NATIONAL RESEARCH UNIVERSITY

Higher School of Economics,
faculty of Computer Science.

FACEBOOK SOCIAL NETWORK ANALYSIS

Network Science Project By Fares Ghazzawi

Moscow, 2022

OUTLINE

- Data Preparation
- Basic attributes and Network summary
- Closest Random Graph model to the Real Data
- Centralities
- Node structural equivalence/similarity.
- Clique Search
- Agglomerative clustering

DATA PREPARATION

- The Graph nodes were obtained from my facebook profile using Sekenuim and chromedriver.
- From the total number of my friends on facebook I selected 106 friends the represent a good summary of my social network.
- Each node has 3 attributes (id , name , sur_name).
- my social network is divided into multiple basic categories:

  Family : which contains my close family members (mother ,father , siblings) and far family members (cousins , uncles , aunts)
  Middle School Friends (keep in mind that me and my little sister attended the same middle school for one year)
  Neighborhood friends (keep in mind that my brother also know them)
  High school friends: which was a private school called "national center of distinguished" (contains overall 225 students so most students knew each other)
  University friends.
  People i met in Moscow.

DATA REPRESENTATION

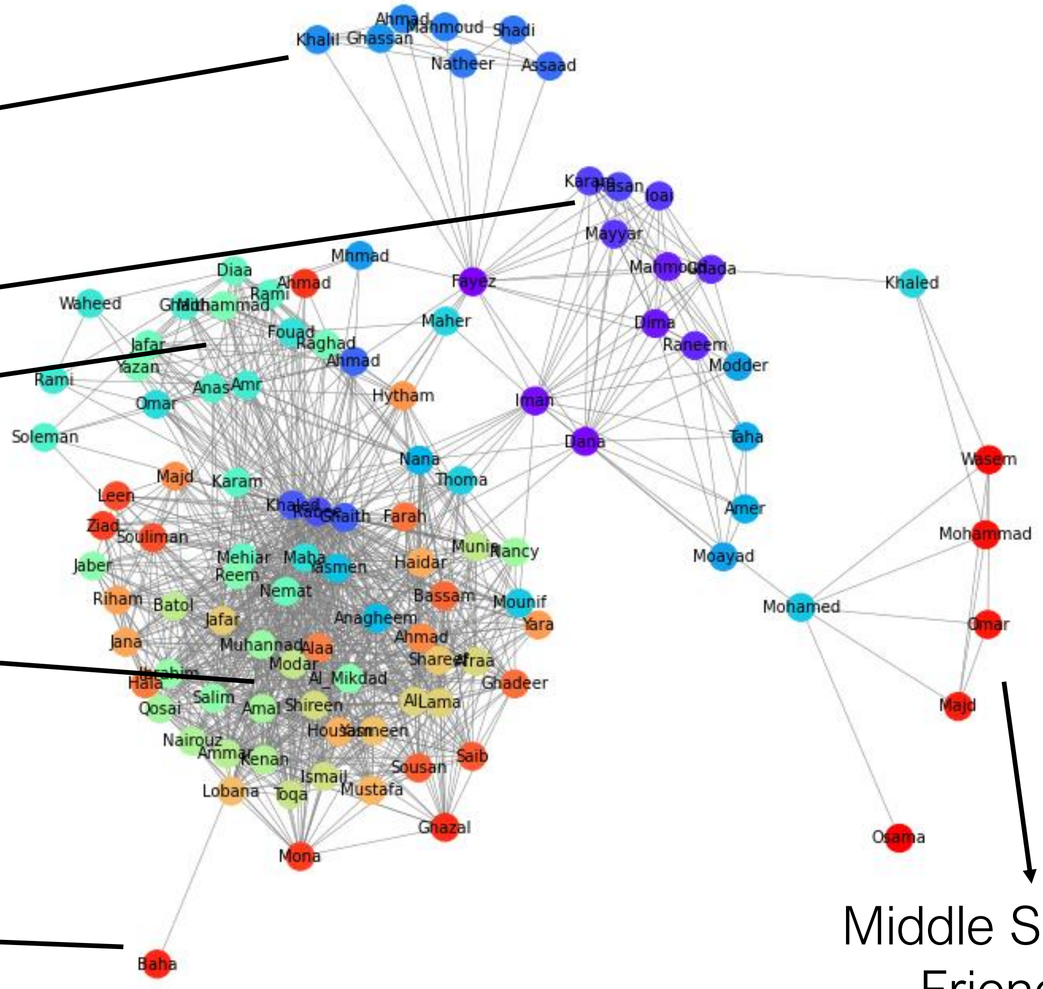
Neighborhood friends

Family

University Friends

High School Friends

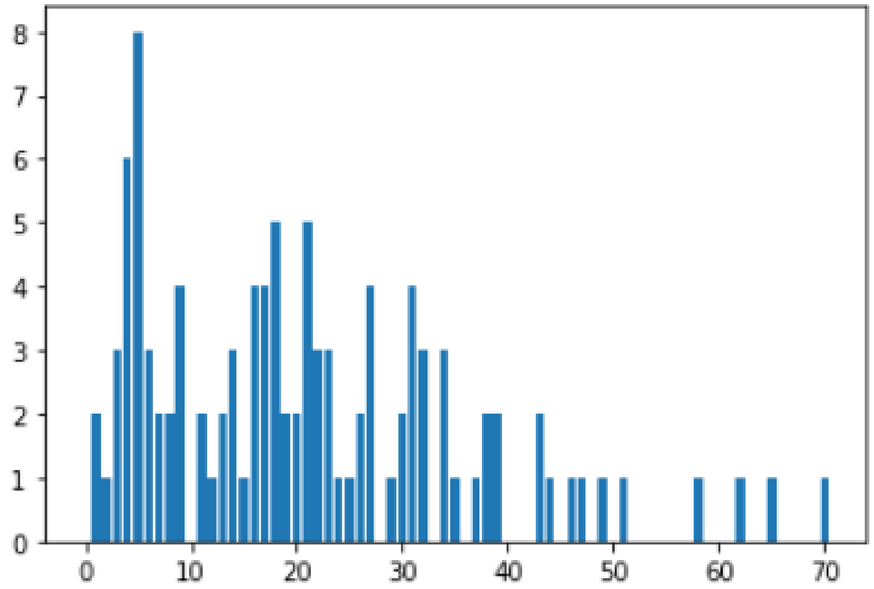
<u>People i met in Moscow</u>



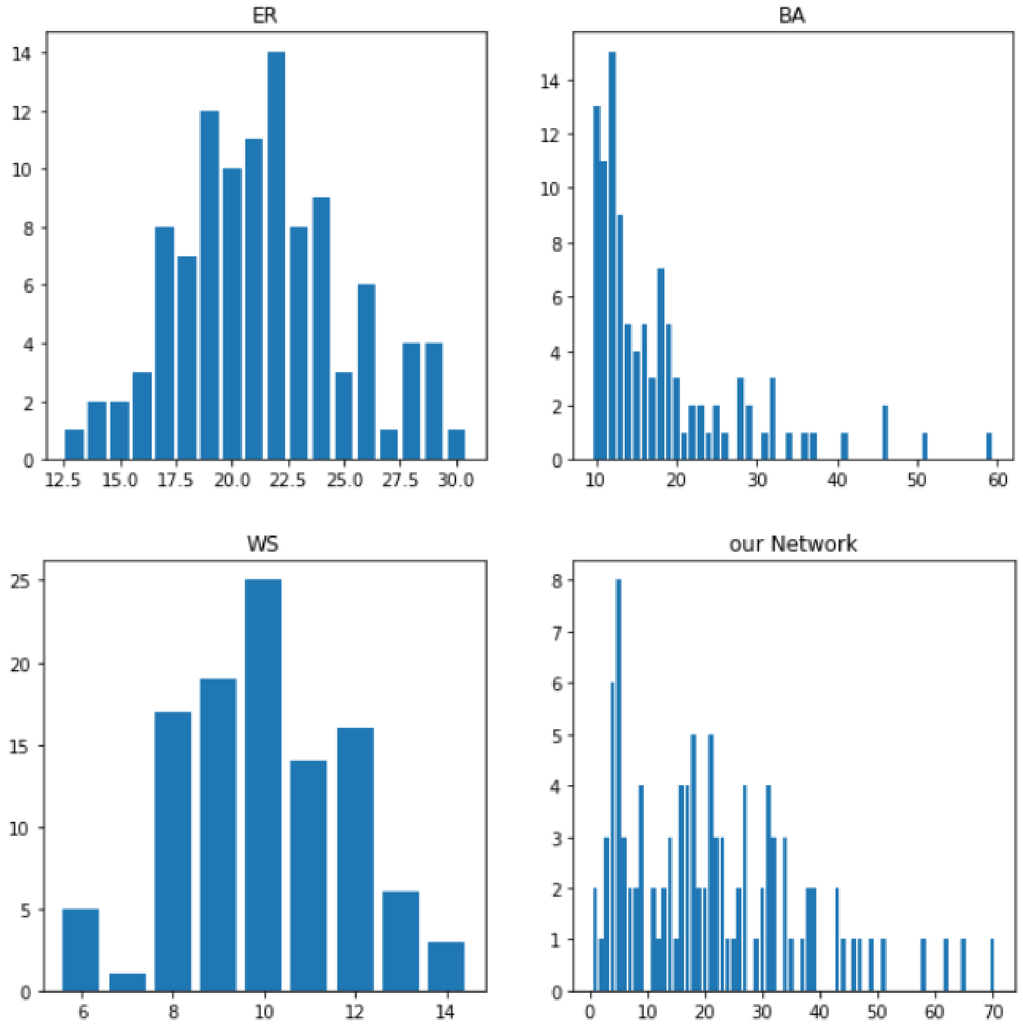
Middle School Friends

BASIC ATTRIBUTES AND NETWORK SUMMARY

Number of Nodes	106
Number of Edges	1121
Number of Connected Components	1
Diameter	5
Radius	3
Average Clustering	0.685822617296087
Global Clustering coefficient	0.5862058993637941



CLOSEST RANDOM GRAPH MODEL TO THE REAL DATA



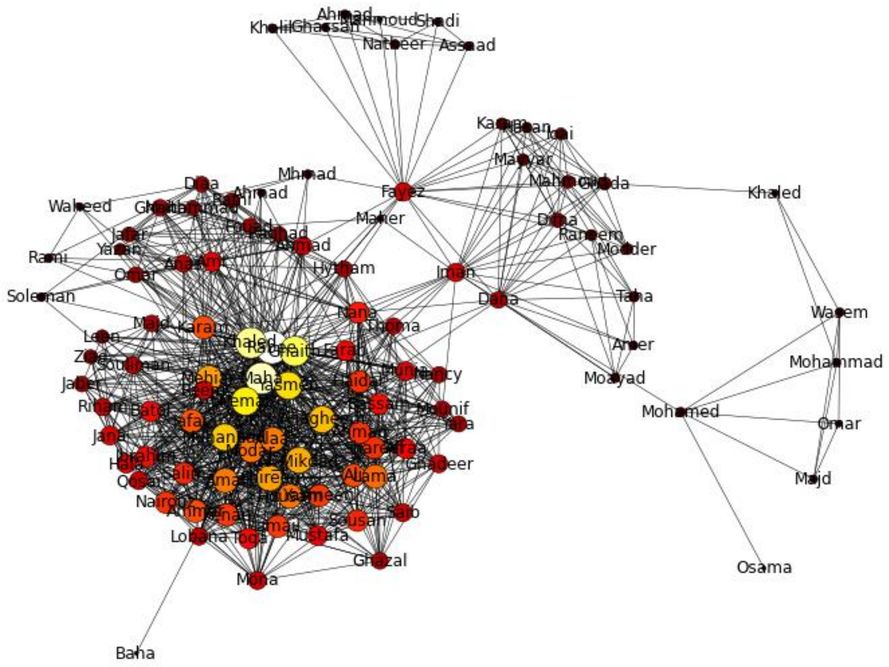
	My Network	ER	BA	WS
Num of Edges	1121	1132	960	530
Diameter	5	3	3	4
Radius	3	2	2	3
Average shortest Path	2.273315363881401 7	1.805031446540880 5	1.877448337825696 3	2.296855345911949 6
Average Clustering	0.685822617296087	0.194023207110519	0.268682395350505	0.139201992975577
Transitivity	0.586205899363794	0.194656006732589	0.252867619405791	0.133993532740501

- we have found that BA with m = 10 is the closest to our network.
- BA gives as power law-like distribution which is very good.
- The closest value for Global clustering cofficients from the above models is the one for BA model.
- BA model also got relativly close Average shortest path value to our network

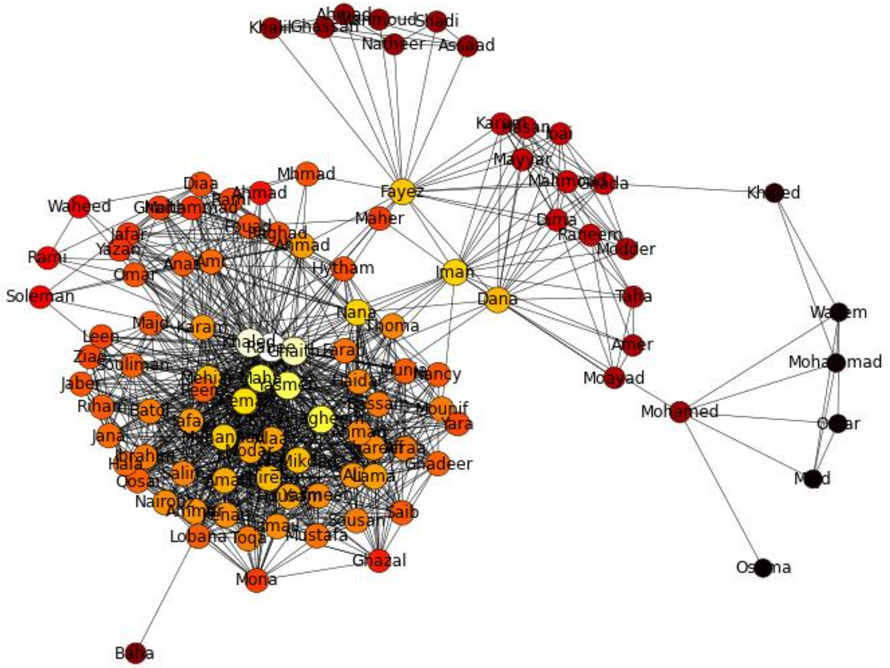
CENTRALITIES

Top 5 Degree Centrality	Top 5 Closeness Centrality	Top 5 betweenness Centrality	Top 5 Eigenvector Centrality
Rabee_AlFazza	Rabee_AlFazza	Fayez_Ghazzawi	Rabee_AlFazza
Maha_Quarqout	Khaled_Ali_Deeb	Dana_Ghazzawi	Maha_Quarqout
Khaled_Ali_Deeb	Ghaith_Abtah	Khaled_Ali_deeb	Khaled_Ali_Deeb
Ghaith_Abtah	Yasmen_Al_shelle	Rabee_AlFazza	Ghaith_Abtah
Nemat_Aloush	Anagheem_Ibrahim	Iman_Mardini	Nemat_Aloush
Those are my friends who studied with me in both high school and in university, thus they know people from both categories and thus they have a bigger degree centrality	Those are my friends who are more socially active , thus they know a lot of people making distance between them and any other friend is relatively low thus higher closeness centrality value	My brother got the highest value since he connects my neighborhood friends with the rest of the Graph (lots of shortest paths going through the node). my sister is second because she connects my middle school friends with the rest of the graph	those friends are important in the graph since they are connected to each other and connected to multiple other nodes (based on the concept that eigenvector centrality presents which is "the importance of a node depends on the importance of its neighbors")

CENTRALITIES

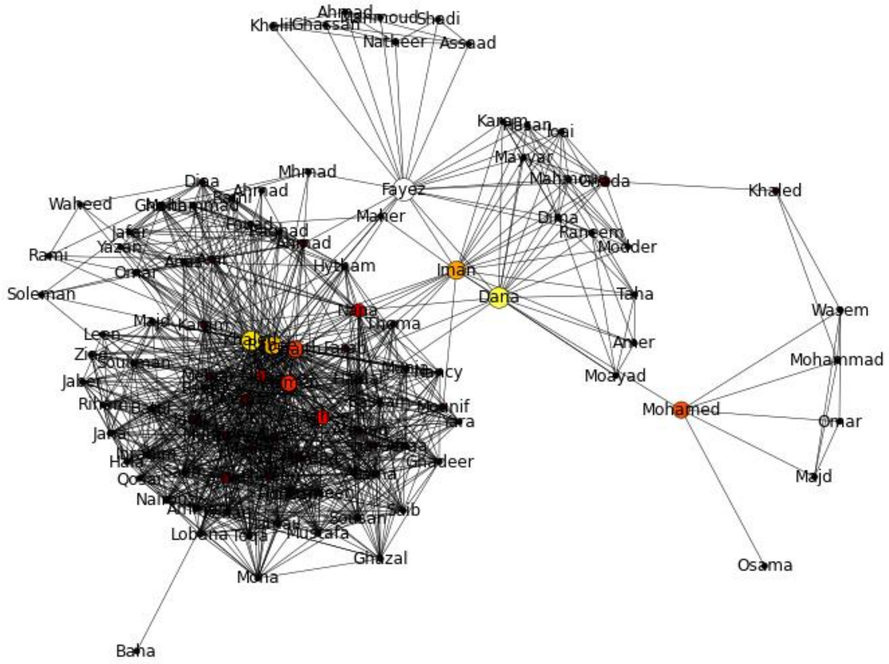


Degree

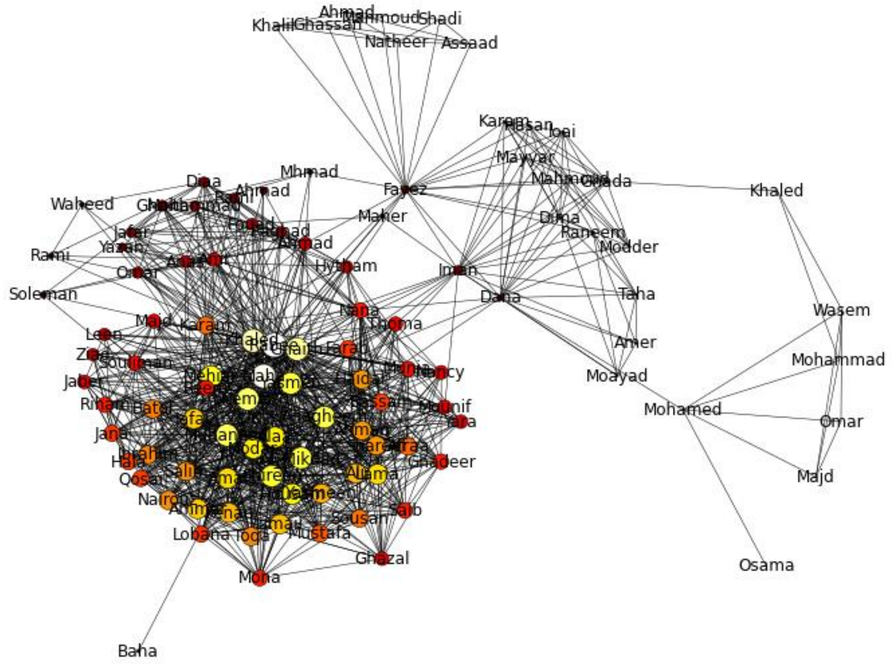


Closeness

CENTRALITIES

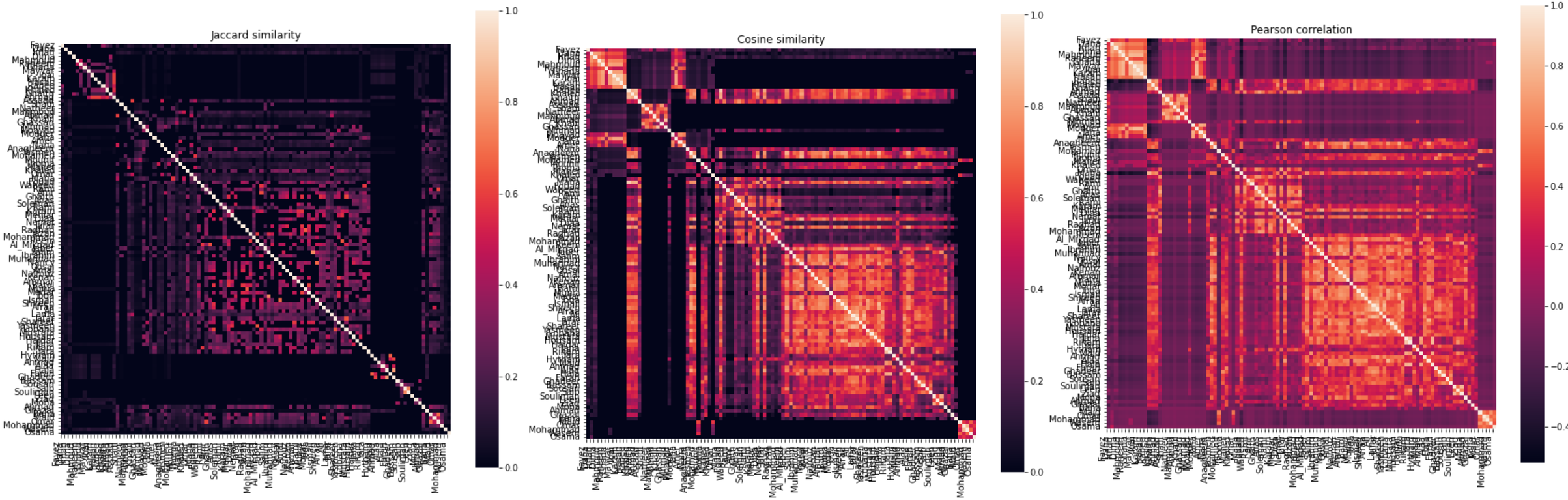


betweenness



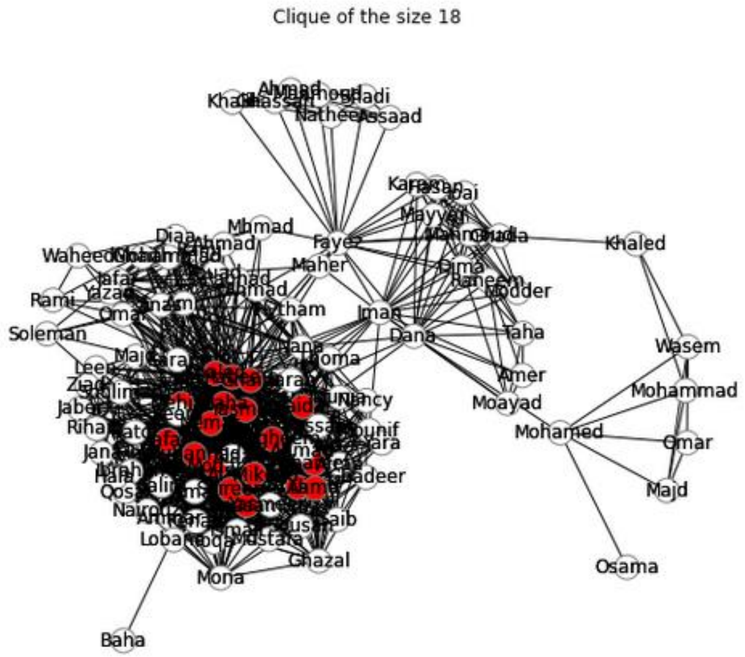
Eigenvalue

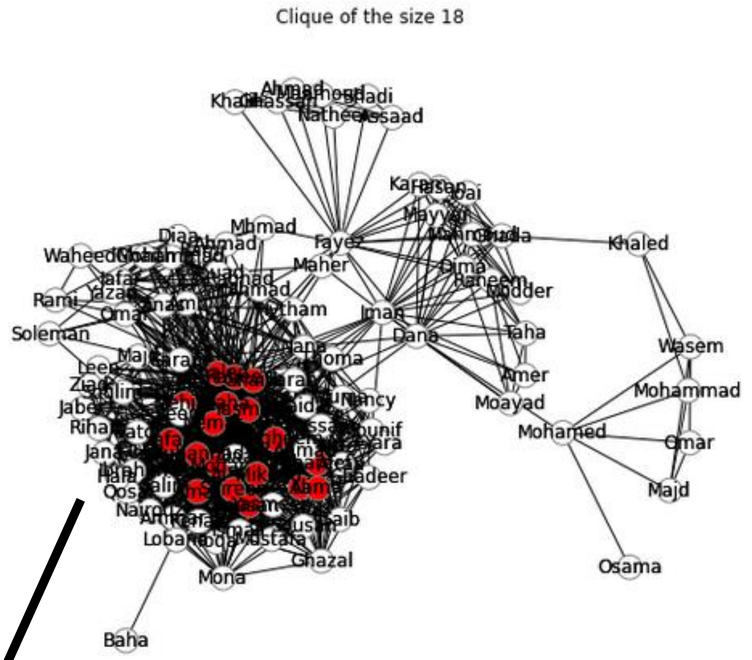
NODE STRUCTURAL EQUIVALENCE/SIMILARITY



CLIQUE SEARCH

- **We have 2 cases of largest Complete subgraph**
- **The largest complete subgraph in both cases contains my colleagues from high school(since as I explained there was a really small number of students and mostly everyone knew each other)**



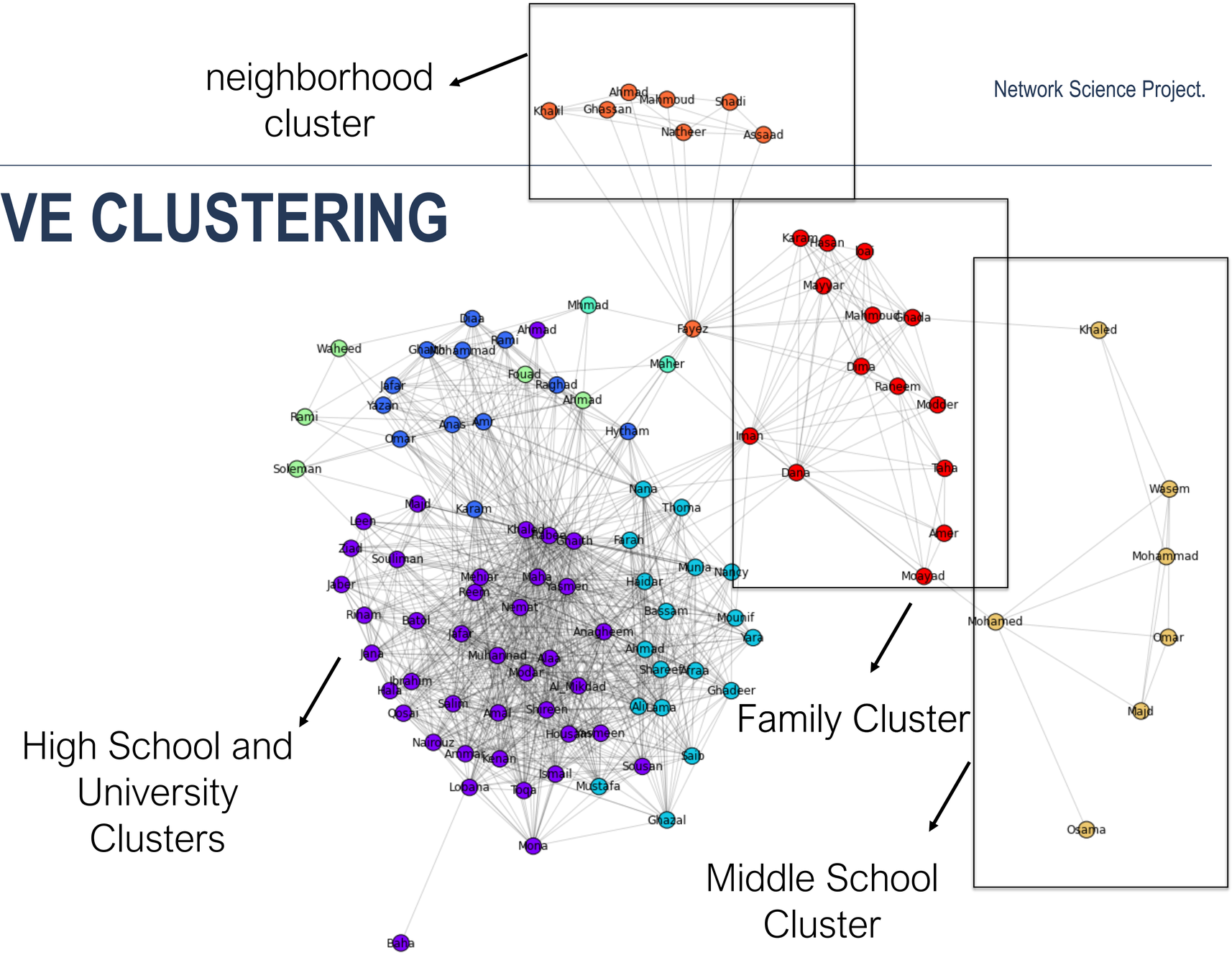


contains these nodes ['Rabee' 'Khaled' 'Ghaith' 'Anagheem' 'Yasmen' 'Maha' 'Mehiar' 'Nemat' 'Al_Mikdad' 'Muhannad' 'Modar' 'Shireen' 'Ali' 'Lama' 'Jafar' 'Shareef' 'Housam' 'Haidar']

contains these nodes ['Rabee' 'Khaled' 'Ghaith' 'Anagheem' 'Yasmen' 'Maha' 'Mehiar' 'Nemat' 'Al_Mikdad' 'Muhannad' 'Amal' 'Modar' 'Shireen' 'Ali' 'Lama' 'Jafar' 'Shareef' 'Housam']

AGGLOMERATIVE CLUSTERING

-family cluster is almost the same (missing fayez node).
- neighborhood cluster is the same as the defined one
- middle school cluster is the same as the defined one
- university cluster has been divided into two clusters (that should have been one) and finally, high school cluster has been also divided into two clusters(that can be explained that the cluster with dark purple where either my colleagues in the same year or younger and the one with the sky color is the cluster with my elder colleagues)



NATIONAL RESEARCH
UNIVERSITY