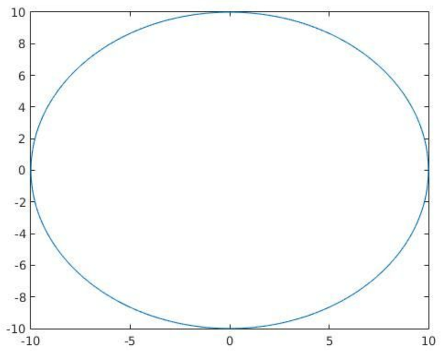
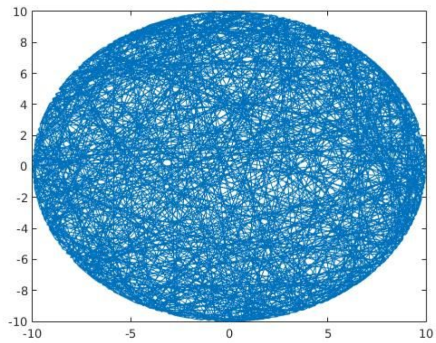
# Ανάλυση Κοινωνικών Δικτύων

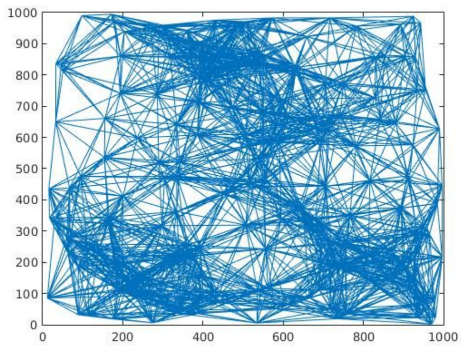
## Άσκηση 1

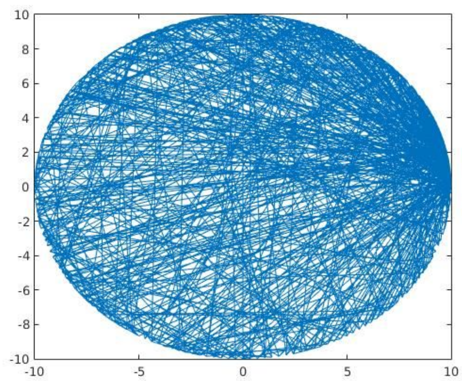
**Νικόλαος Ζαρίφης**
**03112178**

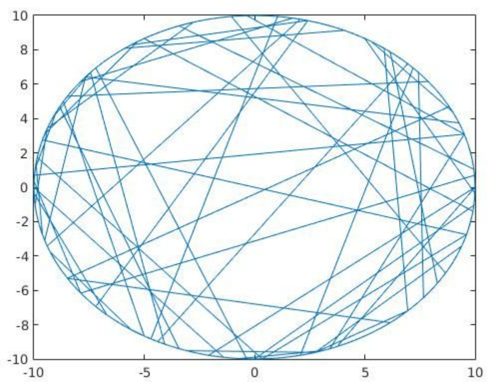
## A) Δημιουργία και οπτικοποίηση σύνθετων τύπων δικτύου







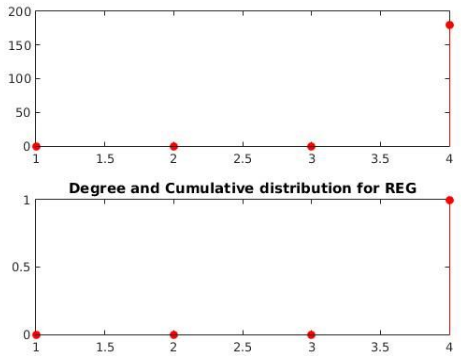


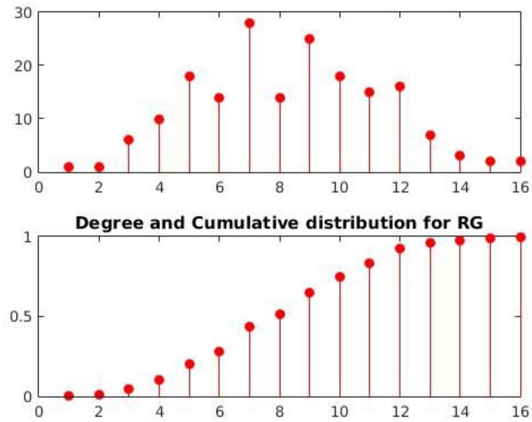


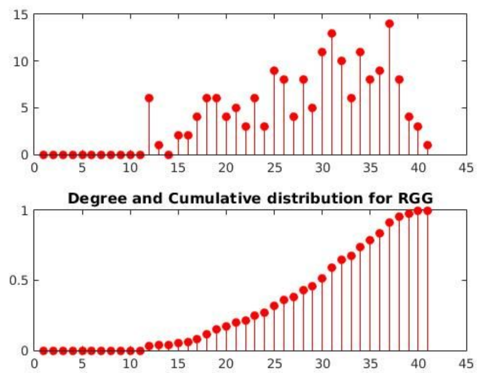
Βλέπουμε με την σειρά REG ,RG,RGG,SF,SW με 180 κόμβους ο καθέ ένας.
Βλέπουμε πως στον REG κάθε κόμβος έχει 4 ακμές, επίσης ο SW μοιάζει με τον REG με την διαφορά ότι έχουν γίνει rewire κάποιες ακμές. Βέβαια λόγο της χαμηλής πιθανότητας που έχουμε η εικόνα μας μοιάζει αραιή απο rewired ακμές.

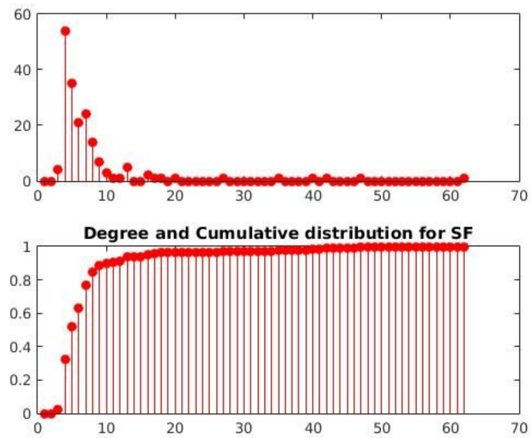
Ο RG γράφος μοιάζει ποιο πυκνός σχηματικά , πράγμα λογικό αφού έχει 750 συνδέσεις . Στον SF βλέπουμε πως έχει αρκετές μακρινές συνδεσεις αλλά λιγότερες απο τον RGG.

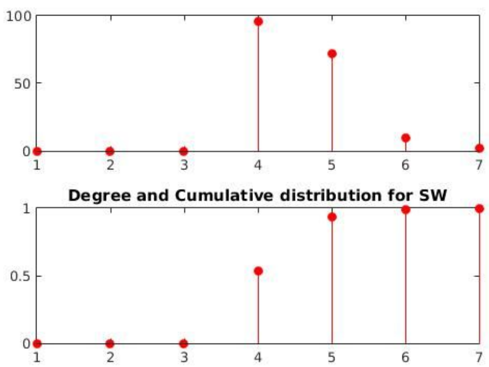
## Μελέτη Βαθμού Κόμβων




| **Γράφος** | **Mean** | **Var** |
|---|---|---|
| REG | 4 | 0 |

| | | |
|---|---|---|
| RG | 8.3333 | 9.0391 |
| RGG | 28.7556 | 53.8394 |
| SF | 7.3889 | 53.9038 |
| SW | 4.5444 | 0.4282 |

Προφανώς τα αποτελέσματα για τον REG ήταν αναμενόμενα. Αφού κάθε κορυφή έχει ακριβώς 4 ακμές.

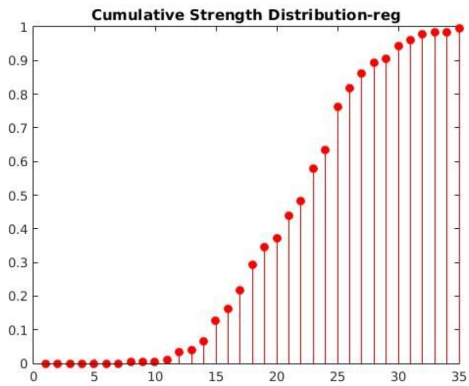
Ο RG επίσεις είναι αναλλοίωτος ως προς το άθροισμα των βαθμών , Αφού σύμφωνα με τον euler το άθροισμα των βαθμών είναι 2 φόρες οι ακμές., που εδώ είναι σταθερές (750) . Έτσι ξέρουμε ότι όσες φορές κι να τρέξουμε τον κωδικά μας, η μέση τιμή θα μείνει σταθερή. Επίσης βλέπουμε πως το πλήθος κορυφών με συγκεκριμένο βαθμών συγκεντρώνεται κοντά στο 8 .Πράγμα αναμενόμενο λόγο της μέσης τιμής. Θεωρητικά έχουμε πει πως επίδη η κατανομή πιθανότητας είναι binomial για μεγάλες τιμές γίνεται poisson . Πέρα λοιπόν απο την γραφική που μοιάζει με poison , η μικρή διαφορά ανάμεσα στο mean κι στο var είναι ένα άλλο στοιχείο αφού σε poison ισχυεί : mean(Poison)=var(Poison)=λ .

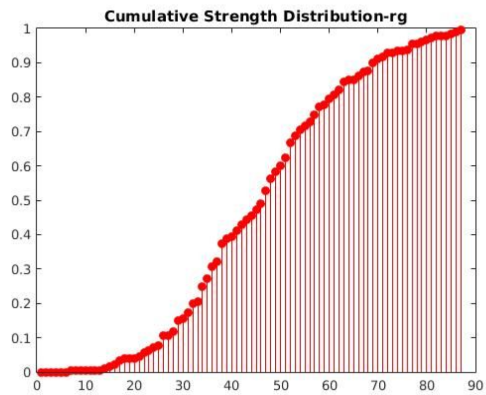
Για τον RGG έχουμε ότι θεωρητικά κάθε κόμβος έχει αναμενόμενη μέση τιμή: $\frac{\pi r^2 n}{L^2}$ = 35.3
Εμείς βγάλαμε ως μέση τιμή 28.75 που απέχει απο την θεωρητική τιμή της , βέβαια αυτό είναι αναμενόμενο λόγο της τεράστιας διασποράς που έχουμε.

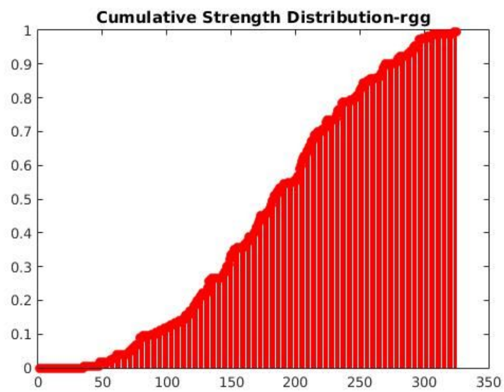
Για τον SF παρατηρούμε ότι έχουμε πάλι τεράστια διασπορά όπως κι πριν, επίσης βλέπουμε πως η μέση τιμή της είναι μικρή, το μοντέλο scale free χαρακτηρίζεται απο το γεγονός ότι είναι μικρή πιθανότητα να υπάρχει κόμβος με μεγάλο βαθμό, κι το περισσότερο πλήθος έχει μικρό βαθμό.

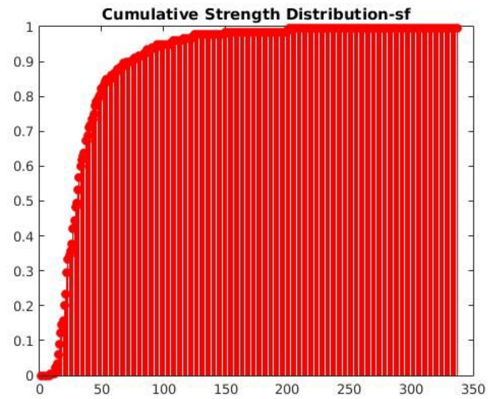
Στον SW βλέπουμε πως ο μέσος βαθμός είναι κοντα στο 4,5 αλλα κι ότι η απόκλιση είναι χαμηλή. Λόγο της μικρής τιμής της πιθανότητας που έχουμε , βλέπουμε πως τα αποτελέσματα είναι πολύ κοντά στο REG. Επιβεβαιώνεται ότι οι περισσότερες κορυφές έχουν τον ίδιο βαθμό.

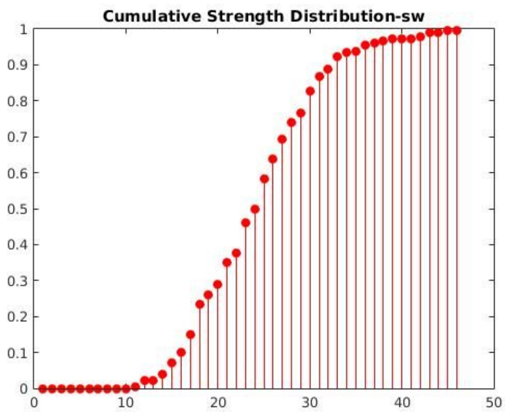
## Δίκτυα με βάρη




| **Γράφος** | **Μέση Δύναμη** |
|---|---|
| REG | 21.5871 |
| RG | 46.3422 |
| RGG | 183.3481 |
| SF | 39.8301 |

| | |
|---|---|
| SW | 24.0227 |

Βλέπουμε πως ο REG κι ο SW έχουν αρκετά μικρότερη δύναμη απο τους άλλους γράφους. Οι RG-RGG έχουν την μεγαλύτερη μέση δύναμη .Κι τέλος αν κι ο SF έχει σχετικά μικρή μέση δύναμη έχει μεγαλύτερα άκρα απο τα άλλα.

## Υπολογισμός Μέσου μήκους μονοπατιού

| **Γράφος** | Μέσο μήκος μονοπάτιου | Διασπορά |
|---|---|---|
| REG | 22.8771 | 169.8137 |
| RG | 2.6614 | 0.4810 |
| RGG | 2.7736 | 1.4214 |
| SF | 2.6058 | 0.4564 |
| SW | 5.3176 | 3.2497 |

Στον γράφο REG η διασπορά κι το μέσο μήκος μονοπάτιου είναι πολύ υψηλές σε σχέση με τους άλλους γράφους, αυτό συμβαίνει γιατί κάθε κόμβος συνδέετε με τον i+1,i+2,i-1,i-2 , κι έτσι η διάμετρος τους γράφου είναι αρκετά μεγάλη.
Στην θεωρία είδαμε πως το SF κι το SW εχούν μικρό average lenth path. Το SF επιβεβαιώνεται αλλά το SW έβγαλε σχετικά μεγαλύτερο απο τα αλλά βέβαια εξυγήτε εξαιτίας της μεγαλής διασποράς που έχει. Ο RGG ξέρουμε ότι έχει μεγάλο μέσο μονοπάτι, αν κι εδώ είναι μεγαλύτερο απο τα RG,SF δεν είναι όμως όυτε αρκετά μεγάλο κι είναι μεγαλύτερο απο το SW . Αυτο λογικά συμβαίνει γιατί έχουμε σχετικά λίγους κόμβους, αφού τα περισσότερα αποτελέσματα στους random graphs τα έχουμε ως ασυμπτωτικά

## Υπολογισμός Συντελεστή ομαδοποιήσης

Το τοπικό ΣΟ του i υπολογίζεται ως : $\frac{2 * \text{Άκμες γειτόνων μεταξύ γετόνων του } i}{\text{μέγιστο πλήθος ακμών μεταξύ γειτώνων του ι}}$

Στον πρώτο γράφο, σε όλες τις κορυφές εκτώς απο την μεσσαία, ο LCC δεν ορίζεται, γιατί το μέγιστο πλήθος ακμών μεταξύ γειτώνων είναι 1(1-1)=0 αφού όλες έχουν 1 γείτονα.
Στην μεσαία κορυφή έχουμε 4 γείτονες αλλά 0 ακμές μεταξυ τους . Άρα: 2*0/4*3 = 0 .

Στον δεύτερο γράφο, αριθμίζω από αριστερά προς δεξία κι προς τα κάτω τις κορυφές . (1,2 οι πάνω, 3 η μεσσαία, 4,5 οι κάτω) .
Λόγο συμμετρίας προφανώς οι 1,2 ,4,5 θα έχουν ίδιο LCC.

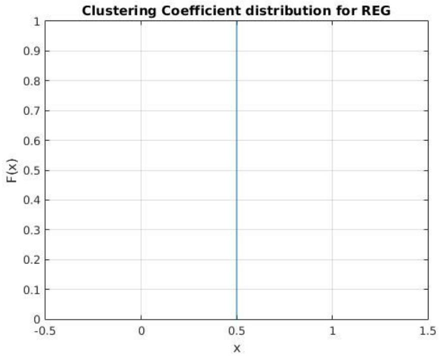
LCC(1)=2*1/(2*1) =1
LCC(2)=2*1/(2*1) =1
LCC(4)=2*1/(2*1) =1
LCC(5)=2*1/(2*1) =1
αφού όλες αυτές έχουν 2 γείτονες, κι μεταξύ τους υπάρχει μια ακμή.
LCC(4)=2*2/(4*3)=⅓
Γιατί έχει 4 γείτονες, κι οι 1,2 κι 4,5 συνδεονταί με μια ακμή.
Μέσος όρος: (1+1+1+1+⅓ )/5 =13/15 .

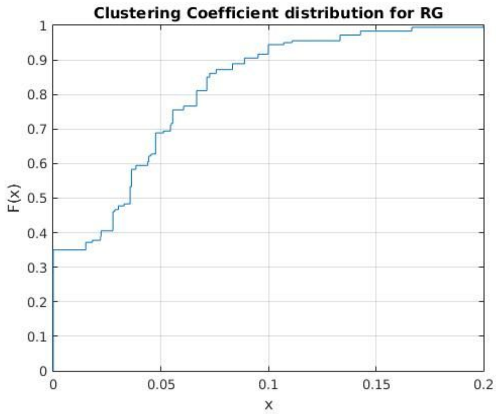
Στον τρίτο γράφο αριθμό όπως πριν άρα: η πάνω 1, 2,3,4 οι μεσαίες κι 5 η τελευταία.Λόγο συμμετρίας αναμένουμε πάλι όλες εκτώς απο την μεσσαία να έχουν το ίδιο.
Έχουμε λοιπόν:
LCC(1)=2*2/(3*2)=⅔ 3 γείτονες: 2,3,4 κι 2 ακμές 2-3 και 3-4.
LCC(2)=2*2/(3*2)=⅔ 3 γείτονες: 3,1,5 κι 2 ακμές 3-1 και 3-5.
LCC(4)=2*2/(3*2)=⅔ 3 γείτονες: 1,5,3 κι 2 ακμές 1-3 και 3-5.
LCC(5)=2*2/(3*2)=⅔ 3 γείτονες: 2,3,4 κι 2 ακμές 2-3 και 3-4.

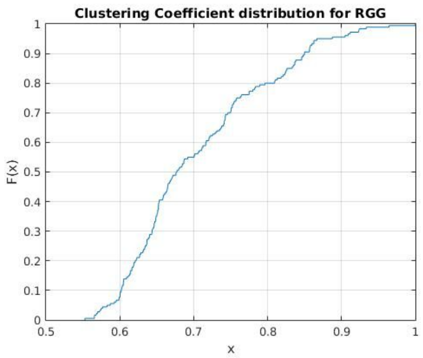
LCC(3)=2*4/(4*3)=⅔ 4 γείτονες: 1,2,4,5 κι 4 ακμές 2-1,2-5 και 1-4,4-5.

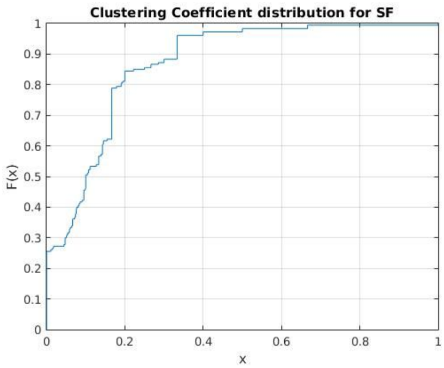
Και ο μέσος όρος είναι προφανώς ⅔

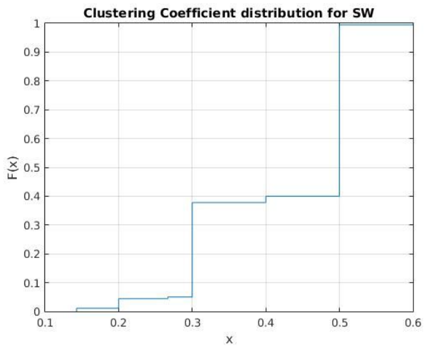
## Υπολογισμός ΣΟ σε μεγαλύτερες συνθετικές τοπολογίες με χρήση Matlab




| **Γράφος** | **Mean CC** |
|---|---|
| REG | 0.5 |
| RG | 0.038 |
| RGG | 0.7064 |
| SF | 0.1317 |
| SW | 0.4175 |

Στον REG όλες οι κορυφές έχουν LCC 0.5 ,γιατί κάθε κόμβος είναι ενωμένος με τον i+1,i+2,i-1,i-2. Κι έτσι με έναν υπολογισμό βγαίνει ότι είναι 0.5.

Στον RG βλέπουμε τον χαμηλότερο συντελεστή ομαδοποιήσεις .

Στον RGG βλέπουμε τον μεγαλύτερο συντελεστή ομαδοποιήσεις.

Ο SF έχει μεγαλύτερο ασυντέλεστη απο τον RG αλλά μικρότερο από όλα τα άλλα.

Ο SW έχει συντελεστή ομαδοποιήσεις κοντά στον REG , βέβαια είναι μικρότερος κι αυτό συμβαίνει εξαιτίας του rewire των ακμών.Επίσεις χαρακτηρίζεται απο σχετικά μεγάλο CC πράγμα που δείχνουν κι οι μετρήσεις μας.

## Υπολογισμός Κεντρικότητα κόμβων

**Αναλυτικός υπολογισμός κεντρικότητας**

Γραφος 1:Κεντρικός κόμβος 1, δεξιοστροφα: 2,3,4

Για τον κόμβο 1 έχουμε:
Degree Centrality: 3 αφού έχει 3 ακμές
Closeness Centrality: ενώνεται με όλες της κορυφές άρα η απόσταση από όλες είναι 1. Άρα 1/(1+1+1)=$\frac{1}{3}$
Betweenness Centrality: Είναι ενδιάμεσος απο τον 2-3,2-4 και 3-4 άρα 3.

Για τον κόμβο 2 έχουμε:
Degree Centrality: 1 αφού έχει 1 ακμή
Closeness Centrality: με την 1 απέχει απόσταση 1, ενώ με τις άλλες απόσταση 2 άρα 1/(1+2+2)=$\frac{1}{5}$
Betweenness Centrality: Δεν είναι ενδιάμεσος κανενός άρα: 0

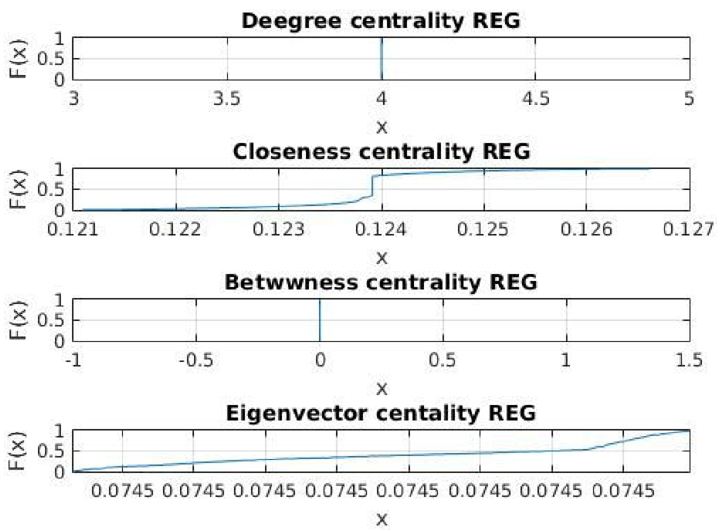
Για τον κόμβο 3 έχουμε:
Degree Centrality: 1 αφού έχει 1 ακμή
Closeness Centrality: με την 1 απέχει απόσταση 1, ενώ με τις άλλες απόσταση 2 άρα 1/(1+2+2)=$\frac{1}{5}$
Betweenness Centrality: Δεν είναι ενδιάμεσος κανενός άρα: 0

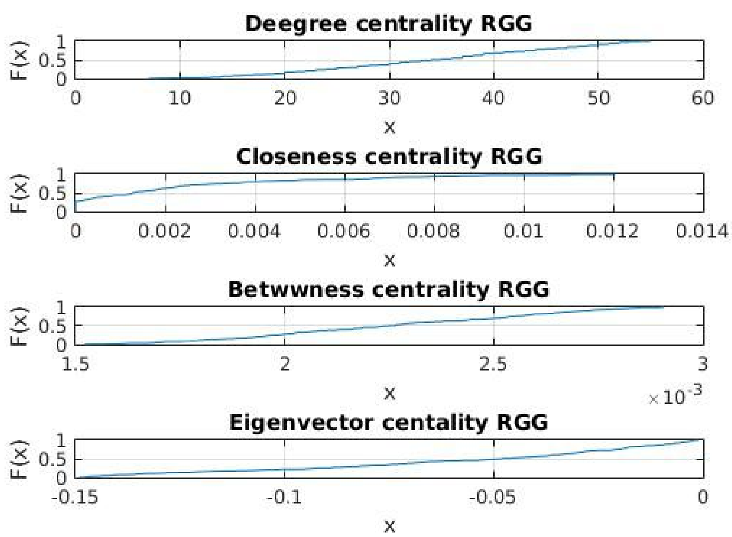
Για τον κόμβο 4 έχουμε:
Degree Centrality: 1 αφού έχει 1 ακμή
Closeness Centrality: με την 1 απέχει απόσταση 1, ενώ με τις άλλες απόσταση 2 άρα 1/(1+2+2)=$\frac{1}{5}$
Betweenness Centrality: Δεν είναι ενδιάμεσος κανενός άρα: 0

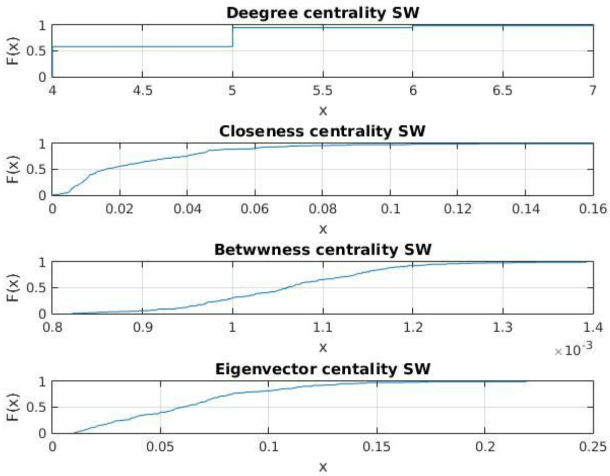
Γράφος 2:
Αριθμό απο αριστερά στα δεξιά 1,2 κι απο κάτω 3,4.

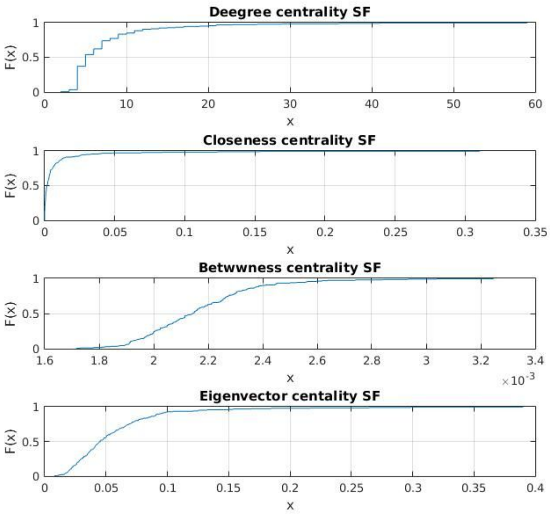
Επίδη ο γράφος είναι κλίκα ότι ισχύει για έναν κόμβο ισχυεί για όλους.
Για τον κόμβο 1 πχ. έχουμε:
Degree Centrality: 3 αφού έχει 3 ακμές
Closeness Centrality: ενώνεται με όλες της κορυφές άρα η απόσταση από όλες είναι 1. Άρα 1/(1+1+1)=⅓
Betweenness Centrality: 0 γιατί κανένας κόμβος δεν είναι ενδιάμεσος κανενός αφού είναι κλίκα.

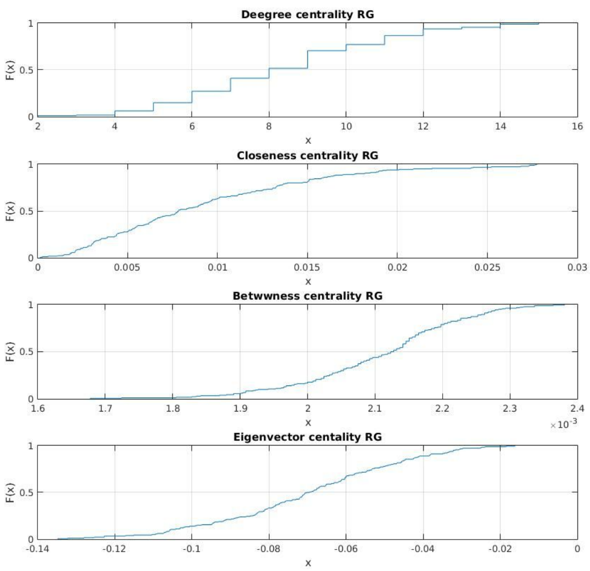
## Υπολογισμός Κεντρικότητας σε μεγαλύτερες συνθετικές τοπολογίες με χρήση Matlab




Deegree centrality SW
F(x)
x
Closeness centrality SW
F(x)
x
Betwwness centrality SW
F(x)
x
$\times 10^{-3}$
Eigenvector centality SW
F(x)
x

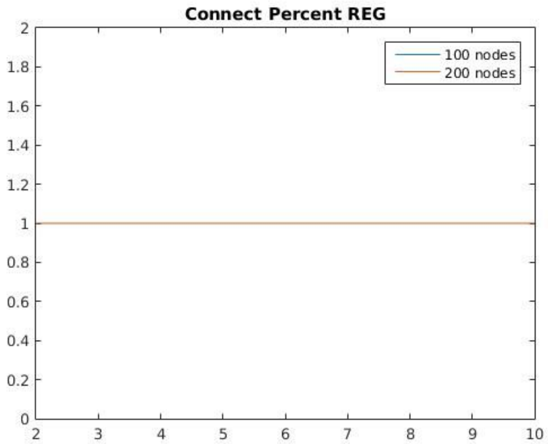

Deegree centrality SF
F(x)
x
Closeness centrality SF
F(x)
x
Betwwness centrality SF
F(x)
x
$\times 10^{-3}$
Eigenvector centality SF
F(x)
x

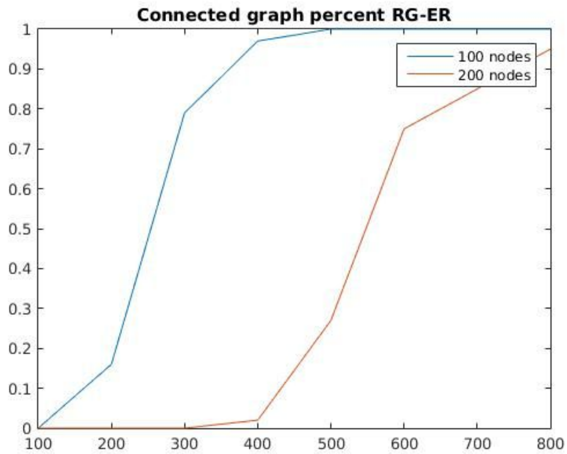

Deegree centrality RG
F(x)
x
Closeness centrality RG
F(x)
x
Betwwness centrality RG
F(x)
x
$\times 10^{-3}$
Eigenvector centality RG
F(x)
x

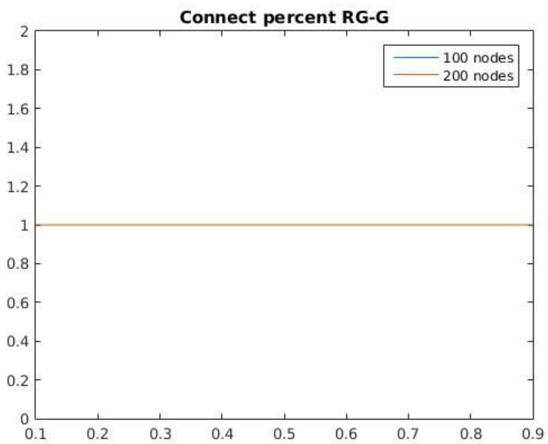
| **Γράφος** | mean Deg.Cent | mean Closs cent | mean Betw. cent | mean Eigen cent |
|---|---|---|---|---|
| REG | 4 | 2.4420e-04 | 0.1238 | 0.0745 |
| RG | 8.333 | 0.0021 | 0.0093 | -0.0704 |
| RGG | 28.7556 | 0.0023 | 0.0023 | -0.0597 |
| SF | 7.3889 | 0.0022 | 0.0085 | 0.0574 |
| SW | 4.5444 | 0.0011 | 0.0258 | 0.0639 |

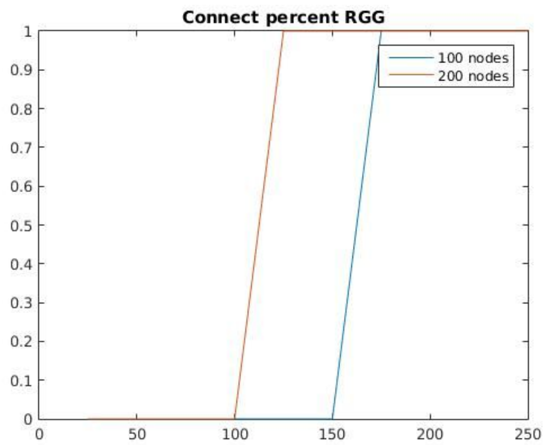
Για τον REG τα αποτελέσματα μας δείχνουν το αναμενόμενο ότι το closscent centrality είναι αρκετά μικρό, αφού η απόσταση προς κάθε κόμβο είναι μεγάλη, αλλά το betweeness centrality είναι μεγαλύτερο από όλα τα άλλα αφού για να πάω σε έναν κόμβο αναγκαστικά θα πρέπει να περάσω απο αρκετούς λόγο του τρόπου που έχουν γίνει οι συνδέσεις
Στον RG και στον RGG οι κεντρικότητες είναι σχεδόν ίδιες εκτός απο του βαθμού που είναι λογικό, πράγμα που ξέρουμε ότι θεωρητικά ισχύει αφού ειναι uniform. Ο SW έχει χαμηλό closscent λόγο του rewire αλλά ακόμα έχει αρκετά πιο ψηλό Betweenses απο το RG,RGG,SF.Στον small world βλέπουμε αυτά, εξαιτίας του small world phenomenon.

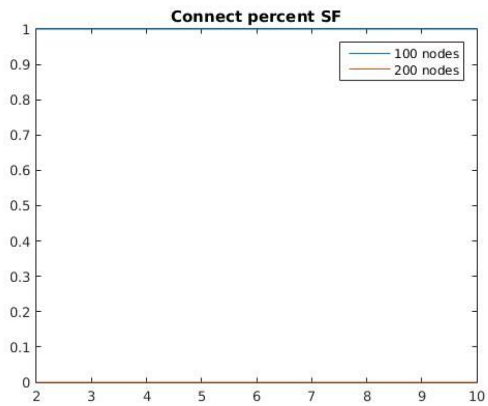
# Μελέτη συνεκτικότητας και συμπεριφορά κατωφλίων

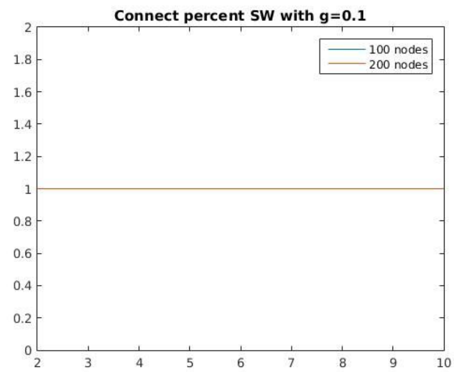


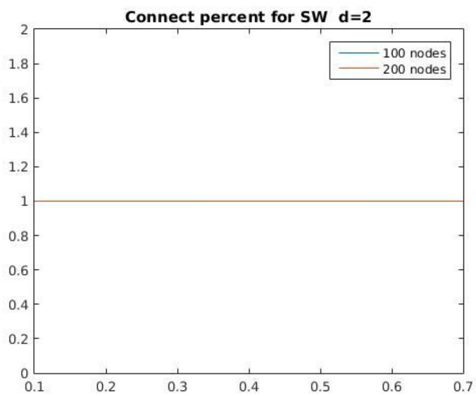






Σε όσα φαίνεται μια γραμμή συμβαίνει γιατί είναι η μια γραμμή πάνω στην άλλη.

Ξεκινώντας με το REG βλέπουμε ότι για κάθε τιμή είναι connected (εξ-ορισμού λογικό) άρα δεν έχει φαινόμενα κατοφλίου.

Στον ER-RG έχουμε φαινόμενα κατοφλίου , βλέπουμε πως στις 100 κορυφές μετά τις 400 ακμές αρχίζει να είναι πάντα connected,ενώ στα 200 είναι μετά τις 800 ακμές. Κι οι δυο μεταβιβάσεις είναι sharp.

Στον RG-G γράφο , βλέπουμε πως είναι συνέχεια connected, αυτό συμβαίνει γιατί η πιθανότητα ξεκινά απο 0.1 > logn/n=0.02 (100), 0.01(200) , όπου είναι το κατώφλη για connected.

Στον RGG  γράφο βλέπουμε φαινόμενο κατωφλίου: στα 175 για 100 κορυφές κι στα 125 για 200 κορυφές.Σε αυτά τα 2 σημεία έχουμε sharp αλλαγή φάσης.Αυτό σημβαίνει γιατί η αναμενόμενη τιμή γειτόνων είναι ανάλογη τις ακτίνας κι τον κορυφών.
BA-SF βλέπουμε πως για 100 ακμές είναι πάντα connected, ενώ για 200 ακμές δεν είναι. Δεν παρατηρούμε κανένα κατώφλη, στα διαγράματα μας.
Τέλος κι στα 2 SW βλέπουμε πως είναι πάντα connected, κι αυτό σημβαίνει γιατί ξεκινάμε απο ένα REG κι κάνουμε διαδοχικά rewires.

## Μελέτη μοντέλων τυχαίων γράφων

| Τοπολογία | 100 | 10^3 | 10^4 | 10^5 | 10^6 |
|---|---|---|---|---|---|
| RG (G) | 0.1 | 10^-2 | 10^-3 | 10^-4 | 10^-5 |
| RG (ER) | 495 | 4995 | 49995 | 499995 | 4999995 |

Στο μοντέλο του GIlbert ξέρουμε πως αν np είναι σταθερό κι αν το   pn^2 είναι αρκετά μεγάλο τότε συμπεριφέρεται σαν τον ER μοντέλο με M=n(n-1)*p/2 . Απο αυτό το τύπο συμπληρώνουμε τον πινακά μας.

## Μελέτη της εξελικτικής μετατροπής δικτύου REG σε δύκτιο SW και RG(ER)

| Πιθανότητα | Μέσο μήκος Μονοπατιού | Mean CC |
|---|---|---|
| 0 | 22.8771 | 0.5 |
| 0.1 | 8.4028 | 0.4739 |
| 0.2 | 5.7520 | 0.4307 |
| 0.3 | 5.1165 | 0.4069 |
| 0.4 | 4.5428 | 0.3710 |
| 0.5 | 4.3970 | 0.3649 |
| 0.6 | 3.9083 | 0.3115 |
| 0.7 | 3.8151 | 0.3091 |
| 0.8 | 3.6352 | 0.2744 |
| 0.9 | 3.4580 | 0.2468 |
| 1 | 3.4192 | 0.2433 |

Για πιθανότητα ίση με 0 το δύκτιο συμπεριφέρεται σαν REG , ενω για πιθανότητα 1 ως random graph.
Συμφώνα με την wikipedia βλέπουμε πως το μέσο μήκος μονοπατιού για p=1 είναι lnN/lnd στην περίπτωση μας N=180,d=4 άρα είναι περίπου ίσο με 3,745 , αρκετά κοντά στην δικιά μας τιμή .Επίσεις για p=0 πρέπει να είναι ίσο με N/2K =22,5 που είναι πάλι αρκετά κοντά στις τιμές μας.
Στα CC πρέπει να ισχυεί ότι για p=0, C=3(K-2)/(4(K-1)) =0.5 ( το έχουμε υπολογίσει κι πιο πριν). Και προσσεγγιστικά για p=1 C=K/N=0.02 εδώ δεν ταιρίαζει με την τιμή που έχουμε βρει μιας κι η δικιά μας τιμή είναι αρκετά μεγαλήτερη.

Βλέπουμε πως για το p=0.1-0.4 το average path πέφτει γρηγόρα ενώ ο CC πιο αργά(small world phenomenon ). Πράγμα λογικό αφού εμπειρικά το small world λεει πως μπορούμε να πάμε από έναν κόμβο σε έναν άλλο σε πολύ λίγα βήματα αν κι οι περισότεροι κόμβοι δεν ειναι μεταξύ τους γείτονες.