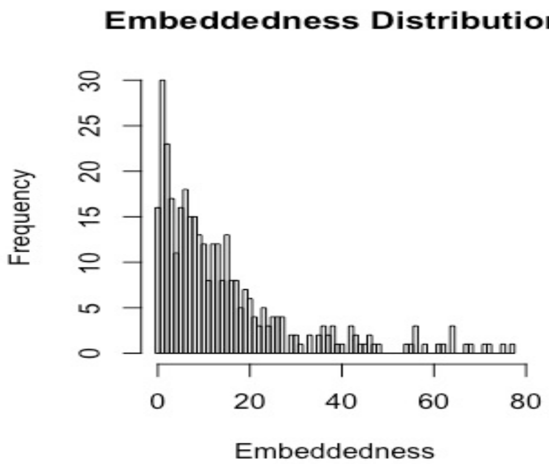

Fig 5.1 The Embeddedness Distribution

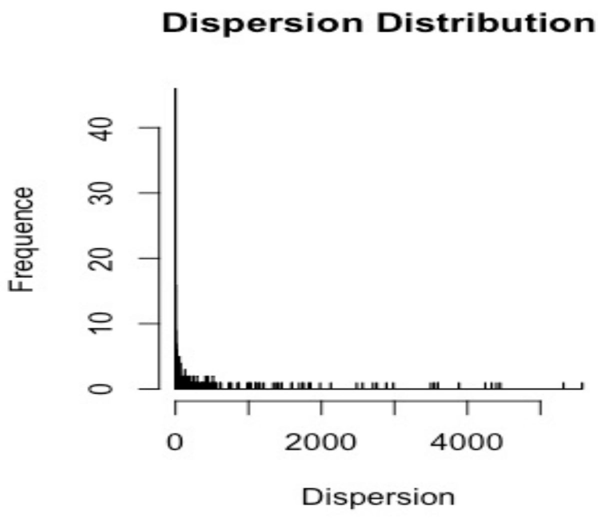

Fig 5.2 The Dispersion over Distribution

Then we randomly pick 3 core nodes, node 1, node 10 and node 15, for further analysis in this problem. For each node, we generate the personal network respectively, and calculate the embeddedness and dispersion over each personal network. Then we plot 3 personal network structures with highlight of maximum dispersion node, maximum embeddedness node and maximum dispersion/embeddedness node.

For node 1:

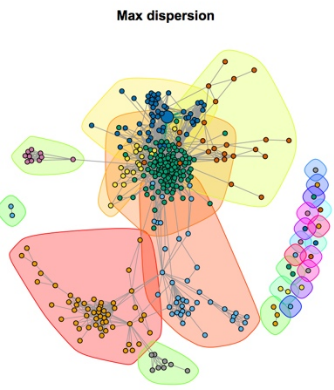

Fig 5.3.1 Network Structure of Node 0 with Maximum Dispersion Highlighted

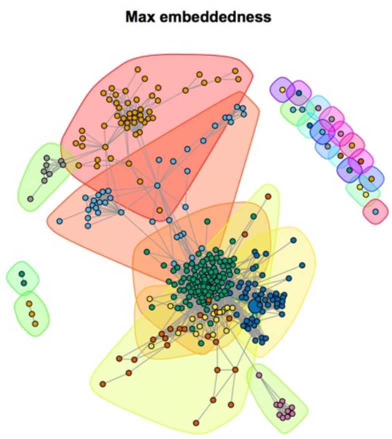

Fig 5.3.2 Network Structure of Node 0 with Maximum Embeddedness Highlighted

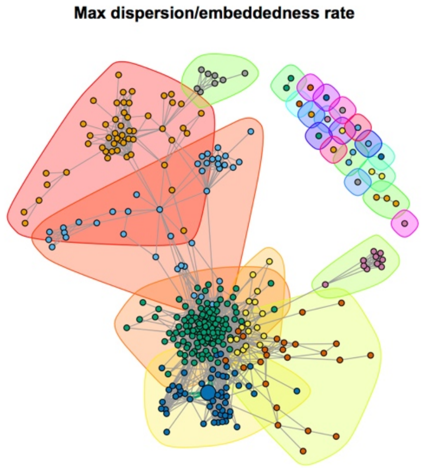

Fig.5.3.3 Network Structure of Node 0 with Maximum Rate Highlighted

For node 10:

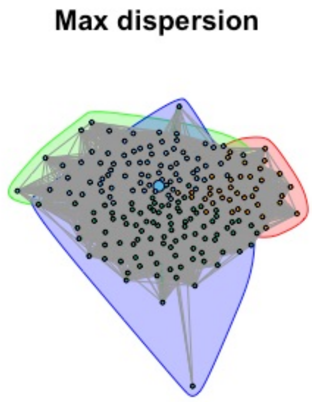

Fig 5.4.1 Network Structure of Node 10 with Maximum Dispersion Highlighted

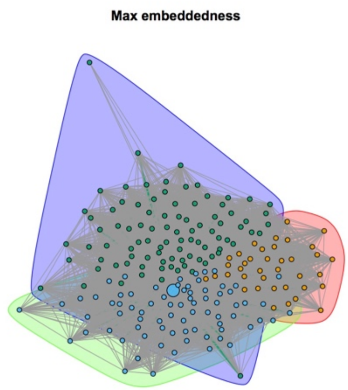

Fig 5.4.2 Network Structure of Node 10 with Maximum Embeddedness Highlighted

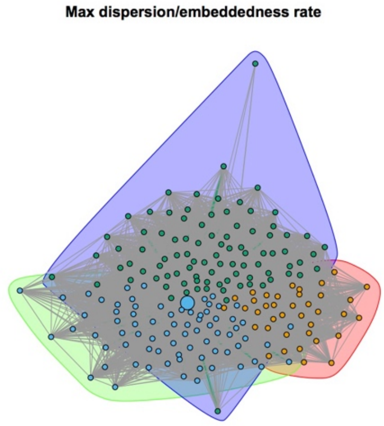

Fig 5.4.3 Network Structure of Node 10 with Maximum Rate Highlighted

For node 15:

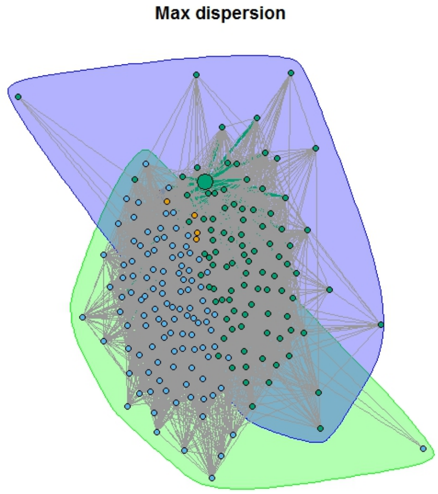

Fig 5.5.1 Network Structure of Node 30 with Maximum Dispersion Highlighted

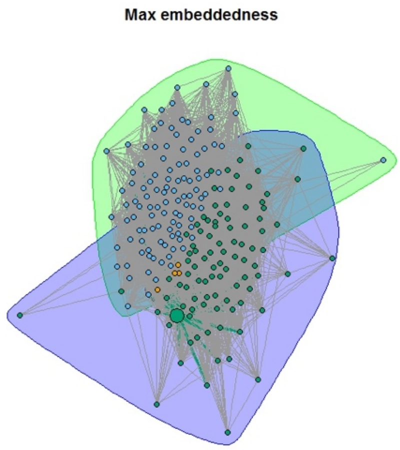

Fig 5.5.2 Network Structure of Node 30 with Maximum Embeddedness Highlighted

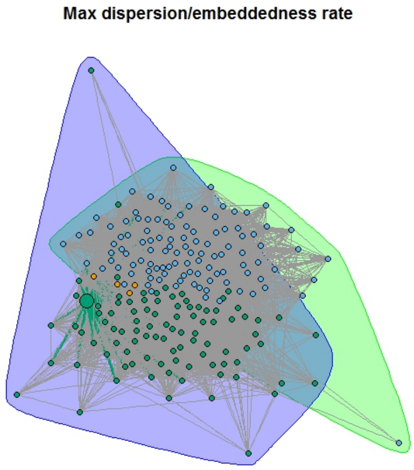

Fig 5.5.3 Network Structure of Node 30 with Maximum Rate Highlighted

Analysis:

From the result above, it could be concluded that the embeddedness indicates the closeness between two people. Dispersion means that how far away between your friends. Larger dispersion means that your mutual friends are not likely to know each other. The dispersion/embeddedness rate is the combination of the previous two features which could more accurate.

6. Community

Communities in personal network can translate into different aspects of one's life. For example, friends in your Facebook can be characterized as high school friends, college friends, and colleagues, etc. So, in this part of the project, we continue analyzing the Facebook data, we try to find out features to determine the communities with size larger than 10 that belongs to certain kind of types. Here we decide to use intimacy to distinguish two different communities. We choose to use degree, clustering coefficient and density to determine the intimacy. Below are the results:

	Type 1			Type 2		
Core-Node	Max Degree	Max Clustering Coefficient	Max Density	Min Degree	Min Clustering Coefficient	Min Density
1	0.8047337	0.8909091	0.8717949	0.1911357	0.3701419	0.1963016
2	0.7202216	0.8878505	0.7818182	0.0822	0.3956786	0.0830303
3	0.6371191	0.8001407	0.6725146	0.2231445	0.4891128	0.2266865
4	0.4943464	0.7102949	0.5014085	0.3525762	0.5441022	0.356327
5	0.6082231	0.7332656	0.6149068	0.6008652	0.7060839	0.6085686
6	0.6780508	0.8099441	0.6921769	0.56375	0.7063263	0.5782051
7	0.7063712	0.8479592	0.75	0.5330161	0.6548621	0.5463415
8	0.6017287	0.7258199	0.606089	0.585851	0.7100395	0.5939878
9	0.6943483	0.7748131	0.7082353	0.6522445	0.7541058	0.6677741
10	0.8555556	0.8987867	0.8850575	0.6325	0.7122071	0.6487179
11	0.7990488	0.8694466	0.8275862	0.4623757	0.6304599	0.4687976

12	0.7534626	0.8698594	0.7953216	0.09947665	0.3177334	0.1001143
13	0.7039239	0.7832022	0.7121091	0.5419403	0.7031564	0.5494673
14	0.7469008	0.8138012	0.7642706	0.5761317	0.7216312	0.5982906
15	0.5460414	0.6867534	0.5498867	0.4602361	0.6494837	0.4672094
16	0.641484	0.744532	0.6478354	0.4460425	0.6237254	0.4492059
17	0.5762167	0.7361703	0.5850816	0.5700601	0.6959259	0.5743463
18	0.5269901	0.6837864	0.5308368	0.4876908	0.6400903	0.4918951
19	0.7146814	0.8296767	0.754386	0.1568696	0.4015702	0.1583495
20	0.8322902	0.8850978	0.840293	0.6037709	0.7547022	0.6088876
21	0.845679	0.9088146	0.8954248	0.4013841	0.6357616	0.4264706
22	0.7926743	0.8430165	0.7978215	0.6955111	0.7975296	0.7120709
23	0.8888889	0.9710145	0.969697	0.4146939	0.6174561	0.4268908
24	0.8827526	0.9071845	0.8903625	0.6122449	0.6588486	0.6593407
25	0.8238379	0.86417	0.828771	0.5636147	0.7104485	0.5837438
26	0.7943099	0.8430899	0.7995356	0.6771228	0.7834786	0.6932447
27	0.802361	0.8551278	0.8073758	0.7045455	0.7772587	0.7209302
28	0.8711111	0.9370277	0.9333333	0.7455621	0.8048808	0.7538462
29	0.8550013	0.8913101	0.8618413	0.5612245	0.6774194	0.6043956
30	0.7178218	0.8046302	0.721393	0.7178218	0.8046302	0.721393
31	0.7781178	0.841428	0.7823012	0.6666667	0.8174387	0.7272727
32	0.7535322	0.8195464	0.7576953	0.4142012	0.6386139	0.4487179
33	0.7636031	0.834086	0.7679172	0.664952	0.7707755	0.6711665
34	0.7701048	0.8326006	0.7744557	0.6736111	0.7734894	0.7028986
35	0.8753909	0.9030825	0.8819237	0.5354671	0.6415334	0.5434592
36	0.8546221	0.8945076	0.8631683	0.7426036	0.8040176	0.7664618
37	0.7830744	0.8464978	0.7880941	0.5535754	0.7227248	0.5577376
38	0.7871585	0.8382447	0.7924062	0.7091837	0.8260638	0.7354497
39	0.7779021	0.8783784	0.8030303	0.7361111	0.8329676	0.7777778
40	0.8571429	0.9250535	0.9230769	0.1795482	0.3726708	0.1826981

The maximum values of the above three features indicate that the community share a high intimacy, we believe this kind of community should be “classmate” type. The minimum values of the above three features indicate that the community share a low intimacy, we believe this kind of community should be like “normal friends” or “friends just meet” type.

7. Tagged Relationship

In this part of the project, we analyze another real social network called Google+ ego networks. We created personal network for users who have more than 2 circles and extract the community structure of each personal network using both Walktrap and Infomap algorithms. Due to the scale of our network is very large, in the report we choose to present the structures of two random chosen users.

First user:

The ID of the user with more than 2 circles is 7, with number of circles equals to 4.

Overlap of the 1st circle	
Community 1	0
Community 2	0.01433121
Community 3	0.11053985
Community 4	0
Overlap of the 2nd circle	
Community 1	0.01785714
Community 2	0.35509554
Community 3	0.01542416
Community 4	0.13157895
Overlap of the 3rd circle	
Community 1	0.05357143
Community 2	0.61783439
Community 3	0.02570694
Community 4	0.80701754
Overlap of the 4th circle	
Community 1	0.03571429
Community 2	0.08917197
Community 3	0.37017995
Community 4	0.01754386

Second user:

The ID of the user with more than 2 circles is 12, with number of circles equals to 3

Overlap of the 1st circle	
Community 1	0.05579399
Community 2	0
Community 3	0
Community 4	0.7053292
Overlap of the 2nd circle	
Community 1	0.07296137
Community 2	0
Community 3	0
Community 4	0.3134796
Overlap of the 3rd circle	
Community 1	0.05150215
Community 2	0
Community 3	0
Community 4	0.1128527