

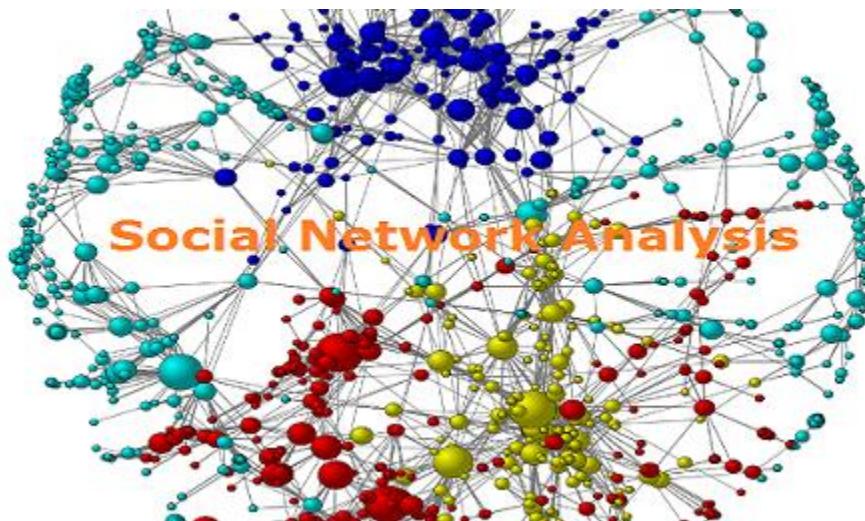
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Ανάλυση Κοινωνικών Δικτύων

1η Εργαστηριακή Άσκηση:

Ανάλυση & Μελέτη Σύνθετων Τοπολογιών Δικτύου



Ονοματεπώνυμο: Σταυρακάκης Δημήτριος

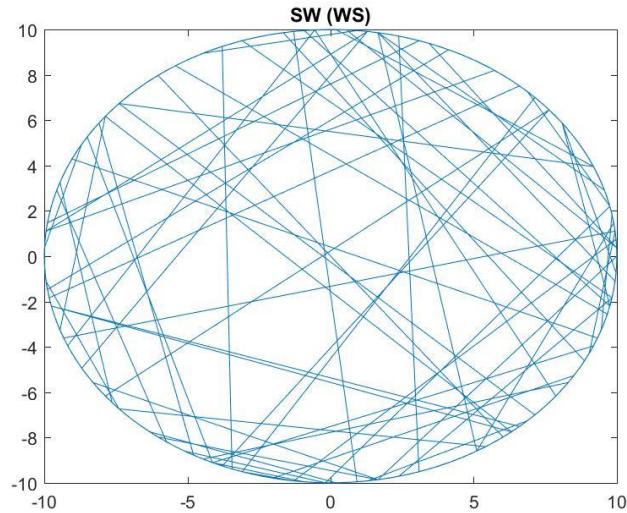
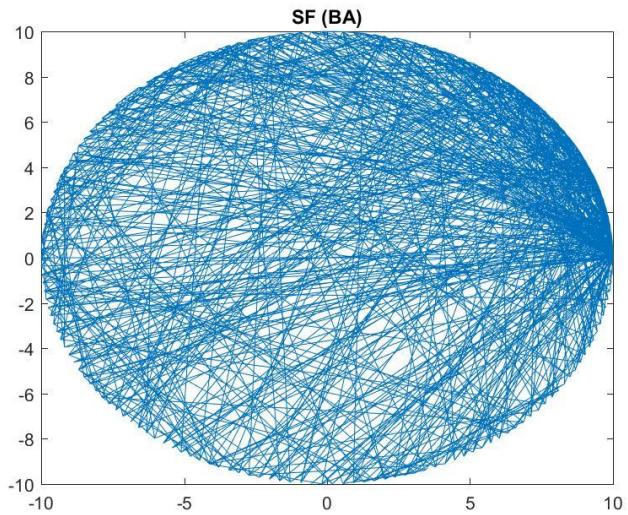
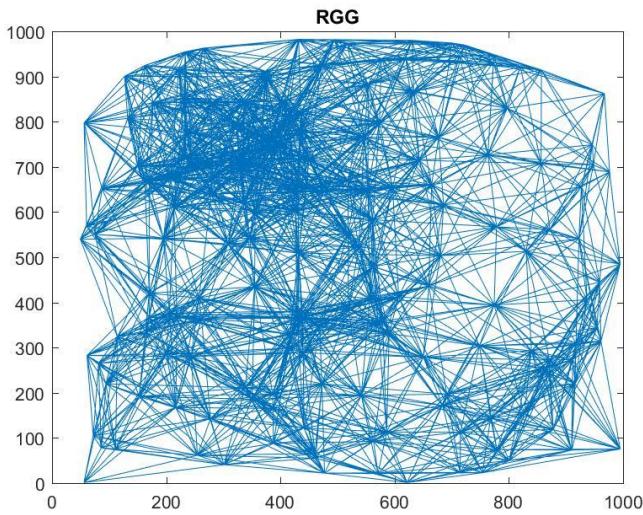
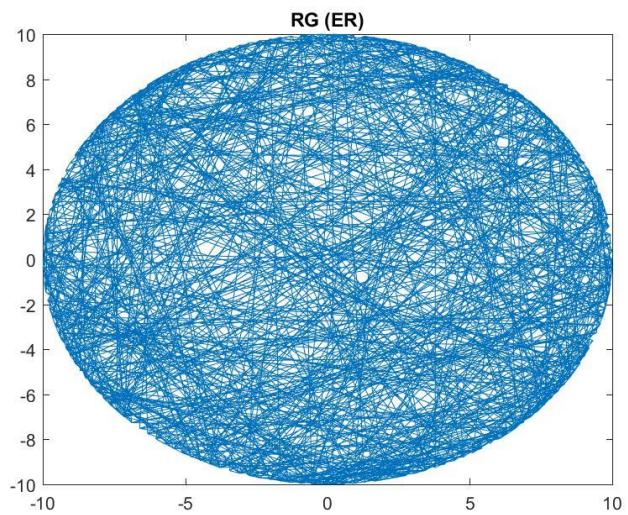
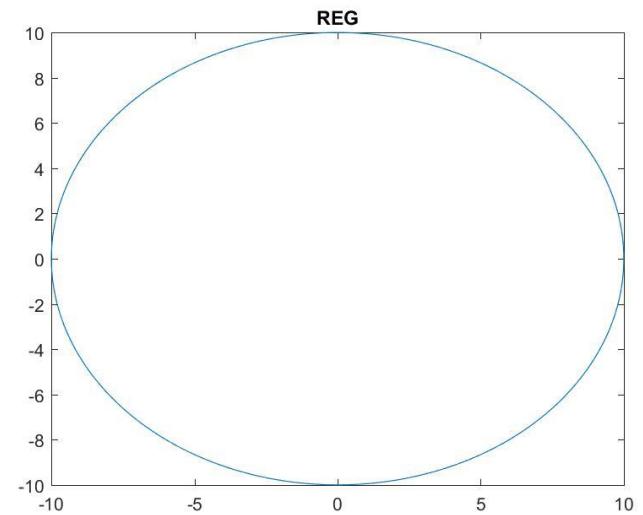
ΑΜ:03112017

Εξάμηνο: 9^ο

Ημερομηνία Παράδοσης: 14/12/2016

A) Ανάλυση και οπτικοποίηση σύνθετων τύπων δικτύου

Στο ερώτημα αυτό δημιουργήθηκαν και οπτικοποιούνται παρακάτω οιτοπολογίες δικτύων με τις παραμέτρους που ζητούνται.(αυτές που αναφέρονται στον Πίνακα 2 της εκφώνησης) Η δημιουργία τους έγινε με χρήση των συναρτήσεων που μας παρέχονται ορίζοντας κατάλληλα τις παραμέτρους που παίρνουν σαν ορίσματα. Ακολουθούν οιτοπολογίες που προέκυψαν:



Β) Μελέτη Βαθμού Κόμβων

❖ **Να υπολογιστεί ο βαθμός κάθε κόμβου**

Παρατίθενται παρακάτω τα αποτελέσματα που προέκυψαν για όλες τις ανωτέρω τοπολογίες:

(Για τον υπολογισμό χρησιμοποιήθηκε η συνάρτηση degrees.m)

✓ REG (n=170 , d=4)

✓ RG (ER) (N=170 , M=750)

✓ RGG ($L \times L = 1000^2$, n=170, R=250)

✓ SF (BA) (n=170 , d=4)

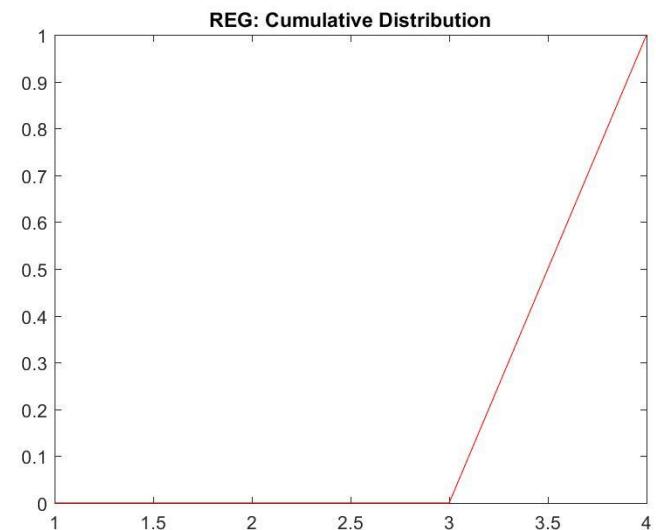
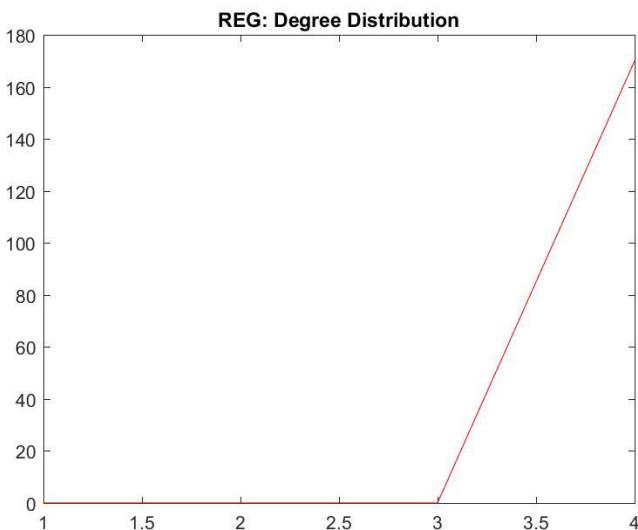
✓ SW (WS) ($n=170$, $d=4$, $g_p=0.3$)

- ❖ **Να αναπαρασταθεί (plot) η κατανομή βαθμών κόμβων του δικτύου (δηλαδή, για κάθε βαθμό πόσοι κόμβοι έχουν αυτό το βαθμό) καθώς και η συγκεντρωτική κατανομή βαθμού κόμβου (δηλαδή, για κάθε βαθμό πόσοι κόμβοι έχουν το πολύ αυτό το βαθμό)**

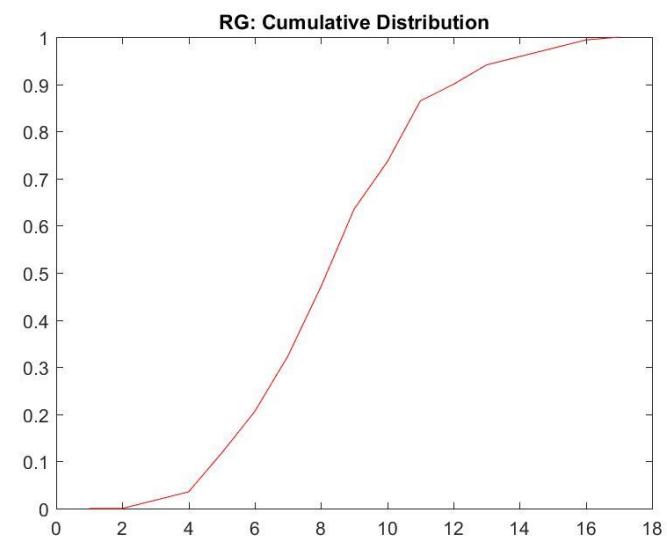
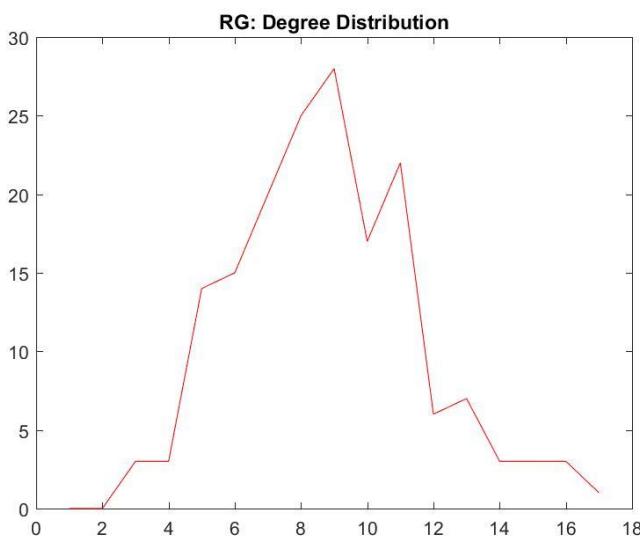
Παρατίθενται παρακάτω αποτελέσματα που προέκυψαν για όλες τις ανωτέρω τοπολογίες:

(Για τον υπολογισμό χρησιμοποιήθηκε η συνάρητση cumulativedist.m)

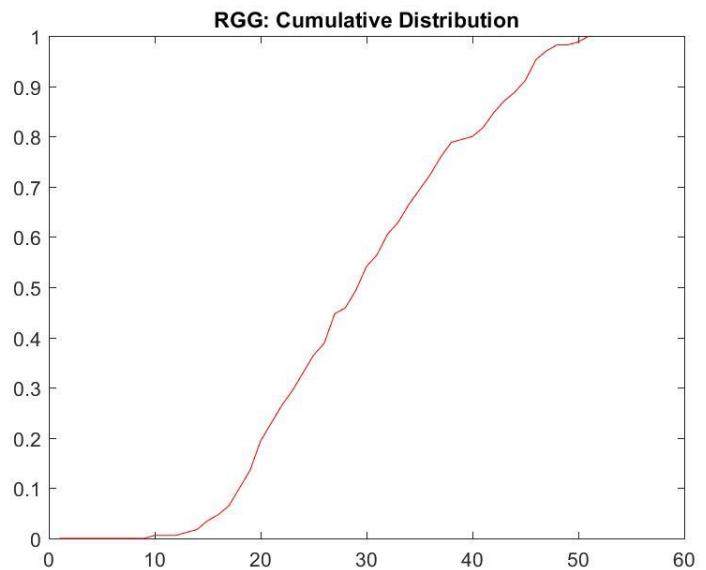
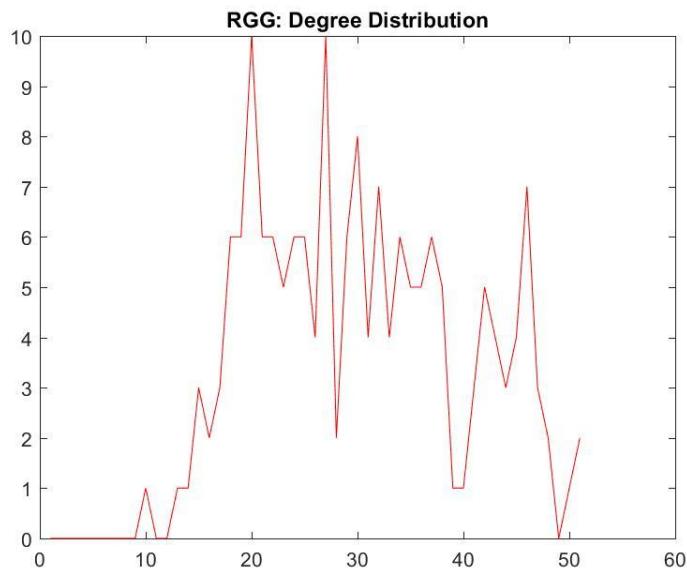
- ✓ REG ($n=170$, $d=4$)



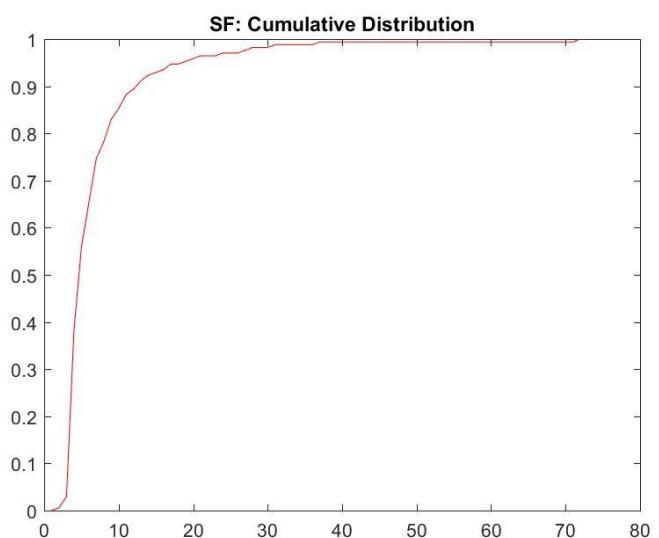
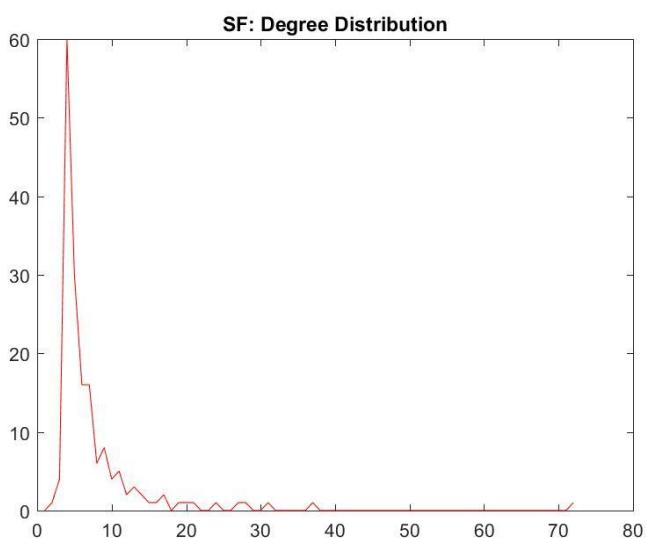
- ✓ RG (ER) ($N=170$, $M=750$)



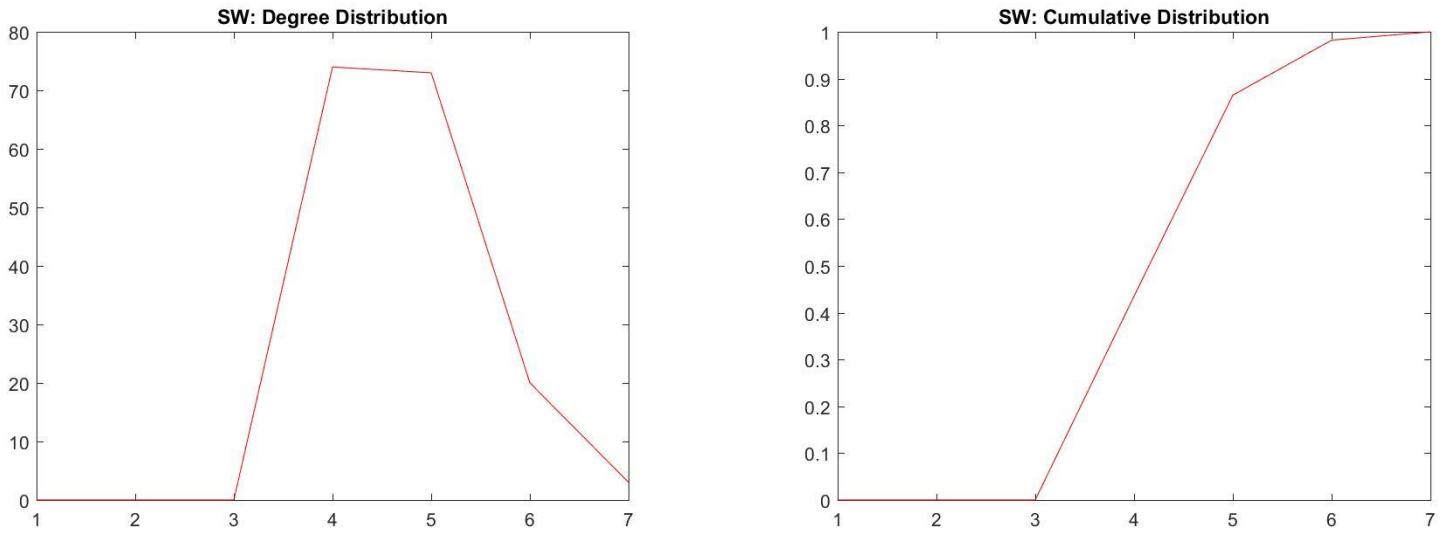
✓ RGG ($L \times L = 1000^2$, n=170, R=250)



✓ SF (BA) (n=170, d=4)



- ✓ SW (WS) ($n=170$, $d=4$, $g_p=0.3$)



- ❖ **Να υπολογιστεί ο μέσος βαθμός κόμβου καθώς και η διασπορά των βαθμών κόμβου για κάθε μια από τις εν λόγω τοπολογίες.**

Παρατίθενται παρακάτω αποτελέσματα που προέκυψαν για όλες τις ανωτέρω τοπολογίες:

(Για τον υπολογισμό χρησιμοποιήθηκαν οι built-in συναρτίσεις του Matlab mean και var)

Τοπολογία Δικτύου:	REG	RG	RGG	SF	SW
Μέση Τιμή Βαθμών	4	8.8235	30.3294	7.4235	4.7176
Διασπορά Βαθμών	0	7.7320	93.7252	52.2456	0.5470

- ❖ **Σχολιασμός αποτελεσμάτων**

- Για τον REG γράφο τα αποτελέσματα που λάβαμε είναι αυτά που περιμέναμε καθώς ο βαθμός κάθε κόμβου είναι 4 ακριβώς, επομένως η μέση τιμή σωστά προκύπτει 4 και η τυπική απόκλιση 0 (αφού δεν υπάρχει απόκλιση από τη μέση τιμή). Η κατανομή είναι προφανώς κανονική.
- Για τον RG (Erdos-Renyi) γράφο αυτό που έρουμε είναι ότι ο μέσος όρος θα παραμείνει πάντα ο ίδιος αφού ορίζουμε να έχουμε 750 ακμές για 170 κόμβους. Επομένως, όσες φορές και να αναπαράγουμε το δίκτυό μας, η μέση τιμή θα παραμείνει: $2 * 750 / 170 = 8.8235$ (ο πολλαπλασιαστικός παράγοντας 2 υπάρχει καθώς το άθροισμα των βαθμών όλων των κόμβων σε ένα μη κατευθυνόμενο γράφημα είναι ίσο με το διπλάσιο των ακμών που βρίσκονται σε αυτό) Αντίθετα, η τυπική απόκλιση μεταβάλλεται κάθε φορά που τρέχουμε τον κώδικα μας. Ωστόσο, παρατηρούμε ότι ο μεγαλύτερος αριθμός των κόμβων έχει βαθμό πολύ κοντά στη μέση τιμή που λαμβάνουμε.
- Για τον RGG γράφο παρατηρούμε ότι έχουμε τόσο αυξημένη μέση τιμή όσο και τυπική απόκλιση. Αυτό οφείλεται στην τυχαιότητα με την οποία δημιουργείται η εν λόγω τοπολογία δικτύου. Η κατανομή των ακμών γίνεται στην τύχη και εξ αυτού δικαιολογείται και η μεγάλη

τιμή στην τυπική απόκλιση. (Άλλοι κόμβοι έχουν πολλές ακμές ενώ άλλοι λίγες) Προφανώς έχουμε πολλούς κόμβους με μεγάλο βαθμό, γι' αυτό προκύπτει και μεγάλη η μέση τιμή.

- Για τον SF (BA) γράφο παρατηρούμε ότι η μέση τιμή λαμβάνει χαμηλή τιμή ενώ η τυπική απόκλιση είναι αρκετά υψηλή (κοντά στην τιμή της περίπτωσης του RGG γράφου) Το ότι η μέση τιμή βρίσκεται σε αυτό το επίπεδο δικαιολογείται από το γεγονός ότι ορίσαμε ως βαθμό αρχικού πλέγματος την τιμή 4. Θα έχουμε τελικά αρκετές κορυφές του γράφου με βαθμό 4.
 - Για τον SW (WS) γράφο παρατηρούμε ότι η μέση τιμή είναι αρκετά κοντά στο 4 και η τυπική απόκλιση λαμβάνει πολύ χαμηλή τιμή. Προσεγγιστικά μοιάζει με την τοπολογία REG. Οφείλεται στην τιμή της πιθανότητας ανασύνδεσης που είναι μικρή. Όσο εκείνη μεγαλώνει, τόσο θα μεγαλώνει και η απόκλιση που θα μας δίνει η συγκεκριμένη τοπολογία (πράγμα λογικό)

Γ) Δίκτυα με βάρη

Αφού χρησιμοποιήσαμε τη built-in συνάρτηση `rand()` του Matlab δημιουργήσαμε τον πίνακα βαρών για τους γράφους μας. Το εύρος των βαρών είναι [1,10] όπως υποδεικνύεται και στην εκφώνηση της άσκησης. Πολλαπλασιάσαμε στοιχείο προς στοιχείο τον πίνακα αυτό με τους πίνακες γειτνίασης που είχαμε ως τώρα στην κατοχή μας και πήραμε έτσι το κάθε δίκτυο με τα κατάλληλα βάρη στις ακμές του. Στη συνέχεια, κάνοντας χρήση της συνάρτησης `strength_graph()` που δημιουργήσαμε, (η οποία στην ουσία κάνει χρήση της `degrees()` και της `cumulative_dist()`) Που χρησιμοποιήσαμε και στα προηγούμενα ερωτήματα) πήραμε τα αποτελέσματα που μας ζητούνται στο συγκεκριμένο κομμάτι της άσκησης μας, τα οποία παρατίθενται στη συνέχεια:

- ❖ **Να υπολογιστεί η δύναμη κάθε κόμβου**
 - ✓ REG (n=170 , d=4)

✓ RG (ER) (N=170 , M=750)

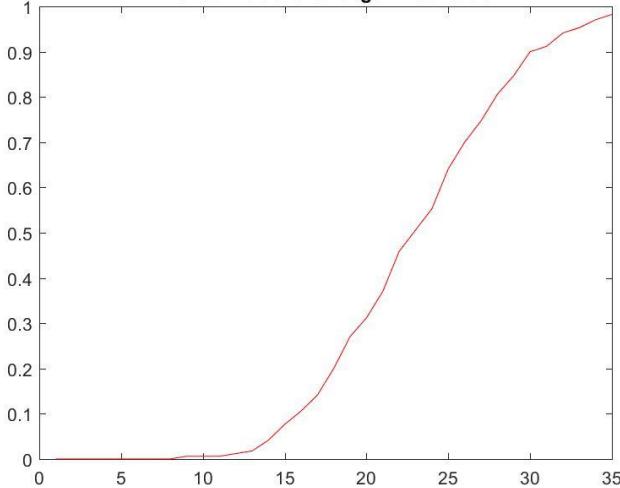
✓ RGG ($L \times L = 1000^2$, n=170, R=250)

✓ SF (BA) (n=170 , d=4)

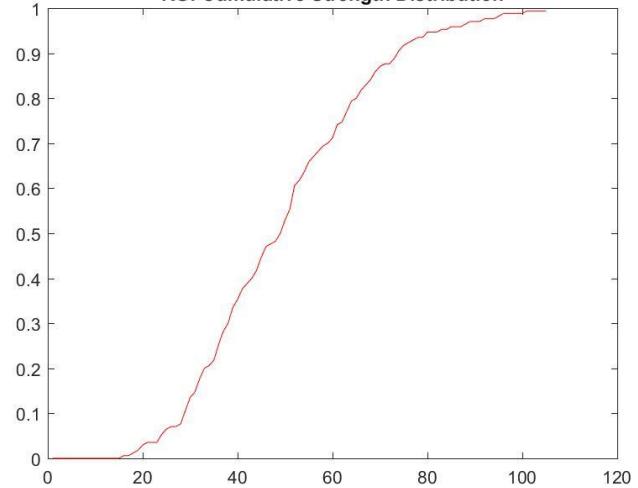
✓ SW (WS) ($n=170$, $d=4$, $g_p=0.3$)

- ❖ ***Να αναπαρασταθεί η συγκεντρωτική κατανομή δύναμης (cumulative strength distribution) των κόμβων του δικτύου (σε κάθε τιμή δύναμης αντιστοιχίζεται το πλήθος των κόμβων με δύναμη μικρότερη ή ίση από αυτή)***

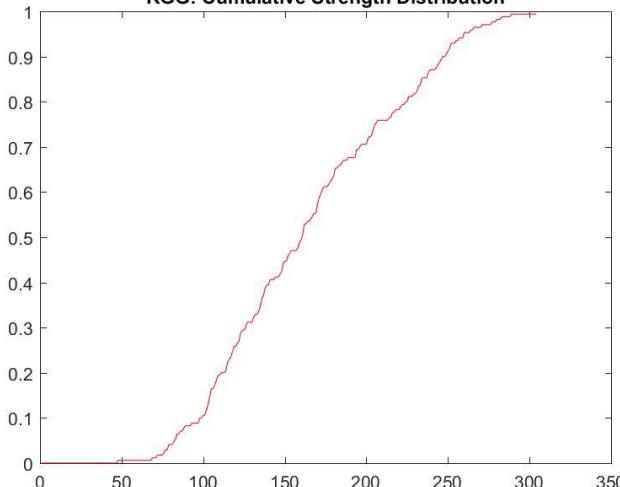
REG: Cumulative Strength Distribution



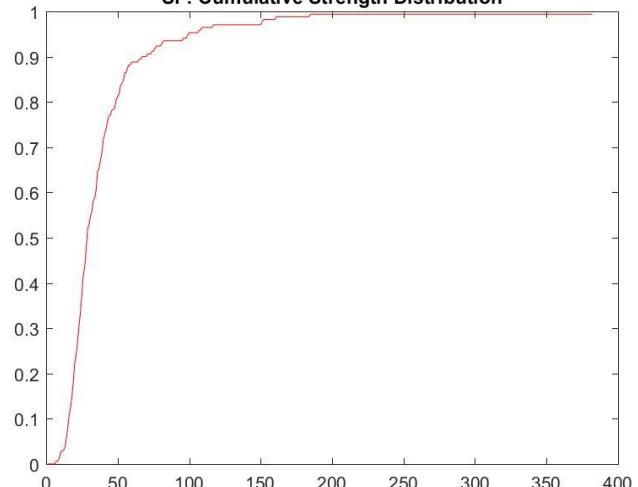
RG: Cumulative Strength Distribution



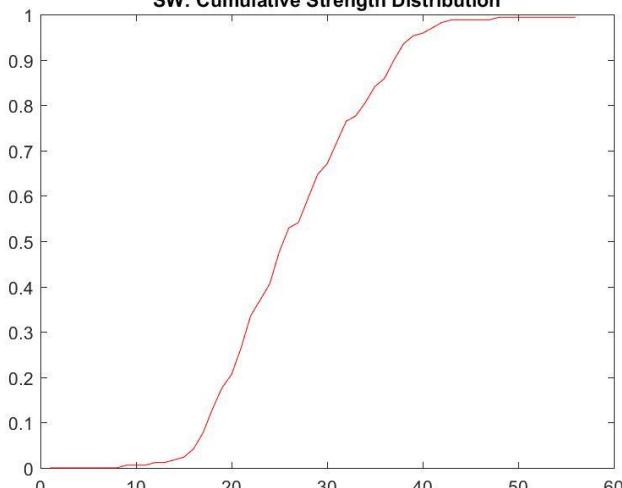
RGG: Cumulative Strength Distribution



SF: Cumulative Strength Distribution



SW: Cumulative Strength Distribution



❖ **Να υπολογιστεί η μέση δύναμη για όλους τους κόμβους**

Τοπολογία Δικτύου:	REG (n=170,d=4)	RG (N=170,M=50)	RGG ($L \times L = 1000^2$, n=170 , R=250)	SF (n=170 , d=4)	SW (n=170 , d=4 , g _p =0.3)
Μέση Δύναμη Κόμβων	21.1090	48.2629	167.0274	40.2137	24.8729

Δ) Υπολογισμός μέσου μήκους μονοπατιού

Στη συνέχεια καλούμαστε για τις τοπολογίες που έχουμε ήδη παράξει, να υπολογίσουμε το μέσο μήκος μονοπατιού και τη διασπορά στο δίκτυο. Για την εύρεση αυτών χρησιμοποιήθηκε η συνάρτηση ave_path_length() που μας παρέχεται. Η συνάρτηση αυτή μας επιστρέφει τόσο το μέσο μήκος μονοπατιού όσο και τη διασπορά. Τα αποτελέσματα που εξάγαμε, παρουσιάζονται συγκεντρωτικά (όπως επισημαίνεται στην εκφώνηση) στη συνέχεια:

Τοπολογία Δικτύου:	REG	RG	RGG	SF	SW
Μέσο Μήκος Μονοπατιού	21.6272	2.5882	2.6679	2.5366	4.4876
Διασπορά Μήκους Μονοπατιού	150.6317	0.4609	1.3095	0.4318	2.0344

Σχολιασμός αποτελεσμάτων:

Ο REG γράφος, όπως βλέπουμε και παραπάνω, έχει πολύ μεγαλύτερες τιμές τόσο στο μέσο μήκος μονοπατιού όσο και στη διασπορά συγκριτικά με τις άλλες τοπολογίες. Αυτό οφείλεται στο γεγονός ότι οι συνδέσεις είναι προκαθορισμένες. Κάθε κόμβος συνδέεται με τους 2 κοντινότερους γείτονες δεξιά και αριστερά (ο ν-οστός κόμβος συνδέεται με τους ν+1, ν+2, ν-1, ν-2). Αυτό κάνει τη διάμετρο του γράφου αυτού μεγάλη και δικαιολογεί τα αποτελέσματα που λαμβάνουμε.

Για τον RGG γράφο, λόγω της μεγάλης τυχαιότητας αναμέναμε μεγάλο μέσο μονοπάτι. Στα αποτελέσματά μας, το μέσο μήκος μονοπατιού του RGG γράφου είναι ναι μεν πιο μεγάλο από αυτό που πήραμε για τους SF (BA) και RG αλλά είναι μικρότερο από αυτό που μας έδωσε η τοπολογία SW (WS). Σε αυτό βοήθησε το γεγονός ότι έχουμε πολλές συνδέσεις. Επίσης, βλέπουμε ότι έχουμε και μια σχετικά μεγάλη τυπική απόκλιση που οφείλεται κυρίως στην τυχαιότητα με την οποία δημιουργείται αυτή η τοπολογία.

Για τον γράφο SF (BA) τα αποτελέσματα είναι όπως αναμέναμε και από τη θεωρία, δηλαδή έχουμε μικρή τιμή στο μέσο μήκος μονοπατιού (λόγω τυχαιότητας, θα υπάρχει κάποια κορυφή που θα κάνει γρήγορα προσβάσμες τις υπόλοιπες).

Παρόμοια αποτελέσματα έχουμε και στον RG γράφο.

Για τον γράφο SW (WS) αναμέναμε να έχουμε χαμηλή τιμή στο μήκος μονοπατιού επίσης. Ωστόσο αυτό δε συμβαίνει καθώς η τιμή που παίρνουμε είναι σχετικά μεγαλύτερη από τις άλλες τοπολογίες (εκτός από τον REG γράφο προφανώς). Επίσης, έχει και αυξημένη τυπική απόκλιση που μας δικαιολογεί ως ένα βαθμό το γεγονός ότι πήραμε λίγο πιο μεγάλο μέσο μήκος μονοπατιού και υποδηλώνει στην ουσία την ύπαρξη μη κεντρικών κορυφών στο δίκτυο μας, που αυξάνουν τόσο το μέσο μήκος μονοπατιού όσο και την τυπική απόκλιση. Γενικά στα SW δίκτυα, τα περισσότερα ζευγάρια κόμβων συνδέονται από τουλάχιστον ένα συντομότερο μονοπάτι, πράγμα που μας κάνει να αναμένουμε αυτη τη χαμηλή τιμή στο μέσο μήκος μονοπατιού. Αυτό προφανώς δε συμβαίνει σε μεγάλο βαθμό στο δικό μας δίκτυο (εξ ού και τα σχετικά «μη αναμενόμενα» αποτελέσματα).

Οι κάποιες αποκλίσεις από τα θεωρητικά αναμενόμενα αποτελέσματα οφείλονται εν μέρει και στον μικρό αριθμό των κόμβων που απαρτίζουν τις τοπολογίες μας.

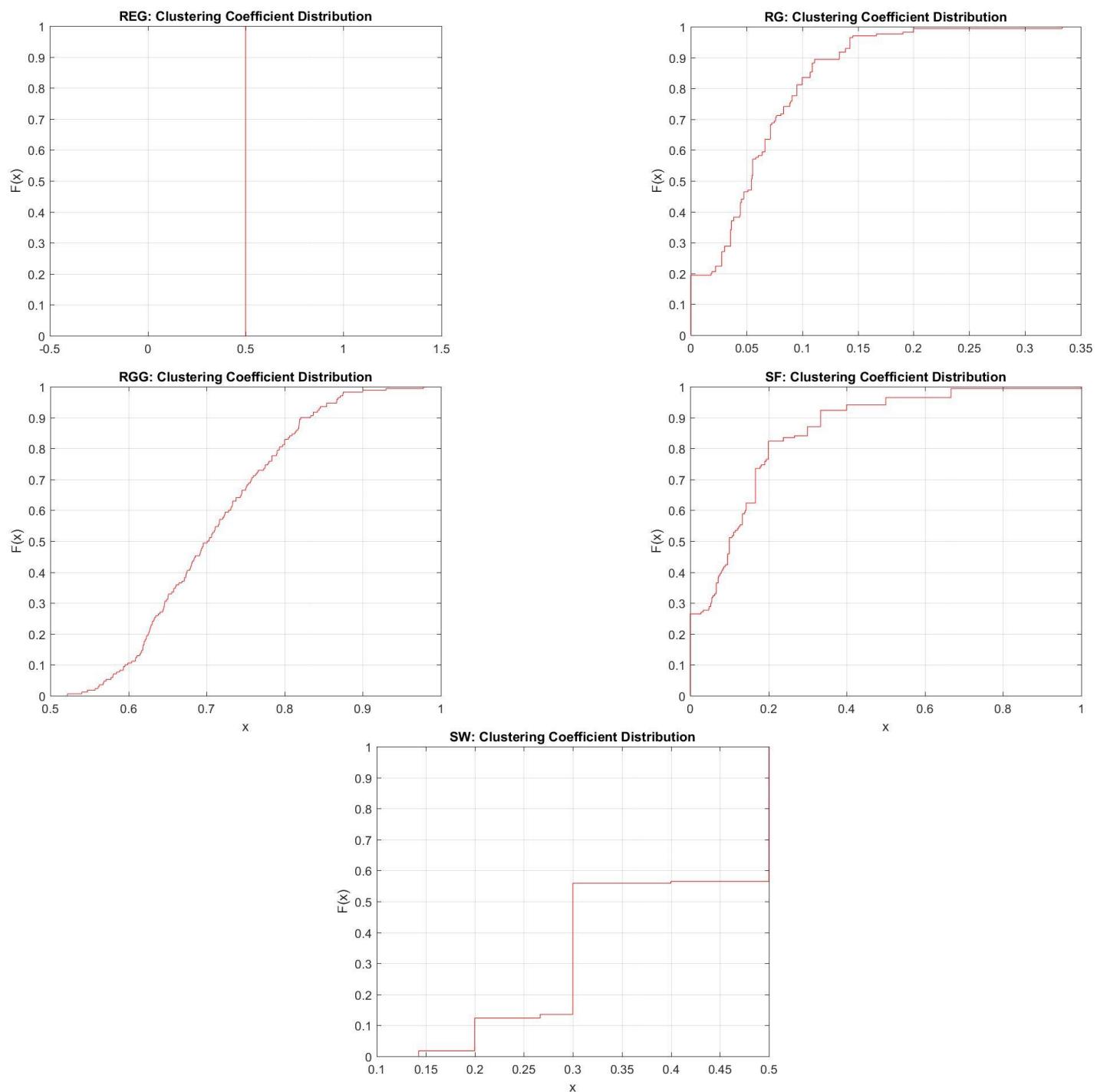
E) Υπολογισμός συντελεστή ομαδοποίησης (ΣΟ)

❖ E.1) Αναλυτικός υπολογισμός του ΣΟ

Ο αναλυτικός υπολογισμός του ΣΟ κάθε κόμβου των τοπολογιών που μας δίδονται βρίσκεται στο τέλος της παρούσας αναφοράς.

❖ E.2) Υπολογισμός ΣΟ σε μεγαλύτερες συνθετικές τοπολογίες με χρήση Matlab

Για τους υπολογισμούς στο ερώτημα αυτό χρησιμοποιήσαμε τη συνάρτηση `clust_coeff()` που μας δίνεται. Η συγκεντρωτική κατανομή του ΣΟ για όλους τους κόμβους των τοπολογιών δικτύου που έχουμε φτιάξει παρατίθεται στη συνέχεια:



Επίσης μας ζητείται και ο υπολογισμός του μέσου ΣΟ κάθε τοπολογίας. Τα αποτελέσματα που λάβαμε από το Matlab (όπως μας τα δίνει η συνάρτηση `clust_coeff()`) είναι τα παρακάτω:

Τοπολογία Δικτύου:	REG (n=170,d=4)	RG (N=170,M=50)	RGG ($L \times L = 1000^2$, n=170 , R=250)	SF (n=170 , d=4)	SW (n=170 , d=4 , g_p=0.3)
Μέσος Συντελεστής Ομαδοποίησης	0.5000	0.0600	0.7084	0.1444	0.3739

Βλέπουμε λοιπόν από τον παραπάνω πίνακα ότι:

Ο REG γράφος έχει σταθερό συντελεστή ομαδοποίησης και ίσο με 0.5 . Αυτό οφείλεται στο γεγονός ότι όλοι οι κόμβοι του έχουν βαθμό σταθερό και ίσο με 4. Αντίστοιχα και οι γείτονές τους.

Επομένως ο συντελεστής ομαδοποίησης θα προκύψει σταθερός και ίσος για όλους τους κόμβους σε μια μέση τιμή , πράγμα που είναι φανερό και στο διάγραμμα παραπάνω.

Στον RG(ER) έχουμε χαμηλότερο συντελεστή ομαδοποίησης. Αυτό μπορεί να ερμηνευτεί αφού η ύπαρξη σύνδεσης μεταξύ 2 κόμβων δε μας εξασφαλίζει και τη σύνδεση μεταξύ των γειτόνων τους.

Στον RGG γράφο παρατηρούμε ότι έχουμε τον υψηλότερο συντελεστή ομαδοποίησης, λόγω της τυχαιότητας της εισαγωγής των ακμών που έχουμε σε αυτήν την τοπολογία , η οποία αυξάνει την τιμή του συντελεστή ομαδοποίησης.

Ο γράφος SF(BA) έχει μεγαλύτερο συντελεστή ομαδοποίησης από τον RG(Erdos-Renyi) , ωστόσο η τιμή του είναι αρκετά χαμηλή. Δεν ξεπερνά την αντίστοιχη τιμή σε κανένα από τα άλλα δίκτυα.

Η τοπολογία δικτύου SW(WS) έχει το δεύτερο μεγαλύτερο συντελεστή ομαδοποίησης, μετά τον γράφο REG. Αυτό οφείλεται στο ότι τα δίκτυα με αυτή την τοπολογία τείνουν να έχουν κλίκες -ή σχεδόν κλίκες-, δηλαδή υποδίκτυα με συνδέσεις μεταξύ σχεδόν όλων των ζευγαριών κόμβων μέσα σε αυτά. Επίσης, ρόλο σε αυτό παίζει και η πιθανότητα ανασύνδεσης των ακμών που υπάρχει.

Z) Υπολογισμός κεντρικότητας κόμβων

❖ Z.1) Αναλυτικός υπολογισμός της κεντρικότητας

Ο αναλυτικός υπολογισμός των κεντρικοτήτων κάθε δικτύου που μας δίνεται βρίσκεται στο τέλος της παρούσας αναφοράς.

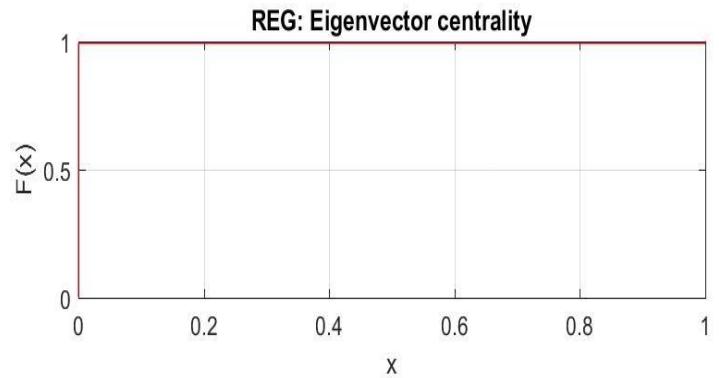
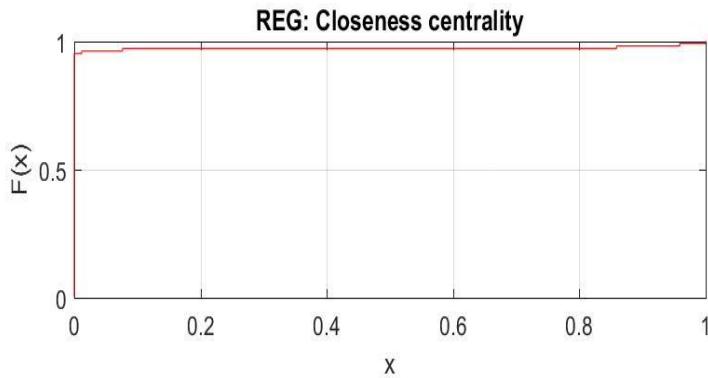
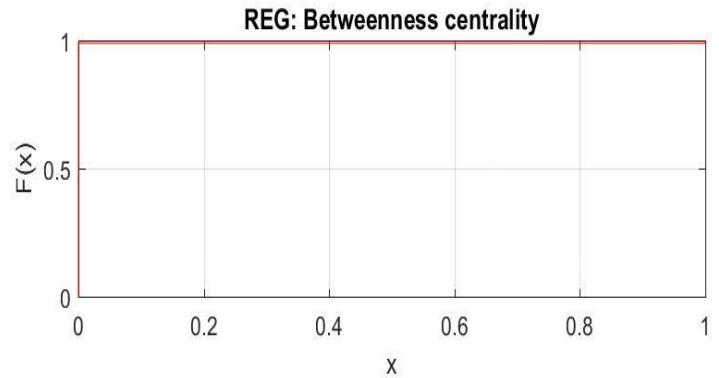
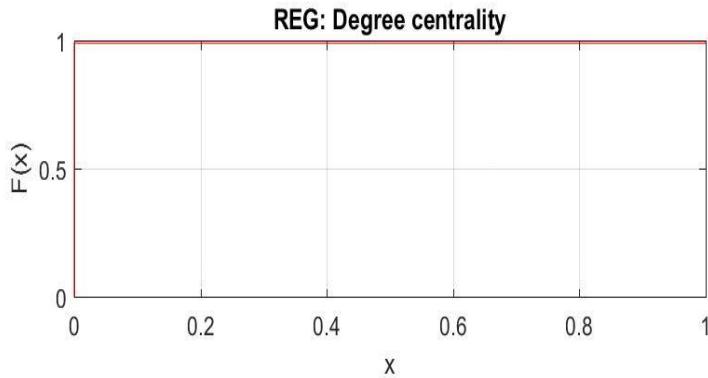
❖ Z.2) Υπολογισμός της κεντρικότητας σε μεγαλύτερες συνθετικές τοπολογίες με χρήση Matlab

Για τους υπολογισμούς στο ερώτημα αυτό χρησιμοποιήσαμε τις συναρτήσεις:

- 1)degrees()
- 2)closeness()
- 3)node_betweenness_faster()
- 4)egeincentrality()
- 5)cumulative_centrality() (για τον υπολογισμό της συγκεντρωτικής κατανομής κεντρικότητας)

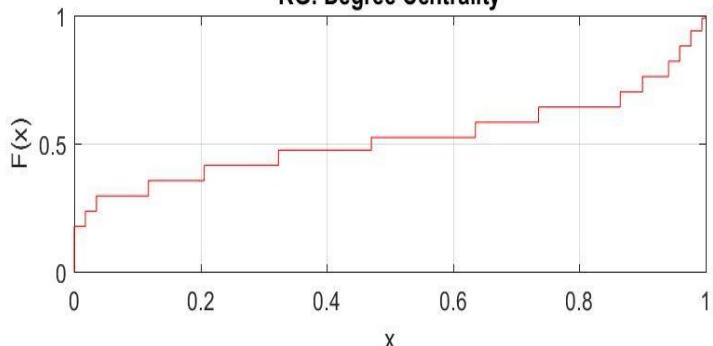
Από τα ονόματα των ανωτέρω συναρτήσεων που μας δίνονται, γίνεται φανερό ποια κεντρικότητα αναλαμβάνει να υπολογίσει έκαστη. Στη συνέχεια, όπως ζητείται, απεικονίζεται η συγκεντρωτική κατανομή της κάθε κατηγορίας κεντρικότητας των κόμβων κάθε τοπολογίας από αυτές που έχετάζουμε:

- ✓ REG (n=170 , d=4)

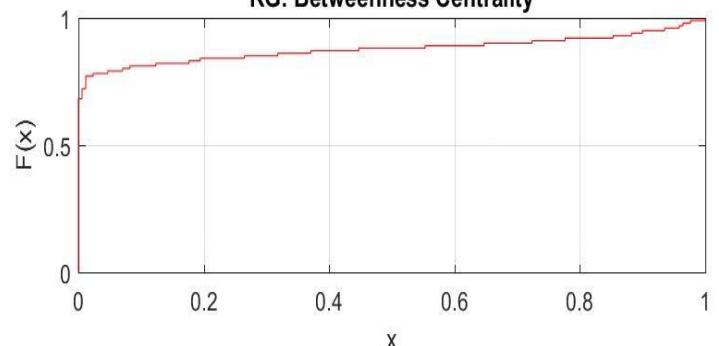


✓ RG (ER) ($N=170$, $M=750$)

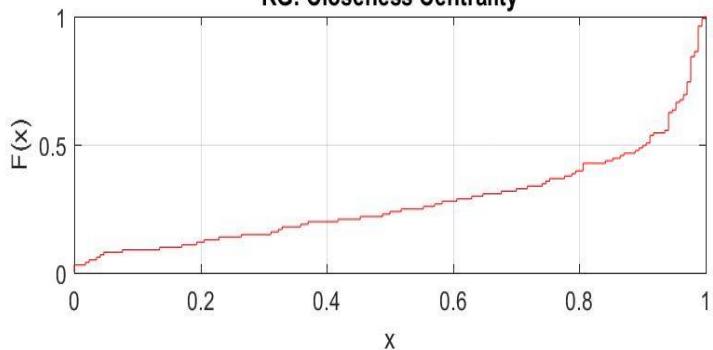
RG: Degree Centrality



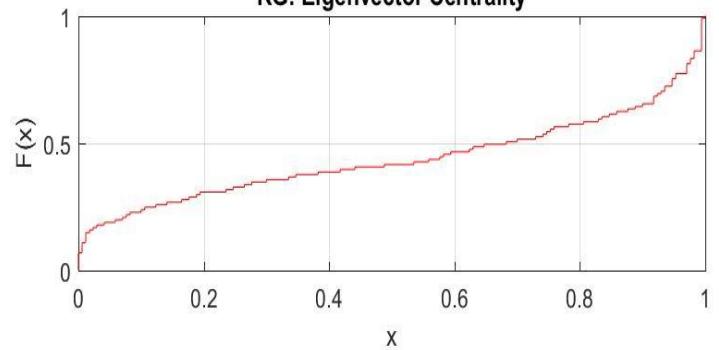
RG: Betweenness Centrality



RG: Closeness Centrality

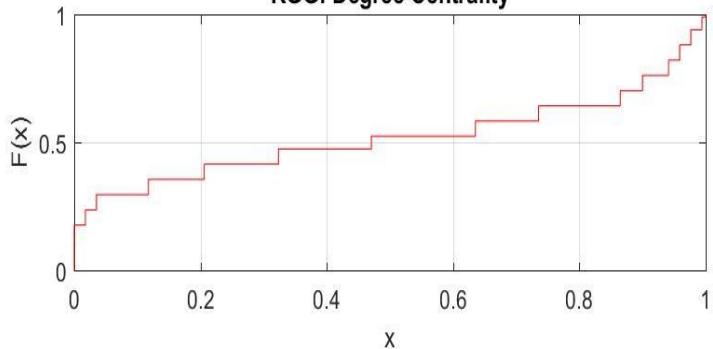


RG: Eigenvector Centrality

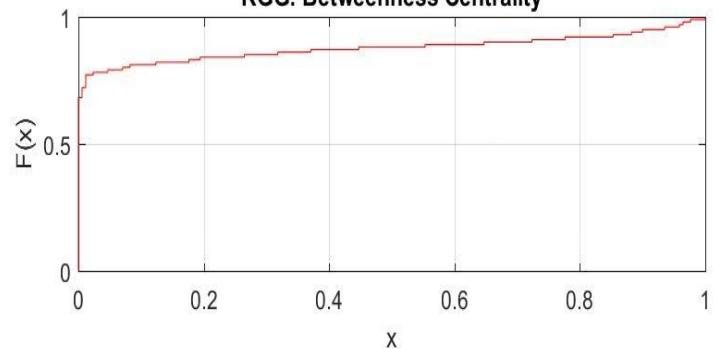


✓ RGG ($L \times L = 1000^2$, $n=170$, $R=250$)

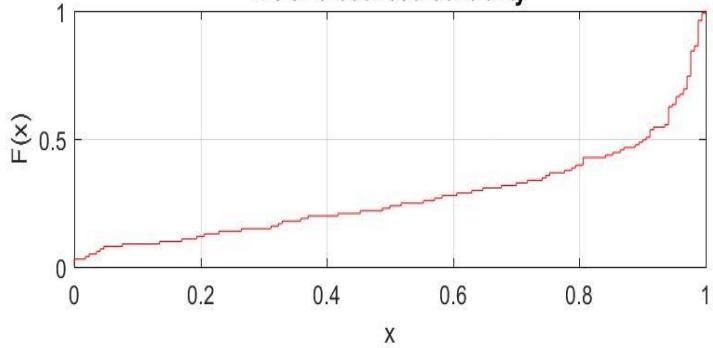
RGG: Degree Centrality



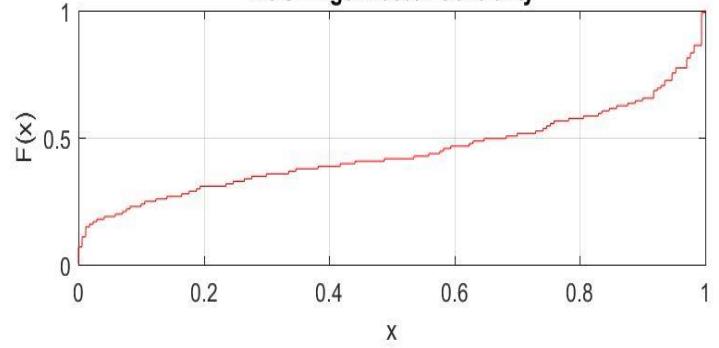
RGG: Betweenness Centrality



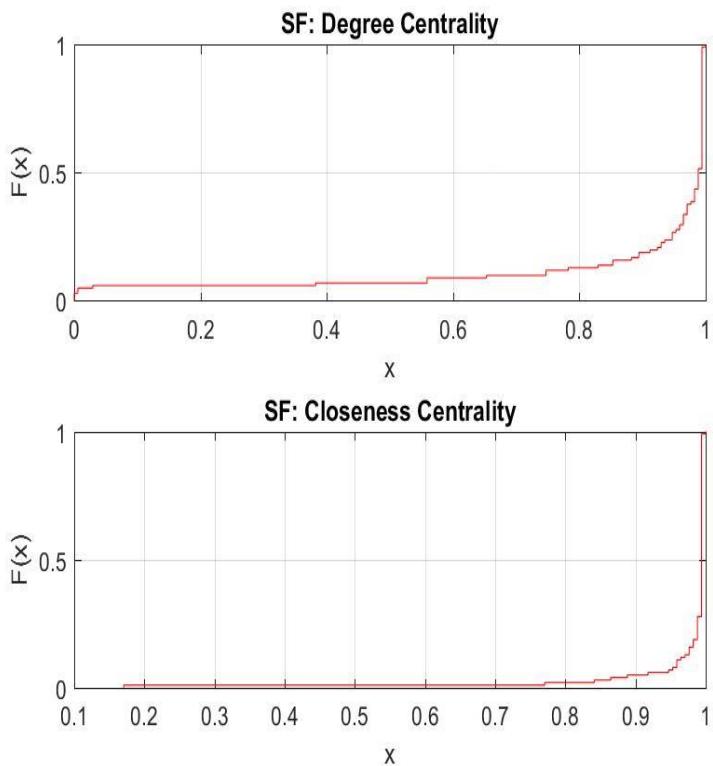
RGG: Closeness Centrality



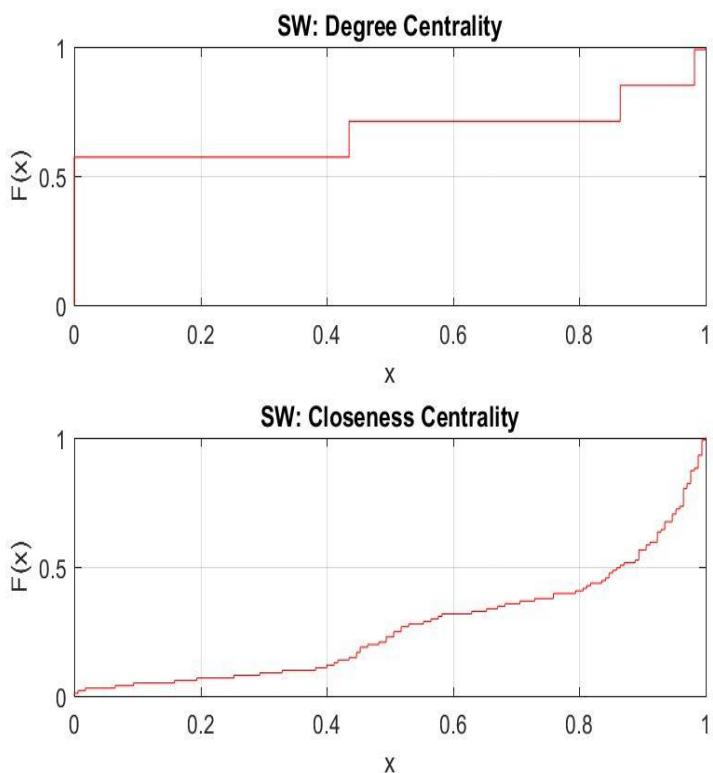
RGG: Eigenvector Centrality



✓ SF (BA) ($n=170$, $d=4$)



✓ SW (WS) ($n=170$, $d=4$, $g_p=0.3$)



Στη συνέχεια, υπολογίζουμε τη μέση κεντρικότητα κάθε τοπολογίας για όλες τις μετρικές:

Τοπολογία Δικτύου:	REG (n=170, d=4)	RG (N=170, M=50)	RGG ($L \times L = 1000^2$, n=170, R=250)	SF (n=170, d=4)	SW (n=170, d=4, g _p =0.3)
Mean Degree Centrality (DC)	4	8.8235	30.3294	7.4235	4.7176
Mean Closeness Centrality (CC)	0.0002736	0.0023	0.0023	0.0024	0.0013
Mean Betweenness Centrality (BC)	0.1252	0.0091	0.0026	0.0084	0.0238
Mean Eigenvector Centrality (EC)	-0.0767	-0.0717	0.0551	0.0597	0.0710

Από τα παραπάνω βλέπουμε ότι στον REG γράφο το closeness centrality έχει μικρή μέση τιμή αφού η απόσταση προς κάθε κόμβο είναι μεγάλη. Επίσης, παρατηρούμε ότι η REG τοπολογία έχει το μεγαλύτερο μέσο betweenness centrality από όλες τις υπόλοιπες. Αυτό συμβαίνει καθώς, όπως είναι συνδεδεμένος ο γράφος αυτός, θα πρέπει να «διέλθω» από πολλούς γείτονες για να μεταβώ από έναν κόμβο σε έναν άλλο.

Στο SW γράφο παρατηρούμε ότι έχουμε αρκετά χαμηλό mean closeness centrality λόγω της ανασύνδεσης που υπάρχει. Ωστόσο, βλέπουμε και μια μεγάλη σε σχέση με τις άλλες τοπολογίες (RG, RGG, SF), τιμή στο mean betweenness centrality.

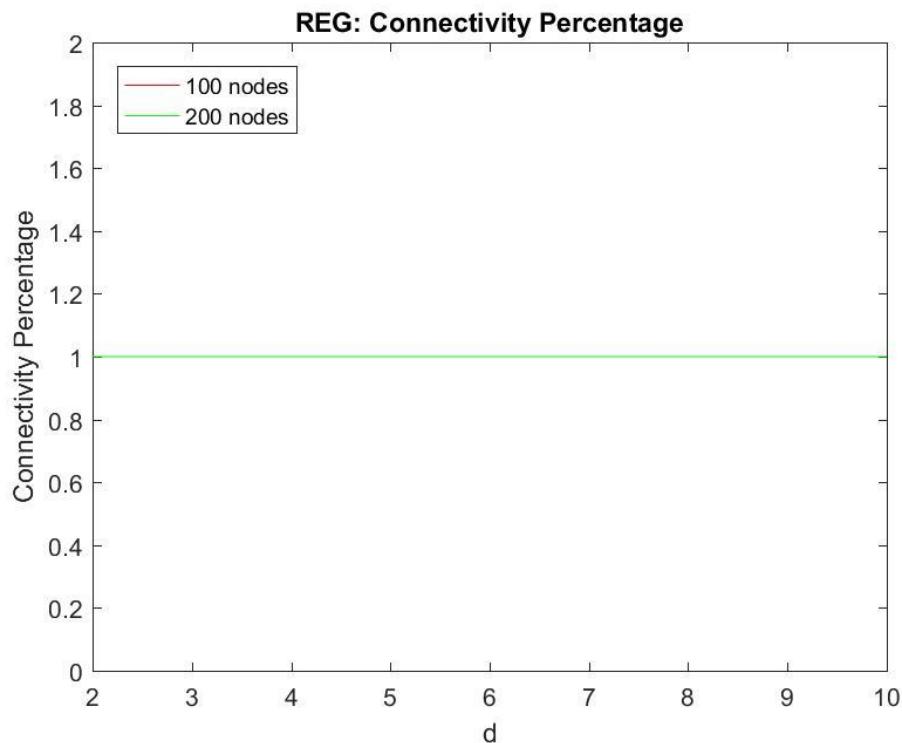
Όσον αφορά τις τοπολογίες RG, RGG μπορούμε να δούμε ότι εξάγαμε παρόμοια αποτελέσματα αν εξαιρέσουμε το mean degree centrality και mean eigenvector centrality.

Στη συνέχεια θα μελετήσουμε τη συνεκτικότητα των δικτύων, την εμφάνιση φαινομένων κατωφλίου (threshold behavior) κατηγορίας εξελεκτική συμπεριφοράς σύνθετων τύπων δικτύων που αναλύθηκαν παραπάνω. Επίσης, θα μελετηθούν οι τοπολογίες με βάση τη μετρική της εγω-κεντρικότητας.

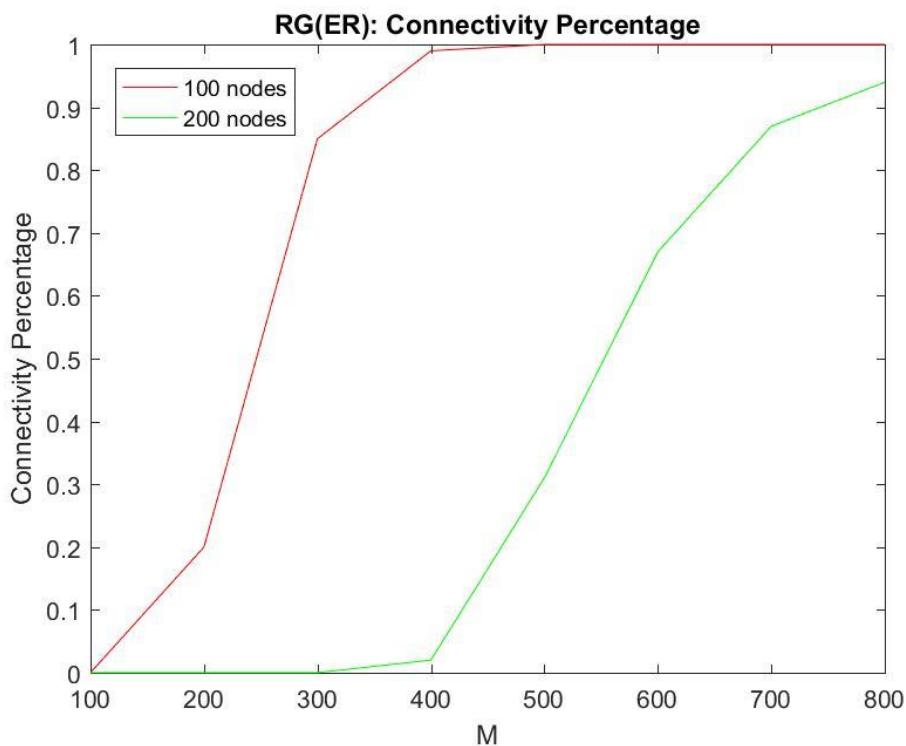
H) Μελέτη συνεκτικότητας και συμπεριφοράς κατωφλίων

Στο ερώτημα αυτό καλούμαστε να υπολογίσουμε το ποσοστό συνεκτικότητας για κάθε είδος δικτύου που έχουμε μελετήσει προηγουμένως. Για να το επιτύχουμε αυτό παράγουμε ένα πλήθος τοπολογιών από κάθε κατηγορία δικτύου με μεταβλητές παραμέτρους (κάνοντας χρήση for loops) όπως αυτές ορίζονται στον πίνακα 4 του συγκεκριμένου ερωτήματος. Στη συνέχεια παρατίθενται τα διαγράμματα ποσοστού συνεκτικότητας σε σχέση με τις καθορισμένες παραμέτρους κάθε δικτύου που προέκυψαν από την ανάλυση που κάναμε για κάθε είδους τοπολογία:

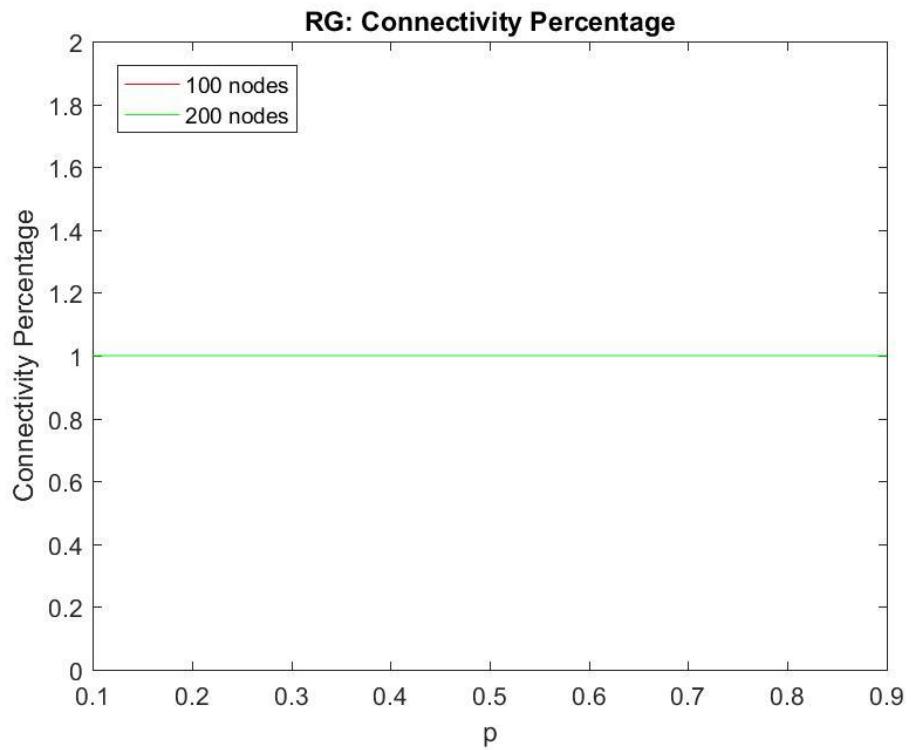
- ✓ REG (Ποσοστό συνεκτικότητας – d)



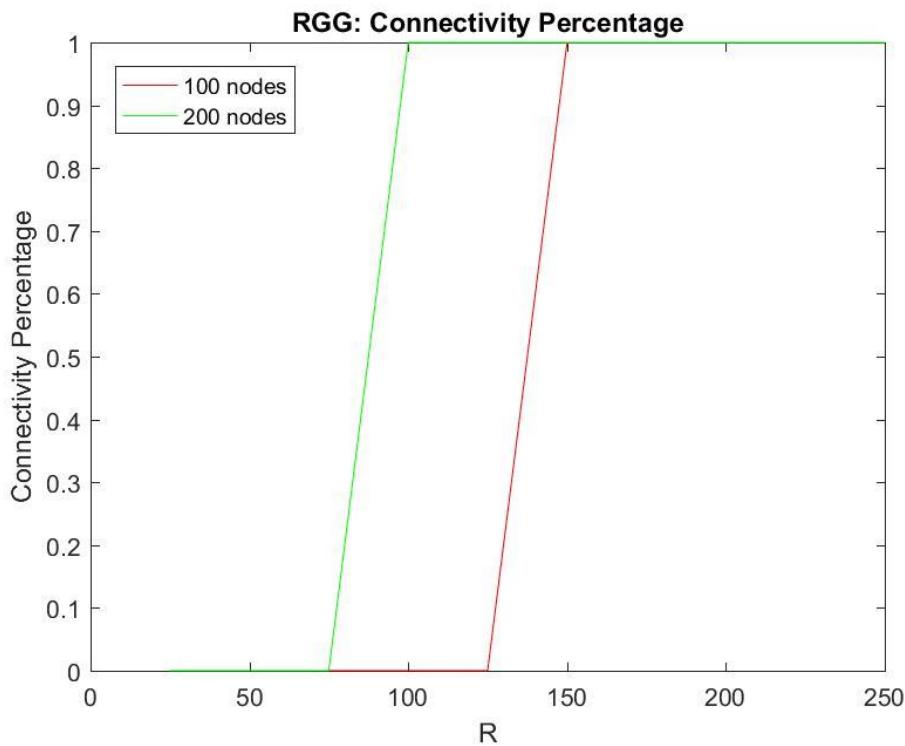
✓ RG (ER) (Ποσοστό συνεκτικότητας – M)



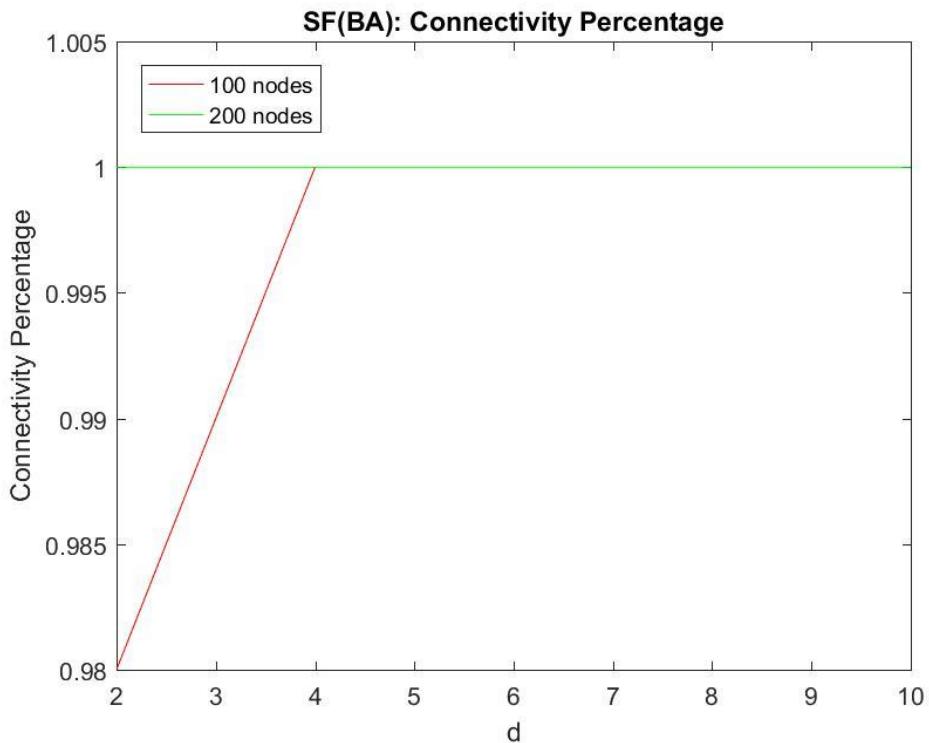
✓ RG (Ποσοστό συνεκτικότητας – p)



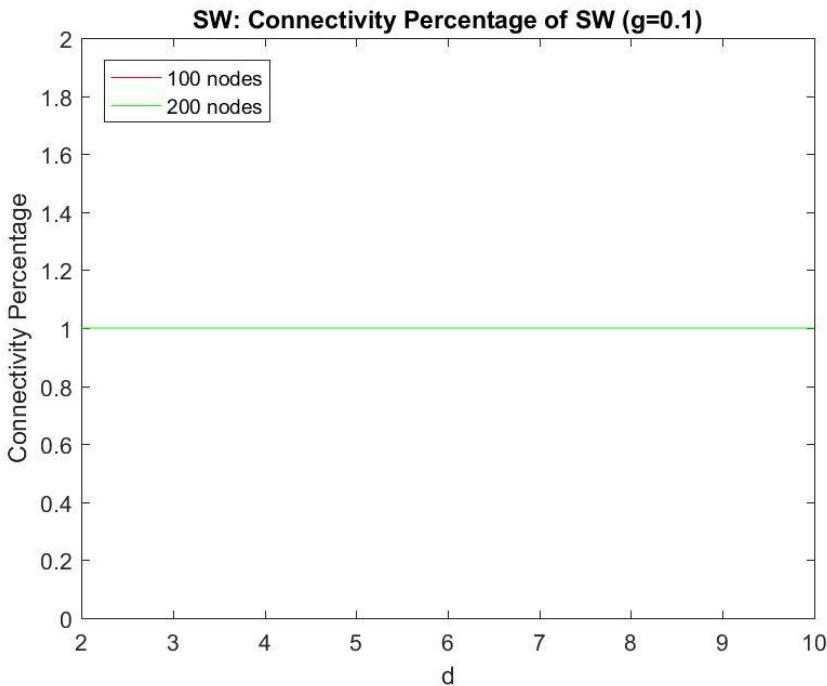
- ✓ RGG (Ποσοστό συνεκτικότητας – R)



- ✓ SF (BA) (Ποσοστό συνεκτικότητας – d)



- ✓ SW (WS) (Ποσοστό συνεκτικότητας – g)



❖ Σε ποιές περιπτώσεις εμφανίζονται φαινόμενα κατωφλίου?

Φαινόμενα κατωφλίου, όπως φαίνεται και από τα παραπάνω διαγράμματα έχουμε στις περιπτώσεις των:

REG: Ο γράφος αυτός είναι συνδεδεμένος πάντα (από τη λογική κατασκευής του) και επομένως δεν παρουσιάζει σημείο κατωφλίωσης.

RG(ER) : Παρατηρούμε ότι έχουμε μια σχετικά ομαλή μετάβαση (απότομη στα κρίσιμα σημεία). Το οριακό σημείο εξαρτάται από τον αριθμό των κόμβων που έχουμε. Όσο αυξάνεται αυτός ο αριθμός, το σημείο αυτό μετατοπίζεται δεξιότερα.

RG: Βλέπουμε από το διάγραμμα ότι ο γράφος RG είναι μόνιμα συνδεδεμένος. Δεν παρουσιάζει σημείο κατωφλίωσης. Αυτό οφείλεται στο γεγονός ότι η πιθανότητα 0.1, τιμή από την οποία ξεκινάμε, είναι μεγαλύτερη από το κατώφλι ώστε ο γράφος να είναι συνδεδεμένος (προκύπτει από πράξεις που εφαρμόσαμε στον παρακάτω τύπο για το γράφο RG(ER)) Άρα ήταν αναμενόμενο να έχουμε connected γράφημα.

RGG: Στο γράφημα αυτό έχουμε φαινόμενο κατωφλίωσης.

SF: Στην τοπολογία αυτή για 200 κόμβους έχω μόνιμα συνδεδεμένο γράφημα ενώ για 100 κόμβους υπάρχει φαινόμενο κατωφλίωσης, ωστόσο είναι θεωρητικά αμελητέο καθώς η μεταπήδηση γίνεται από το 98% στο 100%. Ηδη στο 98% θα μπορούσαμε να το θεωρήσουμε το γράφημα μας συνδεδεμένο. (Αυτό έχει λογικη καθώς σε έναν scale free γράφο, είναι πιθανό ένας κόμβος να συνδέεται με έναν κόμβο με πολλούς γείτονες και μέσω αυτού λοιπόν να συνδέεται με τους υπόλοιπους.)

SW: Παρατηρούμε ότι και για 100 και για 200 κόμβους ο γράφος είναι μόνιμα συνδεδεμένος. Δεν παρουσιάζεται φαινόμενο κατωφλίωσης.

❖ **Ποιές είναι οι κρίσιμες τιμές για κάθε τύπο δικτύου οι οποίες οδηγούν σε μετάβαση φάσης (phase transitions)?**

RG(ER):Στο δικό μας πείραμα, για 100 κόμβους το σημείο αυτό είναι λίγο πριν την τιμή $M=400$ ενώ για 200 κόμβους βρίσκεται περίπου στο σημείο όπου $M=800$. Από το σημείο και μετά οι γράφοι είναι συνδεδεμένοι. Με βάση τη θεωρία το σημείο αυτό βρίσκεται προσεγγιστικά από τον τύπο: $(\text{Αριθμός}_2^{\text{κόμβων}}) * \log(\text{Αριθμός κόμβων})$. Από αυτό το M που δίνει προηγούμενος τύπος και μετά θα έχουμε $\log(\text{Αριθμός κόμβων})$ περίπου ποσοστό συνεκτικότητας.

RGG: Παρατηρώ ότι για 100 κόμβους το κρίσιμο σημείο είναι για $R=125$ ενώ για 200 κόμβους κρίσιμο σημείο έχω για $R=75$. Αυτό συμβαίνει γιατί η αναμενόμενη τιμή γειτόνων είναι ανάλογη τις ακτίνας και των κορυφών του γράφου.

SF:Τυπικά, το σημείο κατωφλίου βρίσκεται για $d=4$ (για τους 100 κόμβους) αν μπορείνα θεωρηθεί τέτοιο (καθώς όπως προείπα η μεταπήδηση από 98% σε 100% δεν αλλάζει και πολύ το γράφο αφού ήδη θεωρείται συνδεδεμένος, έστω και άτυπα)

❖ **Ποιες μεταβάσεις φάσης είναι απότομες (sharp) και ποιές ομαλές (smooth)?**

RG(ER) : Παρατηρούμε ότι έχουμε μια σχετικά ομαλή μετάβαση(απότομη στα κρίσιμα σημεία,εκεί έχουμε μεγάλη κλίση)

RGG: Οι μεταβάσεις μεταξύ των φάσεων, όπως φαίνεται και στο διάγραμμα, είναι αρκετά απότομες.

SF: η μετάβαση είναι αρκετά μικρή για να χαρακτηριστεί απότομη ή ομαλή.(θεωρητικά είναι ομαλή)

Θ) Μελέτη μοντέλων τυχαίων γράφων

Τοπολογία	$n = 100$	$n = 1000$	$n = 10^4$	$n = 10^5$	$n = 10^6$
RG (G)	$p = 0.1$	$p = 10^{-2}$	$p = 10^{-3}$	$p = 10^{-4}$	$p = 10^{-5}$
RG (ER)	$M = 495$	$M = 4995$	$M = 49995$	$M = 499995$	$M = 4999995$

Ο παραπάνω πίνακας συμπληρώθηκε γνωρίζοντας ότι αν το γινόμενο $p \times n$ διατηρείται σταθερό και ο όρος $p \times n^2$ τείνει στο άπειρο, το μοντέλο του Gilbert προσεγγίζει το μοντέλο του Erdos-Renyi με παράμετρο $M = \frac{[n \times (n-1) \times p]}{2}$. Με βάση αυτό και με κατάλληλες πράξεις συμπληρώσα τις τιμές του παραπάνω πίνακα.

I) Μελέτη της εξελικτικής μετατροπής δικτύου REG σε δίκτυο SW και RG(ER)

Στο ερώτημα αυτό μεταβάλλουμε την παράμετρο g_p για την κατασκευή της τοπολογίας SW. Οι τιμές που της δίνουμε είναι από 0 έως 1 με βήμα 0.1. Για κάθε τοπολογία που παράγουμε κρατάμε το μέσο μήκος μονοπατιού καθώς και τον συντελεστή ομαδοποίησης (clustering coefficient) και υπολογίζουμε το μέσο όρο τους. Στη συνέχεια παρουσιάζονται τα αποτελέσματα που μας έδωσε το Matlab:

Πιθανότητα ανασύνδεσης g_p	Μέσο Μήκος Μονοπατιού	Μέσος Συντελεστής Ομαδοποίησης
0	21.6272	0.5000
0.1	8.0879	0.4653
0.2	6.0033	0.4387
0.3	4.8378	0.3940
0.4	4.2162	0.3606
0.5	4.1628	0.3418
0.6	3.8277	0.3140
0.7	3.6505	0.2963
0.8	3.5800	0.2744
0.9	3.4643	0.2566
1	3.3838	0.2418

Σχολιασμός αποτελεσμάτων:

Για $g_p = 0$ το δίκτυο συμπεριφέρεται σαν REG δίκτυο (λογικό επομένως και το μέσο μήκος μονοπατιού και ο συντελεστής ομαδοποίησης).

Για $g_p = 1$ το δίκτυο συμπεριφέρεται σαν RG δίκτυο. (τυχαίος γράφος)

Για $g_p \cong 0.1 - 0.5$ το δίκτυο είναι ένα SW δίκτυο.

Αρχικά, για $g_p = 0$, κάθε κόμβος του δικτύου ήταν συμμετρικά συνδεδεμένος με τους γείτονές του (όπως στον REG γράφο). Όμως, στη συνέχεια, κάθε ακμή έλαβε πιθανότητα ανασύνδεσης g_p . Αν μια ακμή επιλέχτηκε για ανασύνδεση, το ένα άκρος της αποσυνδέθηκε από τον ένα κόμβο (που ήταν ήδη συνδεδεμένο) και ανασυνδέθηκε σε έναν άλλο τυχαία επιλεγμένο κόμβο. Αυτή η ανασύνδεση δημιουργεί συντομότερα μονοπάτια (shortcuts) μέσα στο δίκτυο και πολύ γρήγορα μειώνει το μέσο

μήκος μονοπατιού στο δίκτυο (καθώς μπορούμε να πάμε από έναν κόμβο σε έναν άλλο πιο γρήγορα). Επίσης, αυτό το φαινόμενο της ανασύνδεσης (που αυξάνει, όσο αυξάνουμε την παράμετρο της πιθανότητας ανασύνδεσης στο γράφο μας) μειώνει τον συντελεστή ομαδοποίησης (Clustering Coefficient), καθώς ο αριθμός των γειτονικών κόμβων που συνδέονται μειώνεται με τις ανασυνδέσεις. Ωστόσο, παρατηρούμε ότι υπάρχουν σημεία όπου ο συντελεστής ομαδοποίησης είναι σχετικά μεγάλος και το μέσο μήκος μονοπατιού σχετικά μικρό. (σημαντική «περιοχή» του SW δικτύου)

(reference: http://mathinsight.org/small_world_network)

K) Εγω-κεντρικότητες (Ego-Centralities)

Στο ερώτημα αυτό υπολογίζεται η εγω-κεντρικότητα κάθε κόμβου κάθε είδους τοπολογίας. Ο υπολογισμός της εγωκεντρικότητας έγινε με τη μέθοδο που επεξηγείται στην παρουσίαση.

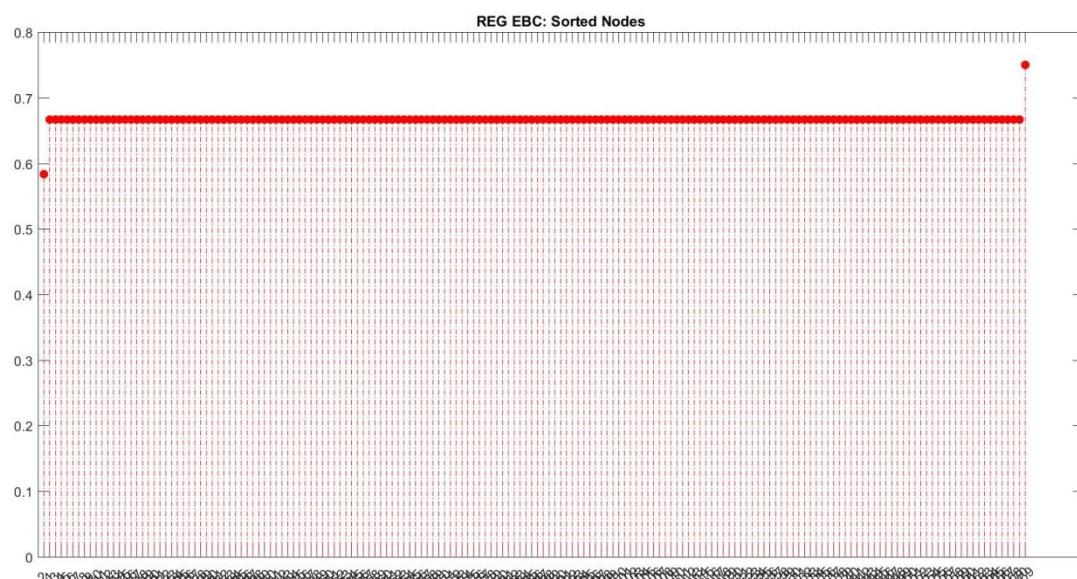
1.εύρεση ego adjacency matrix A κάθε κόμβου με χρήση των συναρτήσεων subgraph() και kneighbors()

2.Υπολογισμός του $A^2[1 - A]$

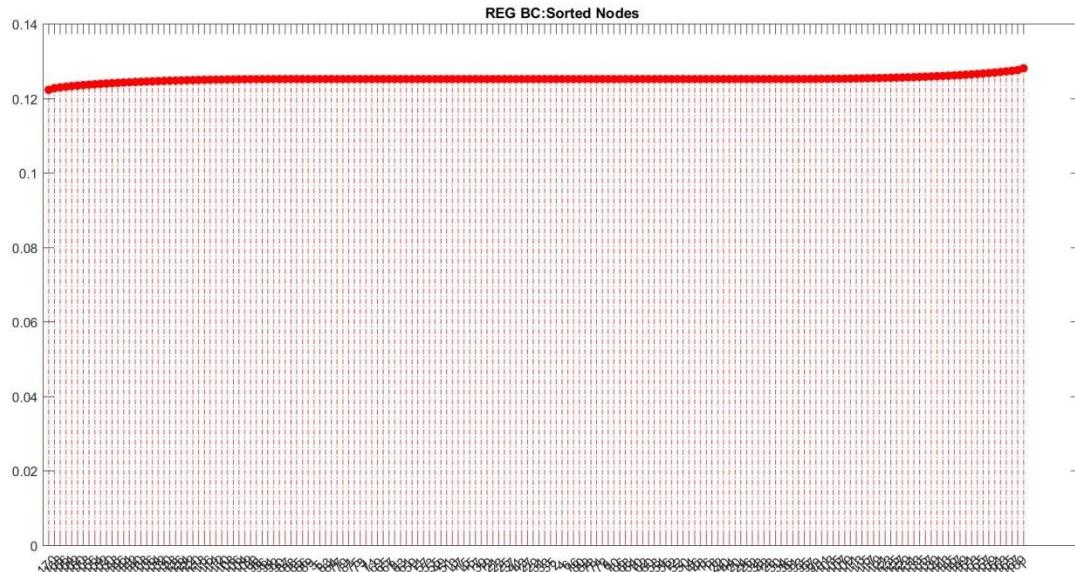
3.Από αυτόν τον πίνακα, κρατάμε μόνο τα στοιχεία που βρίσκονται πάνω από τη διαγώνιο και είναι μη μηδενικά. Το άθροισμα των αντιστρόφων αυτών μας δίνει το ego-centrality του εκάστοτε κόμβου.

Όπως ζητείται στην εκφώνηση, παρουσιάζουμε τη διάταξη των κόμβων με βάση την εγω-κεντρικότητα καθώς και με βάση την κεντρικότητα(betweenness centrality) όπως αυτή υπολογίστηκε προηγουμένως. Τα αποτελέσματα οπτικοποιούνται παρακάτω:

- ✓ REG
 - Εγω-κεντρικότητα

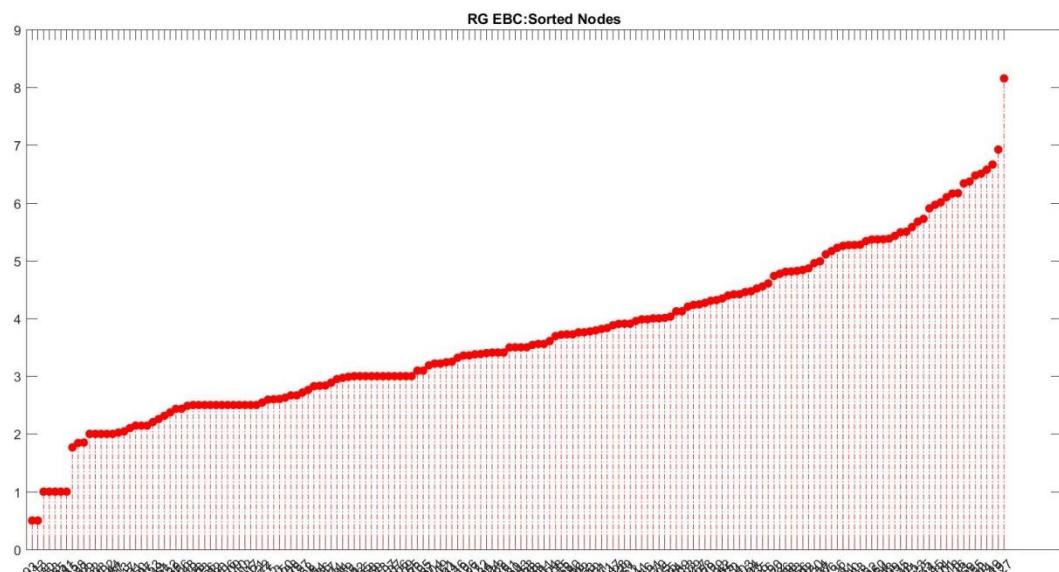


- Κεντρικότητα

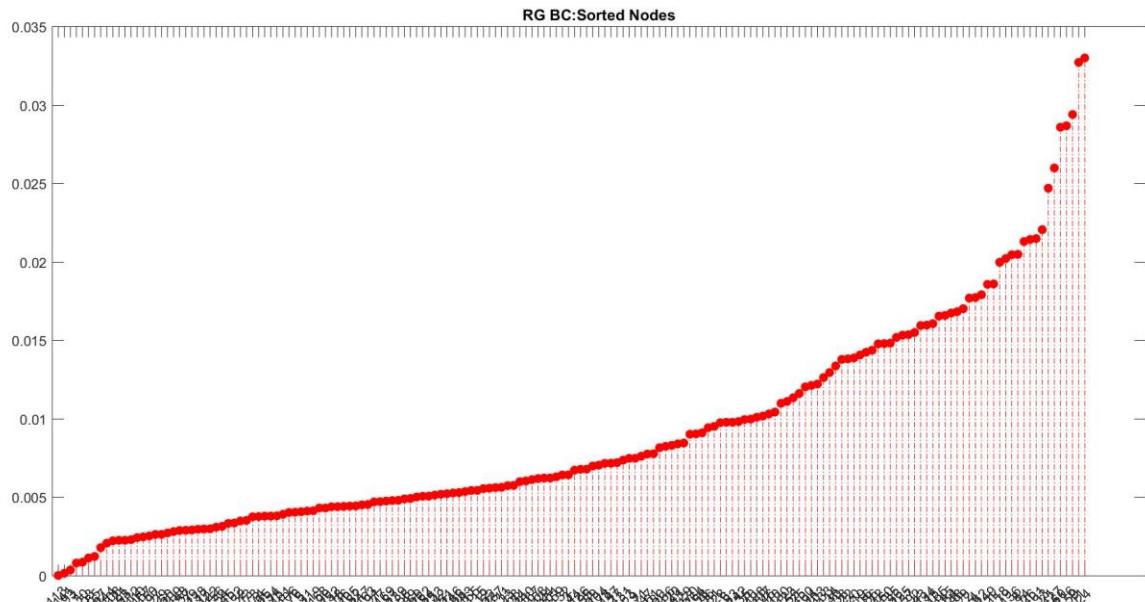


✓ RG (ER)

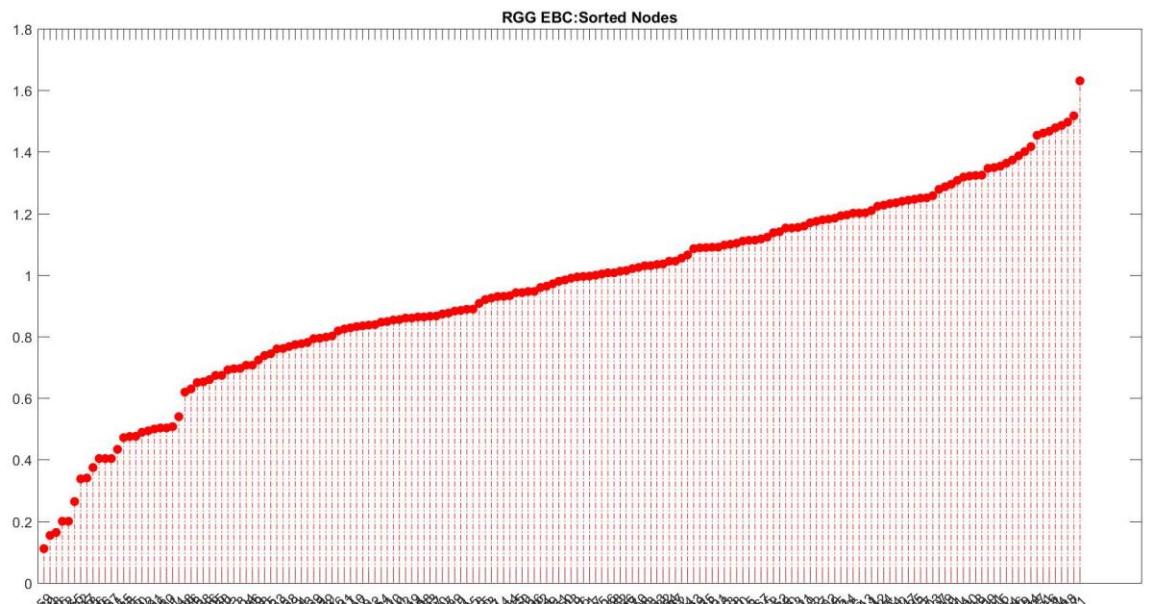
- Εγω-κεντρικότητα



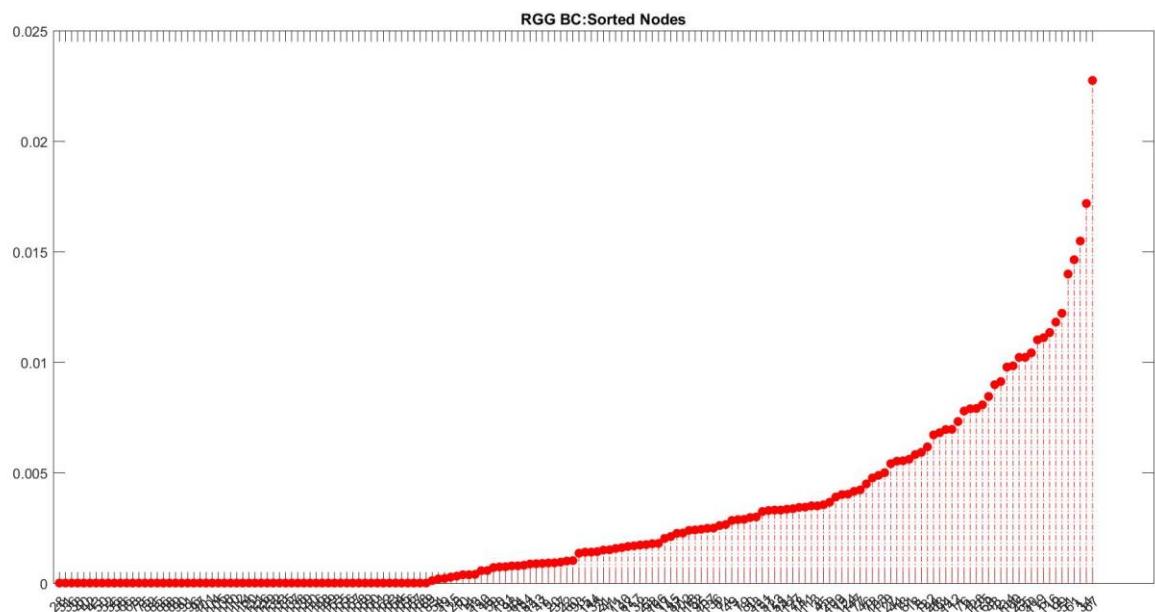
- Κεντρικότητα



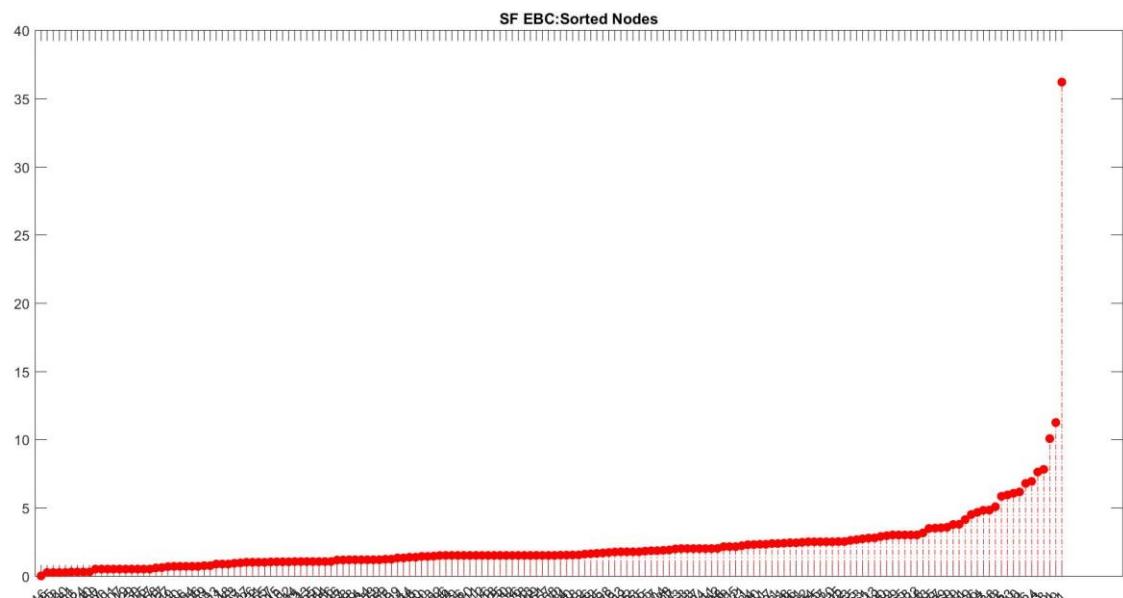
- ✓ RGG
- Εγω-κεντρικότητα



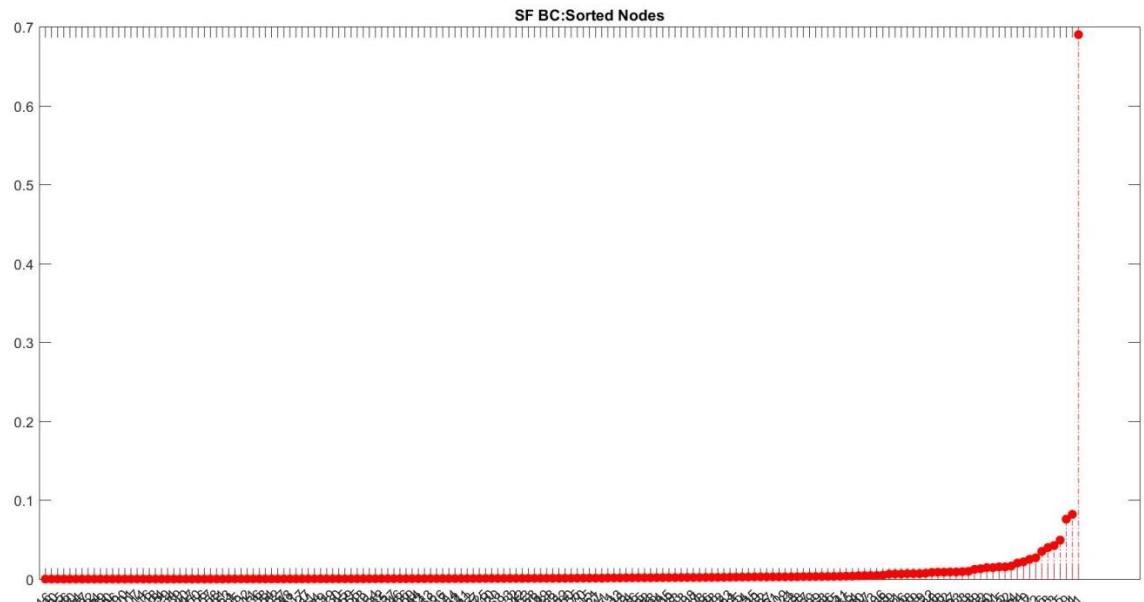
- Κεντρικότητα



- ✓ SF (BA)
- Εγω-κεντρικότητα

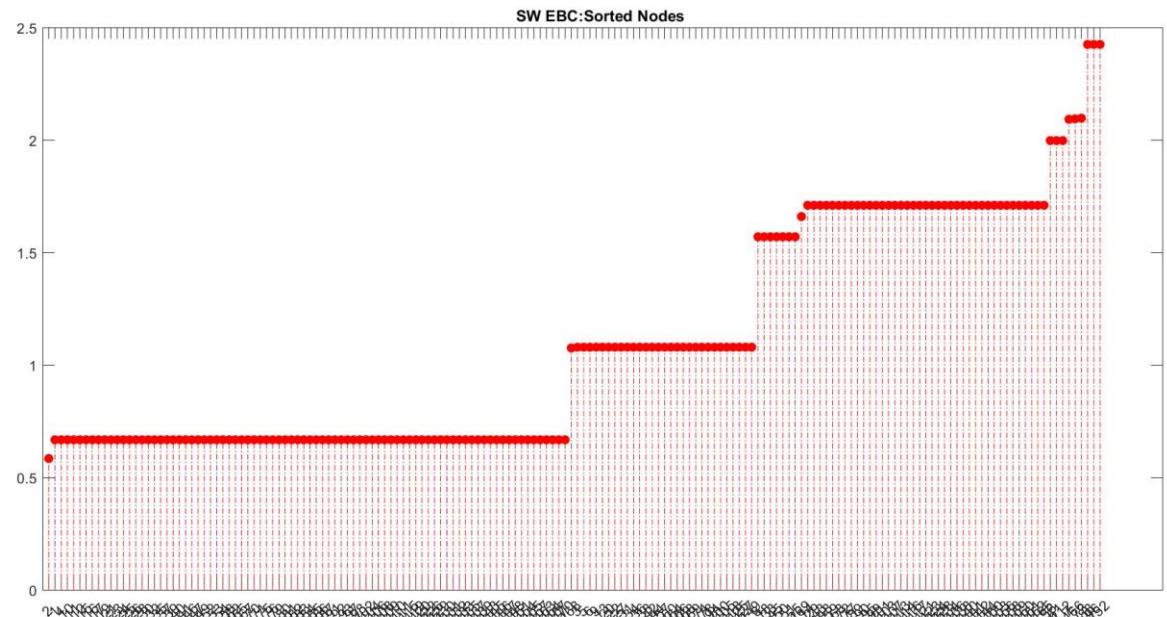


- Κεντρικότητα

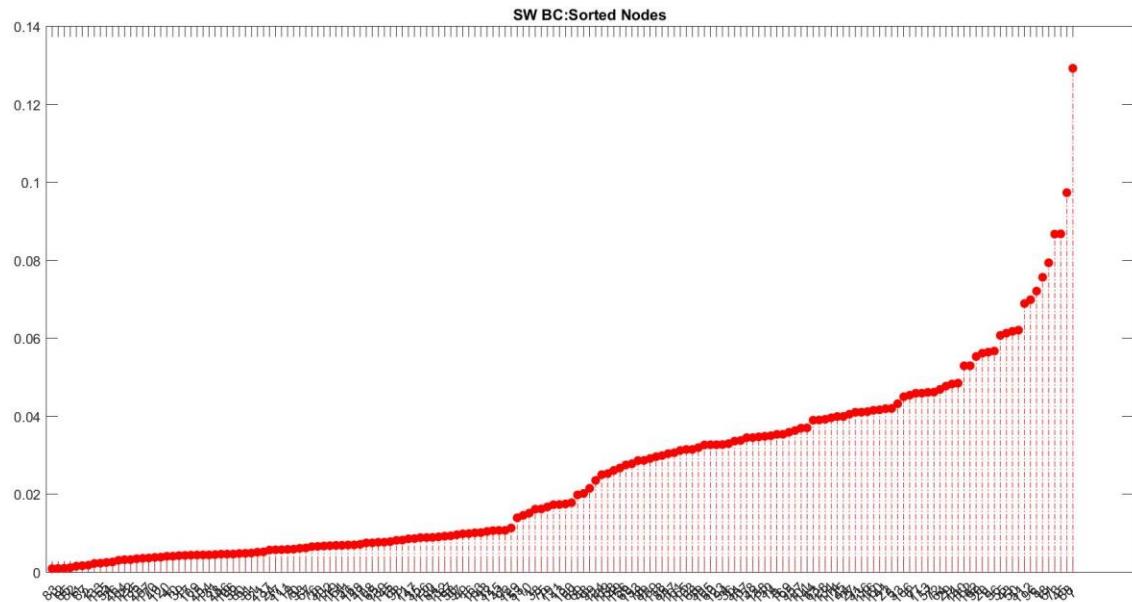


✓ SW (WS)

- Εγω-κεντρικότητα



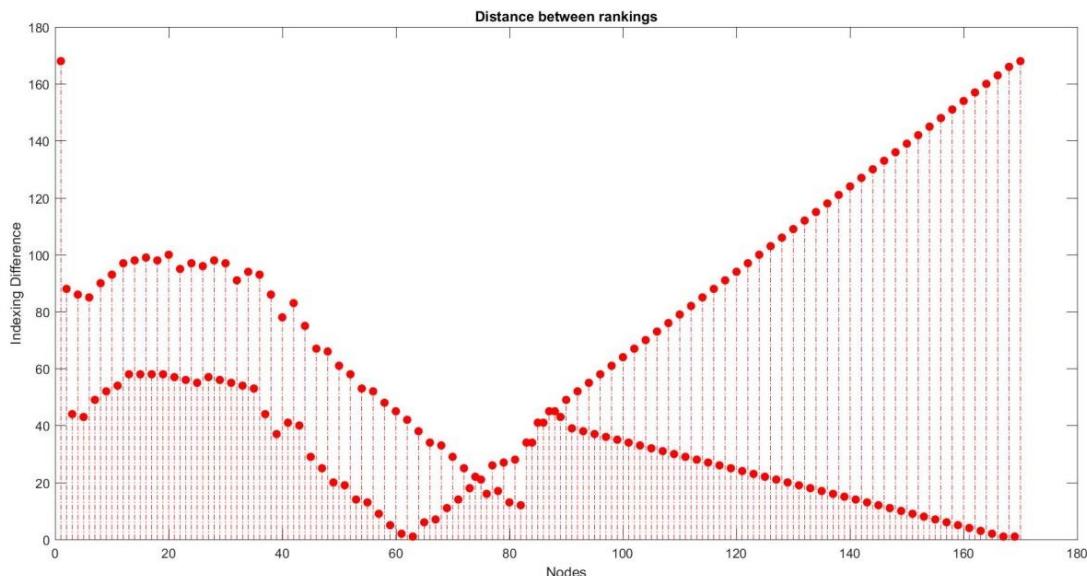
- Κεντρικότητα



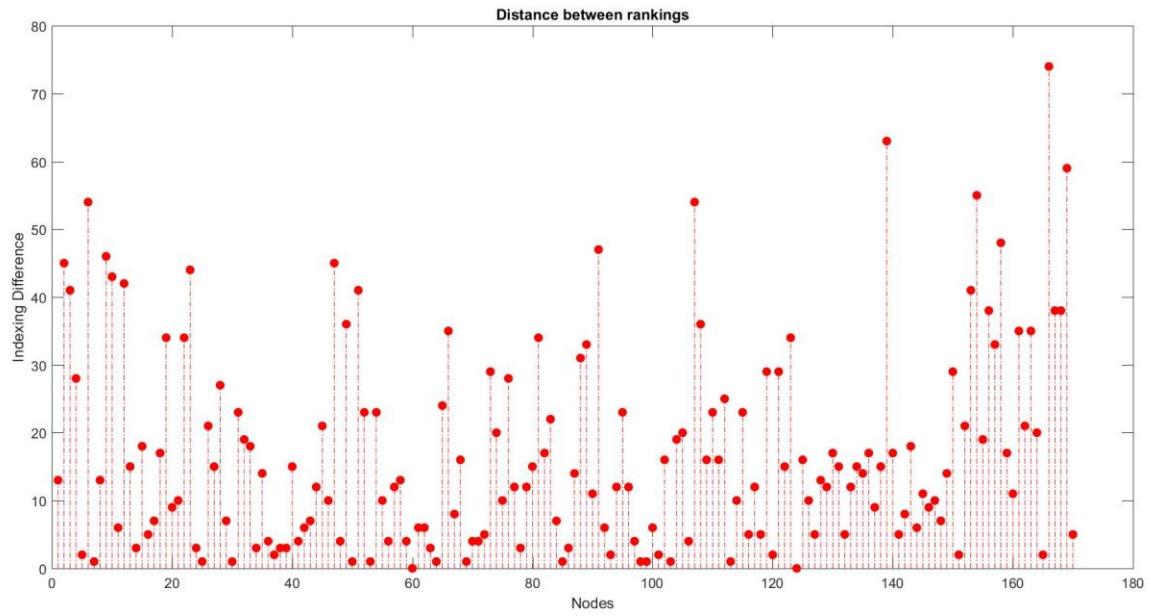
(Η κατάταξη των κόμβων λόγω του πλήθους τους και του μικρού μεγέθους της εικόνας δεν είναι εμφανής)

Για το λόγο αυτό, υπολογίστηκε και η σχετική απόσταση της θέσης κάθε κόμβου μεταξύ των 2 κατατάξεων. (πχ αν ο κόμβος i βρίσκεται στη θέση k στην κατάταξη των κόμβων με βάση το ego-centrality και στη θέση m στην κατάταξη των κόμβων με βάση το betweenness centrality παίρνω και εμφανίζω το $\text{abs}(k-m)$)

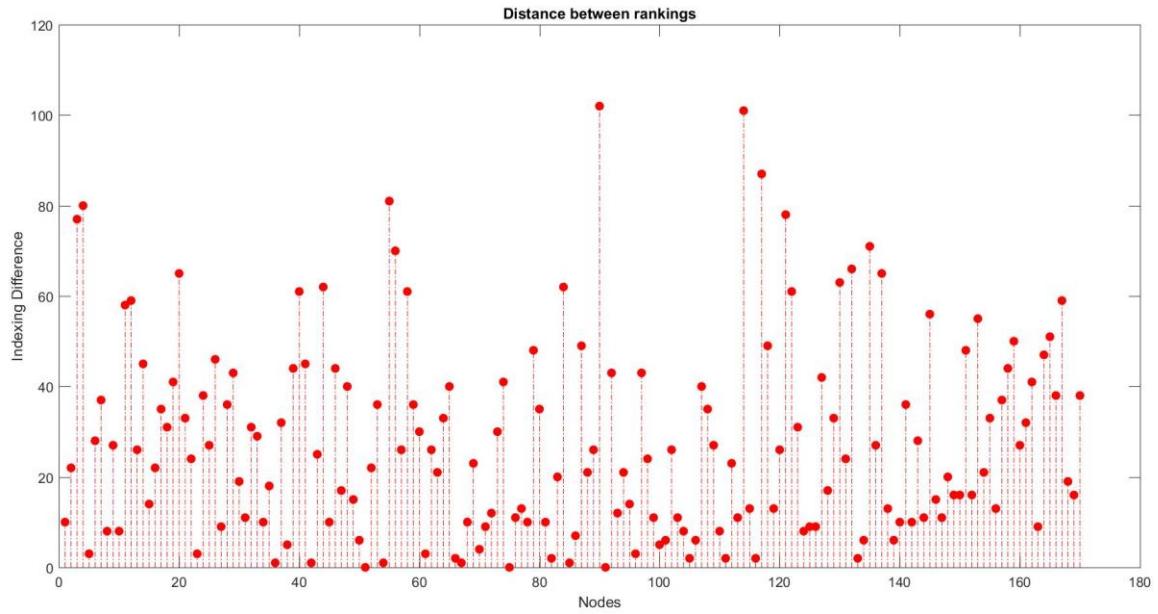
✓ REG



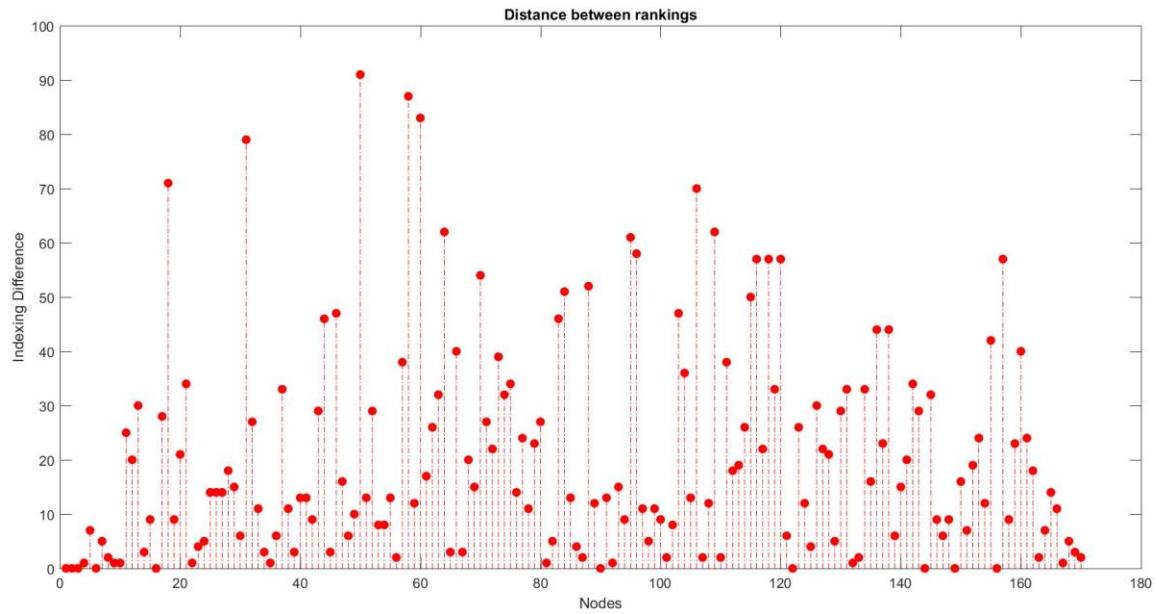
✓ RG



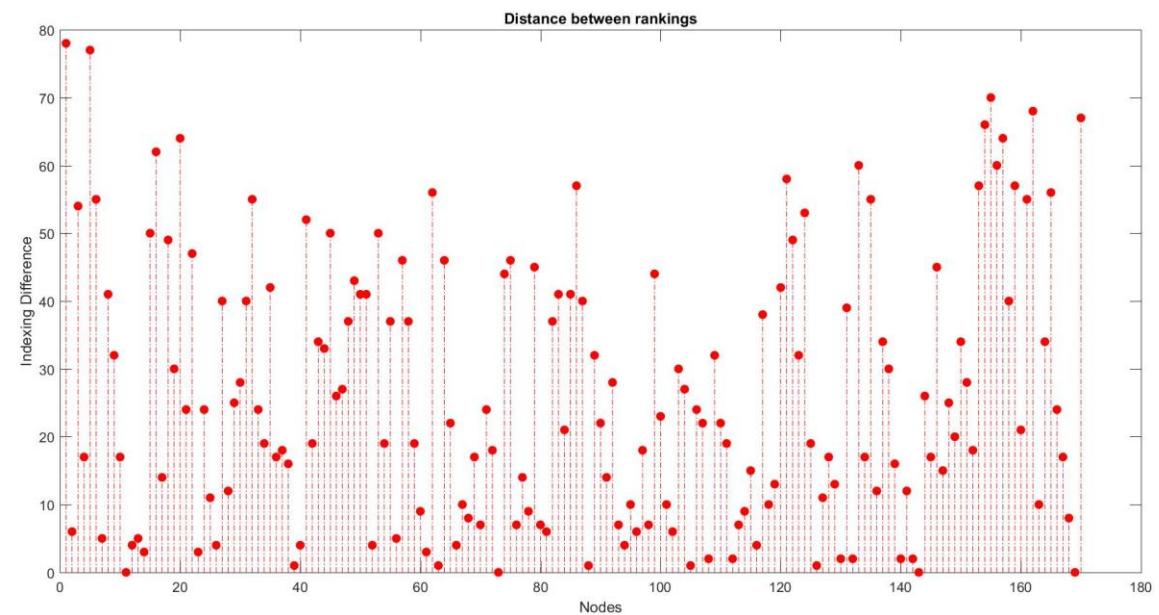
✓ RGG



✓ SF



✓ SW



Σχολιασμός αποτελεσμάτων:

Αυτό που επιθυμούμε να επιτύχουμε, και το κάνουμε ως ένα βαθμό στο ερώτημα αυτό είναι να δούμε πόσο καλή εκτίμηση του betweenness centrality μας δίνει το ego centrality, καθώς η εύρεση του τελευταίου είναι πολύ ευκολότερη και απλούστερη υπολογιστικά. Από τα παραπάνω διαγράμματα που μας δείχνουν την απόσταση στην κατάταξη των κόμβων, όπως τη δίνουν οι 2 υπολογισμένες κεντρικότητες, βλέπουμε ότι επιτυγχάνουμε αυτό που επιθυμούμε ως ένα βαθμό. Πιο αναλυτικά, βλέπουμε ότι στις τοπολογίες SF, SW και RGG πολλοί κόμβοι είναι έχουν την ίδια κατάταξη στο ego και στο betweenness centrality είτε έχουν μικρή διαφορά στη θέση στην οποία βρίσκονται στην κατάταξη μετά το sorting. Τα αποτελέσματα δεν είναι τόσο καλά στην περίπτωση της τοπολογίας REG και RG καθώς εκεί, όπως βλέπουμε και παραπάνω, σημειώνονται σε πολλούς κόμβους μεγάλες διαφορές στις θέσεις που κατατάσσονται με βάση τις 2 μετρικές της centrality. Αναλυτικότερα:

- Στην REG τοπολογία βλέπουμε ότι υπάρχουν κάποιες αποκλίσεις στην κατάταξη κάθε κόμβου. Ωστόσο και εδώ, αν κοιτάξουμε καλά, πολλοί κόμβοι είναι σχετικά σωστά τοποθετημένοι (σε παρόμοια θέση στην κατάταξη που τις έχει το betweenness centrality) με βάση το ego-centrality.
- Στην RG τοπολογία βλέπουμε ότι ότι πολλοί κόμβοι καταλαμβάνουν τις ίδιες θέσεις στο ranking με βάση κάθε μία από τις 2 μετρικές. Λιγότεροι είναι αυτοί με τις μεγάλες αποκλίσεις, επομένως σε γενικές γραμμές, με τον απλό –σχετικά- υπολογισμό του ego centrality έχουμε καλή εικόνα για το betweenness centrality του γράφου μας.
- Στην RGG τοπολογία βλέπουμε ότι πολλοί κόμβοι καταλαμβάνουν τις ίδιες θέσεις στο ranking με βάση κάθε μία από τις 2 μετρικές. Λιγότεροι είναι αυτοί με τις μεγάλες αποκλίσεις, επομένως σε γενικές γραμμές, με τον απλό –σχετικά- υπολογισμό του ego centrality έχουμε καλή εικόνα για το betweenness centrality του γράφου μας. (μοιάζει με την περίπτωση του RG, παρουσιάζει καλύτερα αποτελέσματα)
- Στην SF τοπολογία βλέπουμε ότι υπάρχουν λίγες μεγάλες αποκλίσεις μεταξύ των ranking του κάθε κόμβου με βάση κάθε μετρική. Επομένως βλέπουμε ότι το ego-centrality μας δίνει πολύ καλή πληροφορία που μοιάζει πολύ – προσεγγίζει το betweenness.
- Στην SW τοπολογία βλέπουμε ότι υπάρχουν διαφορές μεταξύ των ranking των αντίστοιχων κόμβων του δικτύου με βάση το ego και το betweenness centrality, ωστόσο σε μεγάλο ποσοστό οι διαφορές αυτές είναι μικρές, επομένως η προσέγγιση που παίρνουμε είναι σχετικά καλή.

Βλέπουμε, λοιπόν, ότι λαμβάνουμε σχετικά καλές προσεγγίσεις και μπορούμε να εξάγουμε πληροφορίες για το betweenness centrality χωρίς να προβούμε στον αναλυτικό-επίπονο υπολογισμό του αλλά με απλές πράξεις για την εύρεση του ego-centrality.

Η εγω-κεντρικότητα μπορούμε να αποφανθούμε ότι δίνει καλά και ασφαλή συμπεράσματα για το betweenness centrality όταν όλοι οι κόμβοι έχουν παρόμοια betweenness centrality's values ή όταν έχουν μεγάλες αποκλίσεις μεταξύ αυτών. (πχ. Δομές όπως αυτή του SF γράφου εμπίπτουν στην κατηγορία που οι κόμβοι έχουν μεγάλες αποκλίσεις στα betweenness centrality's values τους) Για αυτό, θεωρούμε ότι η προσέγγιση του centrality μέσω του ego θα είναι αρκέτα χρήσιμη στα **πραγματικά δεδομένα** και λιγότερο αποτελεσματική στα ομοιόμορφα δημιουργημένα τυχαία δίκτυα.

Για πληρότητα, υπολογίσαμε για την κάθε τοπολογία και το ranking correlation μεταξύ των ego-centrality και betweenness centrality και σας την παρουσιάζουμε στη συνέχεια:

Τοπολογία Δικτύου:	REG	RG	RGG	SF	SW
Mean Ranking Correlation Coefficient EGOCENTRALITY-BETWEENNESS CENTRALITY	0.5884	0.9357	0.7911	0.9570	0.8769

Βλέπουμε, προς επιβεβαίωση των αποτελεσμάτων μας, σχετικά υψηλές τιμές στο mean ranking correlation coefficient για όλες τις τοπολογίες (εκτός του REG γραφήματος)

Reference: <https://pdfs.semanticscholar.org/01eb/fdb48fdf2cd286260db3789a221ffda87a44.pdf>