Fundamentals of Database Management System

End Term Project

Network Analysis of American Football

Submitted to:

Prof. Ashok Harnal

FORE School of Management, New Delhi

Submitted by

Group No: 2

BDA (02) Section- G

Lovy Dhankhar-025019

Devender-025048

Gaurav Sharma- 025051

American Football

Introduction: During the session fall (2000), Football matches were played between American football teams. Here is the network of football game division (IA) where nodes represent individual American Football Teams and edges represent the respective number of games played between them. In this network analysis, communities are formulated and its analyses is carried out to gain insight on the football team division (IA).

Objective: Creating the network and analysing on the basis of

- Important entities from different points of view.
- Communities existing within the network.
- Community characteristics for differentiation with other communities.
- Seeking insight on relationship of nodes within and outside communities.

Network Overview:

This network contains "American football games between Division IA colleges during regular season Fall 2000."

The vertices and edges represent the colleges and games between teams, respectively. The communities in this network represent the conferences to which they belong.

Name	American Football
Data Source	https://www.cc.gatech.edu/dimacs10/archive/clustering.shtml
Category	Interaction Network
Node	Teams
Edge	Games
Format	Undirected
Edge Type	Unweighted, No multiple edges
Multiplicity	No multiple edges

Statistics:

Size	$n =$	115
Volume	$m =$	613
Loop count	$l =$	0
Wedge count	$s =$	5,967
Claw count	$z =$	17,513
Cross count	$x =$	34,475
Triangle count	$t =$	810
Square count	$q =$	3,915
4-Tour count	$T4 =$	56,414
Maximum degree	$dmax =$	12
Average degree	$d =$	10.660 9
Fill	$p =$	0.093 516 4
Size of LCC	$N =$	115
Diameter	$\delta =$	4
50-Percentile effective diameter	$\delta 0.5 =$	2.004 48
90-Percentile effective diameter	$\delta 0.9 =$	2.828 14
Median distance	$\delta M =$	3
Mean distance	$\delta m =$	2.396 81
Gini coefficient	$G =$	0.039 960 3
Balanced inequality ratio	$P =$	0.478 793
Relative edge distribution entropy	$Her =$	0.999 244
Power law exponent	$\gamma =$	3.398 68
Tail power law exponent	$\gamma t =$	8.991 00
Tail power law exponent with p	$\gamma 3 =$	8.991 00
p-valuep =	0.000 00	
Degree assortativity	$\rho =$	+0.162 442
Degree assortativity p-value	$p\rho =$	$1.065\ 06 \times 10{-}8$
Clustering coefficient	$c =$	0.407 240

Spectral norm α = 10.780 6

Algebraic connectivity a = 1.459 00

Spectral separation $|\lambda 1[A] / \lambda 2[A]|$ = 1.161 89

Non-bipartivity bA = 0.579 913

Normalized non-bipartivity bN = 0.557 894

Algebraic non-bipartivity χ = 5.077 40

Spectral bipartite frustration bK = 0.119 066

Controllability C = 3

Relative controllability Cr = 0.026 087 0

Software's Used

- Gephi
- Ms Word
- Adobe Acrobat Reader
- GitHub

Analysis

Pipeline:

- Importing the dataset on Gephi
- Processing the graphs for sensible and meaningful visuals
- Gaining insight from the graphs using Gephi visualization features i.e., statistics, filters, graph layout etc.
- Obtaining the concluded graph based on queries
- Previewing the output graph
- Making report of the entire process with visuals.

Getting started with the dataset

- Opening the dataset in Gephi

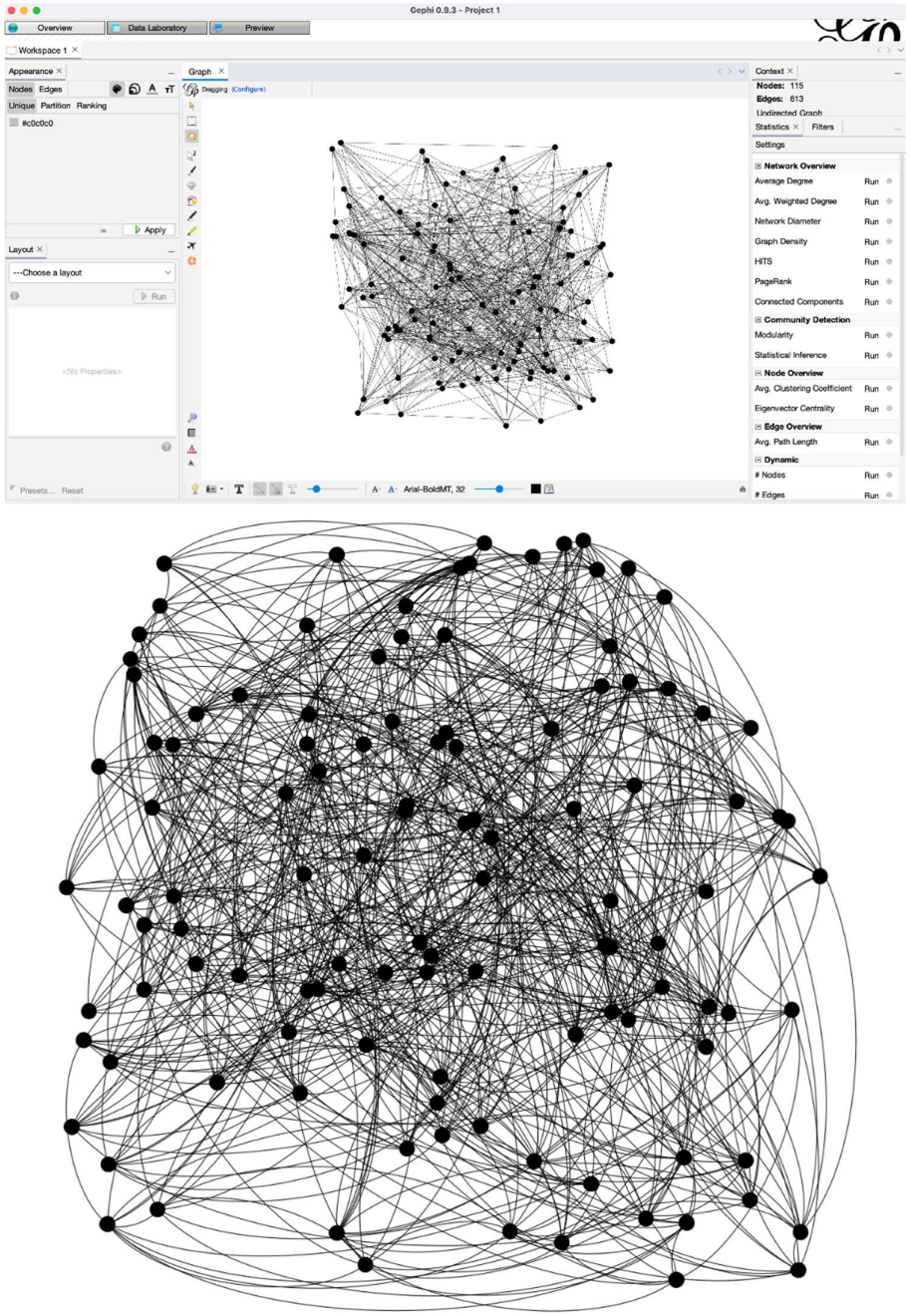


Figure 1: Initial plot after opening the dataset in Gephi

Graph Processing:

Fruchterman Reingold Layout

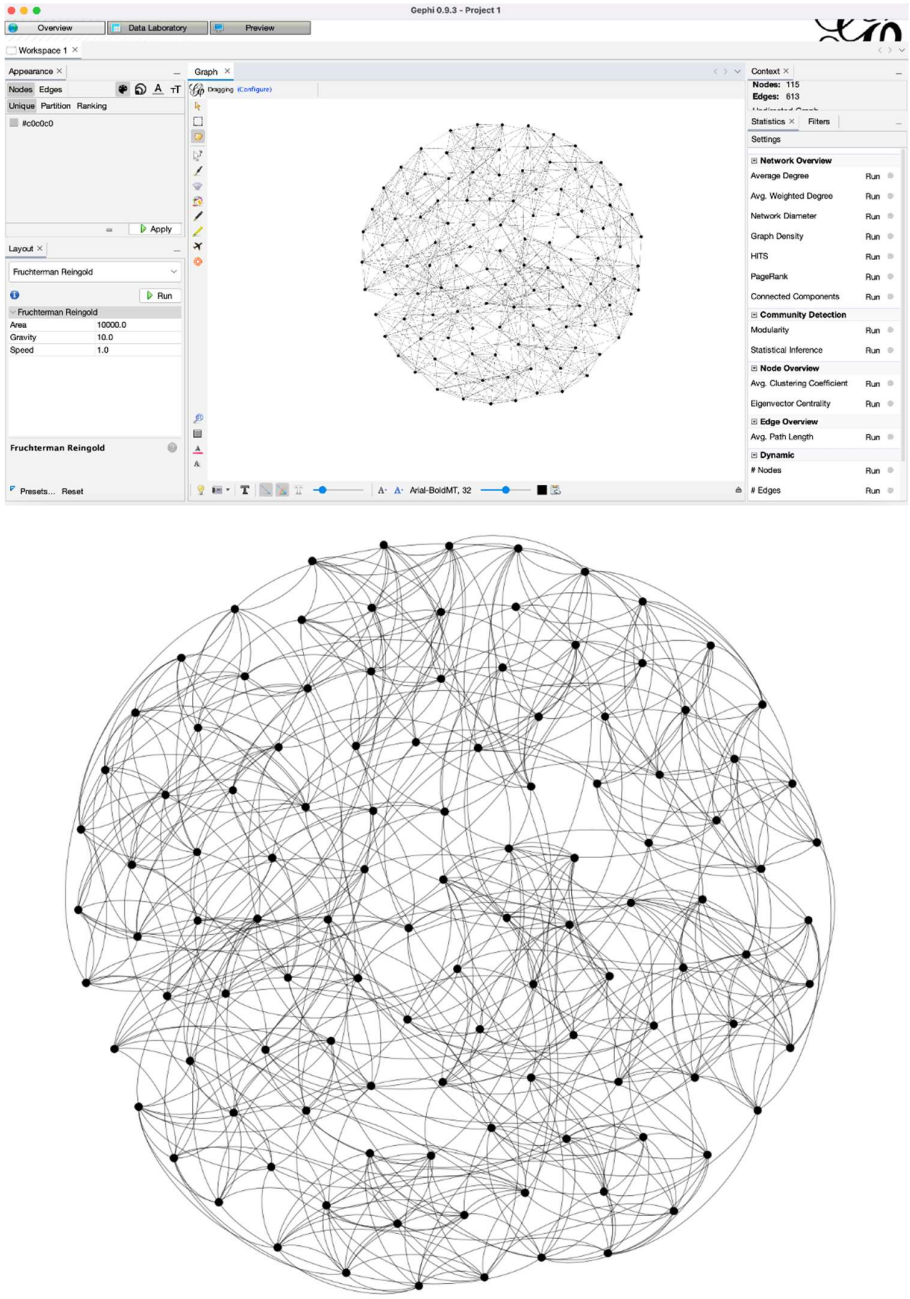
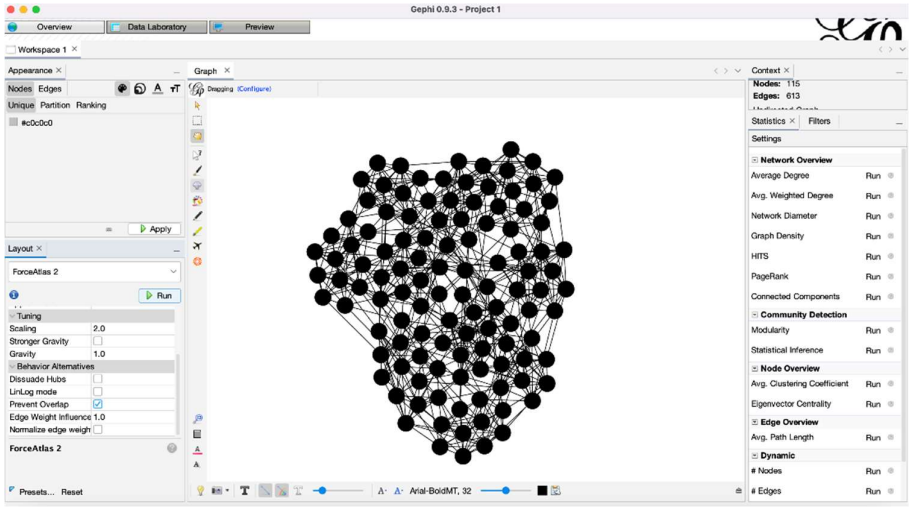


Figure 2: Graph in Fruchterman Reingold Layout

Force Atlas 2 Layout



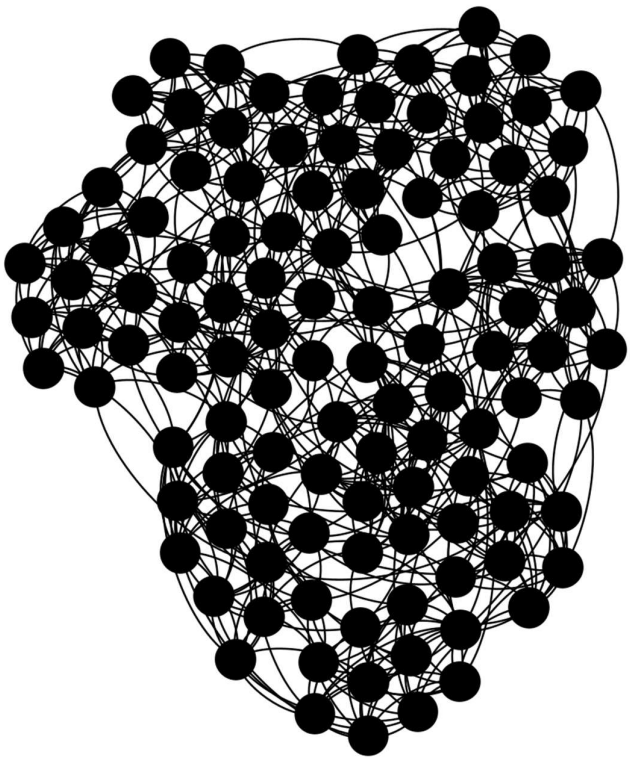


Figure 3: Graph in Force Atlas 2 Layout

Partitioning the nodes with respect different node colour on the basis of node values

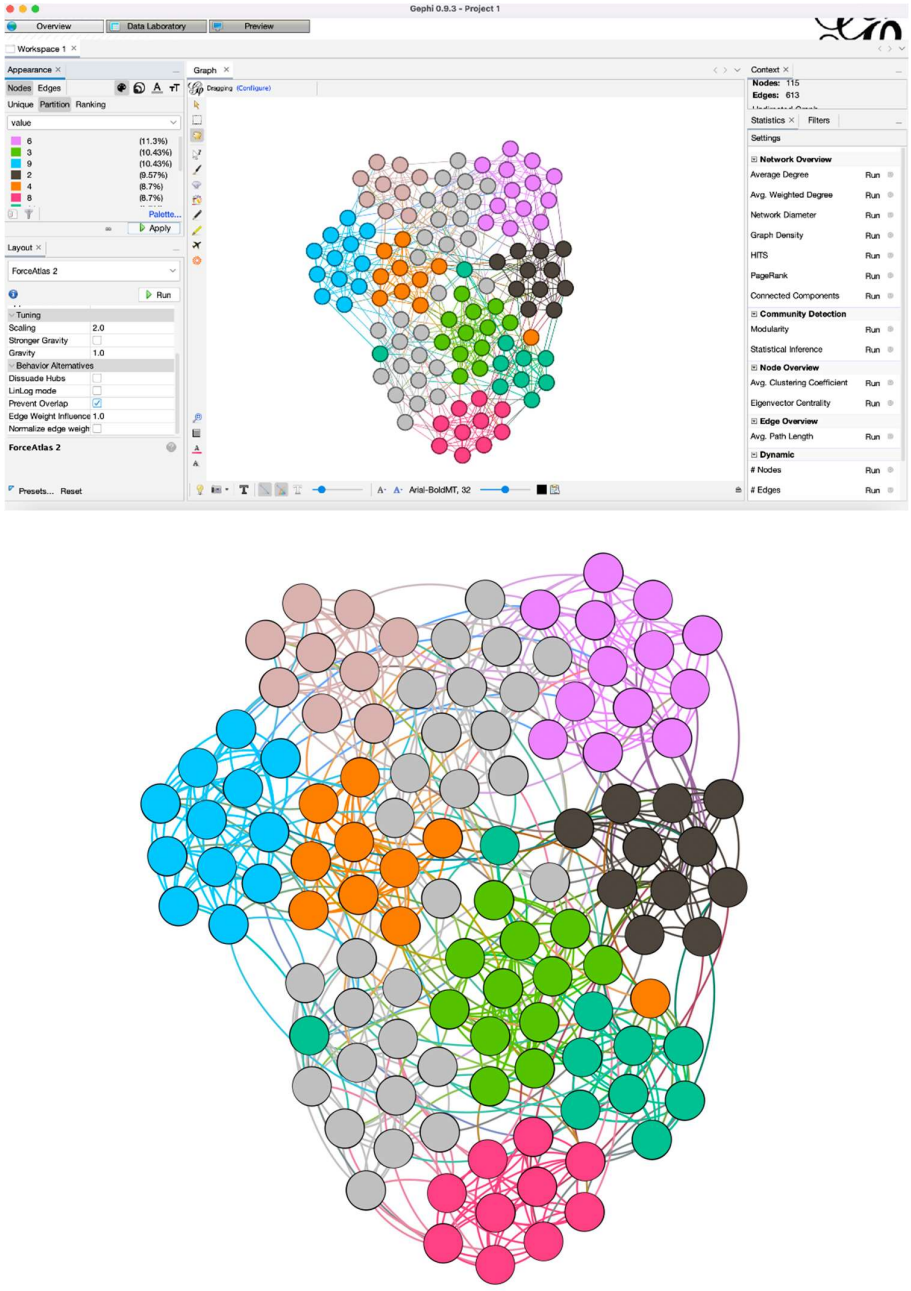


Figure 4: Better visual for different communities (colour differentiation based on node values)

- **Ranking the nodes with respect different node size on the basis of degree**

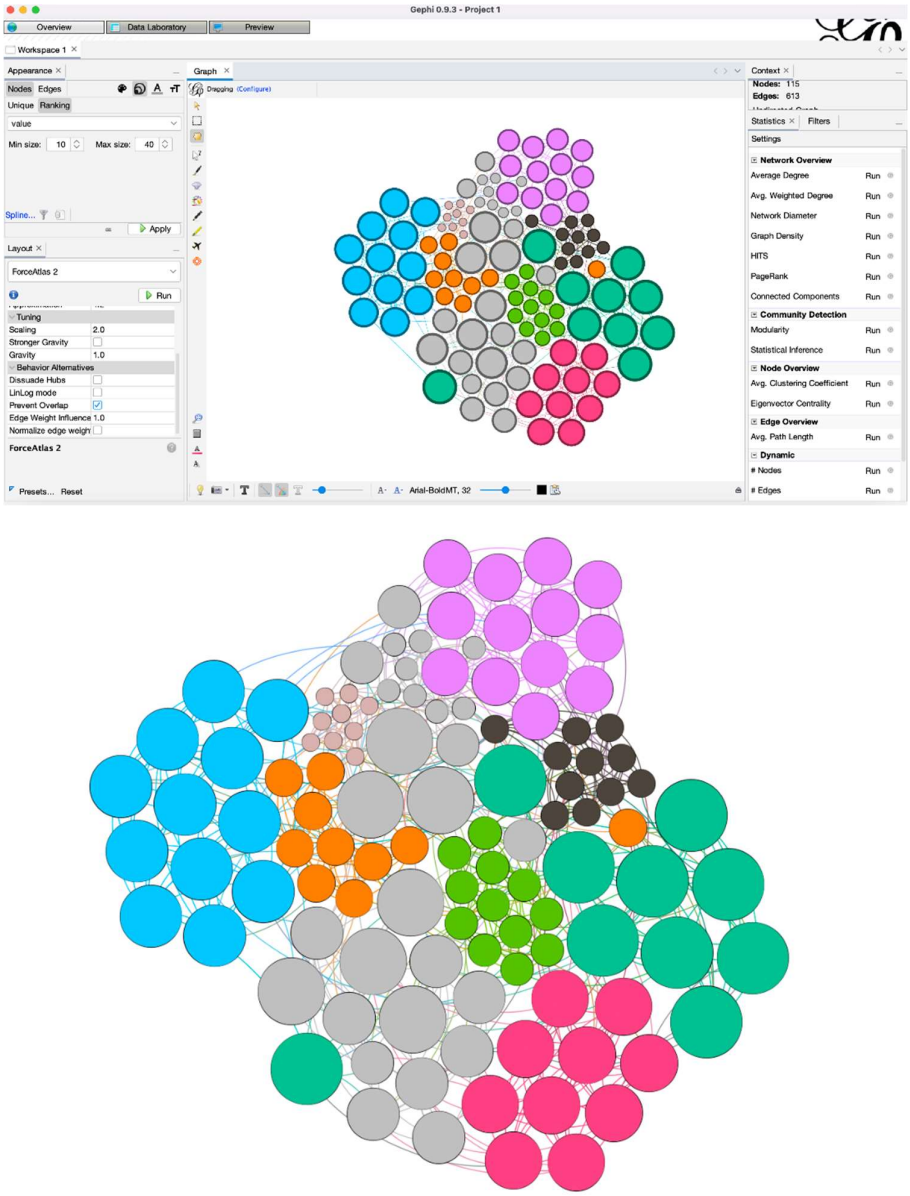


Figure 5: Better visual for different communities (size differentiation based on degree)

- **Ranking the nodes with respect different node size on the basis of value**

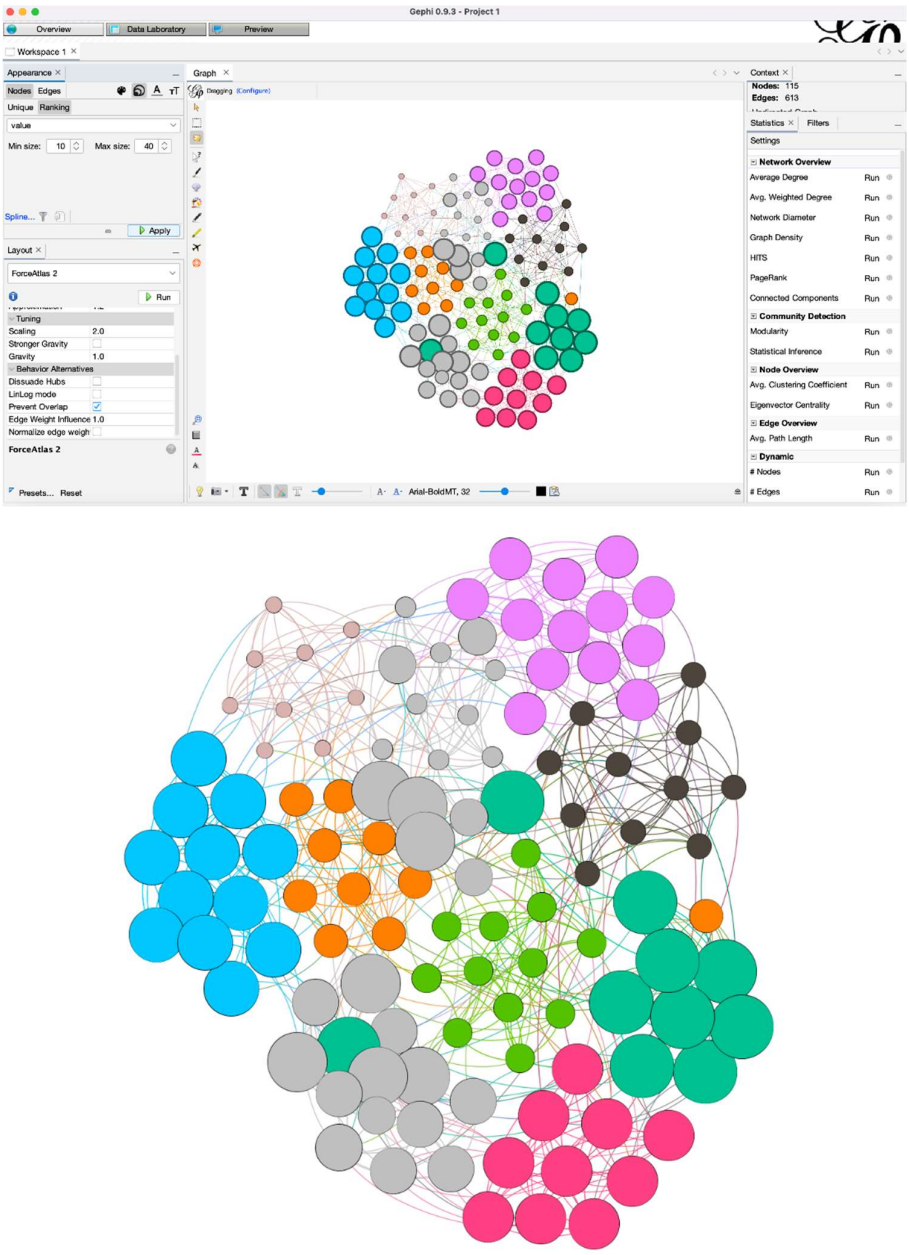


Figure 6: Better visual for different communities (size differentiation based on based on node value)

- **Adding labels to the network graph**

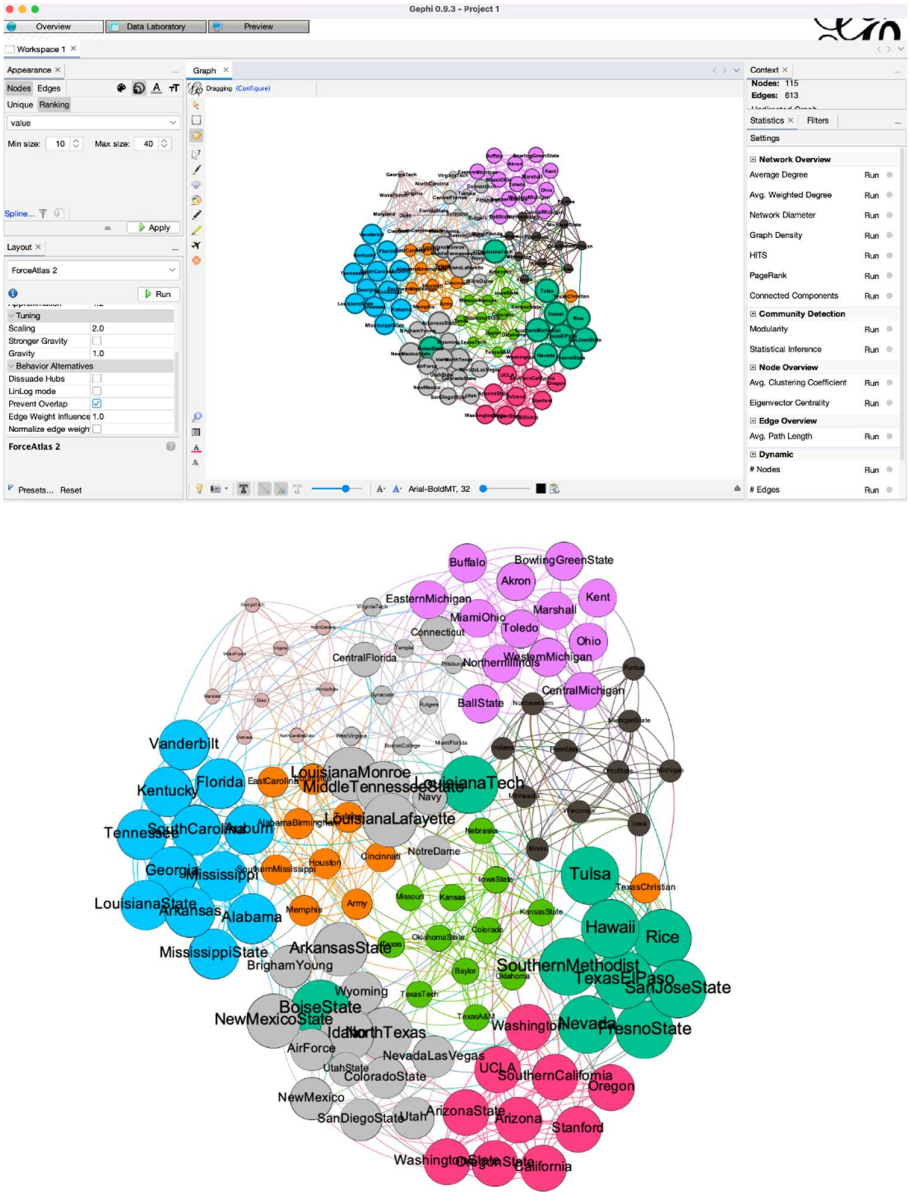


Figure 7: Final graph with labels

Multimode Network for projection: Not required

1. **Communities Detection**
    - Modularity: Obtaining modularity from statistics

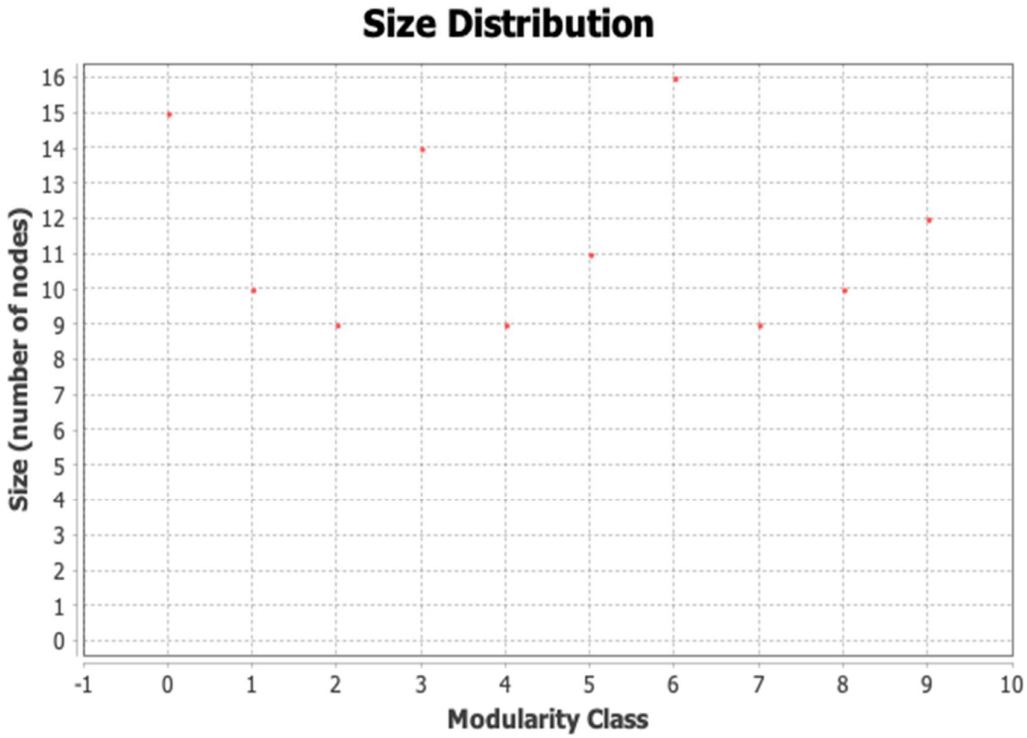
Modularity Report

Parameters:

- Randomize: On
- Use edge weights: On
- Resolution: 1.0

Results:

- Modularity: 0.605
- Modularity with resolution: 0.605
- Number of Communities: 10



Algorithm: Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Resolution: R. Lambiotte, J.-C. Delvenne, M. Barahona Laplacian Dynamics and Multiscale Modular Structure in Networks 2009

Differentiating Communities based on Modularity Ranking

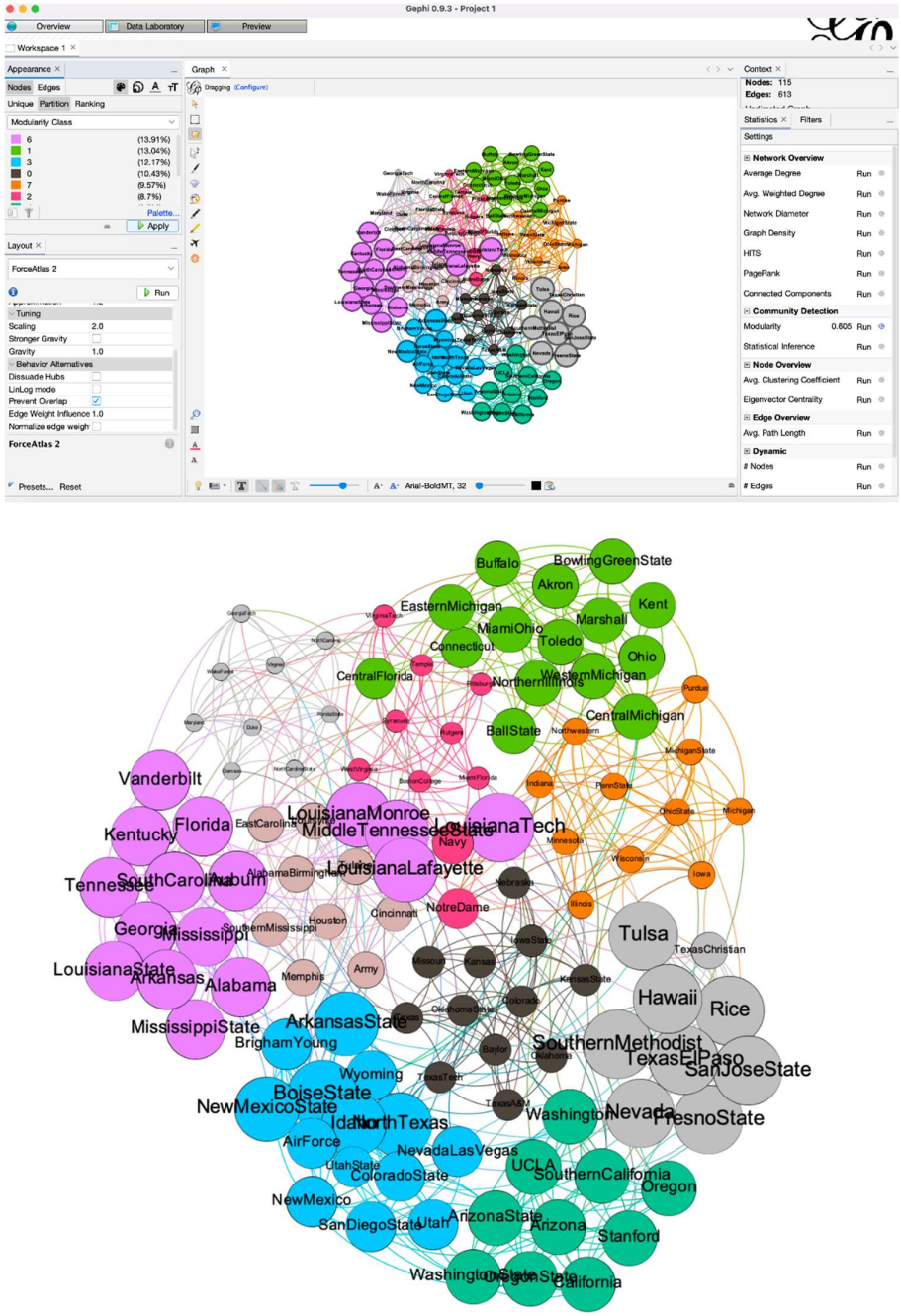


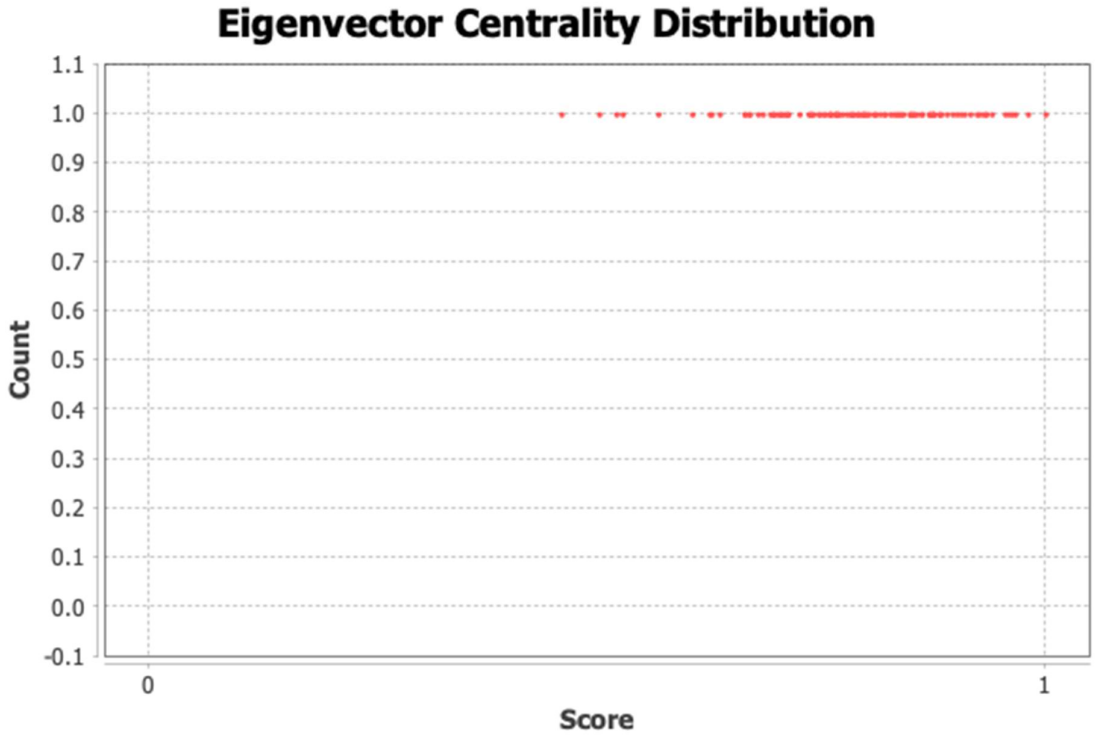
Figure 8: Different communities based on Modularity

Eigenvector Centrality Report

Parameters:

- Network Interpretation: Undirected
- Number of iterations: 100
- Sum change: 0.01573457441361714

Results:



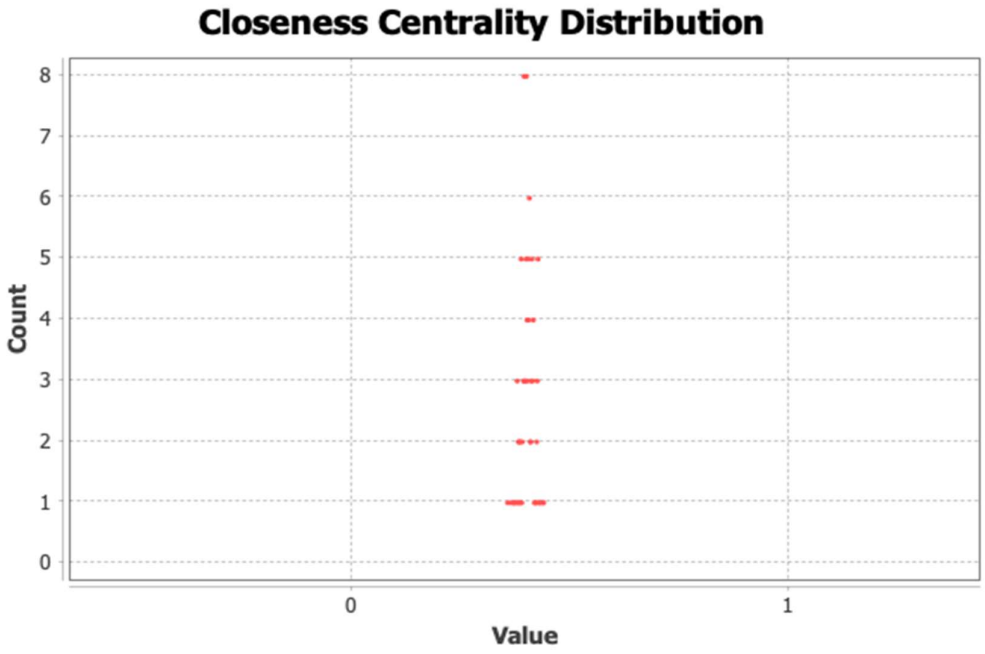
Graph Distance Report

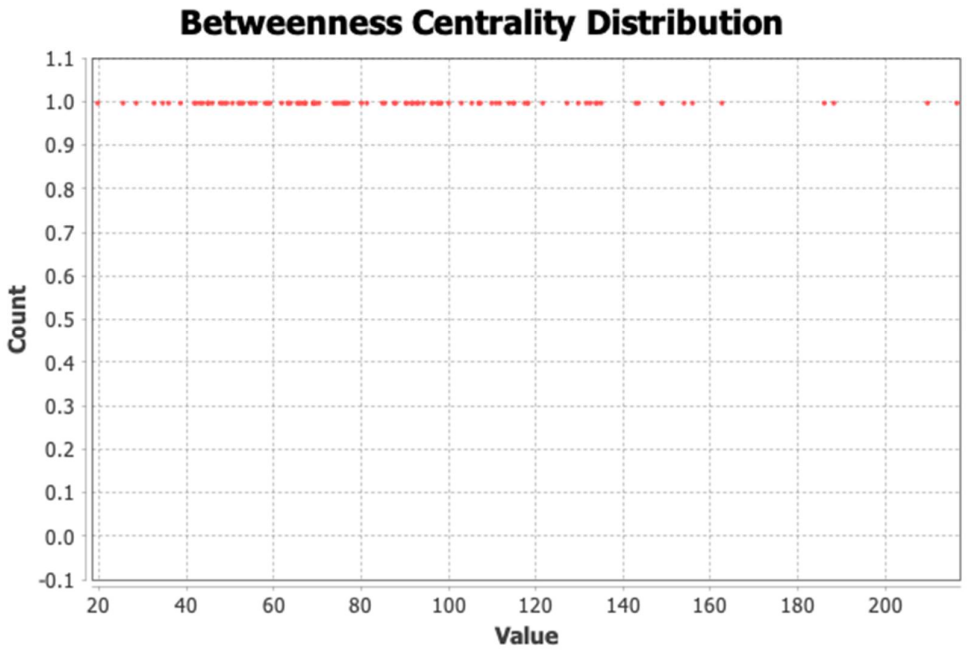
Parameters:

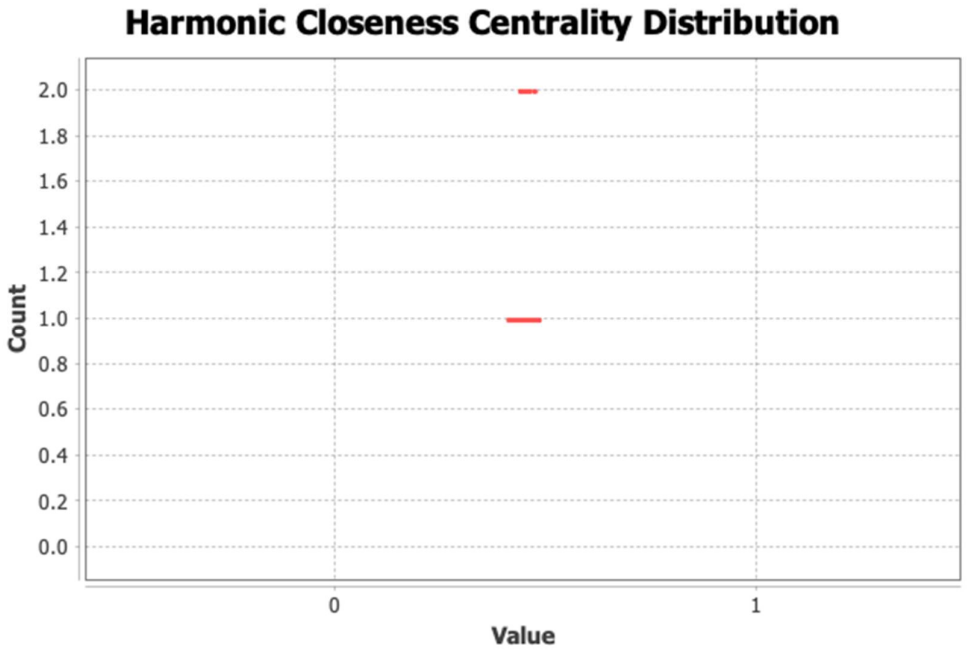
- Network Interpretation: undirected

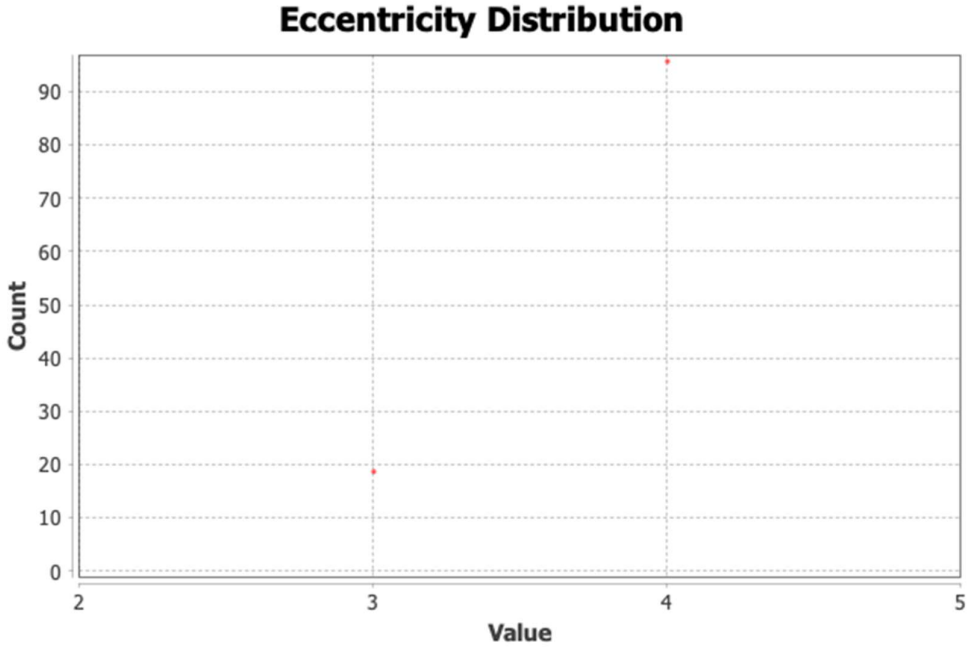
Results:

- Diameter- 4
- Radius- 3
- Average Path length - 2.5081617086193746









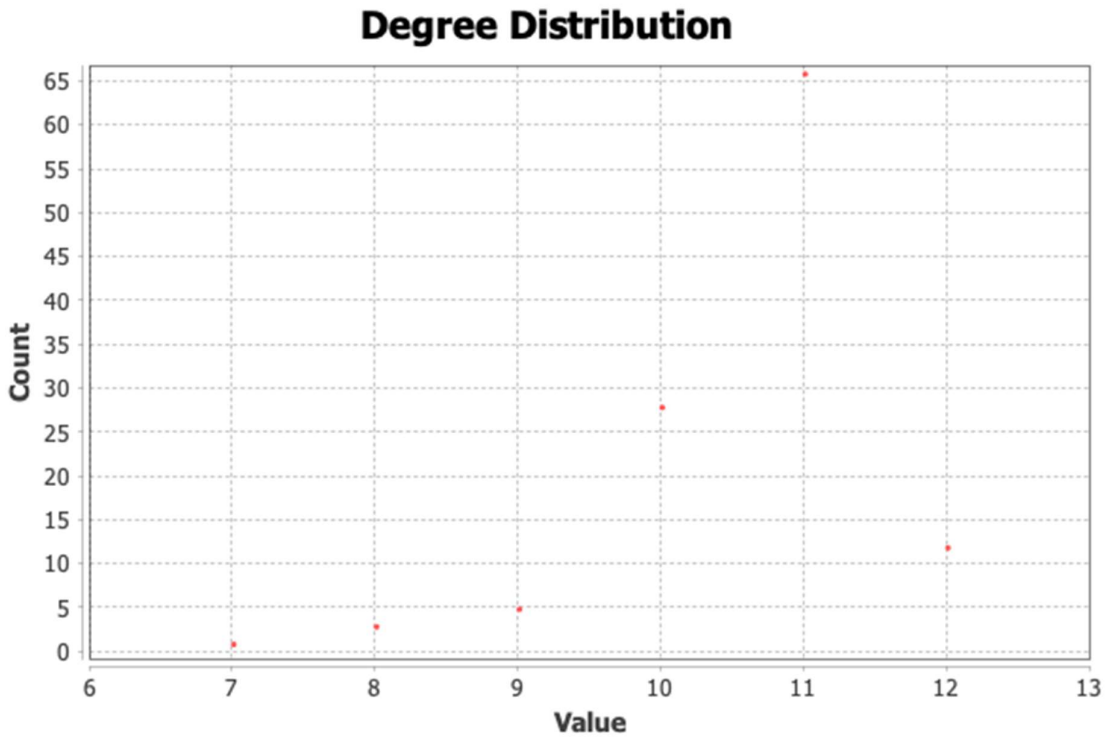
Algorithm:

- Ulrik Brandes, *A Faster Algorithm for Betweenness Centrality*, in Journal of Mathematical Sociology 25(2):163-177, (2001)

Weighted Degree Report

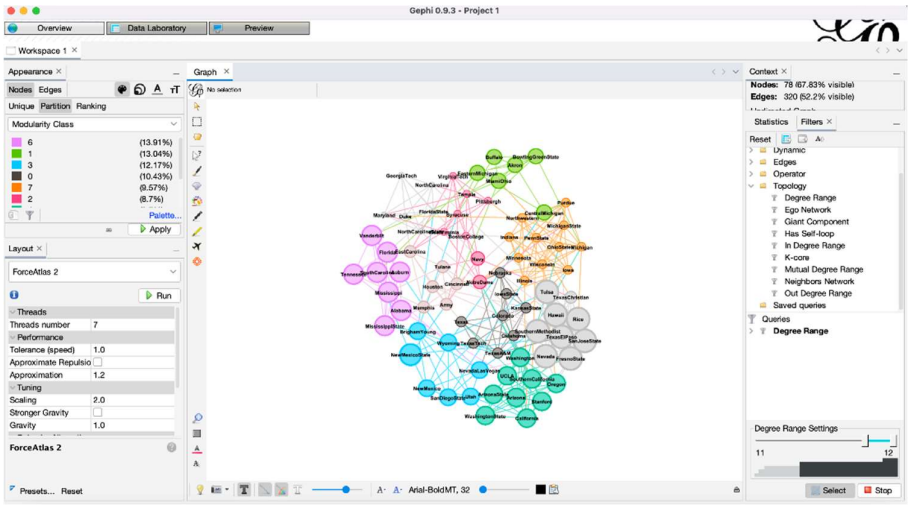
Results:

Average Weighted Degree: 10.661



Identifying important entities:

- **On the basis of degree range**



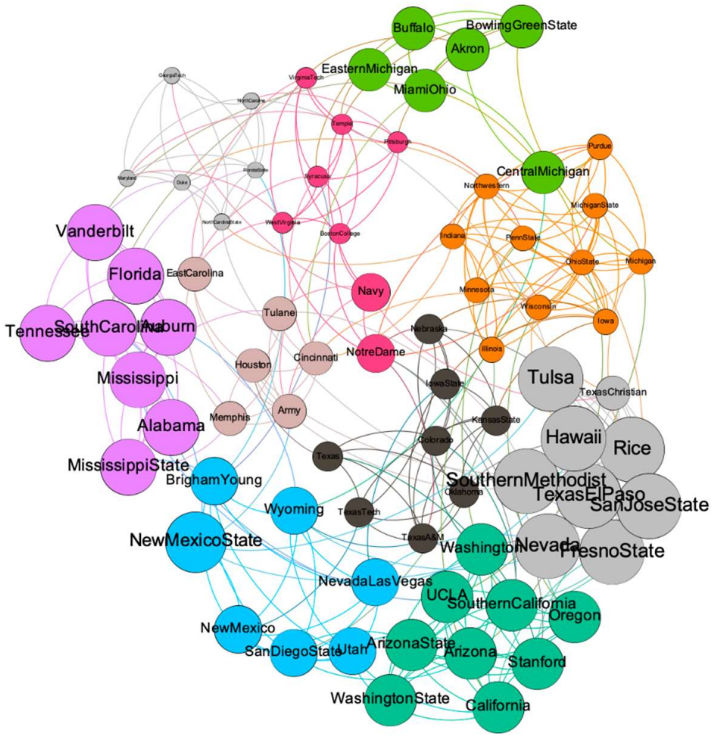


Figure 9: Important Entities based on their Degree Range

Betweenness Centrality

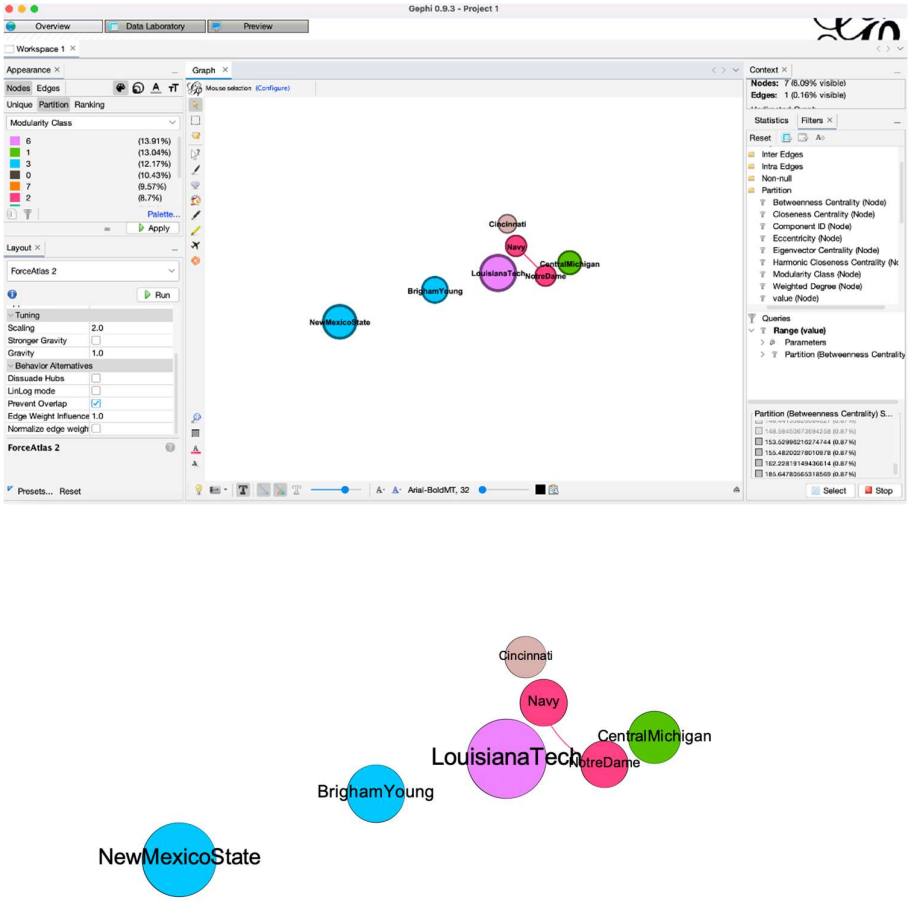


Figure 10: Important entities based on betweenness centrality

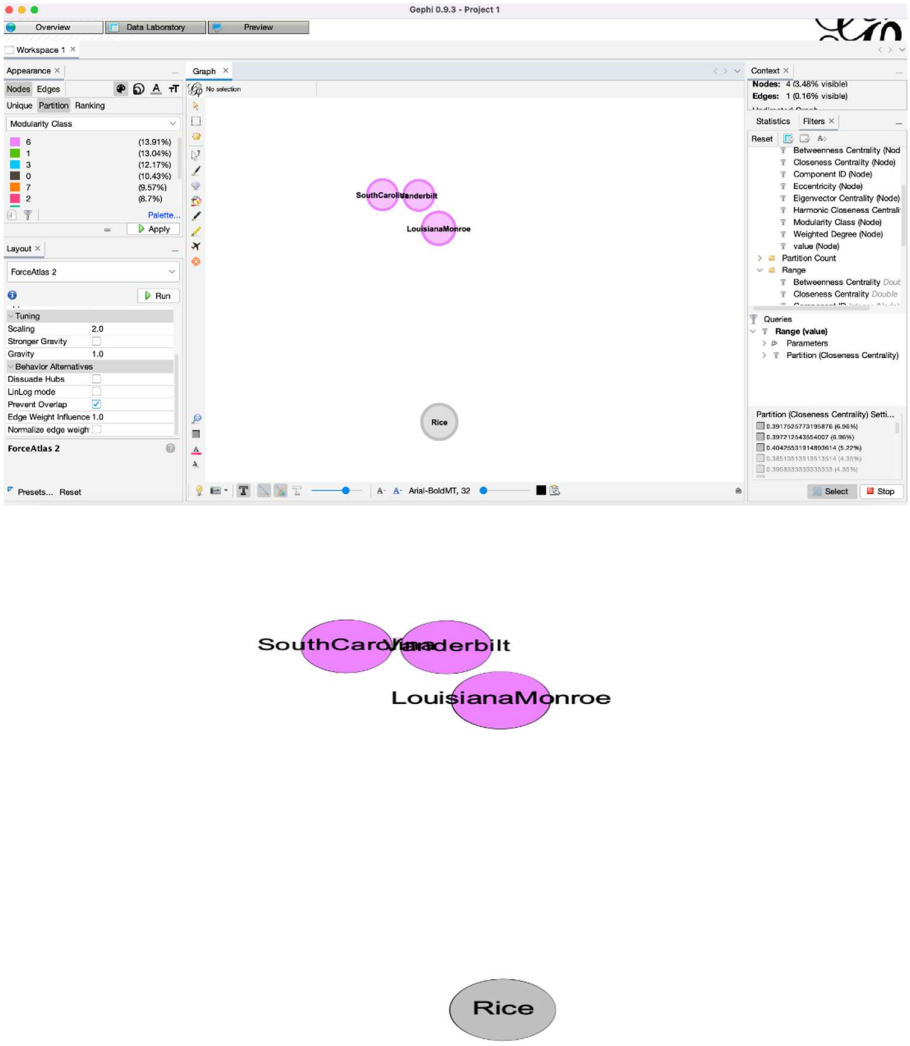
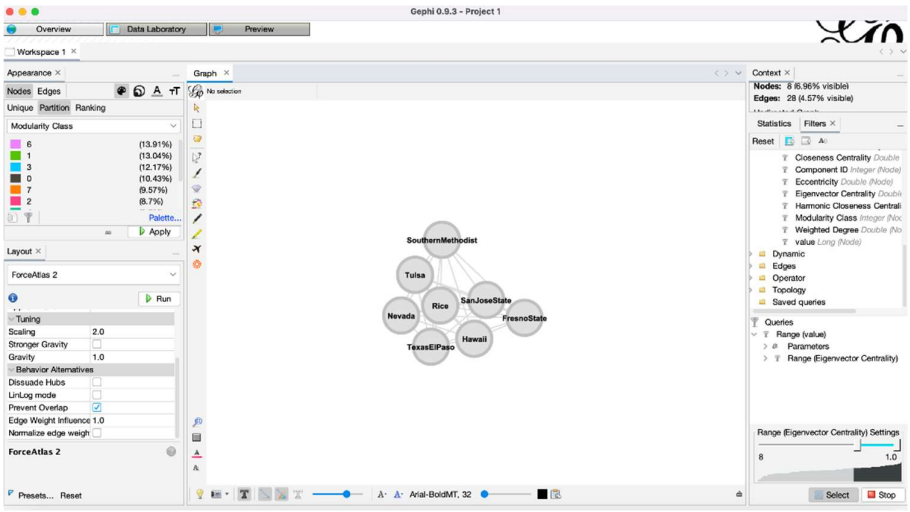


Figure 11: Important Entities based on Closeness Centrality

Eigenvector centrality



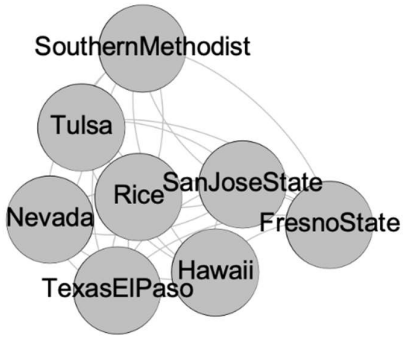


Figure 12: Important Entities based on Eigenvector Centrality

Weighted Degree

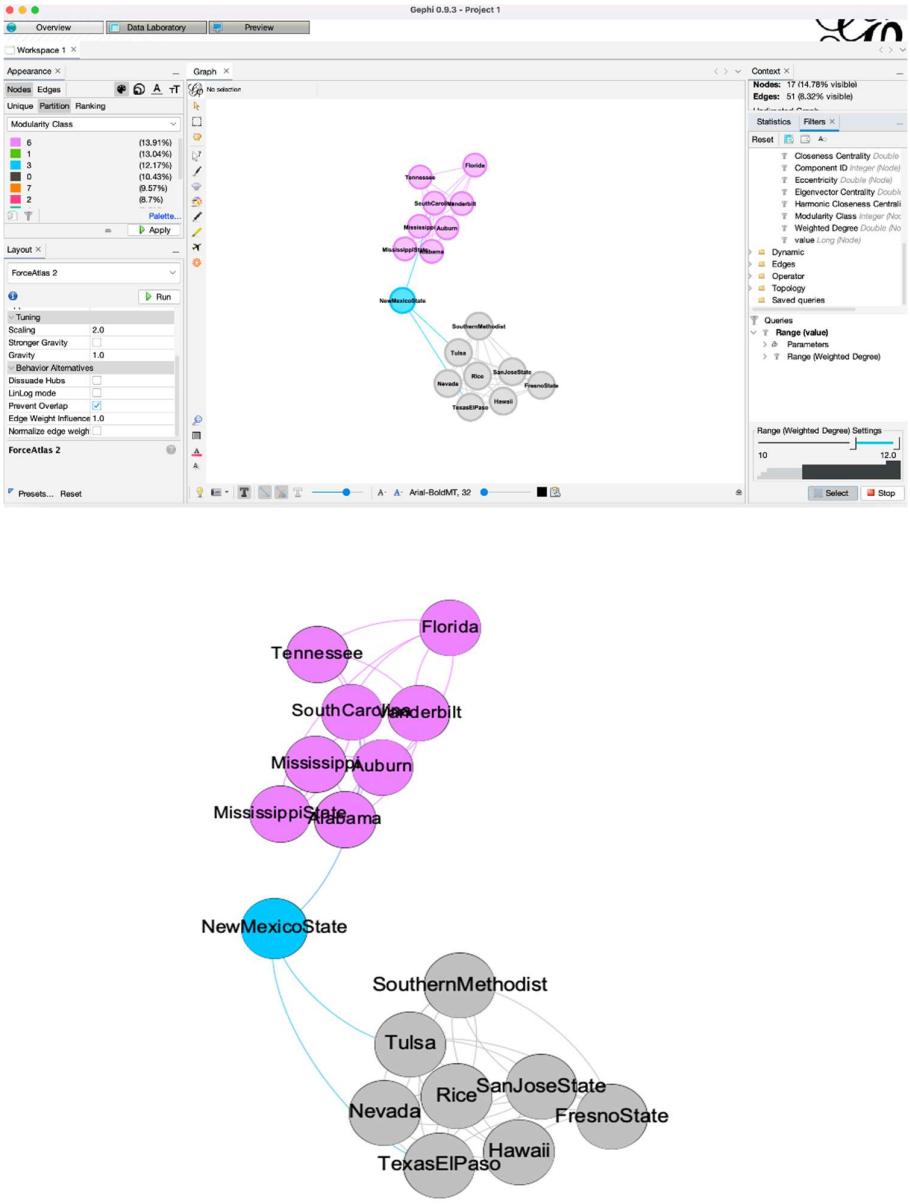
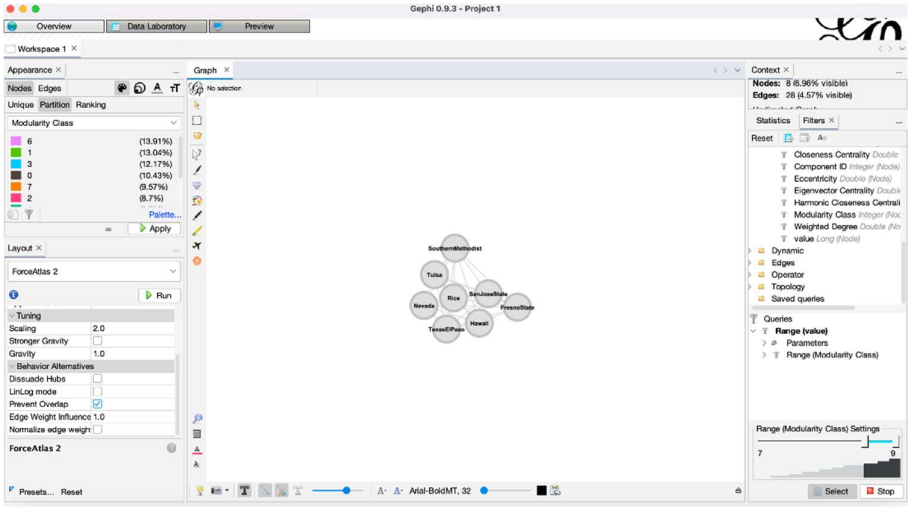


Figure 13: Important Entities based on Weighted Degree

Modularity



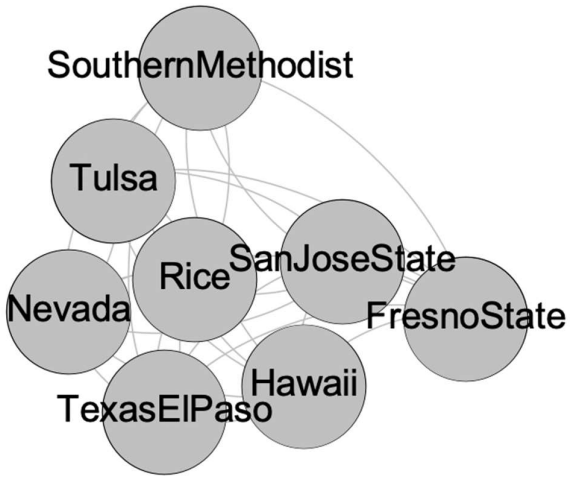


Figure 14 - Important Entities based on Modularity

Clustering coefficient:

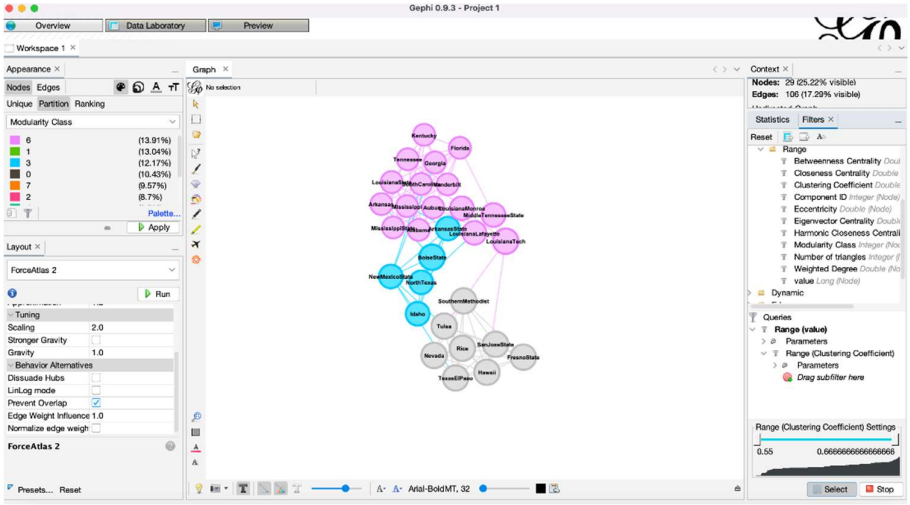
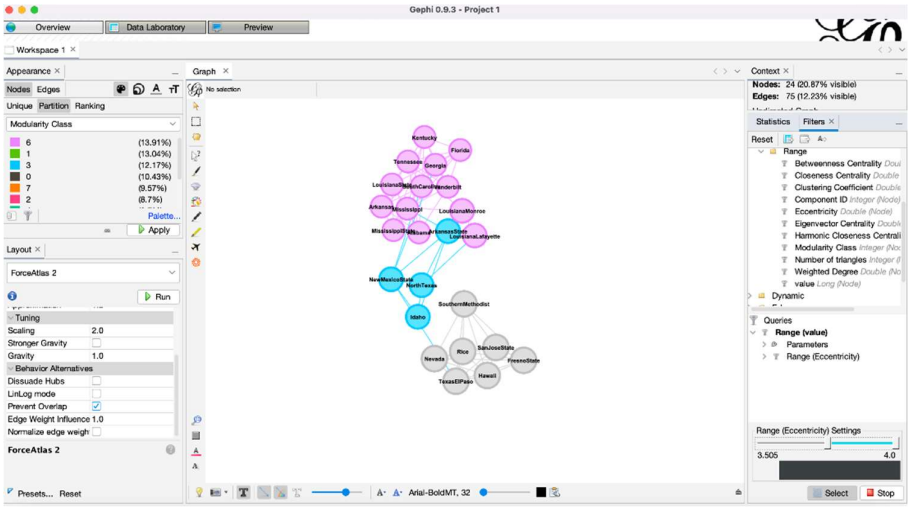


Figure 15 -Important Entities based on Clustering Coefficient

Eccentricity:



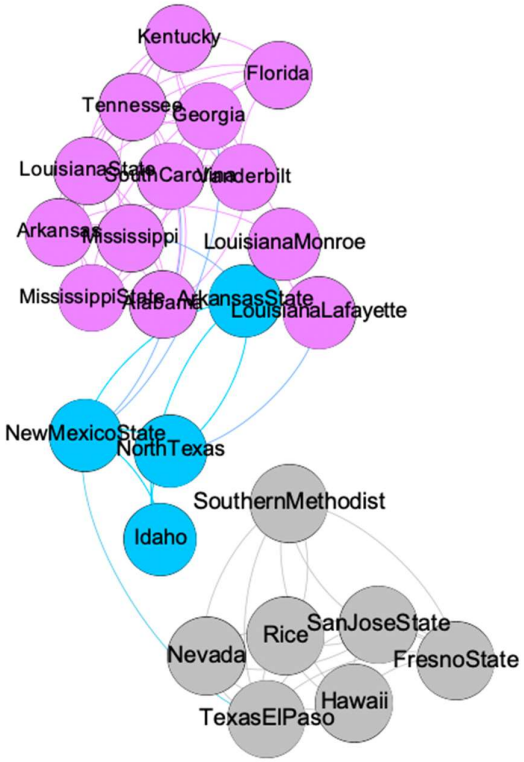
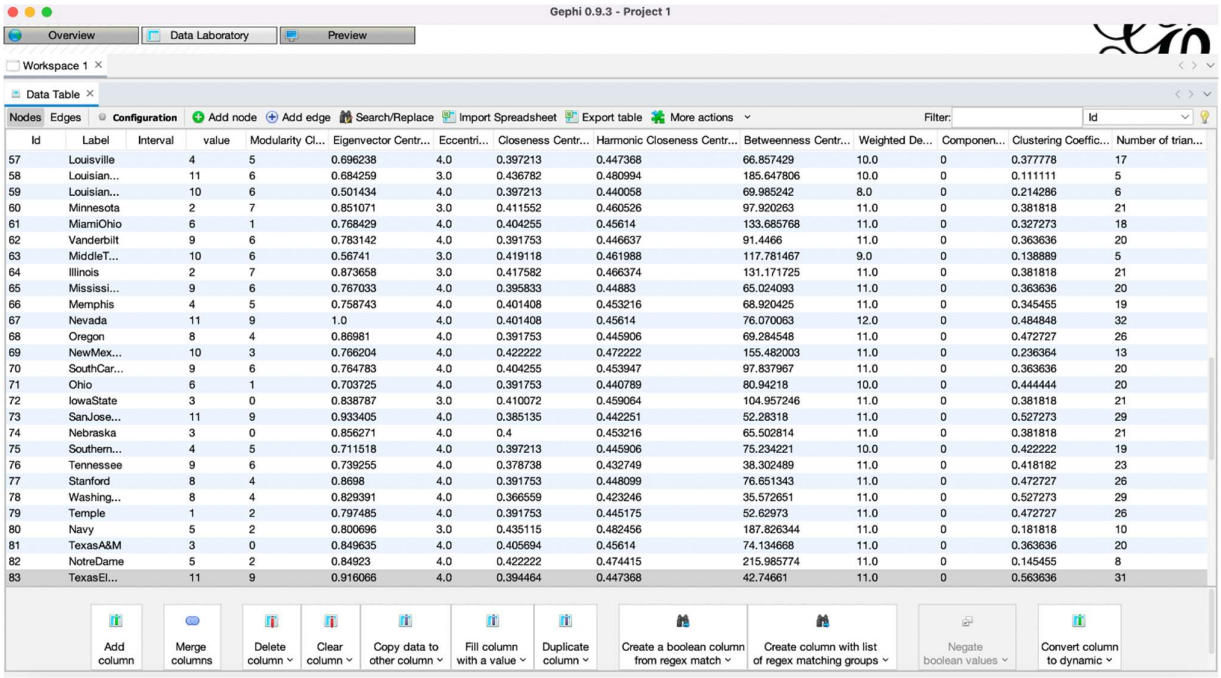


Figure 16- Important Entities based on Eccentricity:

Characteristics of Communities

Data Laboratory Report



Id	Label	Interval	value	Modularity Cl...	Eigenvector Centr...	Eccentri...	Closeness Centr...	Harmonic Closeness Centr...	Betweenness Centr...	Weighted De...	Componen...	Clustering Coeffic...	Number of trian...
57	Louisville		4	5	0.696238	4.0	0.397213	0.447368	66.857429	10.0	0	0.377778	17
58	Louisian...		11	6	0.684259	3.0	0.436782	0.480994	185.647806	10.0	0	0.111111	5
59	Louisian...		10	6	0.501434	4.0	0.397213	0.440058	69.985242	8.0	0	0.214286	6
60	Minnesota		2	7	0.851071	3.0	0.411552	0.460526	97.920263	11.0	0	0.381818	21
61	MiamiOhio		6	1	0.768429	4.0	0.404255	0.45614	133.685768	11.0	0	0.327273	18
62	Vanderbilt		9	6	0.783142	4.0	0.391753	0.446637	91.4466	11.0	0	0.363636	20
63	MiddleT...		10	6	0.56741	3.0	0.419118	0.461988	117.781467	9.0	0	0.138889	5
64	Illinois		2	7	0.873658	3.0	0.417582	0.466374	131.171725	11.0	0	0.381818	21
65	Mississi...		9	6	0.767033	4.0	0.395833	0.44883	65.024093	11.0	0	0.363636	20
66	Memphis		4	5	0.758743	4.0	0.401408	0.453216	68.920425	11.0	0	0.345455	19
67	Nevada		11	9	1.0	4.0	0.401408	0.45614	76.070063	12.0	0	0.484848	32
68	Oregon		8	4	0.86981	4.0	0.391753	0.445906	69.284548	11.0	0	0.472727	26
69	NewMex...		10	3	0.766204	4.0	0.422222	0.472222	155.482003	11.0	0	0.236364	13
70	SouthCar...		9	6	0.764783	4.0	0.404255	0.453947	97.837967	11.0	0	0.363636	20
71	Ohio		6	1	0.703725	4.0	0.391753	0.440789	80.94218	10.0	0	0.444444	20
72	IowaState		3	0	0.838787	3.0	0.410072	0.459064	104.957246	11.0	0	0.381818	21
73	SanJose...		11	9	0.933405	4.0	0.385135	0.442251	52.28318	11.0	0	0.527273	29
74	Nebraska		3	0	0.856271	4.0	0.4	0.453216	65.502814	11.0	0	0.381818	21
75	Southern...		4	5	0.711518	4.0	0.397213	0.445906	75.234221	10.0	0	0.422222	19
76	Tennessee		9	6	0.739255	4.0	0.378738	0.432749	38.302489	11.0	0	0.418182	23
77	Stanford		8	4	0.8698	4.0	0.391753	0.448099	76.651343	11.0	0	0.472727	26
78	Washing...		8	4	0.829391	4.0	0.366559	0.423246	35.572651	11.0	0	0.527273	29
79	Temple		1	2	0.797485	4.0	0.391753	0.445175	52.62973	11.0	0	0.472727	26
80	Navy		5	2	0.800696	3.0	0.435115	0.482456	187.826344	11.0	0	0.181818	10
81	TexasA&M		3	0	0.849635	4.0	0.405694	0.45614	74.134668	11.0	0	0.363636	20
82	NotreDame		5	2	0.84923	4.0	0.422222	0.474415	215.985774	11.0	0	0.145455	8
83	TexasEl...		11	9	0.916066	4.0	0.394464	0.447368	42.74661	11.0	0	0.563636	31