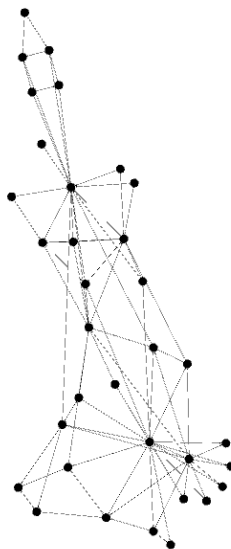


NETWORK OF ZACHARY'S KARATE CLUB

ΕΡΓΑΣΙΑ ΤΗΣ ΠΑΠΑΓΕΩΡΓΙΟΥ ΝΙΚΟΛΕΤΑ Α.Μ.: 3170234

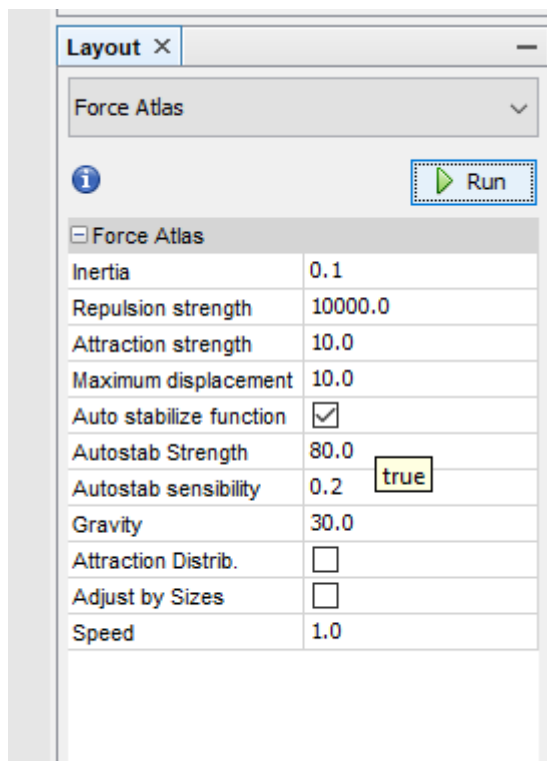
Το κοινωνικό δίκτυο που επέλεξα αποτελεί ένα κοινωνικό δίκτυο φιλιών μεταξύ 34 μελών λέσχης καράτε σε πανεπιστήμιο των ΗΠΑ την δεκαετία του 1970. Στην παρούσα εργασία ,χρησιμοποιώντας εργαλεία ανάλυσης κοινωνικών δικτύων ,συγκεκριμένα μέσω του εργαλείου Gephi , θα αποδειχθούν κι θα αναλυθούν ιδιότητες αυτού του δικτύου.

- **Graphical representation of the network**



Παρατηρούμε ότι κάποιες ακμές είναι πιο πυκνές από κάποιες άλλες. Αυτό συμβαίνει επειδή όσο μεγαλύτερο είναι το βάρος μιας ακμής, τόσο πιο πυκνή είναι.

Το layout του γραφήματος που χρησιμοποίησα είναι το Force Atlas με τις εξής ρυθμίσεις :



- **Basic topological properties, such as numbers of nodes and edges, network diameter, and average path length.**

nodes :34

edges:78

network diameter:3

average path length: 1.274

Η διάμετρος του γραφήματος πρόκειται για την μεγαλύτερη απόσταση ανάμεσα σε όλα τα δυνατά ζεύγη κορυφών, ενώ το μέσο μήκος μονοπατιού για το μήκος των συντομότερων μονοπατιών μεταξύ όλων των ζευγών κορυφών σε ένα δίκτυο.

Έτσι εξηγείται γιατί το average path length είναι πολύ μικρότερο της diameter.

Ένα ακόμη βασικό χαρακτηριστικό του γραφήματος είναι κι η πυκνότητα (density), που μας δείχνει πόσες ακμές έχει το γράφημα αναλογικά με τις κορυφές.

- **Component measures, such as number of connected components, existence of a giant component and component size distribution.**

Number of connected components: 1

Ο αριθμός των συνεκτικών συνιστωσών ισούται με 1 ,καθώς το δίκτυο χρησιμοποιείται συγκεκριμένα στη λήσχη του πανεπιστημίου κι όλοι μπορούν να αλληλεπιδράσουν μεταξύ τους χωρίς να υπάρχουν αποκομμένες οντότητες .

Existence of a giant component:

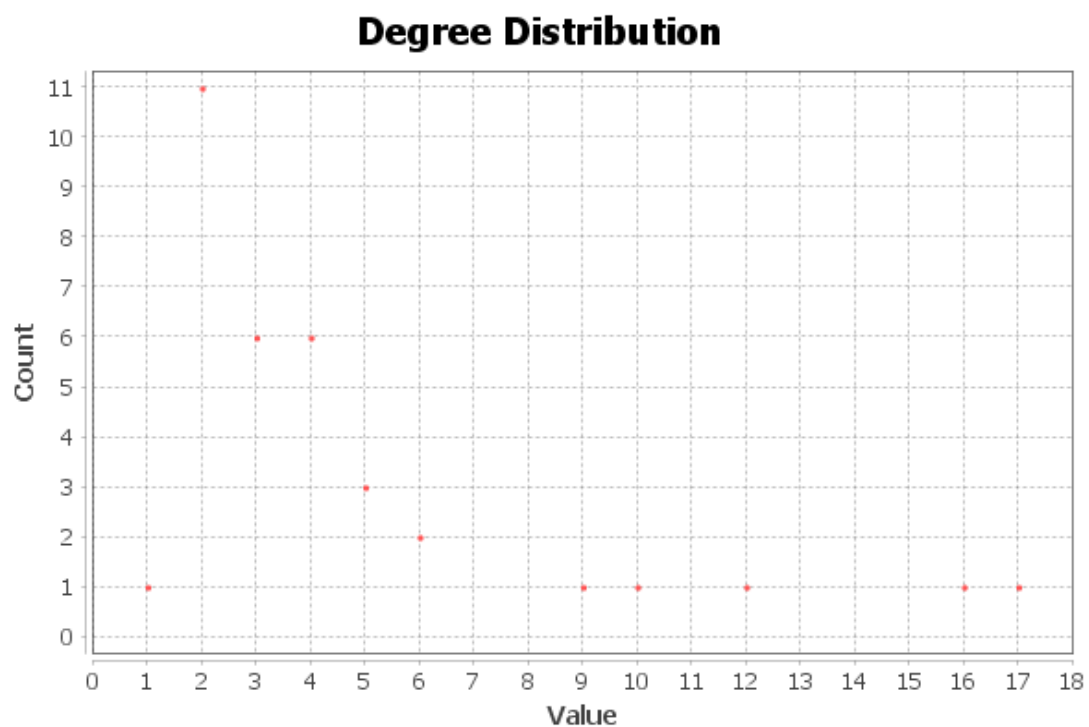
Existence component size distribution:

- **Degree measures, such as maximum and average node degrees, as well as degree distribution.**

Maximum node degrees: 1(value = 17)

Average node degrees: 2,294

Degree distribution:

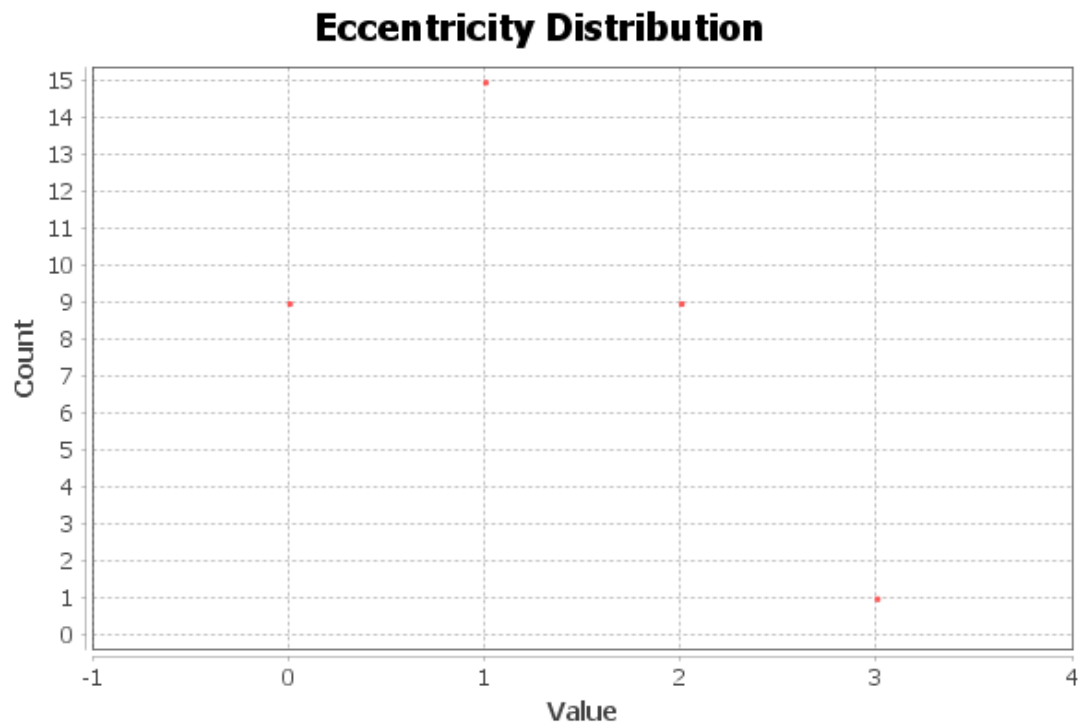


Από το μέσο βαθμό του γραφήματος που ισούται με 2,294 παίρνουμε πώς κάθε μέλος αλληλεπιδρά κατά μέσο όρο με $2,294 \sim 2$ άλλα μέλη.

Παρατηρούμε πως μεγάλο βαθμό έχουν λίγες κορυφές (πχ με βαθμό 17 είναι μόνο μία κορυφή) ενώ σε μικρότερη τιμή έχουμε πιο πολλές κορυφές(βλέπε με βαθμό 2 έχουμε 11 κορυφές).Αυτό μπορεί να δικαιολογήσει και την διαφορά που είδαμε πριν ανάμεσα σε διάμετρο κι μέσο μήκος μονοπατιού.

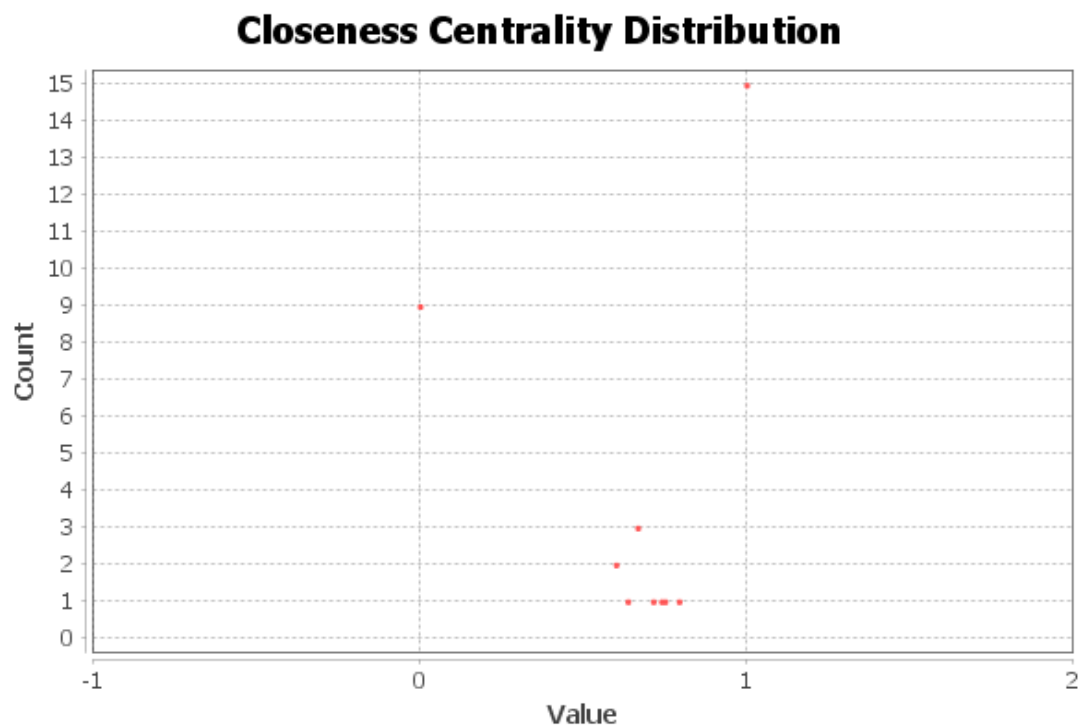
- Centrality measures (degree, betweenness, closeness, eigenvector)

Eccentricity:



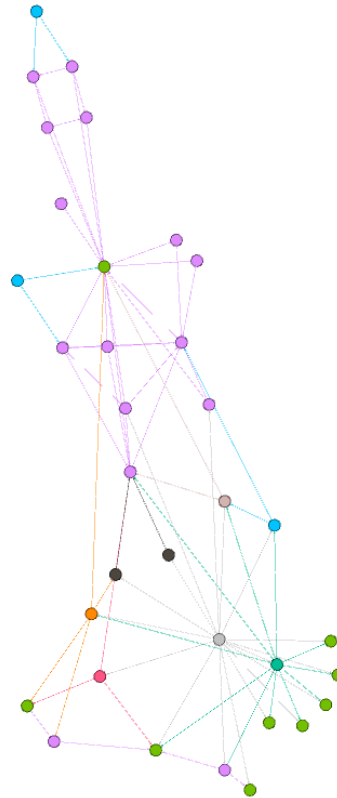
Παρατηρούμε ότι στο σύνολο $\{0,1,2,3\}$ ανήκουν οι τιμές της eccentricity κι αυτό είναι λογικό καθώς γνωρίζοντας ότι η διάμετρος του γράφου μας είναι 3 δεν είναι δυνατόν να υπερβεί κάποια απόσταση αυτή την τιμή.

Closeness:

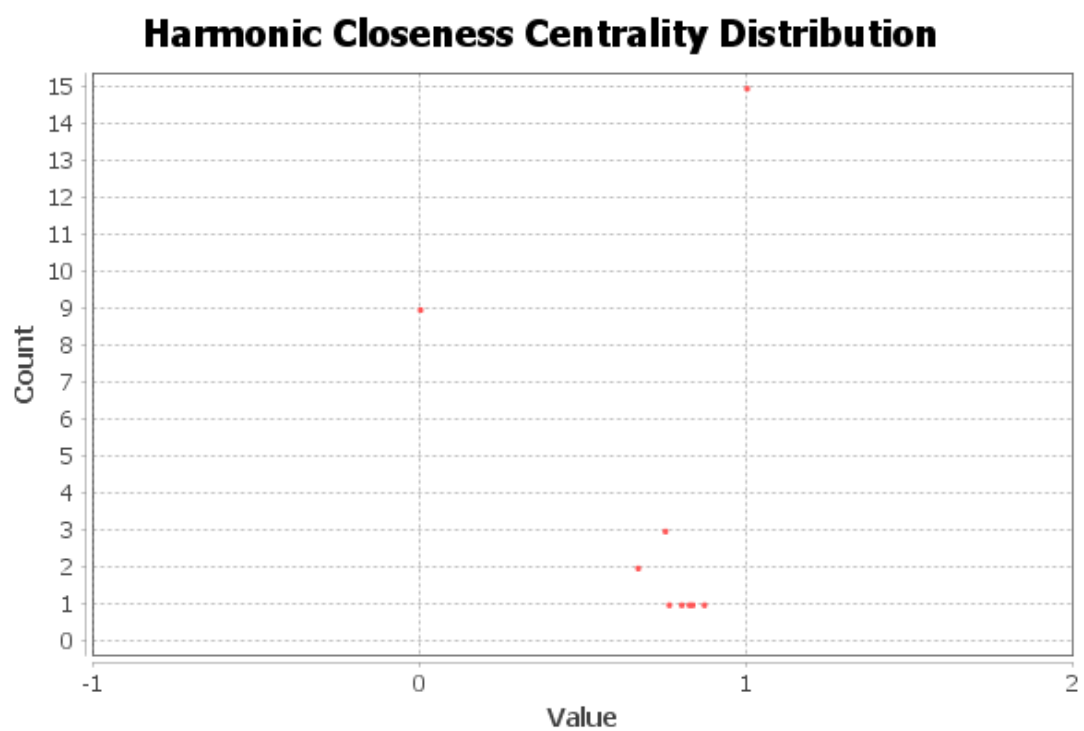


Η closeness centrality είναι ένα μέτρο σπουδαιότητας μιας κορυφής, και μπορεί να θεωρηθεί σαν ένα μέτρο της ταχύτητας επικοινωνίας μιας κορυφής με τις υπόλοιπες, αφού υπολογίζεται βάσει του αθροίσματος των συντομότερων μονοπατιών της εκάστοτε κορυφής με τις υπόλοιπες κορυφές του γραφήματος.

Βλέπουμε από το διάγραμμα πως οι τιμές της closeness centrality βρίσκονται στο διάστημα $[0,1]$, είναι δηλαδή κανονικοποιημένες.



Harmonic Closeness Centrality Distribution:



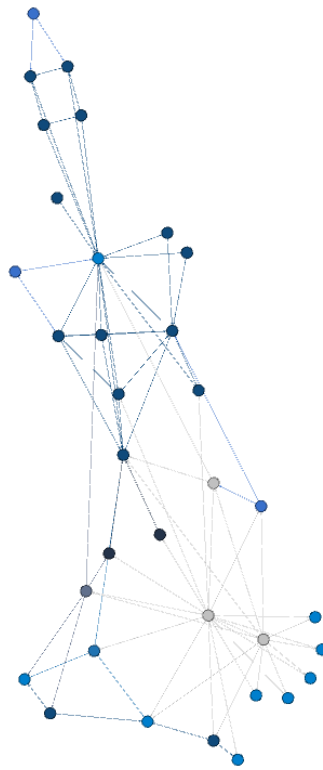
Κι εδώ παρατηρούμε πως οι τιμές της Harmonic Closeness βρίσκονται στο διάστημα $[0,1]$, δηλαδή είναι κανονικοποιημένες κι εδώ. Οι τιμές φαίνεται πως είναι πολύ κοντά

με αυτές της Closeness Centrality κι στον παρακάτω πίνακα πέρα από σύγκριση των διαγραμμάτων τους.

Closeness Centr...	Harmonic Closeness Centr...
0.0	0.0
1.0	1.0
1.0	1.0
1.0	1.0
1.0	1.0
1.0	1.0
1.0	1.0
1.0	1.0
0.75	0.833333
0.6	0.666667
1.0	1.0
1.0	1.0
0.666667	0.75
1.0	1.0
0.0	0.0
0.0	0.0
0.666667	0.75
1.0	1.0
0.0	0.0
1.0	1.0
0.0	0.0
1.0	1.0
0.0	0.0
0.0	0.0

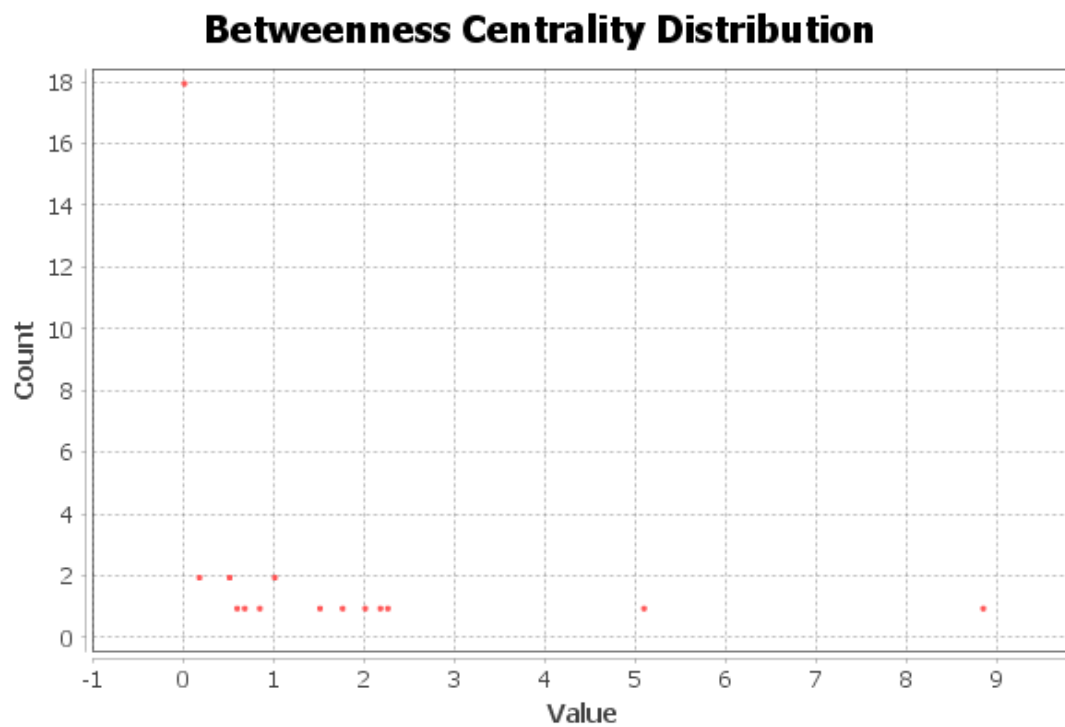
Εικόνα 1Ορισμένα id για σύγκριση

Παρακάτω παραθέτω κι το γράφο της Harmonic Closeness Centrality ,όπου όσο πιο σκούρο μπλε έχει η κορυφή τόσο μεγάλη η τιμή του.



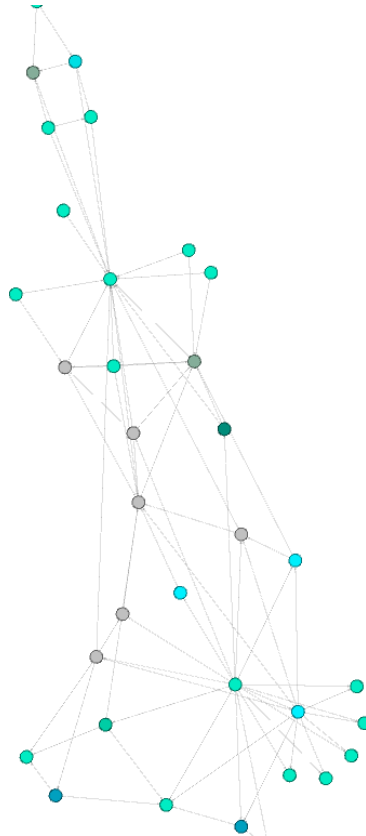
Εικόνα 2 Γράφημα -Harmonic Closeness Centrality

Betweenness Centrality Distribution:



Αναμενόμενο είναι να είναι πιο πυκνό το διάγραμμα για μικρές τιμές της Betweenness Centrality γιατί όπως είδαμε κι νωρίτερα τα περισσότερα μέλη δεν αλληλεπιδρούν με πολλά άλλα μέλη, οπότε είναι κι δύσκολο να υπάρχουν σε πολλά από αυτά συντομότερα μονοπάτια.

Στο γράφημα παρακάτω φαίνεται πως όσο πιο μεγάλη η τιμή του Betweenness Centrality τόσο πιο σκούρο πράσινο είναι το χρώμα των κορυφών.



Εικόνα 3 Γράφημα -Betweenness Centrality Distribution

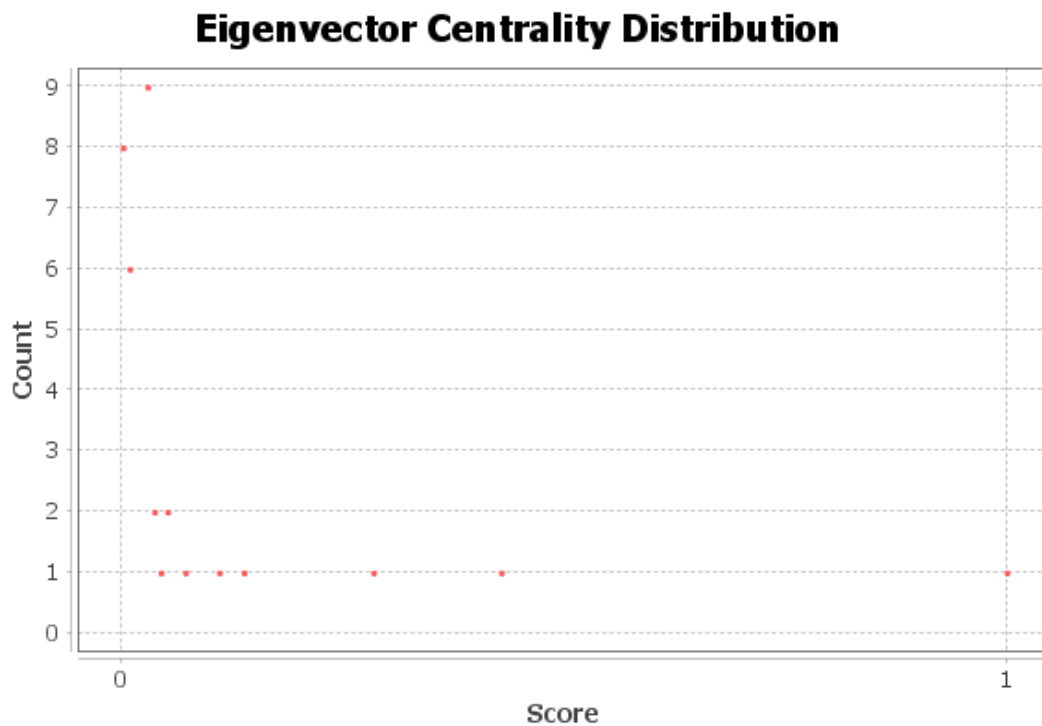
Eigenvector:

Parameters:

Network Interpretation: directed

Number of iterations: 100

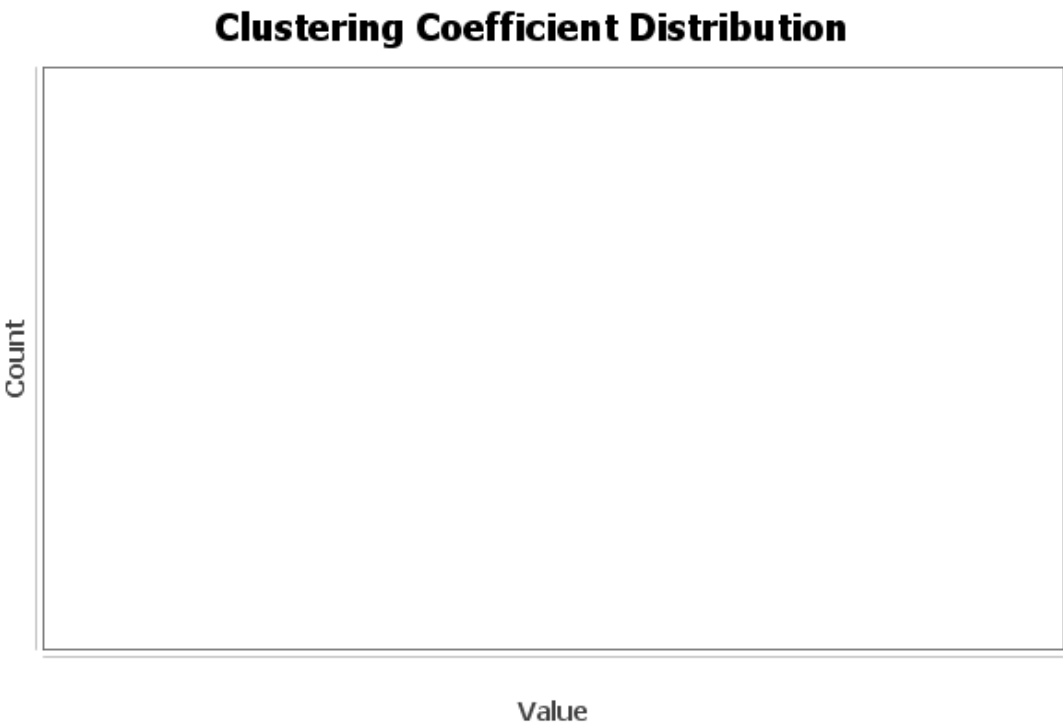
Sum change: 0.003985081431127474



Παρατηρούμε ότι οι τιμές κι εδώ είναι πολύ χαμηλές .

- Clustering effects in the network: average clustering coefficient, number of triangles, clustering coefficient distribution, existence of the triadic closure phenomenon in the friendship neighborhood.

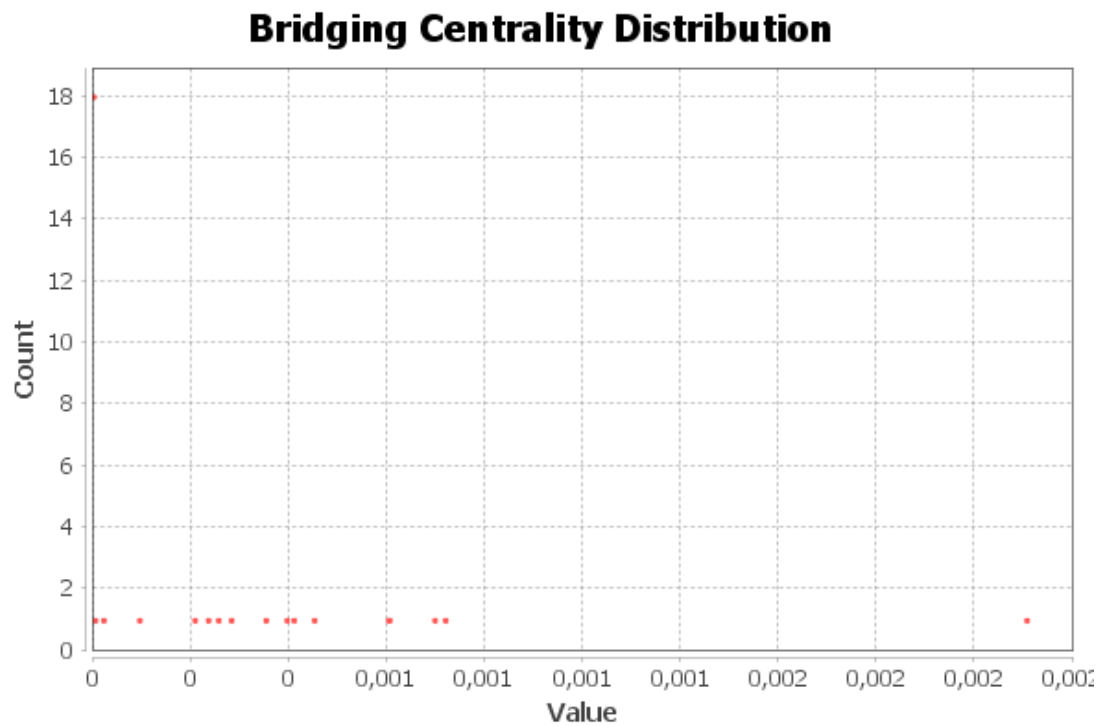
Average clustering coefficient: 0,285



Id	Label	Clustering Coe...
1	1	0.075
2	2	0.166667
3	3	0.122222
4	4	0.333333
5	5	0.333333
6	6	0.25
7	7	0.25
8	8	0.5
9	9	0.25
10	10	0.0
11	11	0.333333
12	12	0.0
13	13	0.5
14	14	0.3
15	15	0.5
16	16	0.5
17	17	0.5
18	18	0.5
19	19	0.5
20	20	0.166667
21	21	0.5
22	22	0.5
23	23	0.5
24	24	0.2

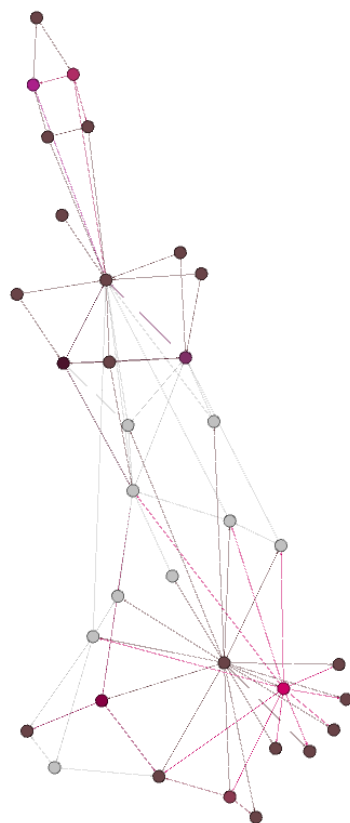
Εικόνα 4Clustering coefficient

- Bridges and local bridges.



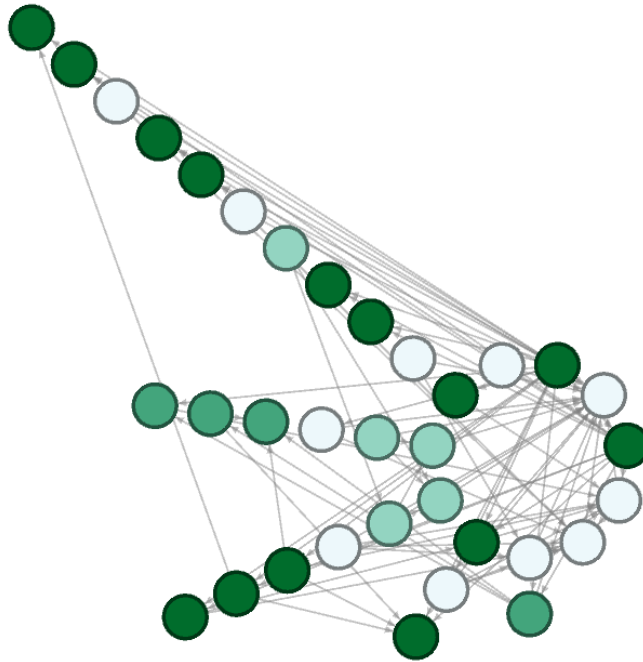
Εικόνα 5 Bridging Centrality Distribution

Κι εδώ βλέπουμε πως οι τιμές είναι πολύ χαμηλές. Το γράφημα μας ως σύνολο ατόμων μιας λέσχης που αλληλεπιδρούν μεταξύ τους είναι δύσκολο να περιέχει ακμή που να ενώνει δύο μεγάλα κομμάτια του. Στο παρακάτω γράφημα φαίνονται ο κορυφές χρωματισμένες ανάλογα με το Bridging Centrality.



Εικόνα 6Γράφος Bridging Centrality

- **Gender and homophily.**



Εικόνα 7 Γράφος homophily

Οι ανοιχτόχρωμες κορυφές, έχουν $\text{betweenness centrality}=0$. Οι πιο σκουρόχρωμες κορυφές έχουν μεγαλύτερη τιμή $\text{betweenness centrality}$, και όσο πιο σκούρο χρώμα έχουν τόσο πιο πολλές οι συνδέσεις τους με κορυφές από άλλες κοινότητες. Όπως φαίνεται στο γράφο είναι λίγες οι κορυφές που δεν έχουν επικοινωνία με κορυφές από άλλες κοινότητες.

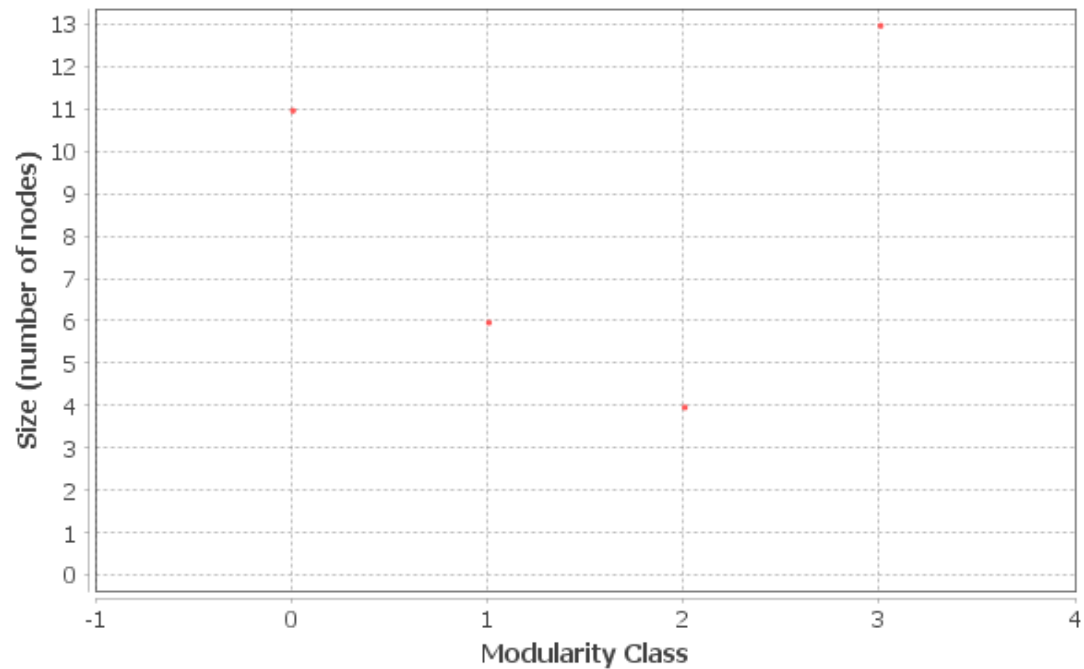
- **Graph density**

Graph Density : 0,07

Η μικρή πυκνότητα μας δίνει το συμπέρασμα πως δεν υπάρχει αρκετή αλληλεπίδραση μεταξύ των μελών του καράτε κλαμπ.

- Community structure (modularity), cliques

Size Distribution



Εικόνα 8 Αριθμός των μελών ανά κοινότητα

Έχοντας πάρει τις εξής παραμέτρους :

HTML Report

Modularity Report

Parameters:

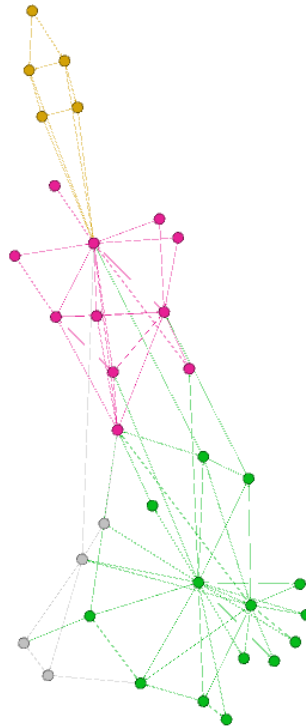
Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0,415
Modularity with resolution: 0,415
Number of Communities: 4

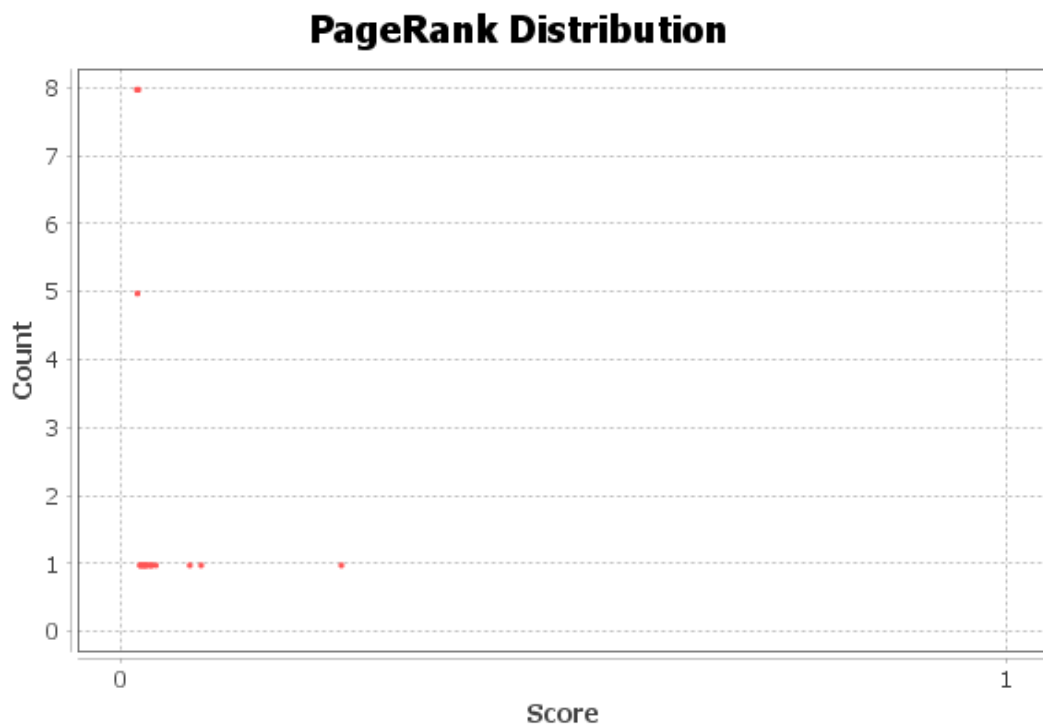
Παρατηρούμε ότι στο γράφημα μας υπάρχουν 4 κοινότητες κι ότι οι κορυφές που ανήκουν σε μια κοινότητα αλληλεπιδρούν μεταξύ τους περισσότερο από ότι με τις υπόλοιπες κοινότητες. Στο γράφημα παρατηρούμε ότι η κοινότητα με ετικέτα 3 περιλαμβάνει τους περισσότερους χαρακτήρες (13).

Παρακάτω φαίνεται ο γράφος χωρισμένος σε κοινότητες με βάση τέσσερα χρώματα.



Εικόνα 9Modularity Class

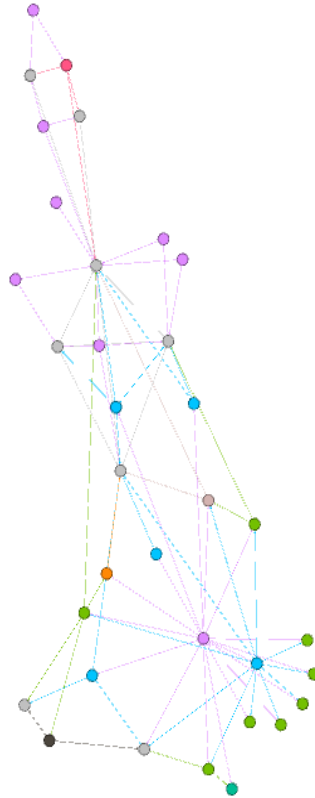
- PageRank



Η μέθοδος Pagerank μετράει το πόσο σημαντική είναι μια ιστοσελίδα βάσει των συνδέσμων (links) που οδηγούν σε αυτήν, συνεπώς σε αυτήν την περίπτωση θα δούμε πως ανταποκρίνεται ο αλγόριθμος σε ένα διαφορετικού είδους γράφημα.

Για τις παραμέτρους έχω αφήσει τις default τιμές του GEPHI : $p = 0.85$, όπου μας δείχνει την πιθανότητα ενός χρήστη που κάνει κλικ σε τυχαίους συνδέσμους να συνεχίσει να κάνει κλικ και σε άλλους συνδέσμους και $\epsilon = 0,001$, είναι το κριτήριο τερματισμού του αλγορίθμου.

Παρατηρώντας ότι το score είναι πολύ μικρό καταλαβαίνουμε πως δεν υπάρχουν σημαντικές κορυφές για αυτή την μέθοδο.



Εικόνα 10 γράφημα PageRank