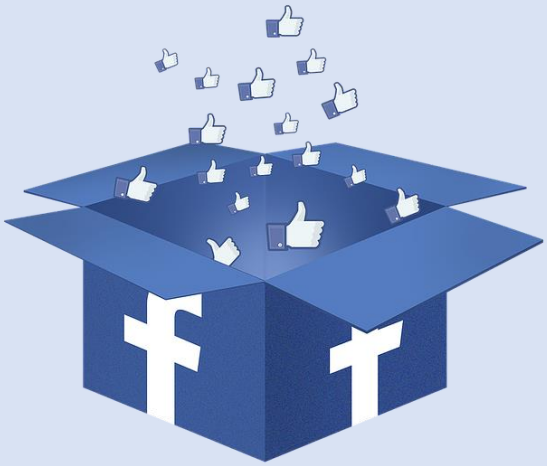
Analysis of Social Network - Facebook friends

Assignment #1



By: Muhammad UMER
SID: 44135149

TABLE OF CONTENTS

Introduction 3

Data Collection Process 3

Steps of Analysing and Visualizing data in Gephi 3

GRAPH - A 4

GRAPH- B 5

Network Diameter: 6

Graph density: 6

Average path length: 6

Clustering statistics 6

Modularity: 6

Number of triangles: 6

Betweenness Centrality: 6

Eigenvector: 6

Conclusions 7

References 7

INTRODUCTION

This report is about my Facebook network analysis with Gephi, it covers the data collection, analysing and visualizing by using different tools and measures in Gephi. The report also compares my Facebook network graphs made from Gephi and its conclusions.

DATA COLLECTION PROCESS

I loaded my Facebook network data into google chrome (lost circles extension), and visualized it. Then I downloaded the file Graphml(nopics).

I used Grpahml(nopics) in Gephi for my network analysis.

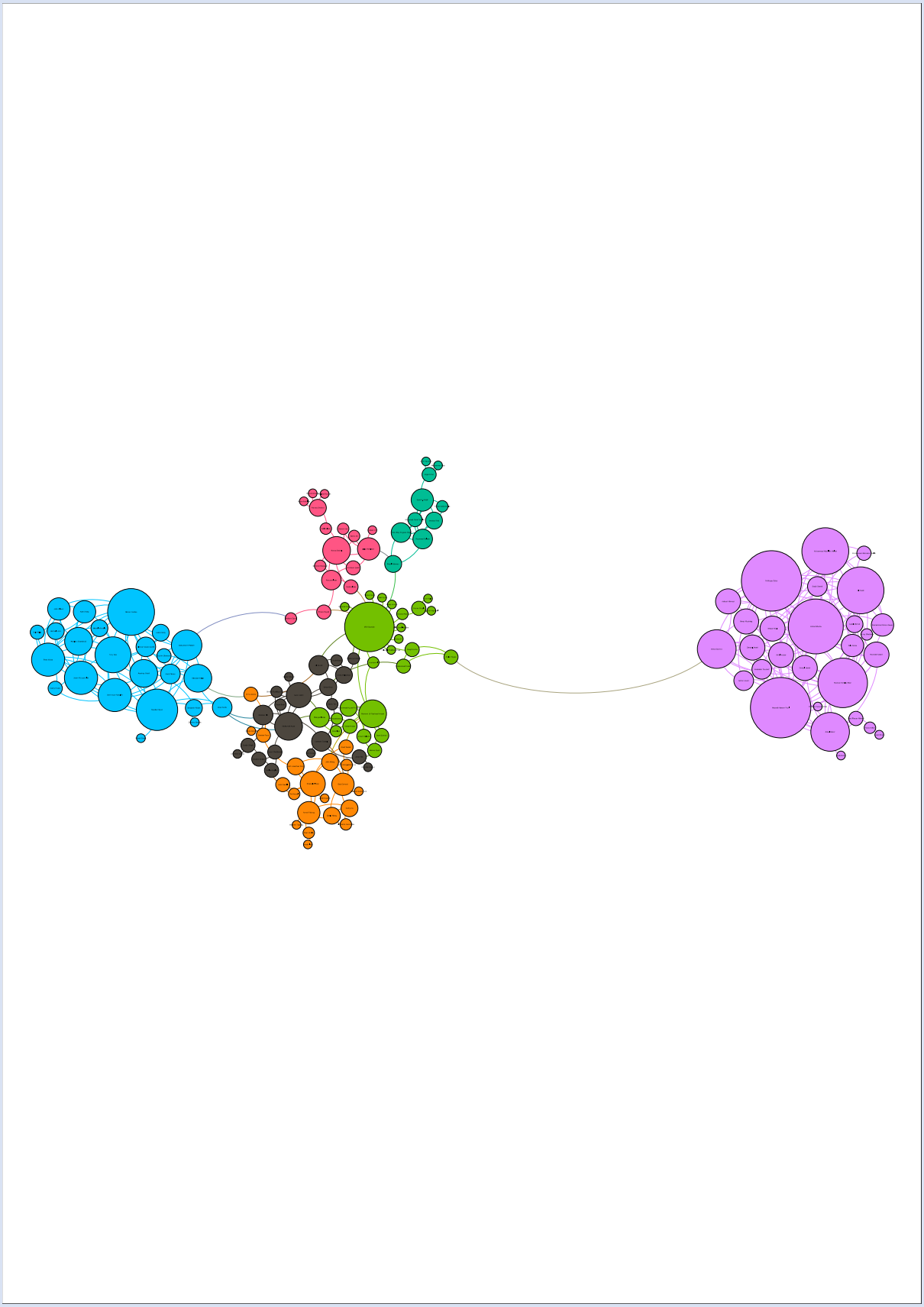
STEPS OF ANALYSING AND VISUALIZING DATA IN GEPHI

I imported my graphML file and opened it in Gephi, I got a big messy bunch of nodes and edges in the start but after applying filters like (Giant Component) to reduce noise and by trying out different algorithms from layout, I started getting an understandable graph. But Force Atlas 2 algorithms got me the Graph in which I got different clusters well defined. I also used appearance section to make it look better. I made two graphs in Gephi and analysed network statistics.

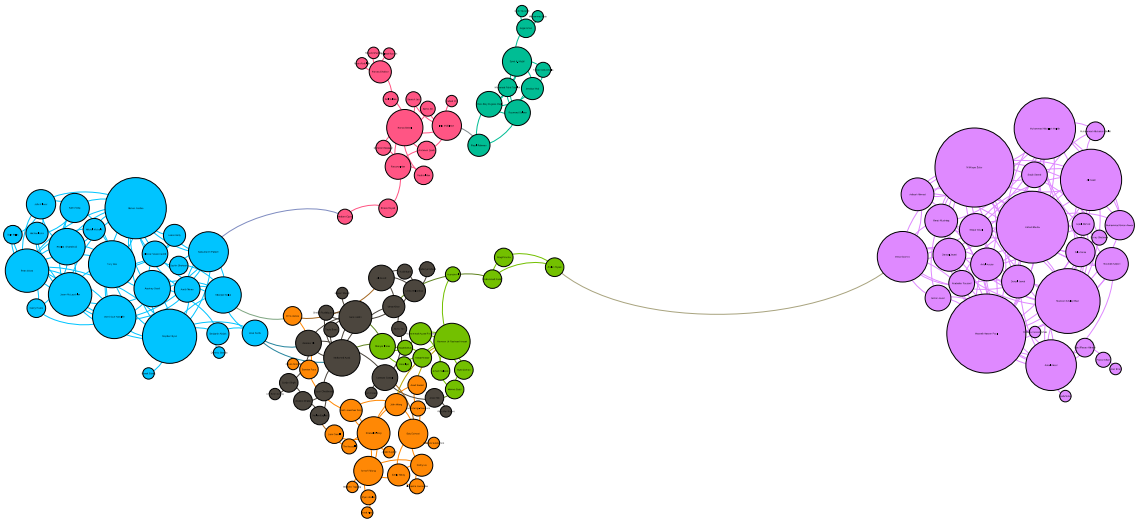
GRAPH-A (shows the Facebook network in different clusters).

GRAPH-B (I removed one of the nodes with the highest betweenness centrality i.e. MD sikander to see its impact on the overall network).

GRAPH - A



GRAPH- B



NETWORK STATISTICS

NETWORK DIAMETER:

The maximum steps required to cross the network in Graph-A is (13) and in Graph-B is (21).

GRAPH DENSITY:

About 3.2% nodes relate to one another based on output value of 0.032 in Graph-A, while 3.6% in Graph-B.

AVERAGE PATH LENGTH:

It takes nodes close to 6 steps on average to reach any other node in the network in Graph-A and 8 steps in Graph-B.

CLUSTERING STATISTICS

MODULARITY:

The modularity statistic splits this graph into 7 different clusters. Which can be differentiated by colours in the graphs.

Clustering coefficient in Graph-A is 42% and Graph-B is 44% which means in both graphs less than half of the available triplets are closed and more than 50% are open.

NUMBER OF TRIANGLES:

The number of triangles in Graph-A and Graph-B are 305 and 304 but most of them as we can see by clustering coefficient are open.

BETWEENNESS CENTRALITY:

There is a rather large spread in betweenness centrality from top to bottom lowest being 0.00 and the highest being 5259 of MD Sikander.

EIGENVECTOR:

M. Waqar Zafar has the highest eigenvector centrality of 1.00 and Derek Alvis is the lowest eigenvector centrality of 0.0024.

CONCLUSIONS

There are 7 main clusters which includes:

1. High School friends
2. University friends
3. Family
4. Squash Club
5. House Mates
6. Work colleagues
7. Friends from First STEP Mentoring Program

Md Sikander is the most well connected person as seen in the Graph-A, He has the highest Betweenness centrality and knows my other friends from 3 clusters (house mates, work colleagues and family members). As we are from the same country and went to the same university in Australia and, we shared accommodation for a while in Australia. And when we removed him in Graph-B we found that the network diameter increased to 21 from 13 and Average path length increased to 8 steps from 6 steps.

M.Waqar Zafar has the highest Eigenvector centrality (1.00) as he is very well connected in one of the biggest clusters which is of my high school friends.

Thankyou.

REFERENCES

CHERVEN, K.

Mastering Gephi network visualization

In-text: (Cherven, 2015)

Your Bibliography: Cherven, K. (2015). *Mastering Gephi network visualization*. Birmingham, U.K.: Packt Pub., pp.89-97,199-215.