ECE232E Project2 Report

Luke Lim   Angela Fan

April 2024

1 Facebook network

1.1 Question 1

1.2 Question 1.1

There are 4039 nodes and 8234 edges
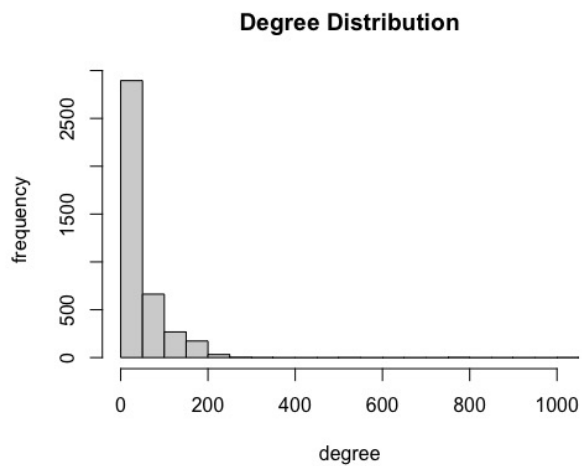
1.3 Question 1.2

The network is connected.
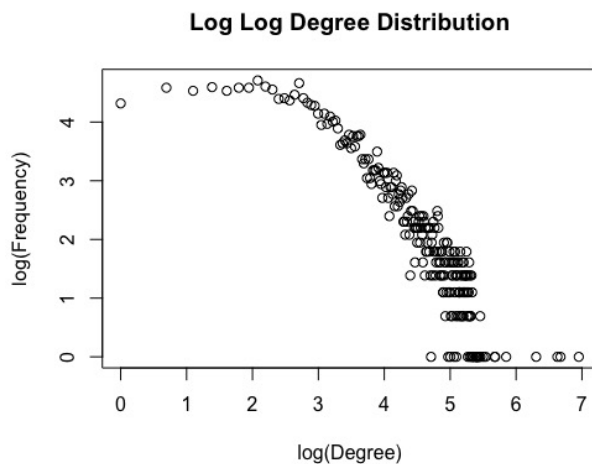
1.4 Question 2

The diameter is 8

1.5 Question 3

Degree Distribution

1.6   Question 4

Log Log Degree Distribution

1.7   Question 5

the ego graph has 348 nodes and 2866 edges

1.8   Question 6

The diameter is 2. The trivial upper bound is 2 and lower bound is 1

1.9   Question 7

upper bound means that the network is not fully connected. lower bound means that the network is fully connected

1.10   Question 8

There are 40 core nodes. The average degree of core nodes is 279

1.11   Question 9

Note: the edge betweeness algorithm is a lot slower than the other two

1.11.1   Node 1

The modularity of node 1 communities fast and greedy is 0.4131014. The modularity of node 1 communities edge betweeness is 0.3533022. The modularity of node 1 communities infomap is 0.3891185. Fast and greedy algorithm produces communities with the highest modularity

2

1.11.2 Node 108

The modularity of node 108 communities fast and greedy is 0.4359294. The modularity of node 108 communities infomap is 0.3891185. Edge betweeness crashed my computer.
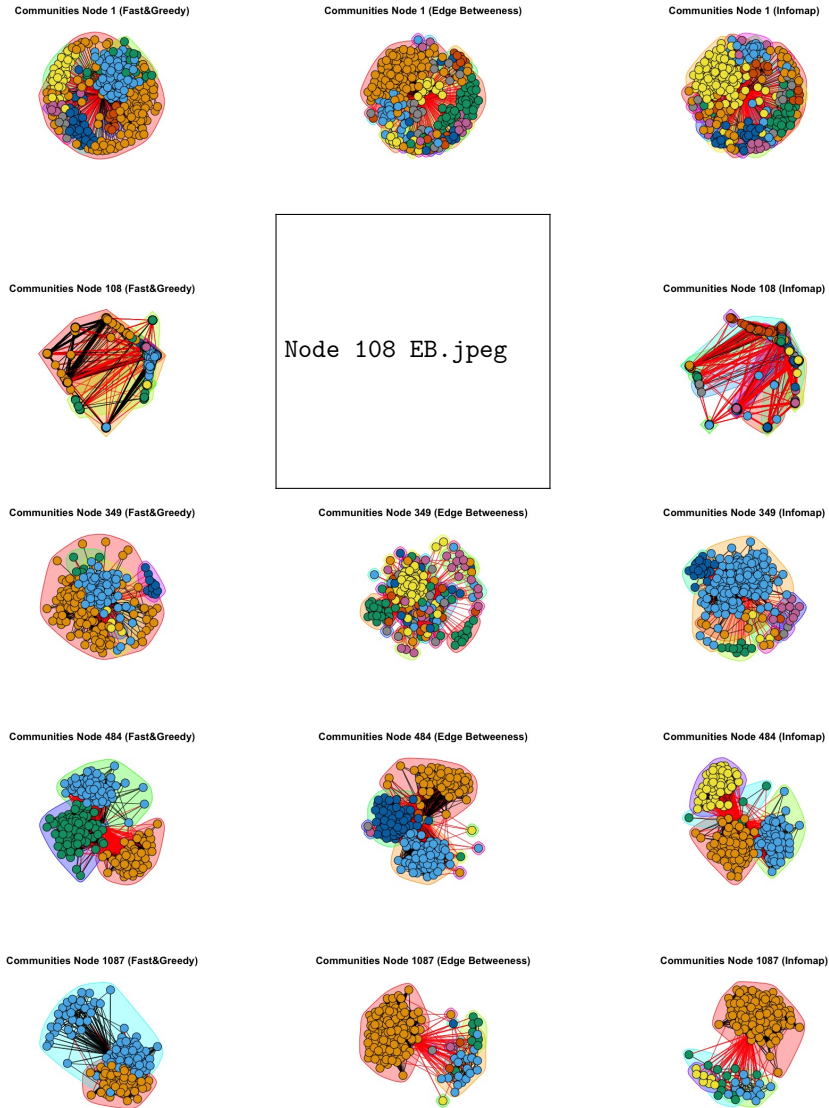
1.11.3 Node 349

The modularity of node 349 communities fast and greedy is 0.2502104. The modularity of node 349 communities edge betweeness is 0.3533022. The modularity of node 349 communities infomap is 0.3891185. Unlike with nodes 1 and 108, infomap algorithm produced communities with the highest modularity

1.11.4 Node 484

The modularity of node 484 communities fast and greedy is 0.5070016. The modularity of node 484 communities edge betweeness is 0.4890952. The modularity of node 484 communities infomap is 0.5152788

1.11.5 Node 1087

The modularity of node 1087 communities fast and greedy is 0.1455315. The modularity of node 1087 communities edge betweeness is 0.02762377. The modularity of node 1087 communities IM is 0.02690662

Communities Node 1 (Fast&Greedy)

Communities Node 1 (Edge Betweeness)

Communities Node 1 (Infomap)

Communities Node 108 (Fast&Greedy)

Node 108 EB.jpeg

Communities Node 108 (Infomap)

Communities Node 349 (Fast&Greedy)

Communities Node 349 (Edge Betweeness)

Communities Node 349 (Infomap)

Communities Node 484 (Fast&Greedy)

Communities Node 484 (Edge Betweeness)

Communities Node 484 (Infomap)

Communities Node 1087 (Fast&Greedy)

Communities Node 1087 (Edge Betweeness)

Communities Node 1087 (Infomap)

1.12 Question 10

1.12.1 Node 1

The modularity of node 1 communities fast and greedy is 0.4418533 The modularity of node 1 communities edge betweeness is 0.4161461. The modularity of node 1 communities infomap is 0.4180077.

1.12.2 Node 108

The modularity of node 108 communities fast and greedy is 0.4580603. The modularity of node 108 communities infomap is 0.5212031. Edge betweeness crashed my computer.

1.12.3 Node 349

The modularity of node 349 communities fast and greedy is 0.2456918. The modularity of node 349 communities edge betweeness is 0.1505663. The modularity of node 349 communities infomap is 0.2448156.

1.12.4 Node 484

The modularity of node 484 communities fast and greedy is 0.5342142. The modularity of node 484 communities edge betweeness is 0.5154413. The modularity of node 484 communities infomap is 0.5434437

1.12.5 Node 1087

The modularity of node 1087 communities fast and greedy is 0.1481956. The modularity of node 1087 communities edge betweeness is 0.0324953. The modularity of node 1087 communities IM is 0.02737159

Communities Node 1 (Fast&Greedy, ego removed)
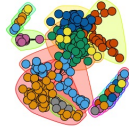
Communities Node 1 (Edge Betweeness, ego removed)
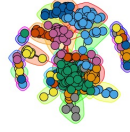
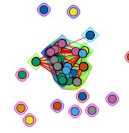Communities Node 1 (Infomap, ego removed)
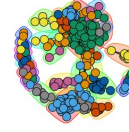
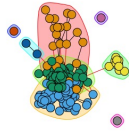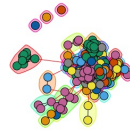Communities Node 108 (Fast&Greedy, ego removed)

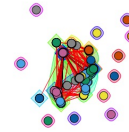rm Node 108 EB.jpeg

Communities Node 108 (Infomap, ego removed)
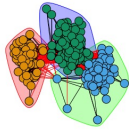
Communities Node 349 (Fast&Greedy, ego removed)
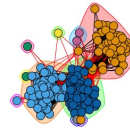
Communities Node 349 (Edge Betweeness, ego remove
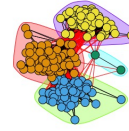
Communities Node 349 (Infomap, ego removed)
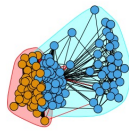
Communities Node 484 (Fast&Greedy, ego removed)
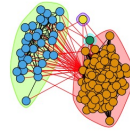
Communities Node 484 (Edge Betweeness, ego remove
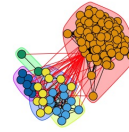
Communities Node 484 (Infomap, ego removed)

Communities Node 1087 (Fast&Greedy, ego removed

Communities Node 1087 (Edge Betweeness, ego remove

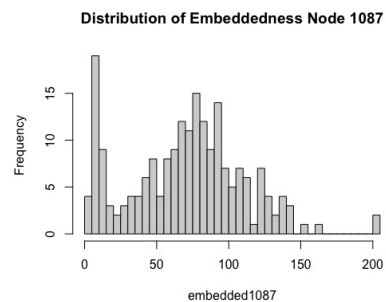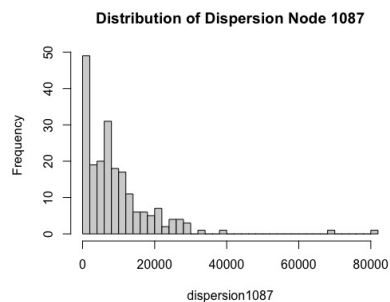Communities Node 1087 (Infomap, ego removed)
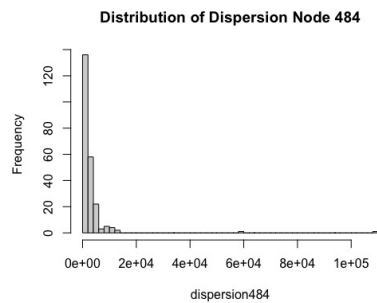
1.13   Question 11

Expression: Degree (in ego network) - 1

1.14 Question 12

Distribution of Dispersion Node 1

Distribution of Embeddedness Node 1

Distribution of Dispersion Node 108

Distribution of Embeddedness Node 108

Distribution of Dispersion Node 349

Distribution of Embeddedness Node 349

Distribution of Dispersion Node 484

Distribution of Embeddedness Node 484

Distribution of Dispersion Node 1087

Distribution of Embeddedness Node 1087

7

1.15 Question 13

Communities Node 1 (Highlight dispersion)
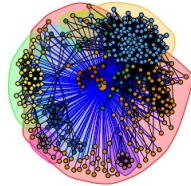
Communities Node 108 (Highlight dispersion)

Communities Node 349 (Highlight dispersion)

Communities Node 484 (Highlight dispersion)

Communities Node 1087 (Highlight dispersion)
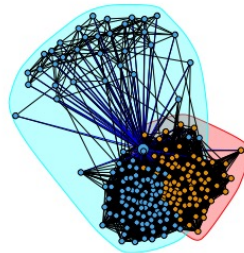
Figure 4: Community Structures with Various Node IDs

1.16 Question 14

Communities Node 1 (highlight ratio)
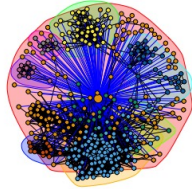
Communities Node 108 (highlight ratio)
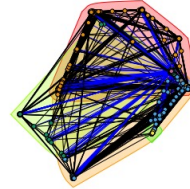
Communities Node 349 (highlight ratio)
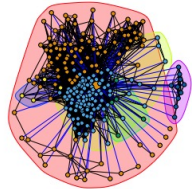
Communities Node 484 (highlight ratio)
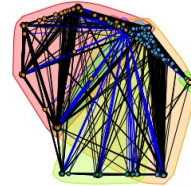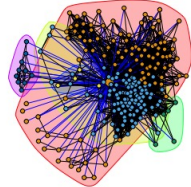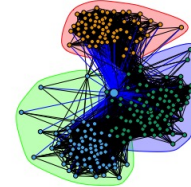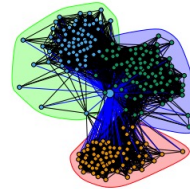
Communities Node 1087 (highlight ratio)
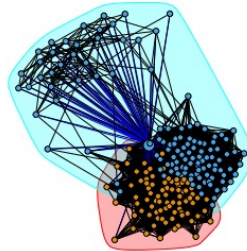
Figure 5: Community Structures with Various Node IDs

1.17 Question 15

The nodes with the highest disperson also have the highest ratio

1.18 Question 16

There are 11 nodes in the list

1.19 Question 17

The average accuracy for common neighbours measure is 0.6832533. The average accuracy for Jaccard measure is 0.6922511. The average accuracy for adamic adar measure is 0.6970664. The adamic adar measure gives the highest accuracy on average

2 Google + network

2.1 Question 18

There are total number of personal network is 189.

2.2 Question 19

In-degree Distribution for node: 10932748047976710849O

Out-degree Distribution for node: 10932748047976710849

In-degree Distribution for node: 115625564993990145546

Out-degree Distribution for node: 11562556499399014554

In-degree Distribution for node: 101373961279443806744

Out-degree Distribution for node: 10137396127944380674
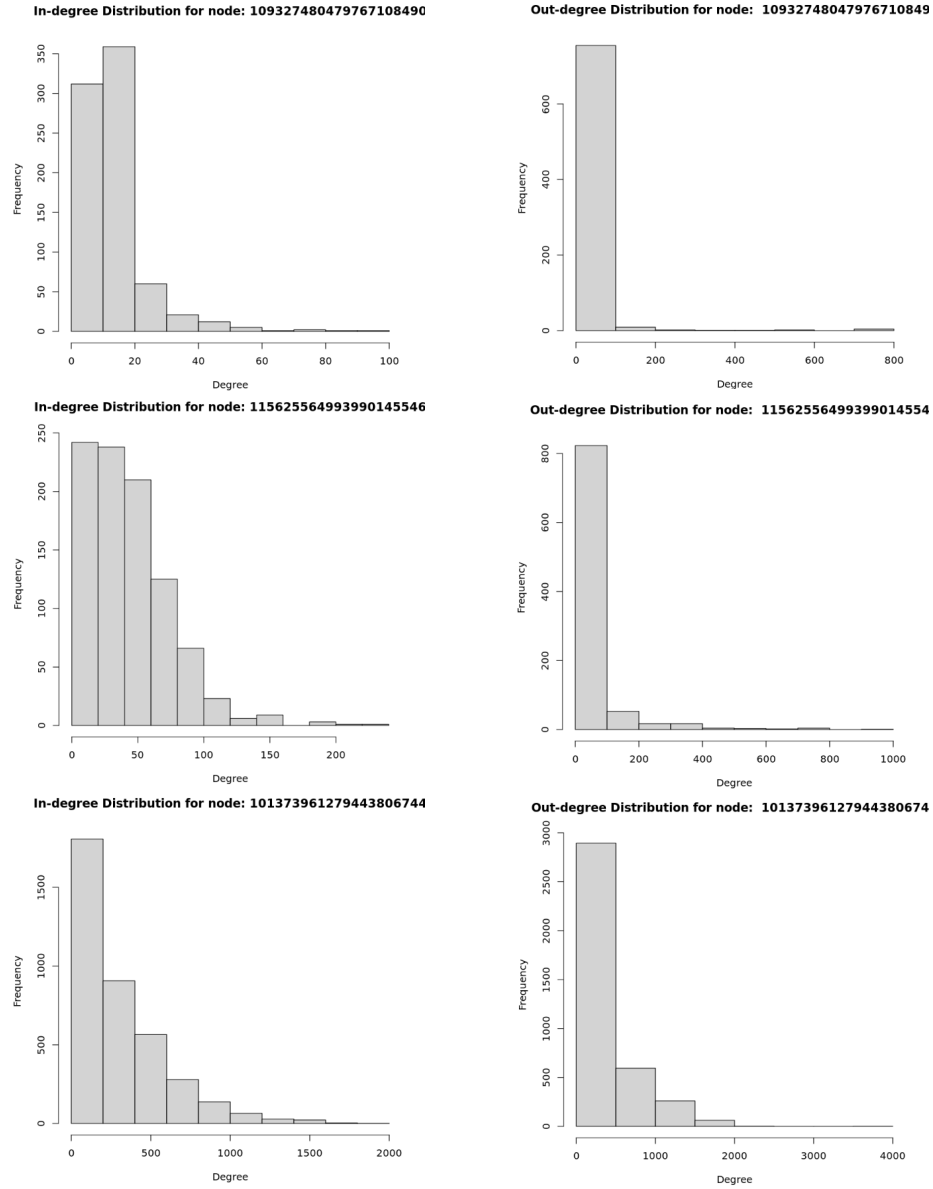
We observe that the in and out degree demonstrates a pattern here. The personal network of the third node includes numerous users. It appears that the indegree of this personal network adheres to a power law. While there are many nodes with a lower indegree, the number of such nodes significantly decreases as the indegree increases.

2.3 Question 20

Following the extraction of community structures for the trio of nodes referred to in Question 19, we employed the walk-trap community detection algorithm to obtain the modularity scores. The resulting modularity scores for the three nodes are presented below:

Node	Modularity
109327480479767108490	0.252765
115625564993990145546	0.319473
101373961279443806744	0.191090

From the above table, the modularity score for the second node is the highest and the score for the third node is the lowest.

A higher modularity score indicates that there is a stronger degree of interconnection among nodes within the same community and less connection between nodes in different communities.

Regarding the modularity of the three nodes in question, the node with the highest modularity demonstrates robust internal community connections and sparse connections between separate communities, leading to superior community clustering. Conversely, the node with the lowest modularity displays an intricate pattern of connections (as detailed in Question 19), including strong inter-community links, culminating in a diminished modularity score.

Community Structure (node: 109327480479767108490)
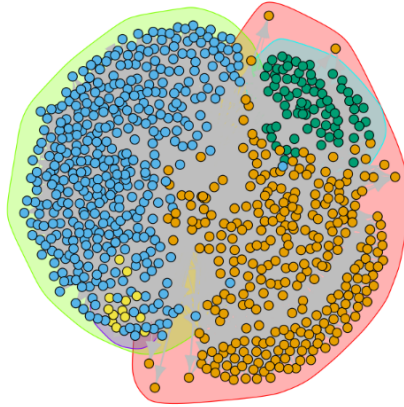
Community Structure (node: 115625564993990145546)
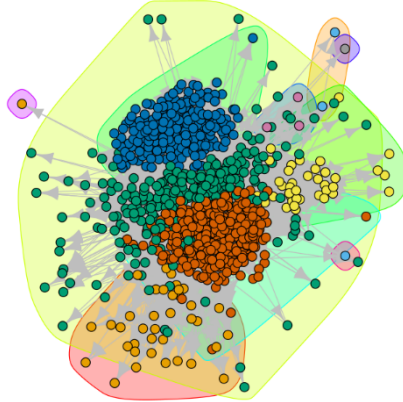
Community Structure (node: 101373961279443806744)
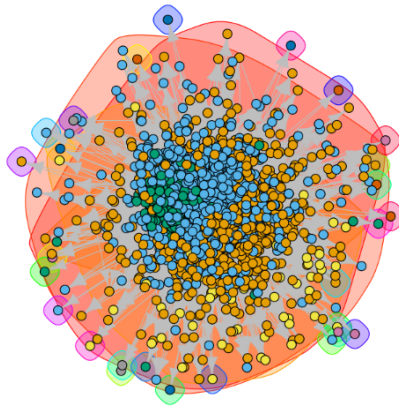
2.4 Question 21

Homogeneity evaluates the uniformity of the community structure. A community achieves a higher homogeneity score if it consists solely of nodes from a single circle.

Completeness assesses the integrity of the circle itself. Higher completeness scores are attained when a circle predominantly allocates its nodes within one community.

2.5 Question 22

The completeness and homogeneity values are:

13

Node	Homogeneity	Completeness
109327480479767108490	0.85189	0.32987
115625564993990145546	0.45189	-3.42396
101373961279443806744	0.00387	-1.50424

From the table, it is evident that the first node exhibits the highest values of homogeneity and completeness, while the third node has the lowest. This observation aligns with the findings in Question 21, indicating that the first node most effectively creates communities within its circle, with members generally belonging to the same community. On the other hand, the third node is the least effective at forming cohesive communities with nodes from its circle, resulting in these nodes forming diverse communities.

3   Cora dataset

3.1   Question 23: Idea 1

Utilizing the Graph Convolutional Neural Network (GCN) with hyper-parameters of 16 hidden channels and 2 layers GCN model we were able to achieve the accuracy of 79.60

3.2   Question 24: Idea 2

Node2vec generates nodes embeddings in graph by performing biased random walks to explore the networks. These walks balance between local exploration and broader graph traversal. This allows for the embeddings to capture both local and global structural properties of the graph.

The choice of classifier we went with was SVM. We were able to achieve accuracy of 84.87% with using only node2vec and achieved accuracy of 71.96% with only using 1433-dimensional text features. This could be possibly due to node2vec being able to capture additional structural information compared to only using 1433-dimensional text features.

When combining the node2vec and text features, we were able to achieve the accuracy of 84.32%. The highest accuracy was 84.87% which was using node2vec + SVM classifier.

3.3   Question 25: Idea 3

When the teleportation probability was set to 0, both the accuracy and the F1 score were observed to be 28%. Increasing the teleportation probability to 0.1 improved both measures to 68%, suggesting that this level of probability achieves a balance between global exploration and local exploitation within the document graph. However, at a teleportation probability of 0.2, there was a

slight decrease in performance, with accuracy and the F1 score dropping to 66%, likely due to too frequent transitions to seed documents that may overshadow the nuanced relationships among local documents. Moderate teleportation seems to be key for the algorithm to effectively navigate and represent the document network.