

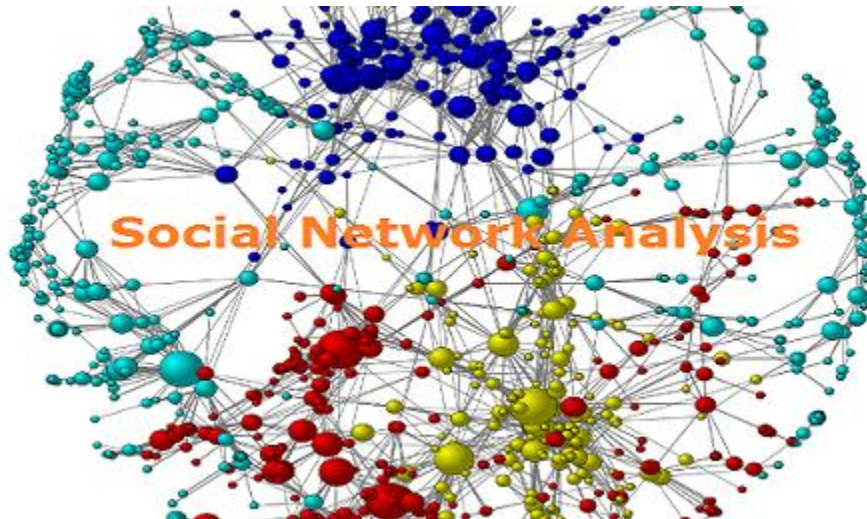
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Ανάλυση Κοινωνικών Δικτύων

2η Εργαστηριακή Άσκηση:

Ανάλυση Κοινωνικής Δομής σε Τεχνητές και Πραγματικές
Σύνθετες Τοπολογίες Δικτύων



Ονοματεπώνυμο: Σταυρακάκης Δημήτριος

AM:03112017

Εξάμηνο: 9^ο

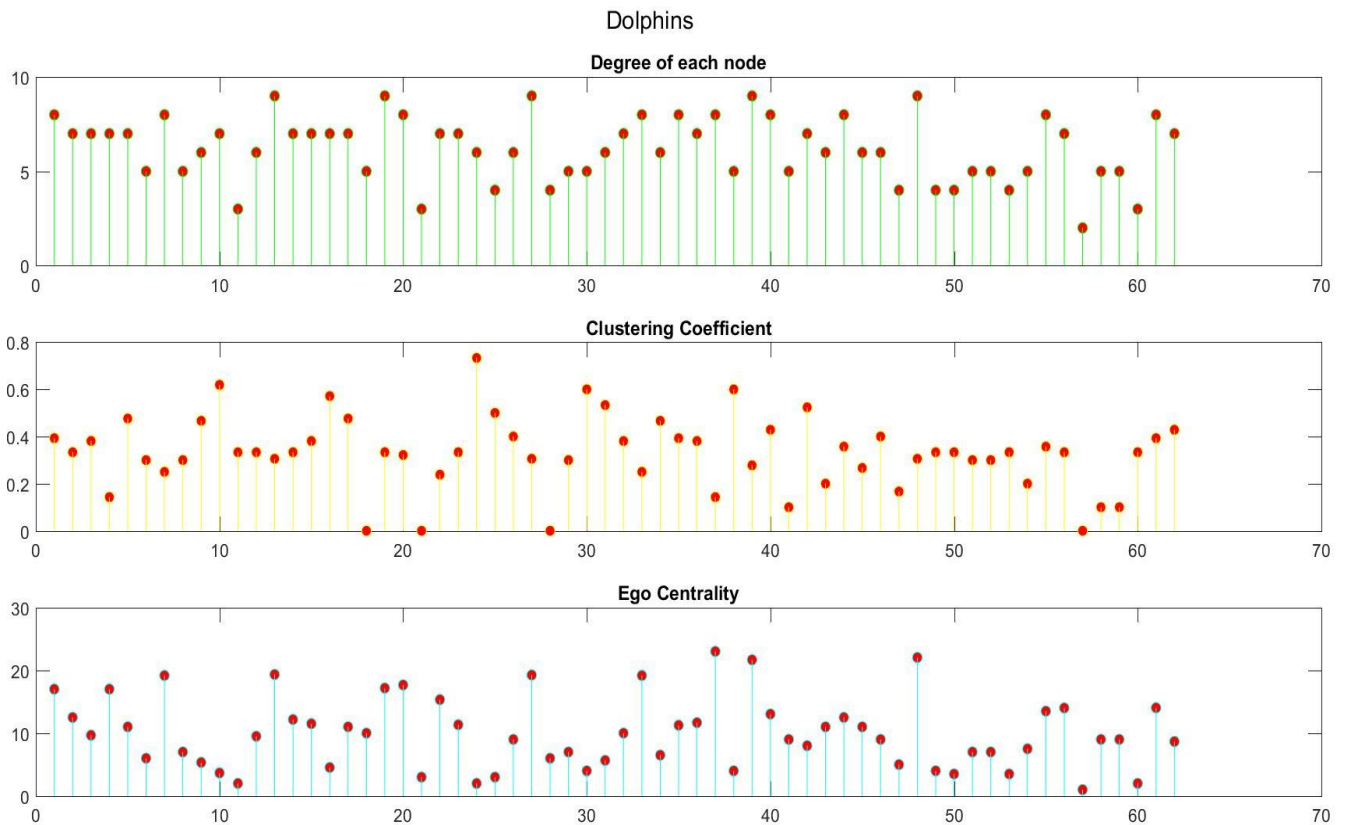
Ημερομηνία Παράδοσης: 11/1/2017

A) Μελέτη Πραγματικών Τοπολογιών

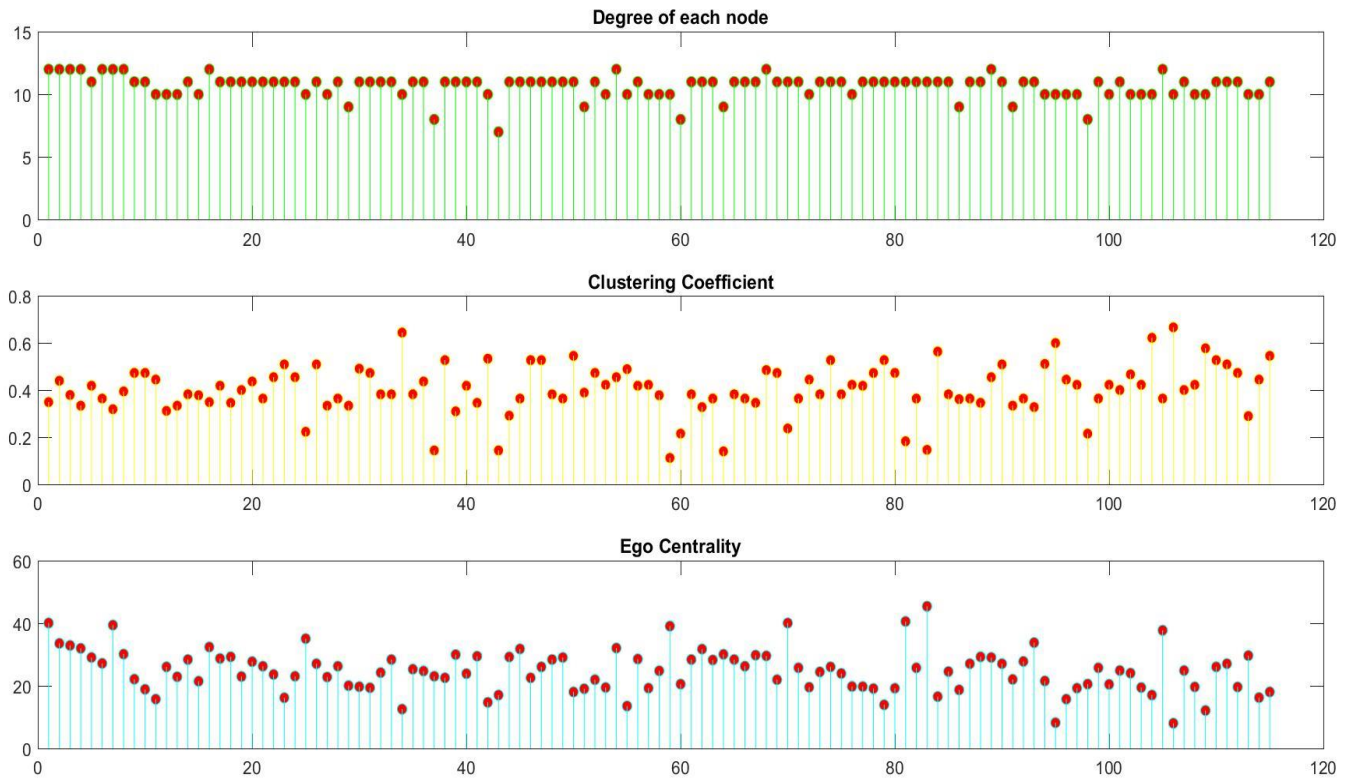
Στο ερώτημα αυτό θα μελετήσουμε τις τοπολογίες του πίνακα 2 της εκφώνησης. Πιο συγκεκριμένα, για τις 3 πραγματικές τοπολογίες υπολογίσαμε τα ακόλουθα:

- Το βαθμό κάθε κόμβου
- Το συντελεστή ομαδοποίησης κάθε κόμβου
- Την ενδιαμεσική εγωκεντρικότητα κάθε κόμβου

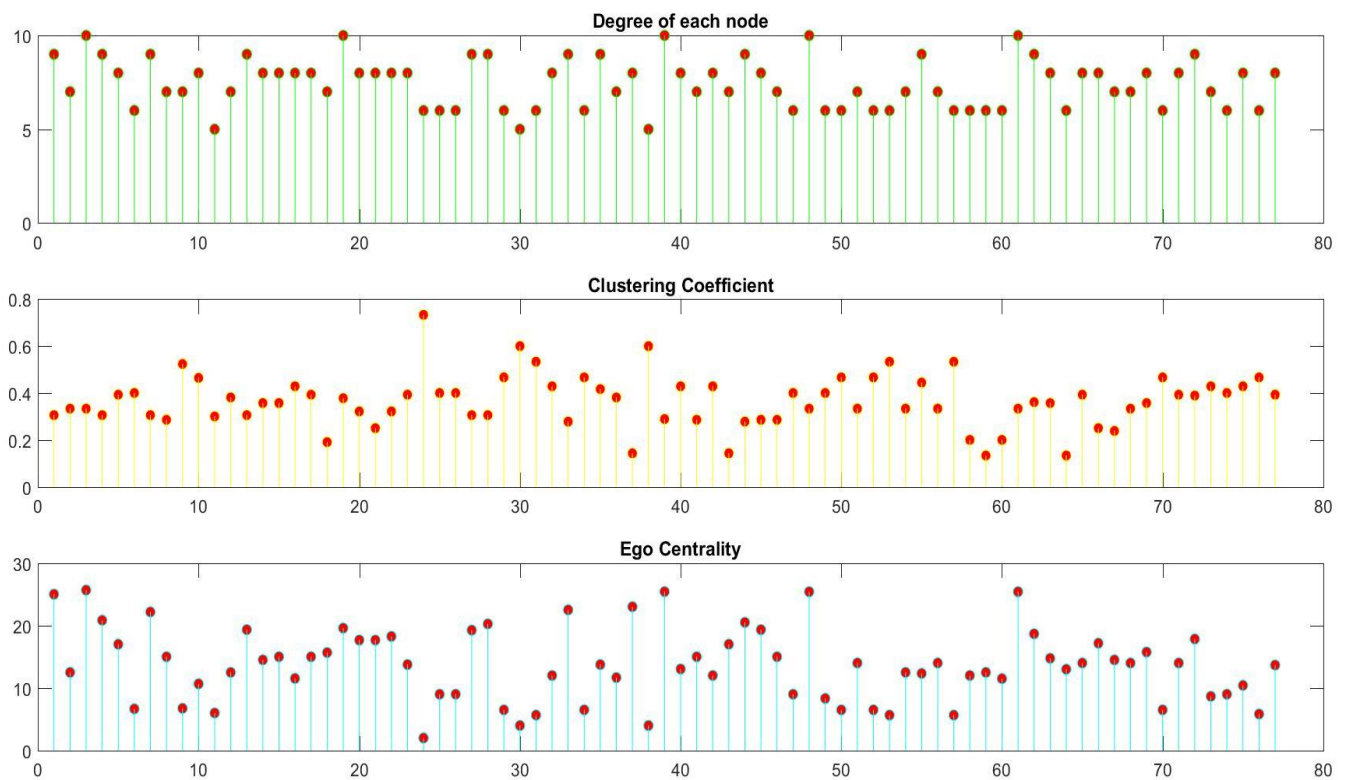
Τα αποτελέσματα φαίνονται στη συνέχεια:



Football



Les Miserables



Πέρα απο τα παραπάνω, υπολογίστηκαν και

- Ο μέσος βαθμός κόμβου της κάθε τοπολογίας
- Ο μέσος συντελεστής ομαδοποίησης κόμβου της κάθε τοπολογίας
- Η μέση ενδιαμεσική εγωκεντρικότητα κόμβου της κάθε τοπολογίας

Τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα:

Τοπολογία→ Μετρική↓	Dophins	Football	Les Miserables
Μέσος Βαθμός Κόμβου	6.225806451612903	10.660869565217391	7.454545454545454
Μέσος Συντελεστής Ομαδοποίησης Κόμβου	0.330325140809012	0.403216011042098	0.365460729746444
Μέση Ενδιαμεσική Εγωκεντρικότητα Κόμβου	10.004838709677419	24.722132505175974	13.689826839826840

Σχολιασμός Αποτελεσμάτων:

Για το δίκτυο **Dolphins** παρατηρούμε ότι η κατανομή των βαθμών προσεγγίζει τοπολογία small world. Επίσης, οι τιμές των συντελεστών ομαδοποίησης καθώς και ο μέσος συντελεστής ομαδοποίησης είναι αρκετά κοντά σε αυτές που υπολογίσαμε για την τοπολογία small world στην προηγούμενη εργαστηριακή άσκηση. (ο μέσος συντελεστής ομαδοποίησης είναι πιο κοντά στη small world τοπολογία με πιθανότητα ανασύνδεσης $g_p = 0.5$, όπως μπορούμε να δούμε με βάση τις μετρήσεις στην προηγούμενη άσκηση) Όσον αφορά το ego-centrality τα αποτελέσματα που λάβαμε για την τοπολογία Dolphins είναι πιο κοντά στην RGG τοπολογία (όμως δε μπορούμε να πούμε ότι η RGG τοπολογία το προσεγγίζει καθώς οι τιμές των clustering coefficients έχουν σημαντικές διαφορές). Γενικότερα όμως το δίκτυο Dolphins προσεγγίζεται καλύτερα απο **small world** τοπολογία. (παρόλο που έχει μικρό αριθμό κόμβων οπότε δε μπορούμε να προβούμε σε κάποιο ασφαλές συμπέρασμα)

Για την τοπολογία **Football** μπορούμε να παρατηρήσουμε με βάση τις τιμές που πήραμε ότι προσεγγίζει την τοπολογία small world (με βάση τις τιμές που λάβαμε για αυτή στην προηγούμενη εργαστηριακή άσκηση). Βλέπουμε ότι οι βαθμοί των κόμβων έχουν μικρές διαβαθμίσεις (μικρή διασπορά) που είναι ένα στοιχείο που είδαμε στη small world τοπολογία. Επίσης, οι τιμές που λάβαμε για το ego-centrality των κόμβων της τοπολογίας αυτής μοιάζουν αρκετά με αυτές της small world τοπολογίας. Επίσης και ο συντελεστής ομαδοποίησης έχει τιμή αρκετά κοντά με την τιμή που βρήκαμε για τη small world τοπολογία στην προηγούμενη άσκηση. Αυτό έχει κάποια λογική αν αναλογιστούμε ότι ο γράφος αυτός αναπαριστά ένα δίκτυο προγραμματισμένων αγώνων American football (δε διεξάγονται τυχαία δηλαδή). Με βάση τα παραπάνω συμπεραίνουμε ότι ο γράφος football προσεγγίζεται καλύτερα απο τη **small world** τοπολογία.

Για το γράφο **Les Miserables** παρατηρούμε ότι παίρνουμε αποτελέσματα με σημαντικές ομοιότητες με το δίκτυο των δελφινιών. Επομένως και εδώ μπορούμε να πούμε ότι το δίκτυο που εκφράζει τις συνδέσεις μεταξύ των χαρακτήρων «των Αθλίων» μπορεί να προσεγγιστεί απο την τοπολογία **small world**, αν και έχει και κάποιες ομοιότητες με την τοπολογία RGG. (όπως εξηγήσαμε και για το δίκτυο των δελφινιών προηγουμένως)

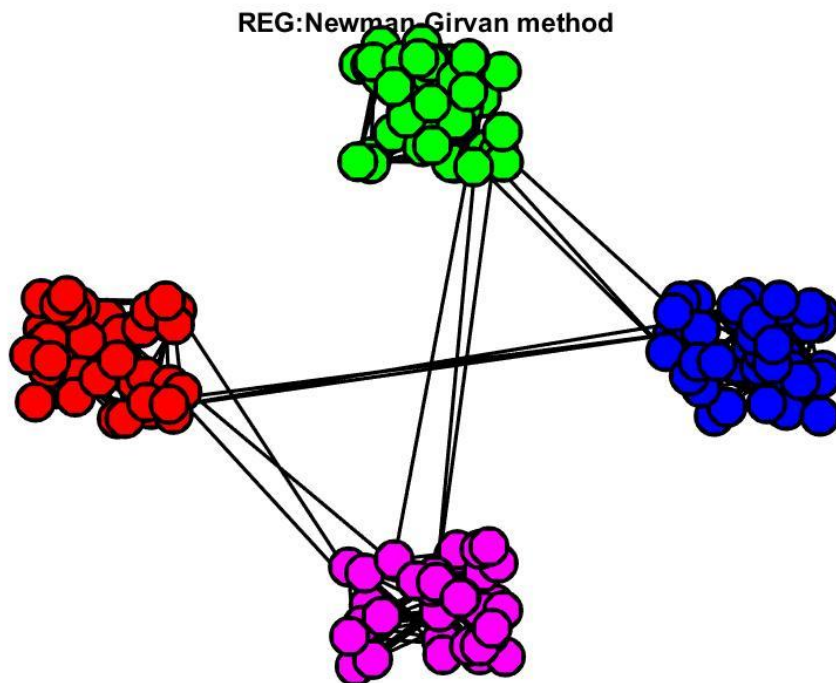
B) Εξοικείωση με τα εργαλεία εντοπισμού κοινοτήτων

Στο ερώτημα αυτό υπολογίσαμε για κάθε τοπολογία (συνθετική και πραγματική) τις κοινότητες αυτής με χρήση 3 διαφορετικών μεθόδων:

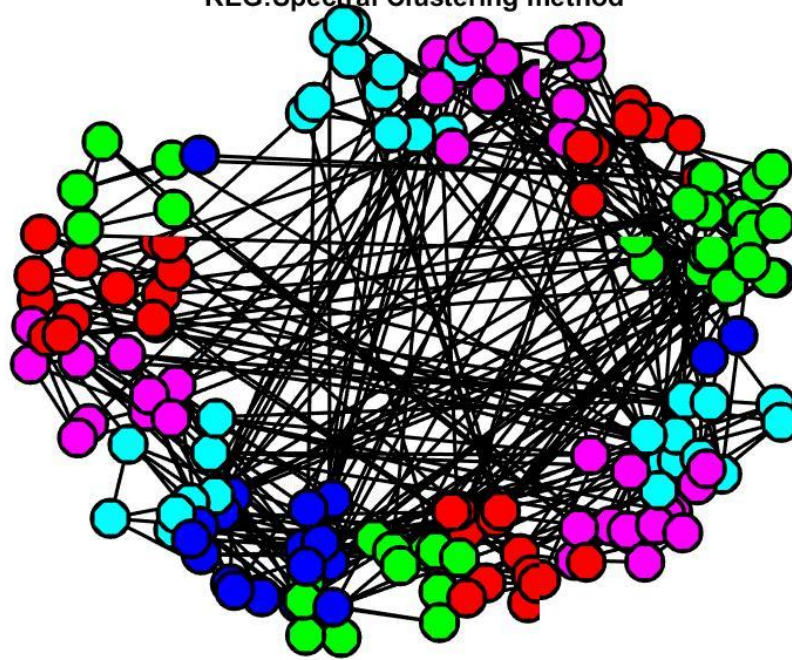
- Newman-Girvan
- Spectral Clustering
- Modularity Maximization

Οπτικοποιήσαμε λοιπόν τις κοινότητες όπως αυτές μας προέκυψαν σε κάθε μία από τις 3 ανωτέρω μεθόδους. Ακόμη υπολογίσαμε και το modularity για κάθε αλγόριθμο εντοπισμού κοινοτήτων για κάθε τοπολογία. Αυτό είναι μια μετρική που στην ουσία μας υποδεικνύει πόσο ικανοποιητικός είναι ο διαχωρισμός σε κοινότητες που κάναμε. Τα αποτελέσματα παρατίθενται και σχολιάζονται στη συνέχεια:

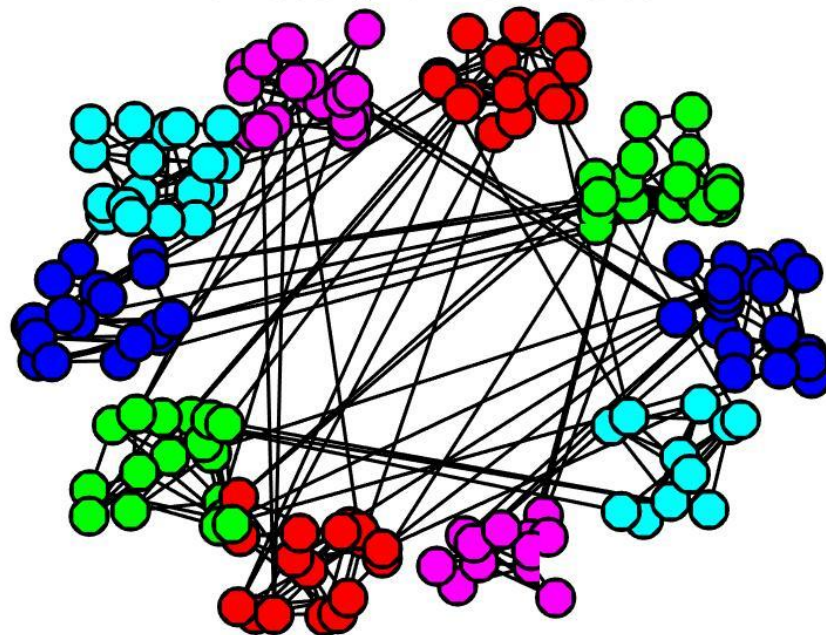
REG:



REG:Spectral Clustering method



REG:Modularity Maximization method

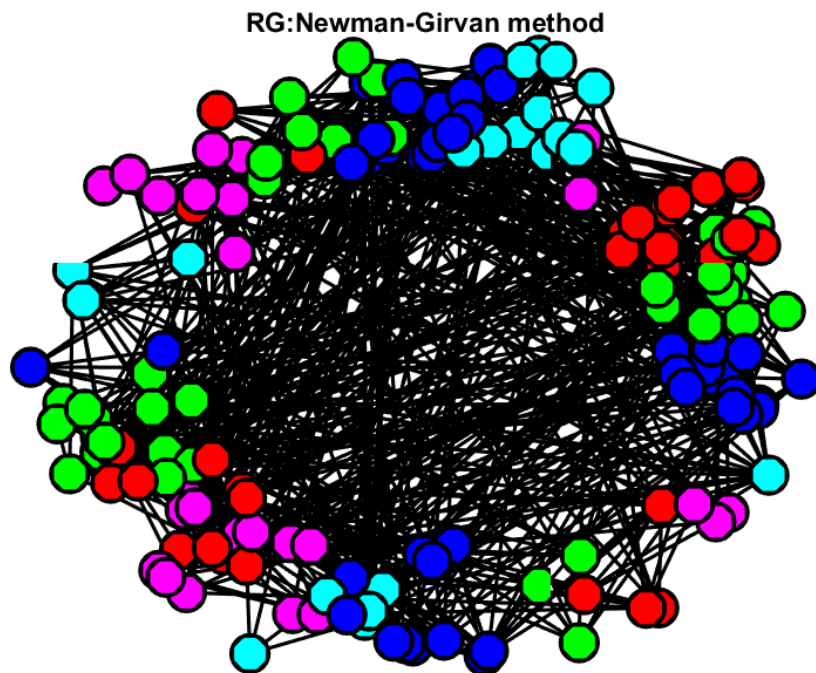


<u>Μέθοδος↓</u>	<u>Modularity↓</u>
Newman-Girvan	0.6958
Spectral Clustering	0.2953
Modularity Maximization	0.7471

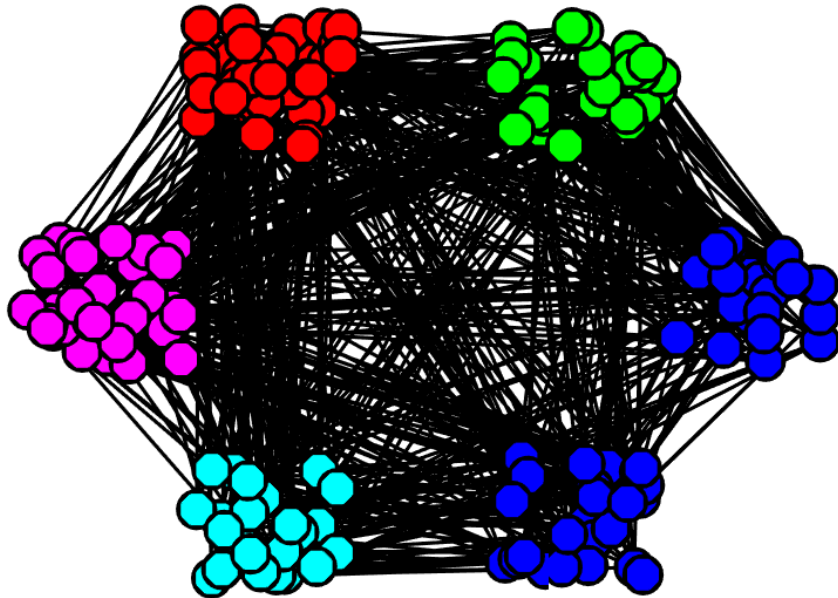
Σχολιασμός Αποτελεσμάτων:

Παρατηρούμε ότι οι μέθοδοι Newman-Girvan και Modularity Maximization δεν απέχουν αρκετά σε απόδοση με βάση τη μετρική που παίρνουμε για την αξιολόγησή τους. Αντίθετα, η μέθοδος Spectral Clustering δίνει σαφώς χειρότερο διαχωρισμό σε κοινότητες (μας έδωσε περισσότερες). Ένας ακόμη τρόπος που μπορούμε και μόνοι μας να δούμε και να αποφανθούμε για την αποδοτικότητα των ντετερμινιστικών αλγορίθμων Newman-Girvan και Modularity Maximization είναι ότι μεταξύ των κοινοτήτων υπάρχουν ελάχιστες συνδέσεις. (ενώ στο Spectral Clustering παρατηρούμε πολλές γέφυρες μεταξύ των κοινοτήτων)

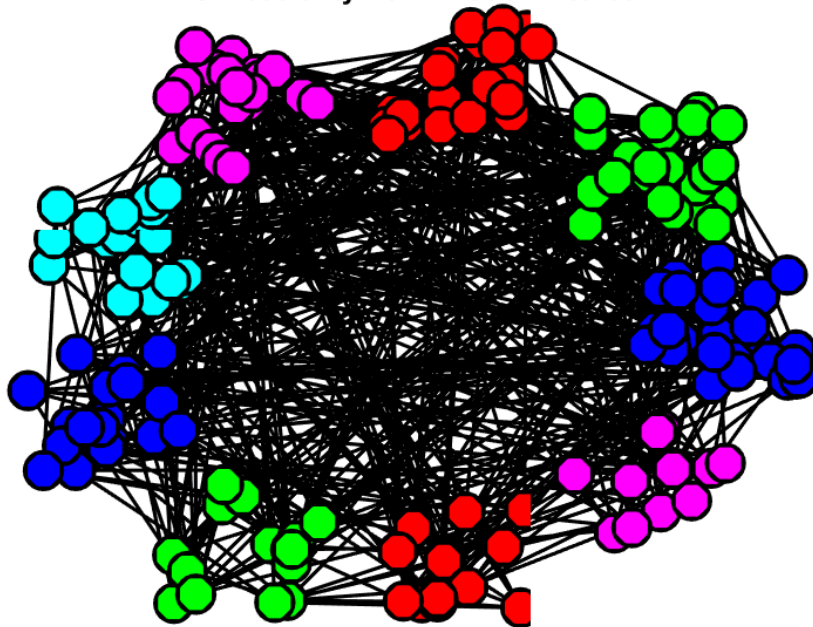
RG:



RG:Spectral Clustering method



RG:Modularity Maximization method

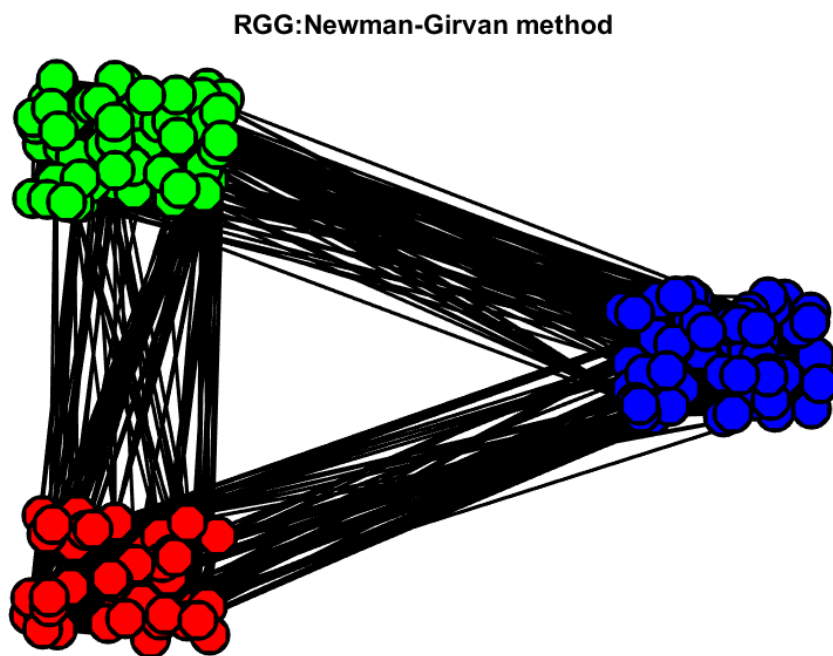


Μέθοδος↓	Modularity↓
Newman-Girvan	0.1522
Spectral Clustering	0.0676
Modularity Maximization	0.2867

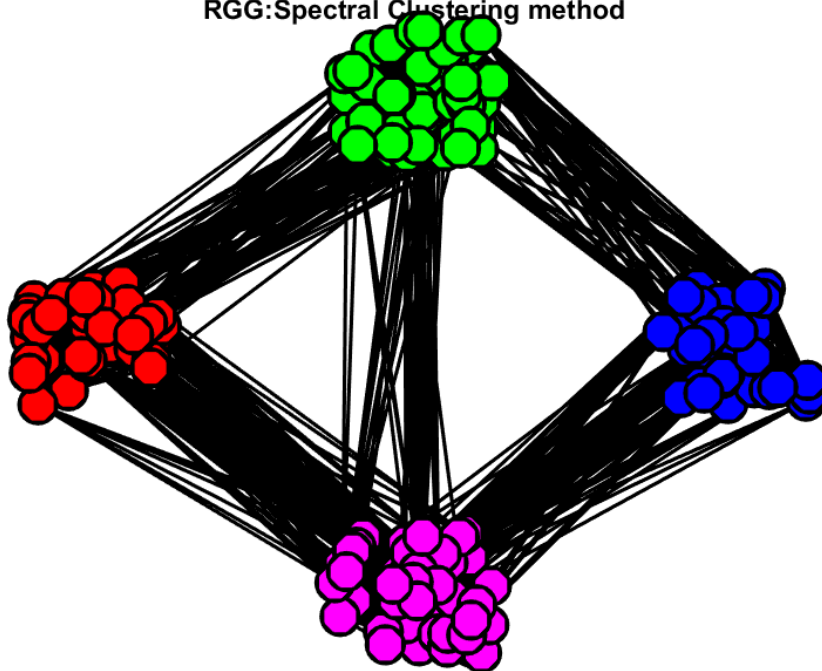
Σχολιασμός Αποτελεσμάτων:

Στην περίπτωση της RG συνθετικής τοπολογίας βλέπουμε με βάση τη μετρική μας ότι καλύτερο διαχωρισμό δίνει η μέθοδος Modularity Maximization. Ωστόσο, οι τιμές που λαμβάνουμε στη μετρική μας είναι αρκετά χαμηλές και δε μπορούμε να πούμε ότι γενικά επιτύχαμε πολύ καλό διαχωρισμό κοινοτήτων στην εν λόγω τοπολογία.

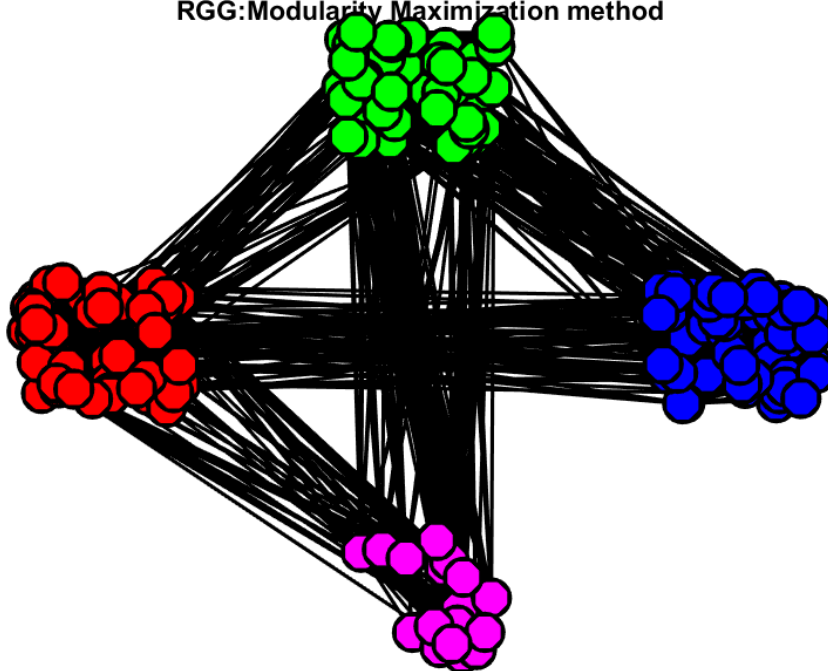
RGG:



RGG:Spectral Clustering method



RGG:Modularity Maximization method

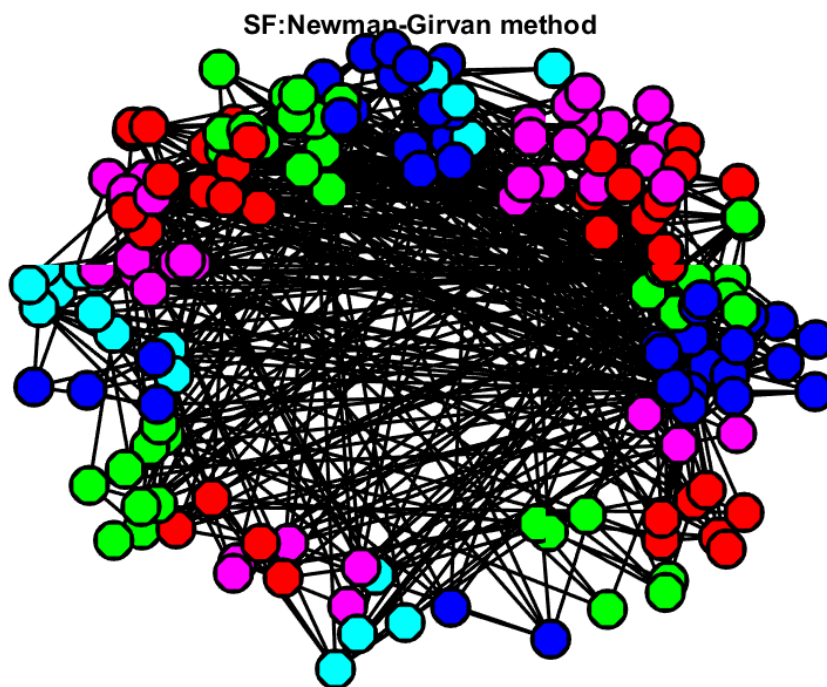


Μέθοδος↓	Modularity↓
Newman-Girvan	0.5401
Spectral Clustering	0.5636
Modularity Maximization	0.5575

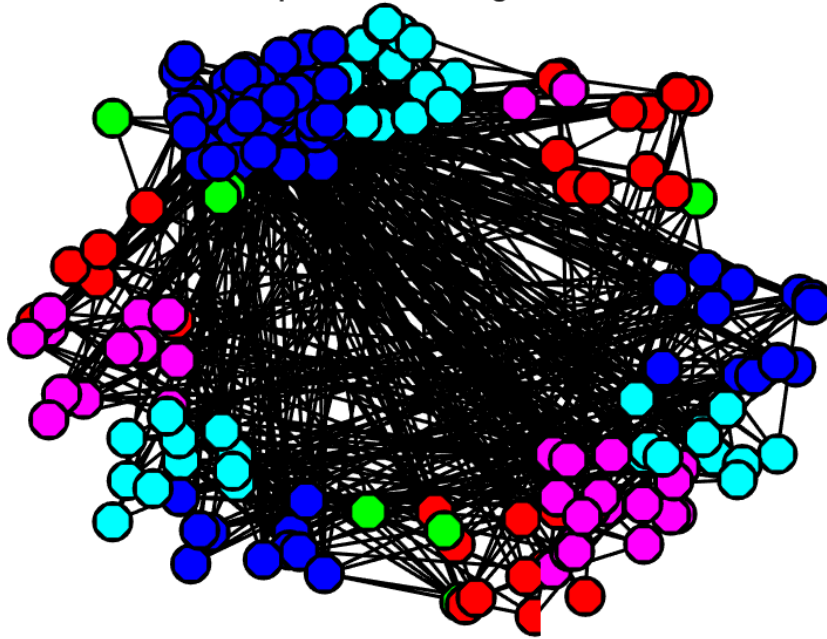
Σχολιασμός Αποτελεσμάτων:

Στην τοπολογία αυτή (RGG) παρατηρούμε ότι και οι 3 μέθοδοι δίνουν αρκετά παρόμοια αποτελέσματα. Χειρότερη, με βάση τη μετρική μας, είναι η Newman-Girvan χωρίς να διαφέρει όμως κατά πολύ από τις υπόλοιπες. Παρατηρούμε ότι εδώ οι κοινότητες που έχουμε, όπως μας τις δίνουν οι 3 μέθοδοί μας, είναι λίγες. Αυτό είναι λογικό, καθώς έχουν σημαντική επίδραση οι συντεταγμένες των κόμβων. (κοντινοί κόμβοι τείνουν να σχηματίζουν κοινότητα) Σε γενικές γραμμές επιτυγχάνουμε ένα μέτριο διαχωρισμό και με τις 3 μεθόδους μας.

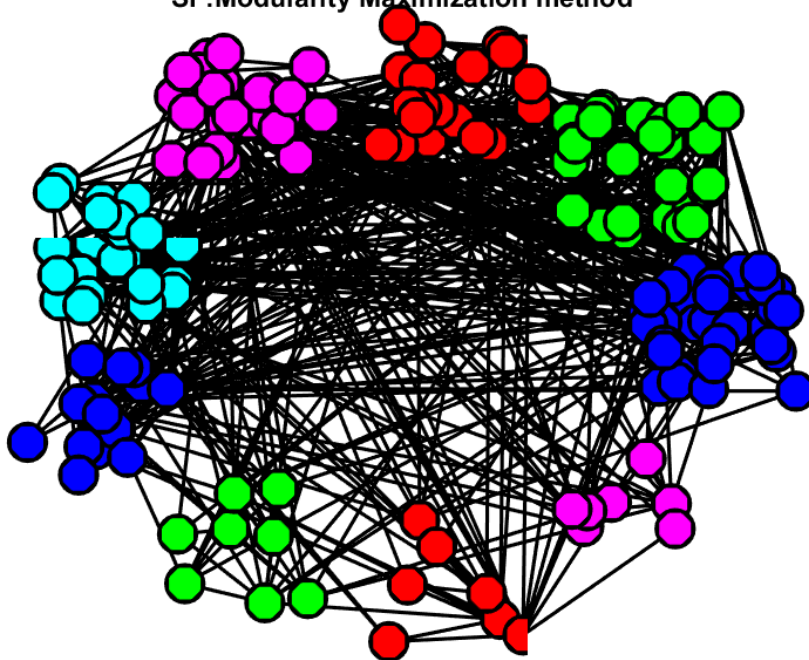
SF:



SF:Spectral Clustering method



SF:Modularity Maximization method

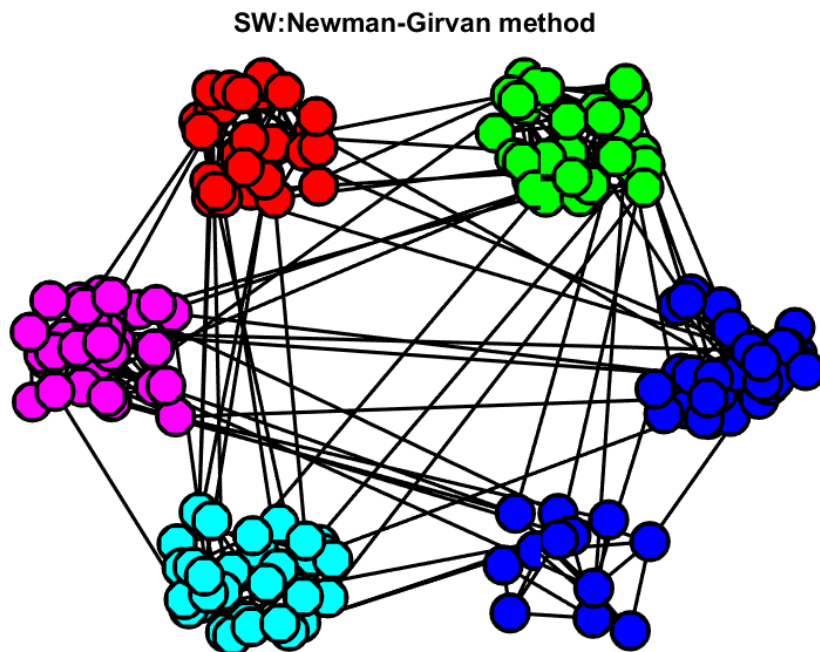


<u>Μέθοδος↓</u>	<u>Modularity↓</u>
Newman-Girvan	0.1728
Spectral Clustering	0.0328
Modularity Maximization	0.3049

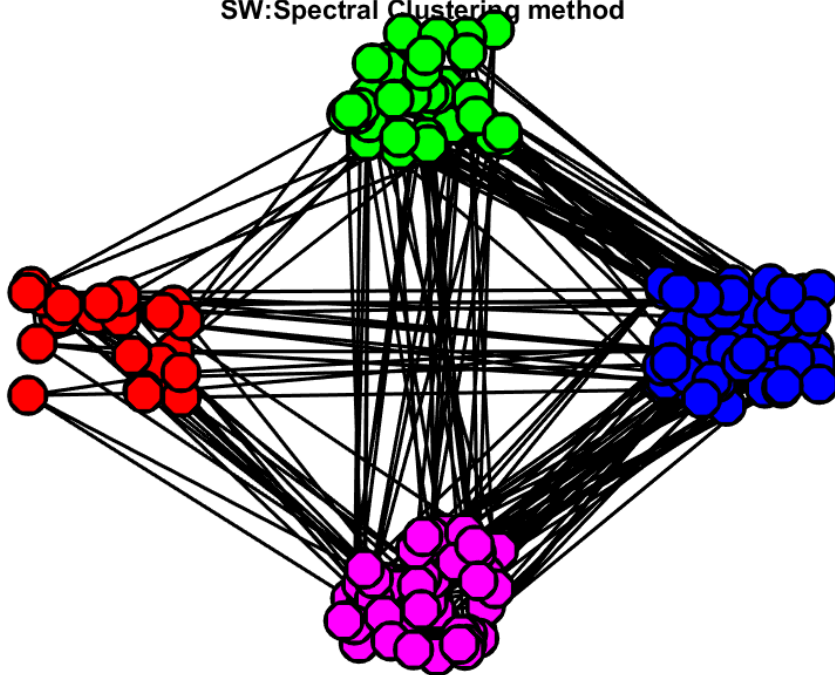
Σχολιασμός Αποτελεσμάτων:

Στην τοπολογία αυτή παρατηρούμε ότι η μέθοδος Modularity Maximization μας δίνει καλύτερα αποτελέσματα σε σχέση με τις άλλες 2 μεθόδους. Ωστόσο, ακόμη και αυτή δεν επιτυγχάνει καλό διαχωρισμό σε κοινότητες με βάση τη μετρική μας. Παρατηρούμε επίσης, ότι η μέθοδος Spectral Clustering δίνει έναν πολύ κακό διαχωρισμό, όπως μας αποδεικνύει και η μετρική μας. Τα συγκεκριμένα κακά αποτελέσματα οφείλονται στο γεγονός ότι στις τοπολογίες scale free κόμβοι με μεγάλο βαθμό τείνουν να συνδέονται μεταξύ τους και να σχηματίζουν κοινότητα. Παίζει σημαντικό ρόλο, λοιπόν, η κατανομή των βαθμών στην τοπολογία.

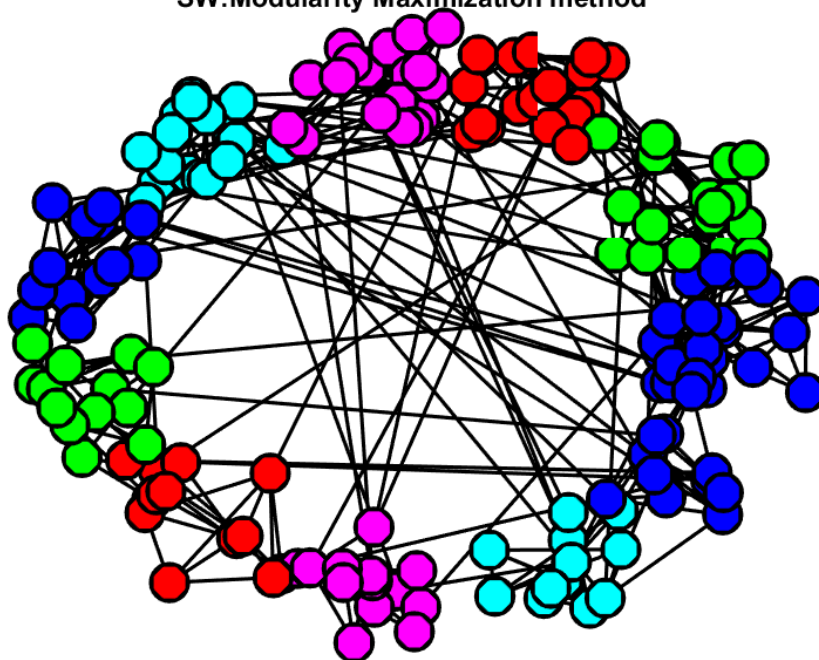
SW:



SW:Spectral Clustering method



SW:Modularity Maximization method



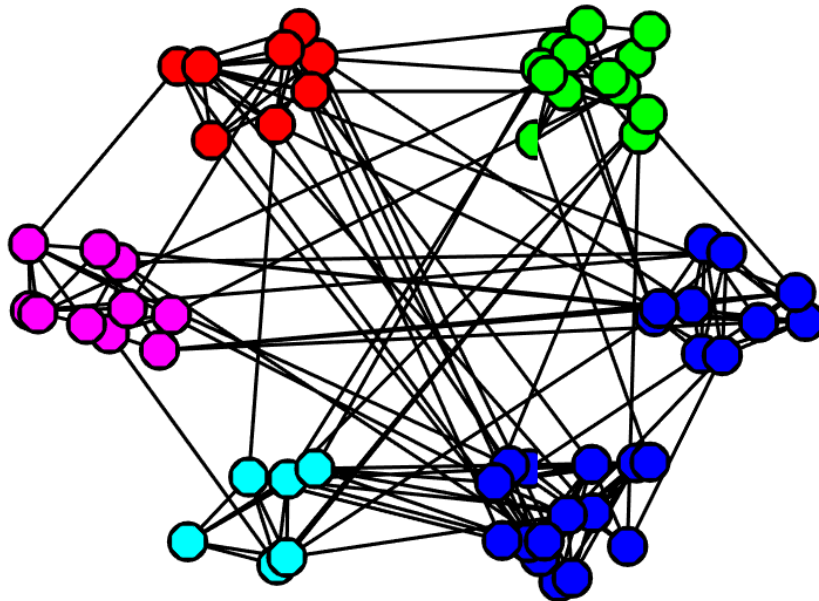
Μέθοδος↓	Modularity↓
Newman-Girvan	0.6403
Spectral Clustering	0.2925
Modularity Maximization	0.7033

Σχολιασμός Αποτελεσμάτων:

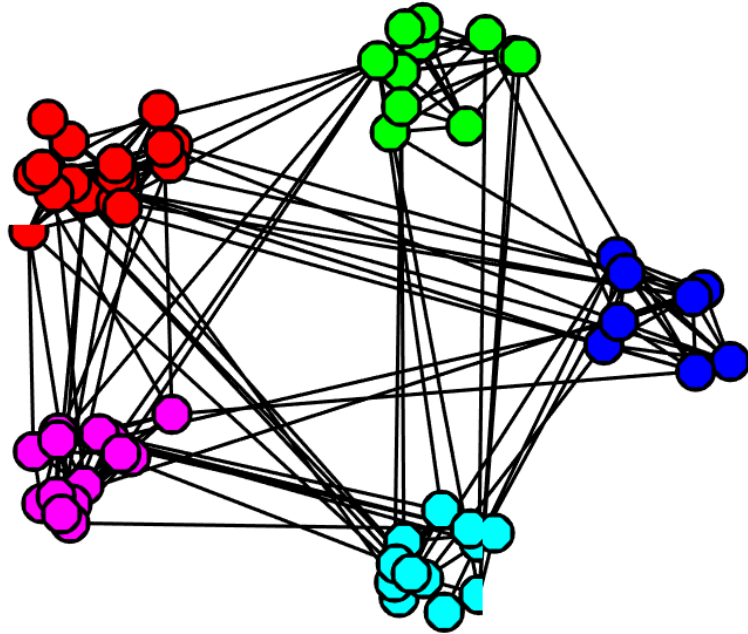
Στην τοπολογία αυτή Small World έχουμε πιθανότητα ανασύνδεσης $g_p = 0.3$ και όπως είδαμε και στην προηγούμενη άσκηση, όσο μικραίνει η πιθανότητα, τείνει η Small World τοπολογία να προσεγγίζει την REG. Επομένως δεν πρέπει να μας παραξενεύει το γεγονός ότι έχουμε εδώ παρεμφερή αποτελέσματα με την REG τοπολογία. Δηλαδή, έχουμε παρόμοια απόδοση με τις μεθόδους Newman-Girvan και Modularity Maximization, ενώ σαφώς χειρότερα αποτελέσματα δίνει η μέθοδος Spectral Clustering. Αν και η τελευταία βρίσκει λίγες σχετικά κοινότητες, όπως φαίνεται και από την ανωτέρω οπτικοποίηση, υπάρχουν πολλές συνδέσεις μεταξύ τους, γεγονός στο οποίο οφείλεται η χαμηλή τιμή της αξιολόγησης του διαχωρισμού που παίρνει.

DOLPHINS:

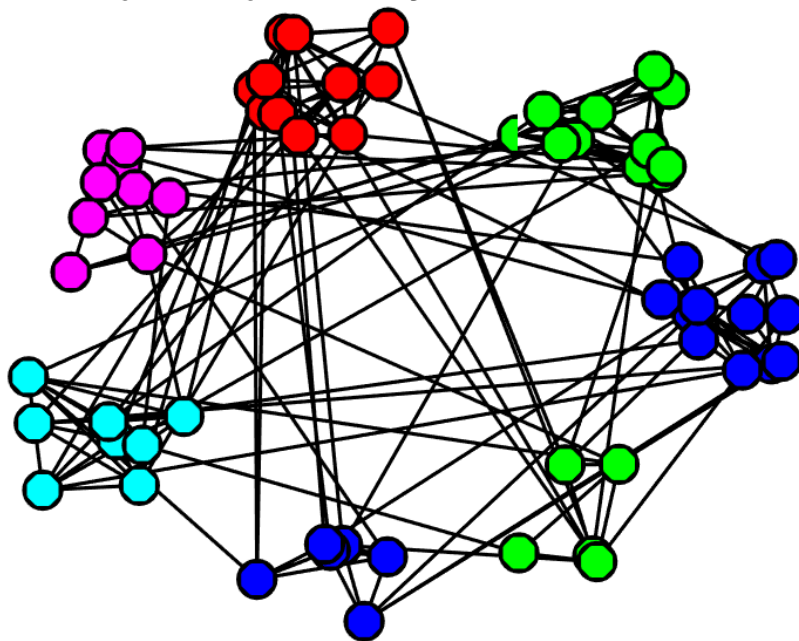
Dolphins Graph:Newman-Girvan method



Dolphins Graph:Spectral Clustering method



Dolphins Graph:Modularity Maximization method



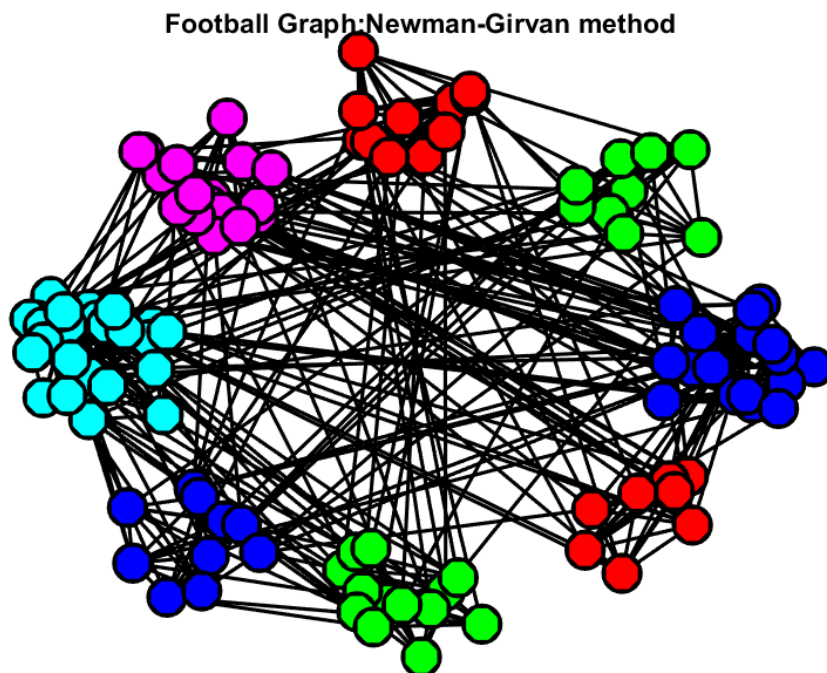
Μέθοδος↓	Modularity↓
Newman-Girvan	0.4163
Spectral Clustering	0.5166
Modularity Maximization	0.5540

Σχολιασμός Αποτελεσμάτων:

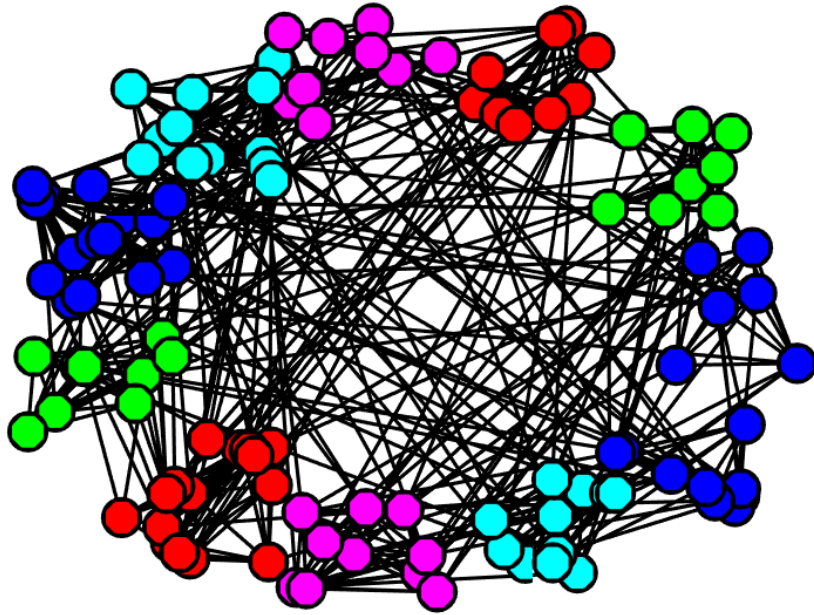
Παρατηρούμε εδώ ότι και οι 3 μέθοδοι δίνουν αποτελέσματα που δεν απέχουν πολύ μεταξύ τους. Ωστόσο, από τη μετρική φαίνεται πως λίγο καλύτερη είναι η μέθοδος Modularity Maximization ενώ υστερεί κάπως η μέθοδος Newman-Girvan. Μπορούμε να πούμε ότι τα αποτελέσματα με βάση τη μετρική μας θυμίζουν την RGG τοπολογία. Είναι σημαντικό να αναφέρουμε ότι και οι 3 μέθοδοι μας έδωσαν περίπου τον ίδιο αριθμό κοινοτήτων

- ✚ Newman-Girvan 6 κοινότητες
- ✚ Spectral Clustering 5 κοινότητες
- ✚ Modularity Maximization 7 κοινότητες

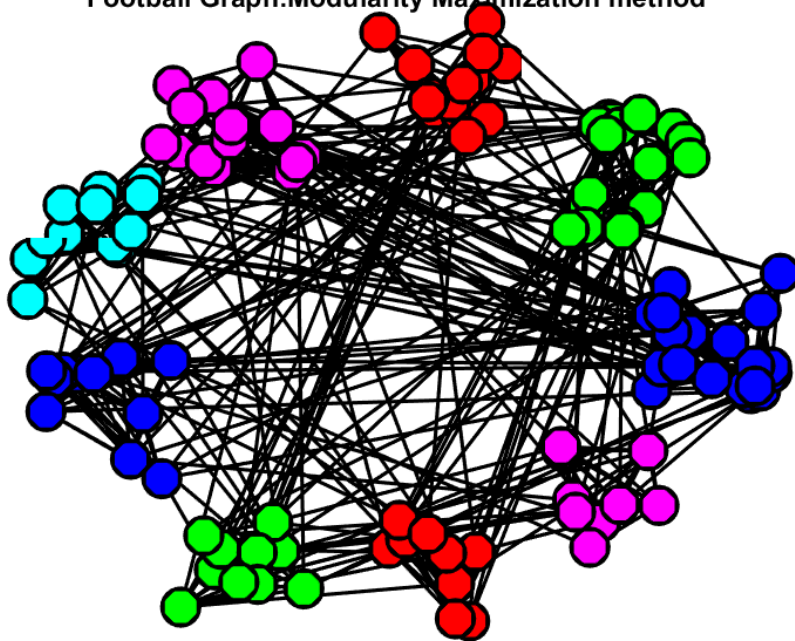
FOOTBALL:



Football Graph:Spectral Clustering method



Football Graph:Modularity Maximization method



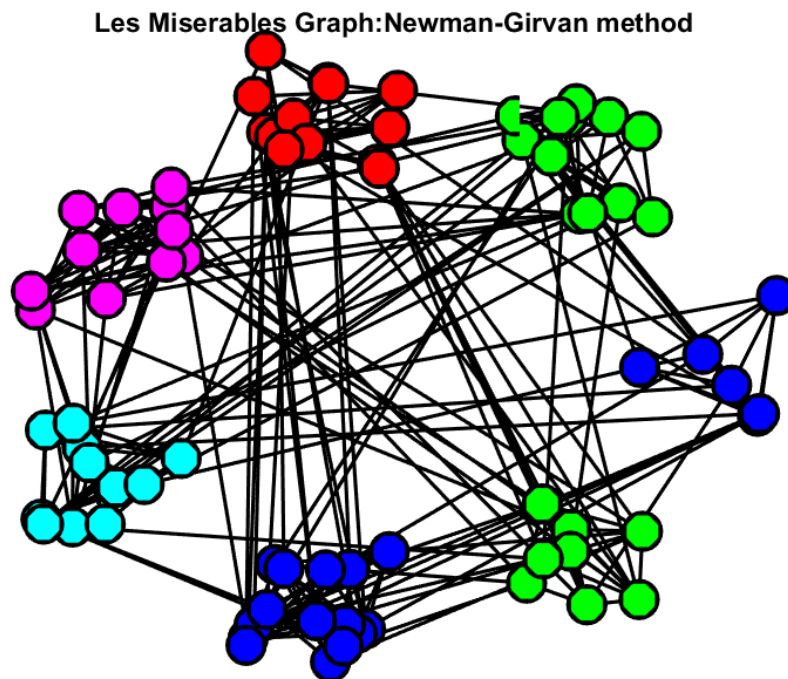
Μέθοδος↓	Modularity↓
Newman-Girvan	0.4914
Spectral Clustering	0.5938
Modularity Maximization	0.5976

Σχολιασμός Αποτελεσμάτων:

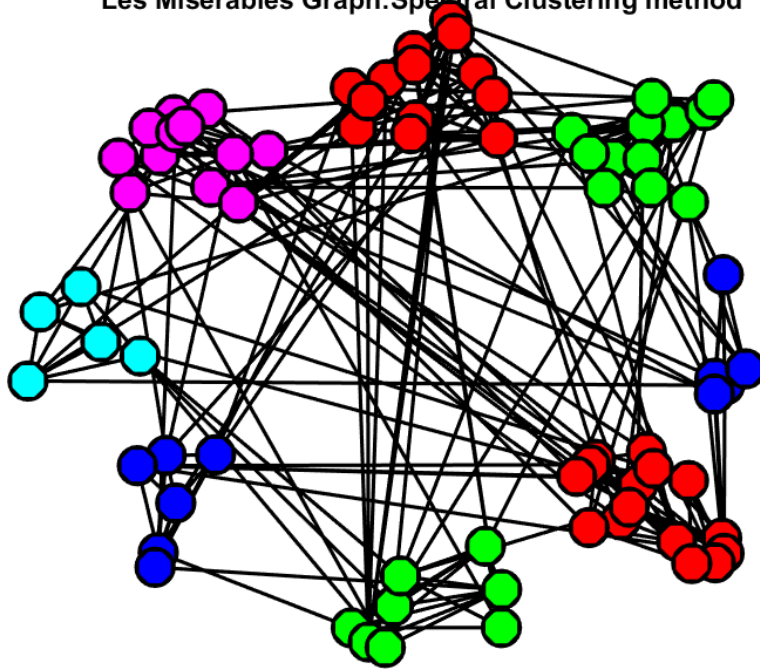
Και σε αυτό το δίκτυο, τα αποτελέσματα μοιάζουν με αυτά της RGG τοπολογίας. Οι μέθοδοι Spectral Clustering και Modularity Maximization δίνουν «παρόμοια» αποτελέσματα και είναι καλύτερες σε σχέση με τη μέθοδο Newman-Girvan. Και στην περίπτωση του δικτύου αυτού, βλέπουμε ότι οι μέθοδοί μας εντοπίζουν κοντινό αριθμό κοινοτήτων.

- ✚ Newman-Girvan 8 κοινότητες
- ✚ Spectral Clustering 10 κοινότητες
- ✚ Modularity Maximization 9 κοινότητες

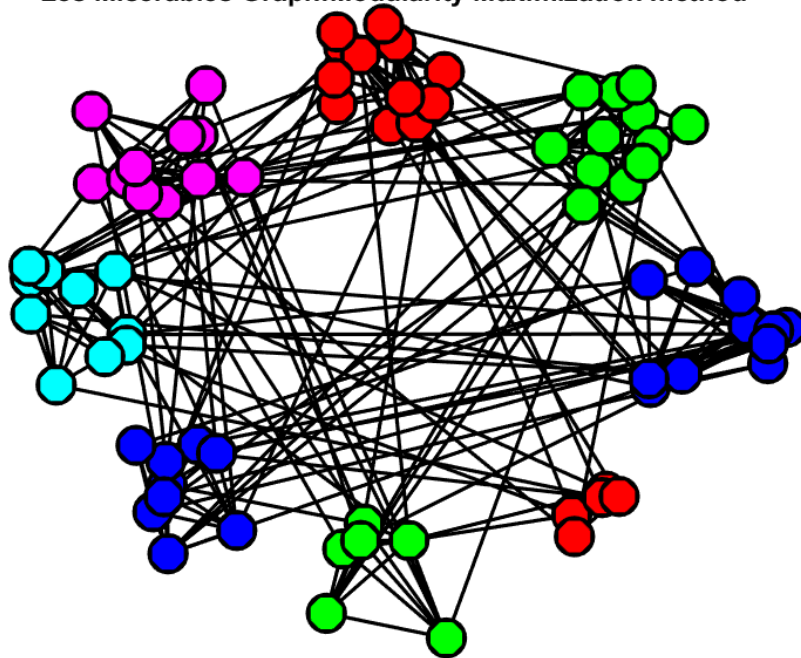
LES MISERABLES:



Les Miserables Graph:Spectral Clustering method



Les Miserables Graph:Modularity Maximization method



<u>Μέθοδος↓</u>	<u>Modularity↓</u>
Newman-Girvan	0.4364
Spectral Clustering	0.5381
Modularity Maximization	0.5670

Σχολιασμός Αποτελεσμάτων:

Και σε αυτό το δίκτυο των «Αθλίων» παρατηρούμε παρόμοια αποτελέσματα με αυτά της RGG συνθετικής τοπολογίας, όπως είδαμε και στο δίκτυο Dolphins και Football. Πάλι, δηλαδή, οι μέθοδοι Spectral Clustering και Modularity Maximization κάνουν παρόμοιο διαχωρισμό σε κοινότητες και είναι καλύτερες σε σχέση με τη μέθοδο Newman-Girvan. Και εδώ οι 3 μέθοδοι εντοπίζουν περίπου τον ίδιο αριθμό κοινοτήτων (Spectral Clustering και Modularity Maximization έρχονται σε απόλυτη συμφωνία)

- ✚ Newman-Girvan 7 κοινότητες
- ✚ Spectral Clustering 8 κοινότητες
- ✚ Modularity Maximization 8 κοινότητες

Σα συνολική παρατήρηση θα μπορούσαμε να πούμε ότι βλέπουμε πως οι πραγματικές τοπολογίες δικτύων μοιάζουν πολύ με την τοπολογία RGG , όσον αφορά το διαχωρισμό σε κοινότητες και στην απόδοση των αλγορίθμων διαχωρισμού που χρησιμοποιήσαμε. Ακόμη, μπορούμε να δούμε ότι η μέθοδος Spectral Clustering στις συνθετικές τοπολογίες δεν έδινε γενικά πολύ καλά αποτελέσματα ενώ στα πραγματικά δίκτυα η απόδοσή της βελτιώνεται σημαντικά. Τέλος, παρατηρούμε ότι στις πραγματικές τοπολογίες η μέθοδος Newman-Girvan κάνει τον (σχετικά) «χειρότερο» διαχωρισμό, πράγμα που δεν ίσχυε στις συνθετικές τοπολογίες που εξετάσαμε παραπάνω, όπου πολλές φορές έδινε ικανοποιητικά αποτελέσματα.