

**ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS**

**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ**

Network of Thrones

Εργασία του φοιτητή Κωνσταντίνου Νικολούτσου, ΑΜ: 3170122.

Περιεχόμενα

1	Το κοινωνικό δίκτυο	2
1.1	Εισαγωγή	2
1.2	Πηγή	2
1.3	Εργαλεία που χρησιμοποιήθηκαν	2
1.4	Δεδομένα	3
2	Υλοποίηση	4
2.1	Αναπαράσταση του γραφήματος	4
2.2	Τοπολογία γραφήματος	5
2.2.1	Γενικά χαρακτηριστικά	5
2.2.2	Πυκνότητα γραφήματος	6
2.3	Συνεκτικές συνιστώσες	6
2.4	Βαθμοί κορυφών	7
2.5	Μέτρα κεντρικότητας	8
2.5.1	Eccentricity	8
2.5.2	Closeness centrality	9
2.5.3	Harmonic closeness centrality	11
2.5.4	Betweenness centrality	12
2.5.5	Eigenvector centrality	14
2.5.6	Bridging centrality	15
2.5.7	Επιπλέον παρατηρήσεις	16
2.6	Clustering coefficient	18
2.7	Pagerank	20
2.8	Modularity	22
2.9	Homophily	24
3	Συμπεράσματα	25
	Βιβλιογραφία	26

Κατάλογος Σχημάτων

2.1	Το γράφημα του κοινωνικού δικτύου μας.	4
2.2	Force Atlas layout.	5
2.3	Degree distribution.	7
2.4	Top 20 χαρακτήρες με τον μεγαλύτερο αριθμό αλληλεπιδράσεων.	7
2.5	Eccentricity distribution.	8
2.6	Σύγκριση betweenness centrality και εκκεντρότητας.	9
2.7	Closeness centrality distribution.	9
2.8	Το γράφημα μας χρωματισμένο βάσει του closeness centrality.	10
2.9	Harmonic closeness centrality distribution.	11
2.10	Σύγκριση closeness centrality και harmonic closeness centrality.	11
2.11	Το γράφημα μας χρωματισμένο βάσει του harmonic closeness centrality.	12
2.12	Betweenness centrality distribution.	12
2.13	Top 5 χαρακτήρες με τις μεγαλύτερες τιμές betweenness centrality.	13
2.14	Το γράφημα μας χρωματισμένο βάσει του betweenness centrality.	13
2.15	Eigenvector centrality distribution.	14
2.16	Το γράφημα μας χρωματισμένο βάσει του eigenvector centrality.	15
2.17	Bridging centrality distribution.	15
2.18	Το γράφημα μας χρωματισμένο βάσει του bridging centrality.	16
2.19	Συγκρίσεις μέτρων κεντρικότητας.	17
2.20	Clustering coefficient distribution.	18
2.21	Clustering coefficient (υψηλές τιμές).	19
2.22	Clustering coefficient (χαμηλές τιμές).	19
2.23	Συγκρίσεις Clustering coefficient και βαθμού κορυφής.	19
2.24	Pagerank distribution.	20
2.25	Γράφημα μεθόδου Pagerank.	21
2.26	Αριθμός ατόμων ανά κοινότητα (modularity).	22
2.27	Οι 7 κοινότητες στο γράφημά μας, χωρισμένες βάσει χρωμάτων.	23
2.28	Αναπαράσταση του homophily.	24

Κεφάλαιο 1

Το κοινωνικό δίκτυο

1.1 Εισαγωγή

Η σειρά Game of Thrones είναι από τις δημοφιλέστερες σειρές του κόσμου και έχει βασιστεί στο μεγαλύτερο κομμάτι της στη σειρά βιβλίων A Song of Ice and Fire. Η σειρά αυτή έδωσε το έναυσμα για τη δημιουργία του κοινωνικού δικτύου το οποίο αποτελείται από τους χαρακτήρες που πρωταγωνιστούν στο τρίτο βιβλίο. Πιο συγκεκριμένα, στο τρίτο βιβλίο της σειράς (οπότε και η πλοκή έχει ξεκινήσει, και υπάρχει νόημα στο να γίνει κάποια ανάλυση) γίνεται μελέτη της σχέσης αλληλεπίδρασης μεταξύ των χαρακτήρων. Η σχέση αυτή βασίζεται στο αν και πόσες φορές υπάρχουν τα ονόματα δύο χαρακτήρων σε απόσταση το πολύ 15 λέξεων. Λόγω της σχέσης που αναλύεται, έχουμε στοιχεία μόνο για το πόσο αλληλεπιδρούν δύο χαρακτήρες μεταξύ τους, και όχι για την μεταξύ τους φιλία, η οποία μπορεί να μην υπάρχει. Στην παρούσα εργασία, χρησιμοποιώντας εργαλεία ανάλυσης κοινωνικών δικτύων, θα αναδειχθούν και θα αναλυθούν ιδιότητες του δικτύου που έχουμε επιλέξει.

1.2 Πηγή

Η πηγή για το κοινωνικό δίκτυο που αναλύεται, βρίσκεται στον ακόλουθο [σύνδεσμο](#). Στον σύνδεσμο αυτόν υπάρχουν άλλα πέντε κοινωνικά δίκτυα με διάφορες θεματολογίες τα οποία έχουν δημιουργηθεί για εκπαιδευτικούς σκοπούς και τα οποία μπορεί κανείς να αναλύσει. Τα δεδομένα είναι έτοιμα και δεν χρειάζονται κάποιου είδους προετοιμασία από εμάς, απλά εισάγουμε τα αρχεία στο εργαλείο της επιλογής μας.

1.3 Εργαλεία που χρησιμοποιήθηκαν

Το εργαλείο που χρησιμοποιήθηκε για την υλοποίηση της εργασίας είναι το περιβάλλον Gephi, το οποίο είναι γραμμένο σε Java και παρέχει πληθώρα έτοιμων εργαλείων για την ανάλυση κοινωνικών δικτύων. Δεν χρειάστηκε κάποιο επιπλέον εργαλείο, ούτε γνώση προγραμματισμού. Οι εικόνες και οι πίνακες της παρούσας εργασίας προέρχονται όλες από το Gephi. Το Gephi εκτός από την οπτική αναπαράσταση του γραφήματος και της ανάλυσης του, όπως για παράδειγμα η εύρεση μετρικών, (καρτέλα Overview) μας επιτρέπει να δούμε στοιχεία του γραφήματος όπως βαθμοί κορυφών, κεντρικότητες, ετικέτες (καρτέλα Data Laboratory).

1.4 Δεδομένα

Τα δεδομένα έχουν αναπαρασταθεί με τη μορφή ενός απλού συνεκτικού γραφήματος με βάρη¹. Κάθε χαρακτήρας του βιβλίου αναπαρίσταται από μία μοναδική κορυφή στο γράφημα. Μία ακμή $u \rightarrow v$ υπάρχει στο γράφημα αν και μόνο αν οι χαρακτήρες u, v εμφανίζονται στο κείμενο σε απόσταση το πολύ 15 λέξεων². Το βάρος της ακμής uv είναι ίσο με το πόσες φορές οι χαρακτήρες u, v έχουν αλληλεπιδράσει μεταξύ τους. Αν δηλαδή οι χαρακτήρες u, v αλληλεπιδράσουν μεταξύ τους 10 φορές, η ακμή uv θα έχει βάρος 10. Επιπλέον, προφανώς το γράφημα δεν είναι κατευθυνόμενο, αφού η αλληλεπίδραση μεταξύ των χαρακτήρων είναι αμφίπλευρη. Τα δεδομένα για το γράφημα έχουν χωριστεί σε τρία αρχεία:

- Το αρχείο `got-edges.csv` περιέχει τις ακμές του γραφήματος.
- Το αρχείο `got-nodes.csv` περιέχει τις κορυφές του γραφήματος.
- Το αρχείο `got-network.graphml` περιέχει το γράφημα σε μορφή XML.

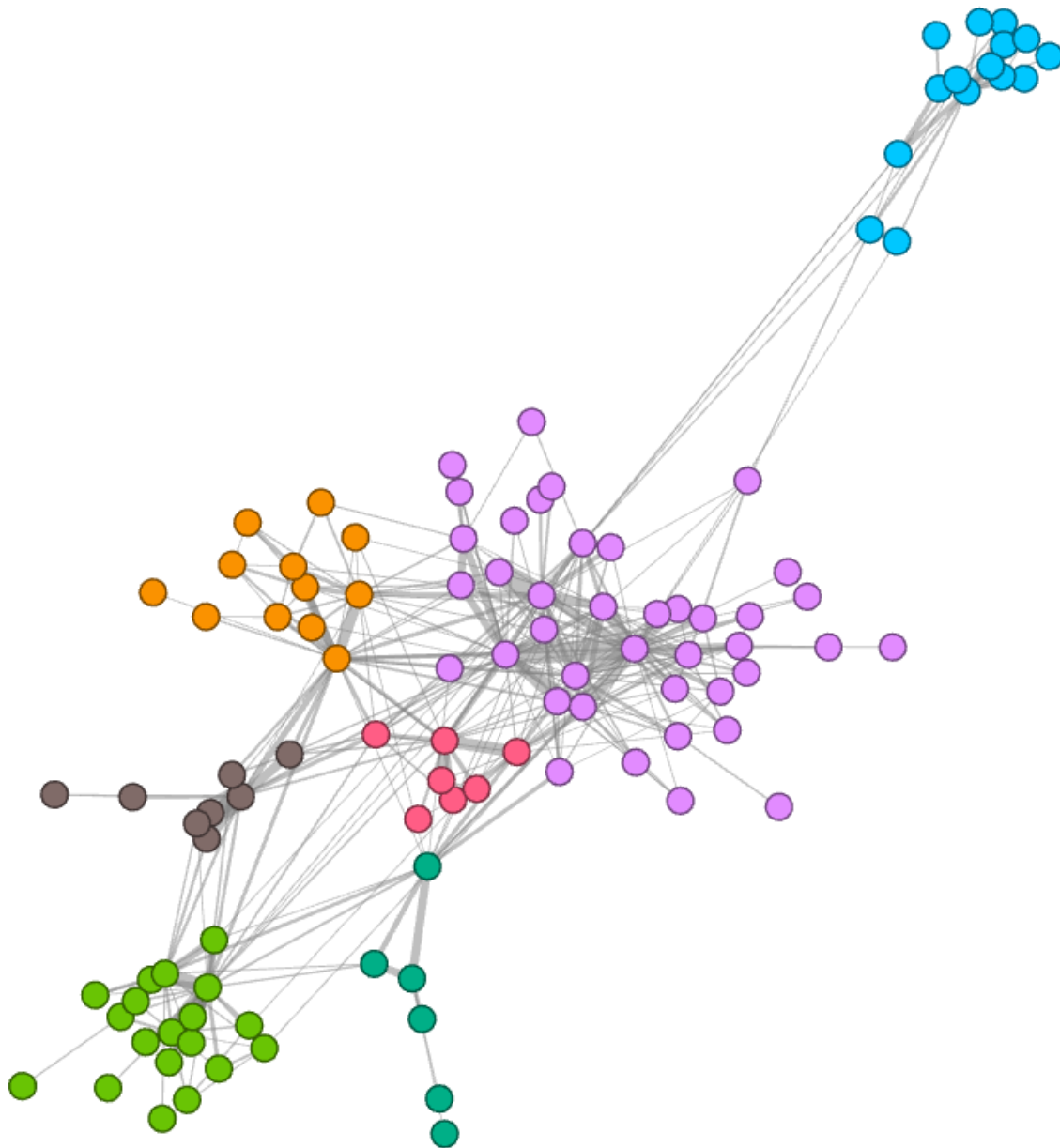
¹Απλό σημαίνει ότι δεν έχει βρόχους.

²Στο εξής θα αναφέρεται ως 'αλληλεπιδρούν'.

Κεφάλαιο 2

Υλοποίηση

2.1 Αναπαράσταση του γραφήματος



Σχήμα 2.1: Το γράφημα του κοινωνικού δικτύου μας.

Όπως βλέπουμε, το γράφημά μας έχει κορυφές με 7 διαφορετικά χρώματα (το γιατί αναλύεται αργότερα). Παρατηρούμε ότι κάποιες ακμές είναι πιο πυκνές από κάποιες άλλες. Αυτό συμβαίνει επειδή όσο μεγαλύτερο είναι το βάρος μιας ακμής, τόσο πιο πυκνή είναι. Δεν έχουν τοποθετηθεί ετικέτες στις κορυφές του γραφήματος καθώς σκοπός είναι να αναδειχθεί η δομή του και όχι τα ονόματα των χαρακτήρων. Οι ετικέτες των κορυφών, αν υπήρχαν, θα αντιστοιχούσαν στο όνομα του χαρακτήρα που σχετίζεται η κάθε κορυφή του γραφήματος.

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

☐ Force Atlas	
Inertia	0.1
Repulsion strength	10000.0
Attraction strength	10.0
Maximum displacement	10.0
Auto stabilize function	<input checked="" type="checkbox"/>
Autostab Strength	80.0
Autostab sensibility	0.2
Gravity	30.0
Attraction Distrib.	<input type="checkbox"/>
Adjust by Sizes	<input type="checkbox"/>
Speed	1.0

Σχήμα 2.2: Force Atlas layout.

Αξίζει να σημειωθεί ότι το Gephi μας παρέχει πληθώρα από layouts με τα οποία μπορούμε να πειραματιστούμε αλλάζοντας τις διάφορες ρυθμίσεις τους, για να αναπαραστήσουμε το γράφημά μας. Εκτός από το layout του γραφήματος μπορούμε να αλλάξουμε τον τρόπο αναπαράστασης των κορυφών (για παράδειγμα αλλάζοντας τους χρώμα ή μέγεθος) και των ακμών.

2.2 Τοπολογία γραφήματος

2.2.1 Γενικά χαρακτηριστικά

Το γράφημά μας έχει 107 κορυφές (πλήθος χαρακτήρων) και 352 ακμές (πλήθος ζευγών χαρακτήρων που αλληλεπιδρούν). Επιπλέον, έχει την εξής τοπολογία:

- Διάμετρος: 6
- Ακτίνα: 3

- Μέσο μήκος μονοπατιού: 2.90

Η διάμετρος ενός γραφήματος είναι η μεγαλύτερη απόσταση ανάμεσα σε όλα τα δυνατά ζεύγη κορυφών, η ακτίνα του γραφήματος είναι η ελάχιστη εκκεντρότητα, ενώ το μέσο μήκος μονοπατιού είναι ο μέσος όρος του μήκους των συντομότερων μονοπατιών μεταξύ όλων των ζεύγων κορυφών σε ένα δίκτυο. Όπως είναι αναμενόμενο, το μέσο μήκος μονοπατιού είναι αρκετά μικρότερο από τη διάμετρο του γραφήματος, καθώς ο αριθμός των χαρακτήρων είναι αρκετά μεγάλος, και υπάρχουν χαρακτήρες οι οποίοι παίζουν μικρό ρόλο στην εξέλιξη της υπόθεσης οπότε δεν υπάρχει μεγάλη αλληλεπίδραση με άλλους παίκτες από μεριάς τους.

2.2.2 Πυκνότητα γραφήματος

Η πυκνότητα του γραφήματος μας δείχνει πόσες ακμές έχει το γράφημα αναλογικά με τον αριθμό των κορυφών του. Το γράφημα μας έχει πυκνότητα 0.062, κάτι αναμενόμενο, καθώς έχει 352 ακμές. Αν το γράφημα ήταν πλήρες θα είχε πυκνότητα 1, που σε αυτήν την περίπτωση θα είχε σύνολο $\binom{107}{2} = \frac{107!}{2!105!} = 5671$ ακμές, δεδομένου ότι το γράφημα δεν έχει βρόχους. Επιπλέον, αν το γράφημα είχε πυκνότητα 1, αυτό θα σήμαινε ότι όλοι οι χαρακτήρες του βιβλίου αλληλεπιδρούν μεταξύ τους ανά δύο. Συνεπώς, η μικρή πυκνότητα του γραφήματος μας υποδεικνύει ότι δεν υπάρχει αρκετή αλληλεπίδραση μεταξύ διαφορετικών χαρακτήρων.

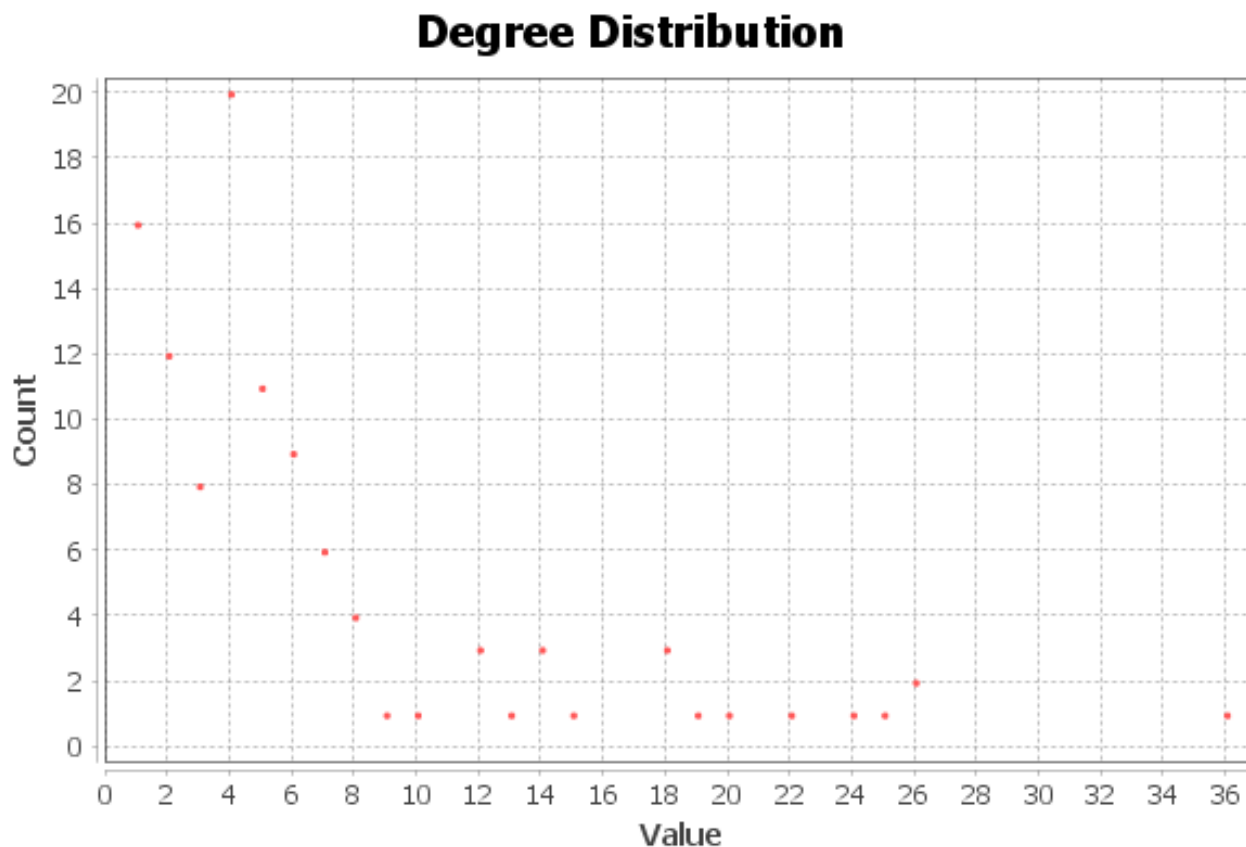
2.3 Συνεκτικές συνιστώσες

Το γράφημα μας αποτελείται από μία μόνο συνεκτική συνιστώσα, αφού η πλοκή διαδραματίζεται σε ένα σύμπαν που όλοι μπορούν να αλληλεπιδράσουν μεταξύ τους και δεν υπάρχουν αποκομμένοι χαρακτήρες, αποκομμένες κοινότητες ή παράλληλα σύμπαντα.

2.4 Βαθμοί κορυφών

Results:

Average Degree: 6,579



Σχήμα 2.3: Degree distribution.

Παρατηρούμε ότι ο μέσος βαθμός στο γράφημα είναι 6.579, δηλαδή κάθε παίκτης αλληλεπιδρά κατά μέσο όρο με $\lceil 6.579 \rceil = 7$ άλλους χαρακτήρες. Επιπλέον, παρατηρούμε ότι οι κορυφές με μεγάλο βαθμό στο γράφημα είναι λίγες (για παράδειγμα, υπάρχει μόνο μία κορυφή βαθμού 36) ενώ υπάρχουν αρκετές κορυφές με μικρό βαθμό (για παράδειγμα, υπάρχουν 20 κορυφές βαθμού 2). Αυτό συνάδει με το γεγονός ότι το μέσο μήκος μονοπατιού είναι αρκετά μικρότερο από τη διάμετρο του γραφήματος. Παρακάτω βλέπουμε τους 20 χαρακτήρες με τον μεγαλύτερο αριθμό αλληλεπιδράσεων.

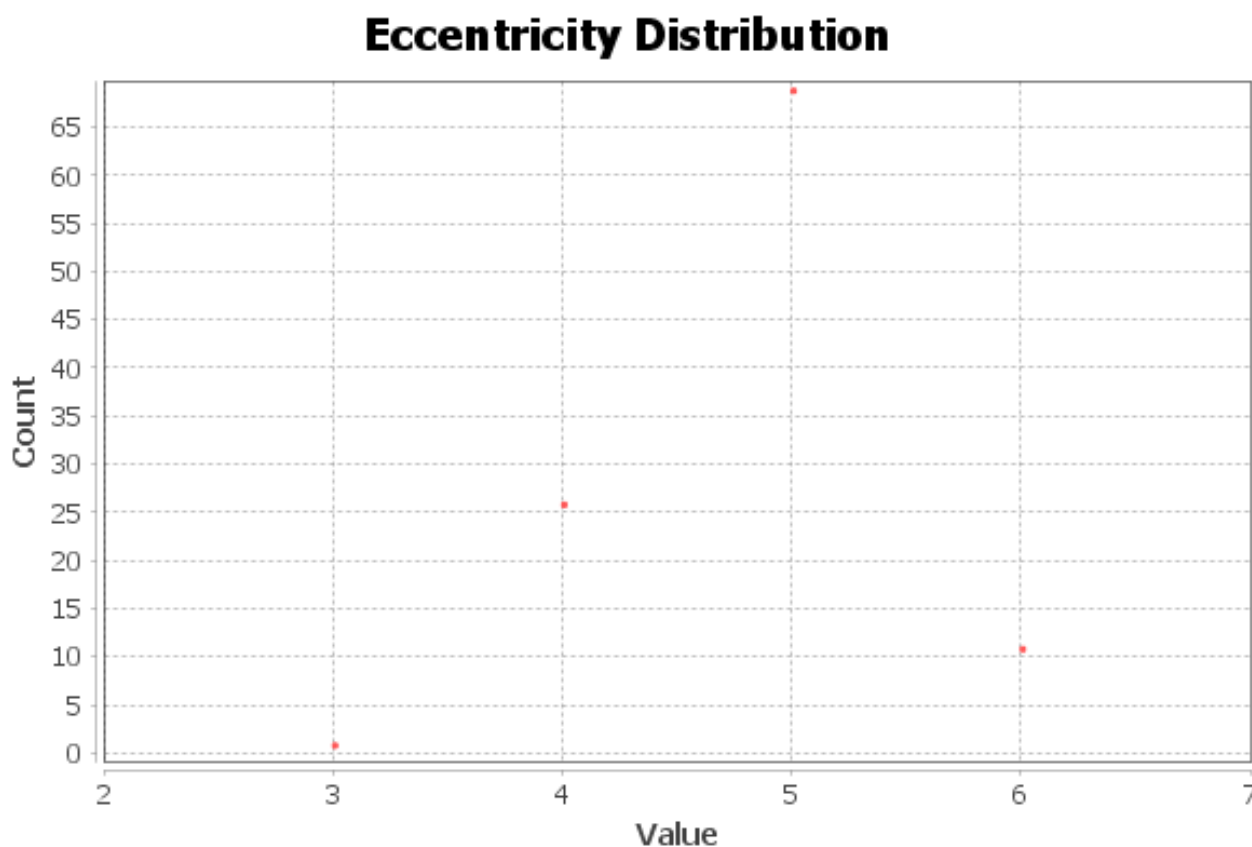
Label	Degree	Cersei	20	Stannis	14
Tyion	36	Arya	19	Daenerys	14
Jon	26	Robert	18	Sandor	13
Sansa	26	Joffrey	18	Mance	12
Robb	25	Catelyn	18	Gregor	12
Jaime	24	Samwell	15	Eddard	12
Tywin	22	Bran	14	Lysa	10

Σχήμα 2.4: Top 20 χαρακτήρες με τον μεγαλύτερο αριθμό αλληλεπιδράσεων.

2.5 Μέτρα κεντρικότητας

Θα χρησιμοποιηθούν τα αγγλικά ονόματα για τα μέτρα κεντρικότητας προς αποφυγή σύγχυσης, καθώς δεν υπάρχουν μεταφράσεις στα ελληνικά για ορισμένα πράγματα που παρουσιάζονται στην παρούσα εργασία.

2.5.1 Eccentricity



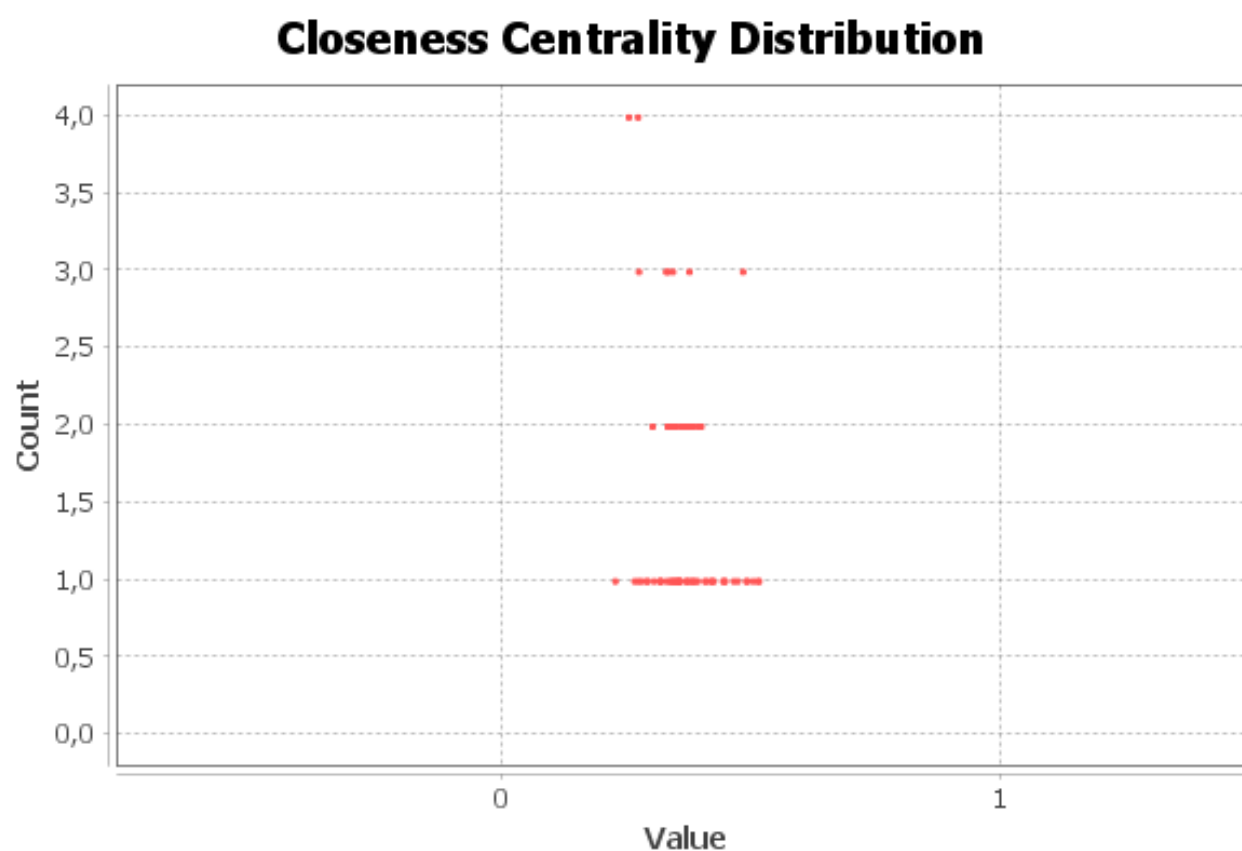
Σχήμα 2.5: Eccentricity distribution.

Παρατηρούμε ότι οι τιμές της eccentricity (εκκεντρότητα) σε αυτήν την περίπτωση ανήκουν στο σύνολο $\{3, 4, 5, 6\}$. Αυτό είναι αναμενόμενο, αφού η τιμή της εκκεντρότητας μιας κορυφής εξαρτάται από την μέγιστη απόσταση της κορυφής αυτής από τις υπόλοιπες κορυφές του γραφήματος. Γνωρίζουμε ότι η διάμετρος του γραφήματος είναι 6, συνεπώς είναι αδύνατον κάποια μέγιστη απόσταση να υπερβαίνει την τιμή αυτή. Προφανώς οι χαρακτήρες με εκκεντρότητα ίση με 6 είναι μακριά από τους υπόλοιπους (δεν συμμετέχουν στα συντομότερα μονοπάτια μεταξύ κορυφών), πράγμα που μας υποδεικνύει ο παρακάτω πίνακας όπου παρατηρούμε ότι 10 από τους 11 χαρακτήρες με εκκεντρότητα=6, έχουν betweenness centrality=0. Το μέτρο betweenness centrality αναλύεται παρακάτω.

Label	Eccentricity	Betweenness
Jojen	6.0	2.261905
Luwin	6.0	0.0
Nan	6.0	0.0
Bowen	6.0	0.0
Lancel	6.0	0.0
Amory	6.0	0.0
Karl	6.0	0.0
Cressen	6.0	0.0
Salladhor	6.0	0.0
Shireen	6.0	0.0
Illyrio	6.0	0.0

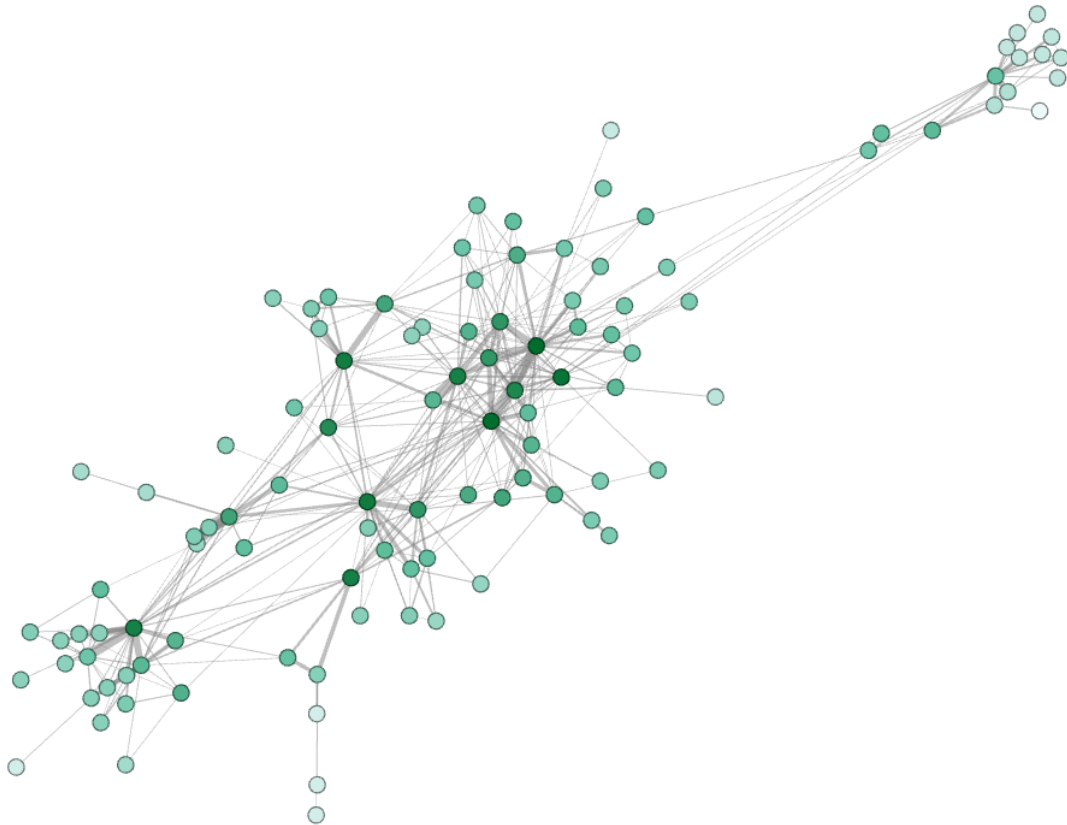
Σχήμα 2.6: Σύγκριση betweenness centrality και εκκεντρότητας.

2.5.2 Closeness centrality



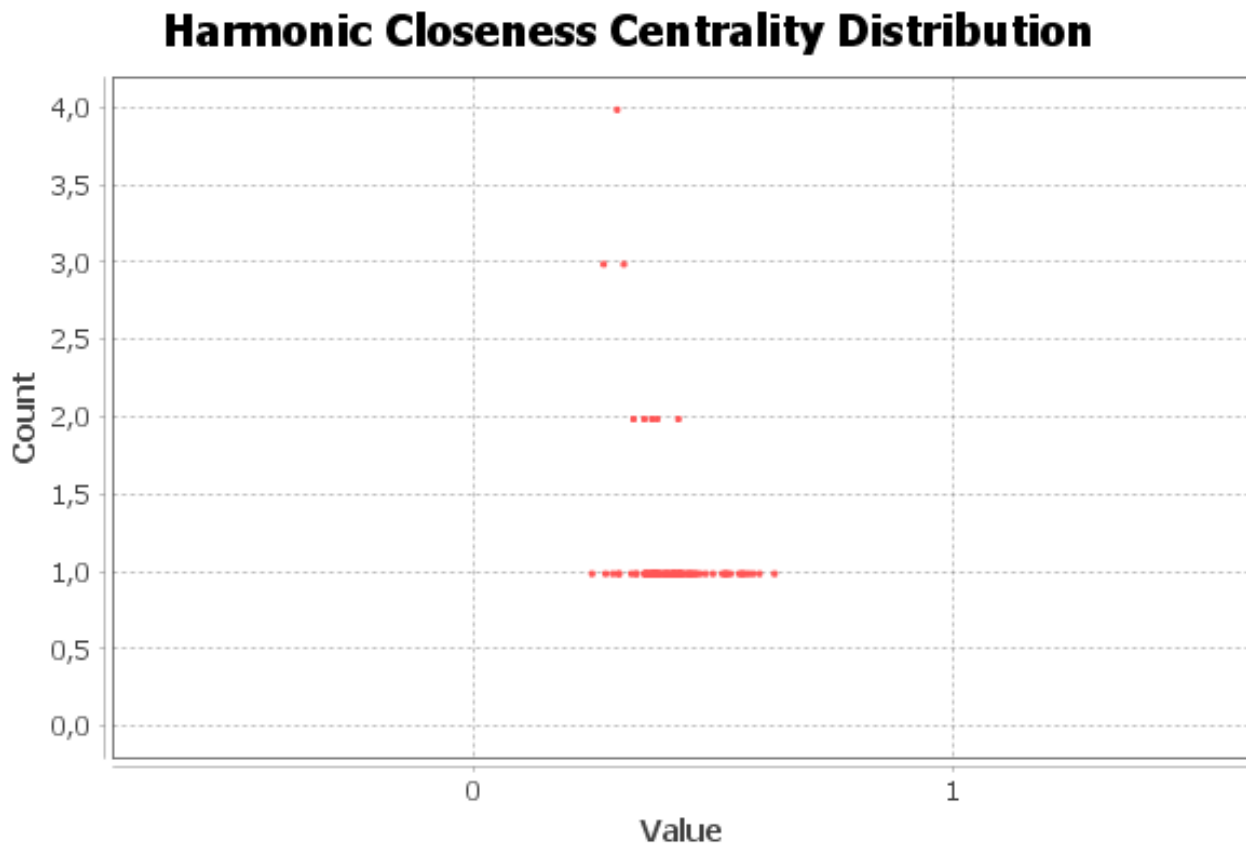
Σχήμα 2.7: Closeness centrality distribution.

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 γράφματος. Παρατηρούμε ότι οι τιμές της closeness centrality βρίσκονται στο διάστημα $[0, 1]$, είναι δηλαδή κανονικοποιημένες. Όπως είναι αναμενόμενο, οι περισσότερες τιμές απέχουν αρκετά από το να έχουν τιμή 1 καθώς στο γράφημα αλληλεπιδρούν 7 διαφορετικές κοινότητες ανθρώπων, οπότε είναι δύσκολο να υπάρχουν πολλές κορυφές που να ανήκουν σε πολλά συντομότερα μονοπάτια. Το αντίθετο θα μπορούσε να επιτευχθεί αν το γράφημα είχε μορφή σαν αυτή του αστεριού ([star graph](#)) όπου εκεί υπάρχει μια κορυφή στη μέση που αλληλεπιδρά άμεσα με όλες τις υπόλοιπες κορυφές. Στο παρακάτω γράφημα, οι κορυφές έχουν χρωματιστεί βάσει του closeness centrality. Όσο πιο μεγάλη η τιμή του closeness centrality, τόσο πιο σκούρο πράσινο χρώμα έχει η κορυφή.



Σχήμα 2.8: Το γράφημα μας χρωματισμένο βάσει του closeness centrality.

2.5.3 Harmonic closeness centrality



Σχήμα 2.9: Harmonic closeness centrality distribution.

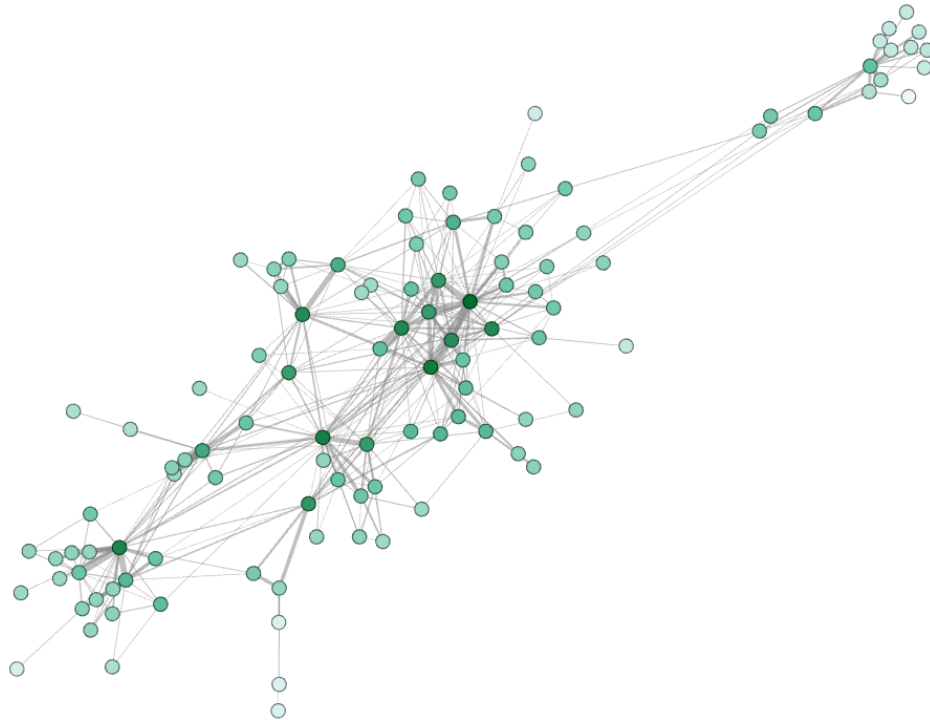
Αρχικά, παρατηρούμε ότι οι τιμές της harmonic closeness centrality βρίσκονται στο διάστημα $[0, 1]$, είναι δηλαδή κανονικοποιημένες. Οι τιμές της harmonic closeness centrality είναι πολύ κοντά με τις τιμές της closeness centrality, όπως φαίνεται τόσο από τις γραφικές παραστάσεις, όσο και από τον παρακάτω πίνακα:

Label	Closeness Centrality	Harmonic Close
Tyrior	0.512077	0.622642
Sansa	0.509615	0.591195
Robert	0.5	0.556604
Robb	0.488479	0.57783
Arya	0.486239	0.553459
Jaime	0.479638	0.562107
Jon	0.479638	0.569182
Stannis	0.479638	0.53066
Tywin	0.469027	0.548742
Eddard	0.46087	0.513365

Σχήμα 2.10: Σύγκριση closeness centrality και harmonic closeness centrality.

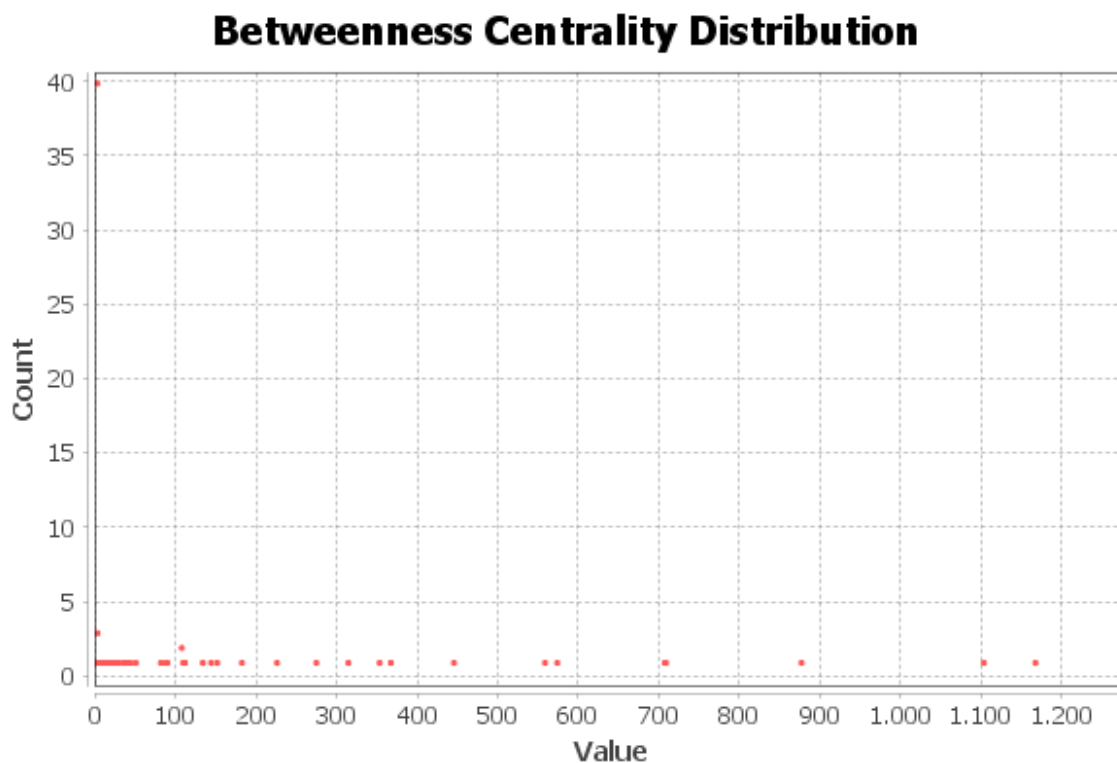
Παρατηρούμε επιπλέον ότι οι διαφορές στις παραπάνω τιμές είναι μέχρι, περίπου, 20%.

Στο παρακάτω γράφημα, οι κορυφές έχουν χρωματιστεί βάσει του harmonic closeness centrality. Όσο πιο μεγάλη η τιμή του harmonic closeness centrality, τόσο πιο σκούρο πράσινο χρώμα έχει η κορυφή.



Σχήμα 2.11: Το γράφημα μας χρωματισμένο βάσει του harmonic closeness centrality.

2.5.4 Betweenness centrality



Σχήμα 2.12: Betweenness centrality distribution.

Η betweenness centrality είναι ένα μέτρο της θέσης μιας κορυφής σε ένα γράφημα. Για μια κορυφή, η betweenness centrality μεγαλώνει ανάλογα με το πλήθος των συντομότερων μονοπατιών του γραφήματος στα οποία βρίσκεται. Παρατηρούμε ότι το μεγαλύτερο ποσοστό συγκέντρωσης υπάρχει στις μικρές τιμές της betweenness centrality, μέχρι την τιμή 200. Αυτό είναι αναμενόμενο, αφού όπως μπορούμε να δούμε και στην αναπαράσταση του δικτύου μας, υπάρχουν αρκετοί χαρακτήρες (κορυφές στο γράφημα) που δεν αλληλεπιδρούν με πολλούς χαρακτήρες, οπότε και είναι δύσκολο να υπάρχουν σε πολλά από τα συντομότερα μονοπάτια. Παρακάτω βλέπουμε τους πέντε χαρακτήρες με τις μεγαλύτερες τιμές betweenness centrality.

Label	Betweenness Centrality
Jon	1279.753353
Robert	1165.602517
Tyrion	1101.384972
Daenerys	874.837211
Robb	706.557283

Σχήμα 2.13: Top 5 χαρακτήρες με τις μεγαλύτερες τιμές betweenness centrality.

Στο παρακάτω γράφημα, οι κορυφές έχουν χρωματιστεί βάσει του betweenness centrality. Όσο πιο μεγάλη η τιμή του betweenness centrality, τόσο πιο σκούρο πράσινο χρώμα έχει η κορυφή.



Σχήμα 2.14: Το γράφημα μας χρωματισμένο βάσει του betweenness centrality.

2.5.5 Eigenvector centrality

Parameters:

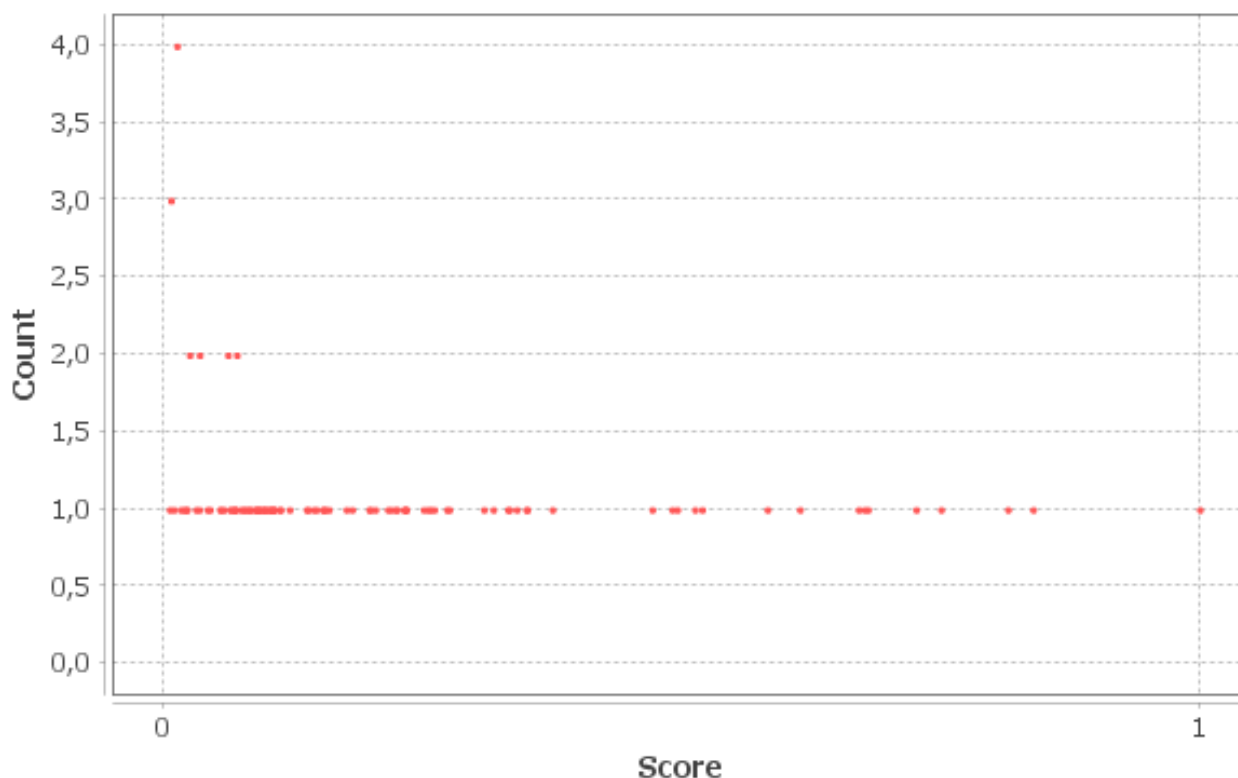
Network Interpretation: undirected

Number of iterations: 100

Sum change: 0.004779113408040032

Results:

Eigenvector Centrality Distribution



Σχήμα 2.15: Eigenvector centrality distribution.

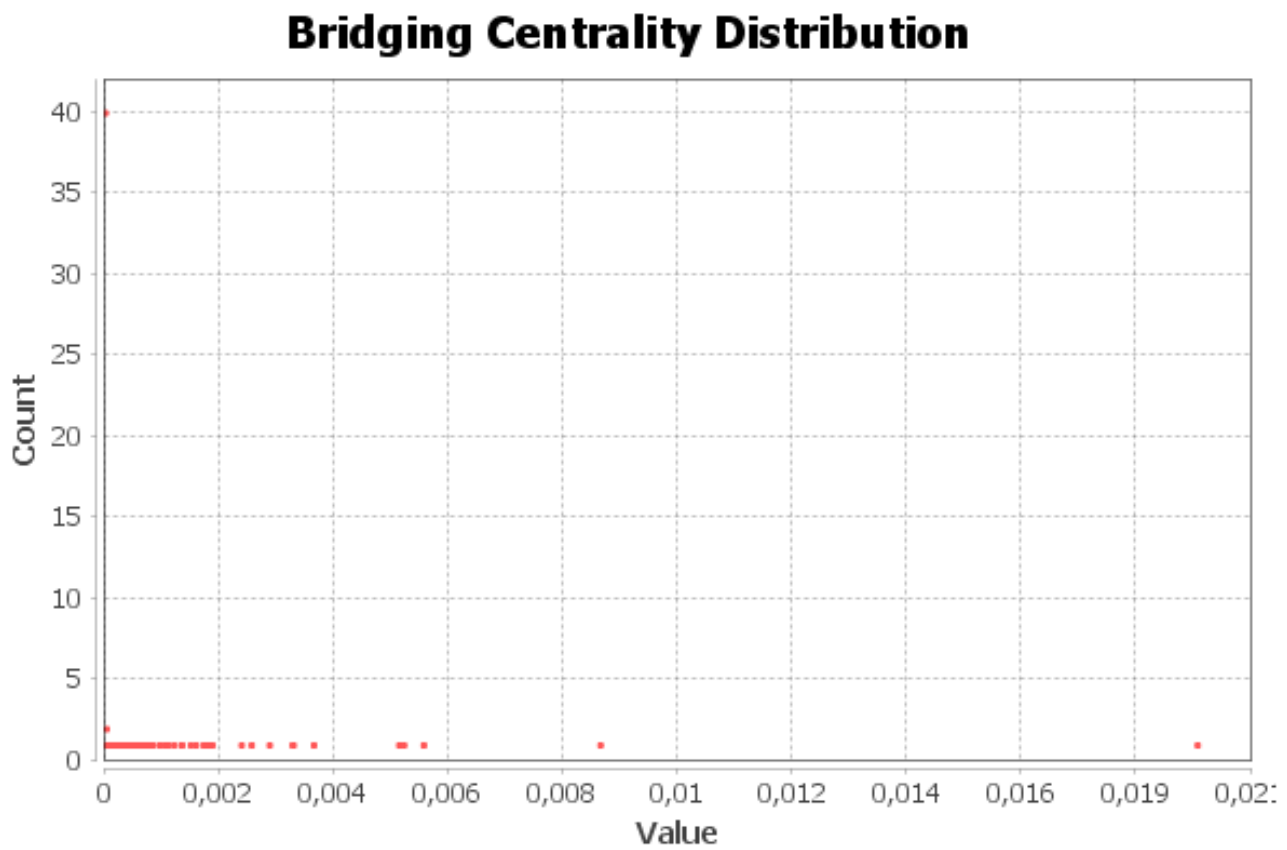
Η έννοια της eigenvector centrality είναι αναδρομική, και είναι γενίκευση της μεθόδου Pagerank και της Katz centrality. Μια κορυφή έχει μεγάλο σκορ ως προς την eigenvector centrality, όταν οι κορυφές με τις οποίες συνδέεται έχουν και αυτές μεγάλο σκορ. Παρατηρούμε ότι η συγκέντρωση των σκορ είναι στις χαμηλές τιμές. Επιπλέον, παρατηρούμε ότι ο Tyrion έχει eigenvector centrality ίση με 1 και αυτό οφείλεται στην πολύ ενεργή του παρουσία στο βιβλίο. Γνωρίζουμε επιπλέον ότι ο Tyrion έχει τον μεγαλύτερο αριθμό αλληλεπιδράσεων, ο οποίος είναι ίσος με 36. Ωστόσο είναι λάθος να συσχετιστεί η eigenvector centrality με τον αριθμό των χαρακτήρων που αλληλεπιδρά ένας χαρακτήρας, όπως θα σκεφτόταν κανείς. Αυτό επιβεβαιώνεται από τον Jon ο οποίος παρότι αλληλεπιδρά με 26 άλλους χαρακτήρες, έχει eigenvector centrality 0.49.

Στο παρακάτω γράφημα, οι κορυφές έχουν χρωματιστεί βάσει του eigenvector centrality. Όσο πιο μεγάλη η τιμή του eigenvector centrality, τόσο πιο σκούρο πράσινο έχει η κορυφή. Παρατηρούμε ότι υπάρχει μια ομοιομορφία στον χρωματισμό των κορυφών, από μέσα (σκούρο πράσινο) προς τα έξω (ανοιχτό πράσινο).



Σχήμα 2.16: Το γράφημα μας χρωματισμένο βάσει του eigenvector centrality.

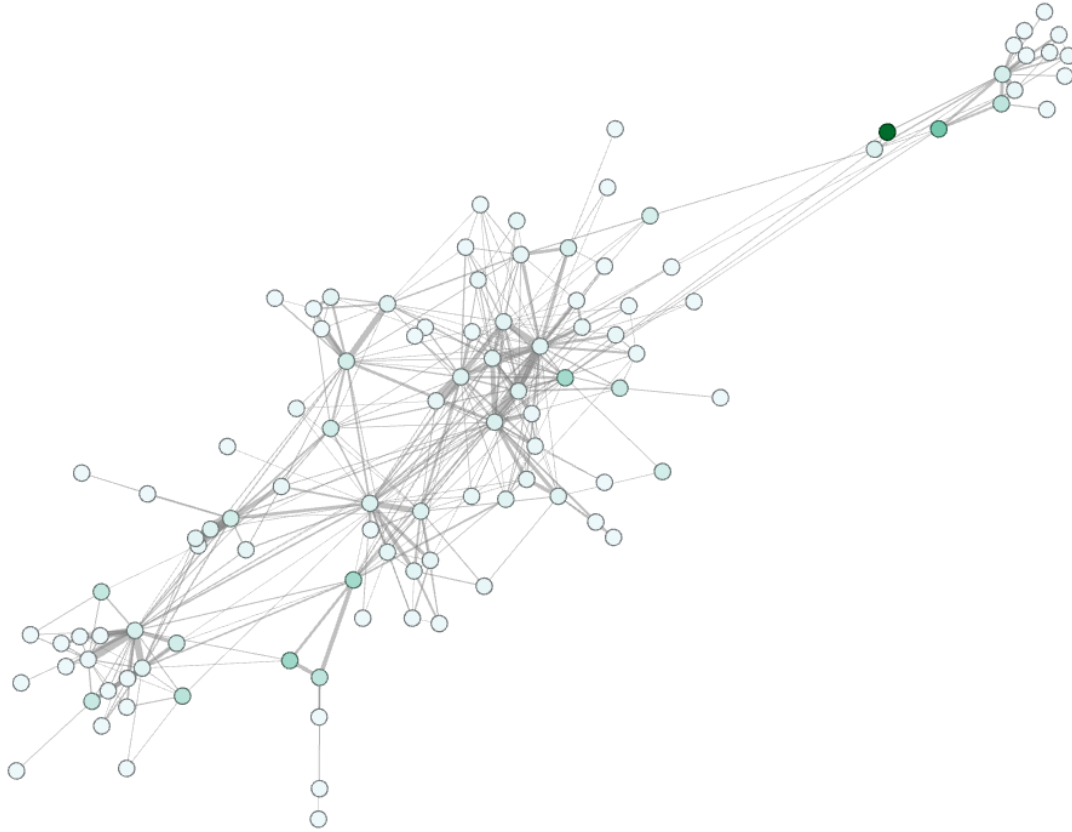
2.5.6 Bridging centrality



Σχήμα 2.17: Bridging centrality distribution.

Αρχικά, παρατηρούμε ότι οι τιμές της bridging centrality είναι αρκετά χαμηλές. Το γράφημά μας αποτελείται από μία συνιστώσα που περιέχει κοινότητες ανθρώπων που αλληλεπιδρούν μεταξύ τους, συνεπώς είναι δύσκολο να υπάρχει μια ακμή του γραφήματος που να ενώνει δύο μεγάλα κομμάτια του (δηλαδή υπογραφήματα με μεγάλο αριθμό κορυφών). Αν, για παράδειγμα, υπήρχαν περιορισμοί στον τρόπο με τον οποίο αλληλεπιδρούν τα άτομα των κοινοτήτων μεταξύ τους, το αποτέλεσμα θα ήταν αρκετά διαφορετικό.

Στο παρακάτω γράφημα, οι κορυφές έχουν χρωματιστεί βάσει του bridging centrality. Όσο πιο μεγάλη η τιμή του bridging centrality, τόσο πιο σκούρο πράσινο χρώμα έχει η κορυφή.



Σχήμα 2.18: Το γράφημα μας χρωματισμένο βάσει του bridging centrality.

2.5.7 Επιπλέον παρατηρήσεις

Στον παρακάτω πίνακα, παρατηρούμε τα εξής: Όσον αφορά τα μέτρα closeness, harmonic closeness, betweenness, eigenvector centrality, υπάρχουν χαρακτήρες οι οποίοι κατέχουν πάντα πολύ υψηλές θέσεις. Για παράδειγμα, ο Tyrion, έχει την πρώτη θέση στα 3 από τα 4 μέτρα που εξετάσαμε. Οι χαρακτήρες Jon, Sansa, Robert έχουν και αυτοί αρκετά υψηλές θέσεις.

Label	Closeness Centrality	Label	Harmonic Closenes
Tyrion	0.512077	Tyrion	0.622642
Sansa	0.509615	Sansa	0.591195
Robert	0.5	Robb	0.57783
Robb	0.488479	Jon	0.569182
Arya	0.486239	Jaime	0.562107
Jaime	0.479638	Robert	0.556604
Jon	0.479638	Arya	0.553459

Label	Betweenness Cen	Label	Eigenvector Centrality
Jon	1279.753353	Tyrion	1.0
Robert	1165.602517	Sansa	0.838784
Tyrion	1101.384972	Jaime	0.814284
Daenerys	874.837211	Robb	0.749776
Robb	706.557283	Cersei	0.725584
Sansa	705.198562	Joffrey	0.679049
Stannis	571.524731	Arya	0.67539

Σχήμα 2.19: Συγκρίσεις μέτρων κεντρικότητας.

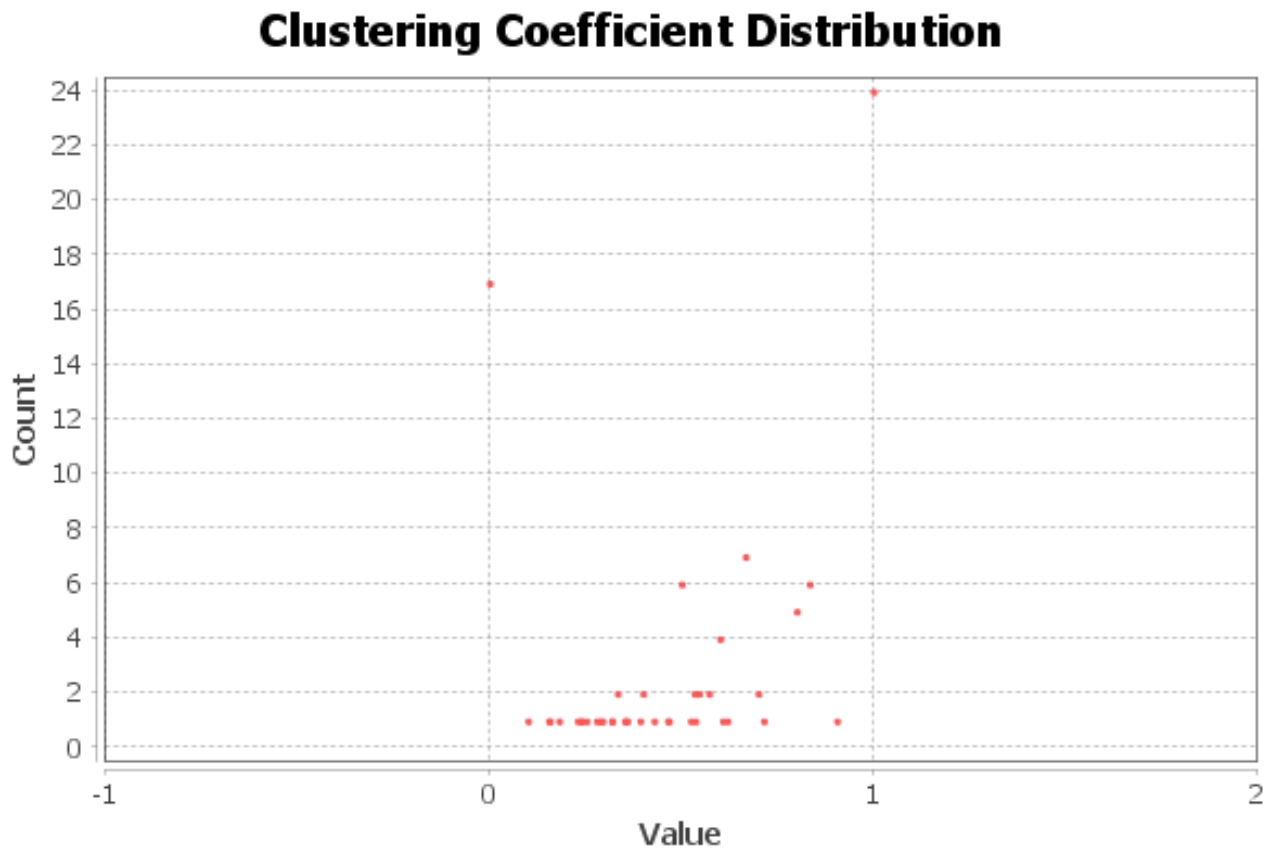
2.6 Clustering coefficient

Results:

Average Clustering Coefficient: 0,648

Total triangles: 469

The Average Clustering Coefficient is the mean value of individual coefficients.



Σχήμα 2.20: Clustering coefficient distribution.

Παρατηρούμε ότι η πλειοψηφία των τιμών για το clustering coefficient συγκεντρώνεται σε τιμές ≥ 0.5 . Επίσης, παρατηρούμε ότι συνολικά στο γράφημα υπάρχουν 469 τρίγωνα, αριθμός αρκετά υψηλός δεδομένου ότι το γράφημά μας έχει 107 κορυφές. Ο μέσος συντελεστής συσταδοποίησης είναι αρκετά μεγάλος, 0.648, και αυτό οφείλεται στην ύπαρξη μεγάλου αριθμού τριγώνων. Πράγματι, όπως θα δούμε στον παρακάτω πίνακα, αρκετές κορυφές έχουν βαθμό συσταδοποίησης 1.

Label	Clustering Coe	Bowen	1.0	Grenn	0.833333
Aerys	1.0	Dalla	1.0	Jojen	0.833333
Alliser	1.0	Rattleshirt	1.0	Theon	0.833333
Anguy	1.0	Styr	1.0	Roslin	0.833333
Gendry	1.0	Ygritte	1.0	Daario	0.833333
Luwin	1.0	Olenna	1.0	Drogo	0.833333
Nan	1.0	Marillion	1.0	Rickon	0.8
Jeyne	1.0	Robert Arryn	1.0	Thoros	0.8
Ilyn	1.0	Ellaria	1.0	Balon	0.8
Pycelle	1.0	Mace	1.0	Lothar	0.8
Missandei	1.0	Rickard	1.0	Qhorin	0.8
Eddison	1.0	Chataya	1.0	Petyr	0.714286
Gilly	1.0	Meryn	0.904762		

Σχήμα 2.21: Clustering coefficient (υψηλές τιμές).

Όπως παρατηρούμε από τα αποτελέσματα του παραπάνω πίνακα, υπάρχουν πολλές κορυφές στο γράφημα με μεγάλη τάση για δημιουργία συστάδων (clusters). Ωστόσο, υπάρχουν 17 κορυφές με τιμή 0, οι οποίες αναπόφευκτα επηρεάζουν το αποτέλεσμα και χωρίς τις οποίες ο μέσος συντελεστής συσταδοποίησης θα ήταν αρκετά μεγαλύτερος. Οι κορυφές αυτές παρουσιάζονται στον επόμενο πίνακα.

Amory	0.0	Qyburn	0.0
Illyrio	0.0	Orell	0.0
Karl	0.0	Jon Arryn	0.0
Aegon	0.0	Lancel	0.0
Kraznys	0.0	Ramsay	0.0
Rakharo	0.0	Shireen	0.0
Worm	0.0	Doran	0.0
Cressen	0.0	Walton	0.0
Salladhor	0.0		

Σχήμα 2.22: Clustering coefficient (χαμηλές τιμές).

Στον παρακάτω πίνακα παρατηρούμε ότι οι κορυφές με μεγάλο συντελεστή clustering έχουν μικρό βαθμό. Αυτό γίνεται καθώς οι κορυφές με μεγάλο βαθμό, αφού αλληλεπιδρούν με πολλές άλλες κορυφές δεν είναι απομονωμένες στο γράφημα, σε αντίθεση με τις κορυφές μικρού βαθμού.

Label	Degree	Clustering			
Ilyn	6	1.0	Ygritte	4	1.0
Aerys	4	1.0	Alliser	3	1.0
Gendry	4	1.0	Eddison	3	1.0
Pycelle	4	1.0	Dalla	3	1.0
Gilly	4	1.0	Marillion	3	1.0
Rattleshirt	4	1.0	Robert A...	3	1.0

Σχήμα 2.23: Συγκρίσεις Clustering coefficient και βαθμού κορυφής.

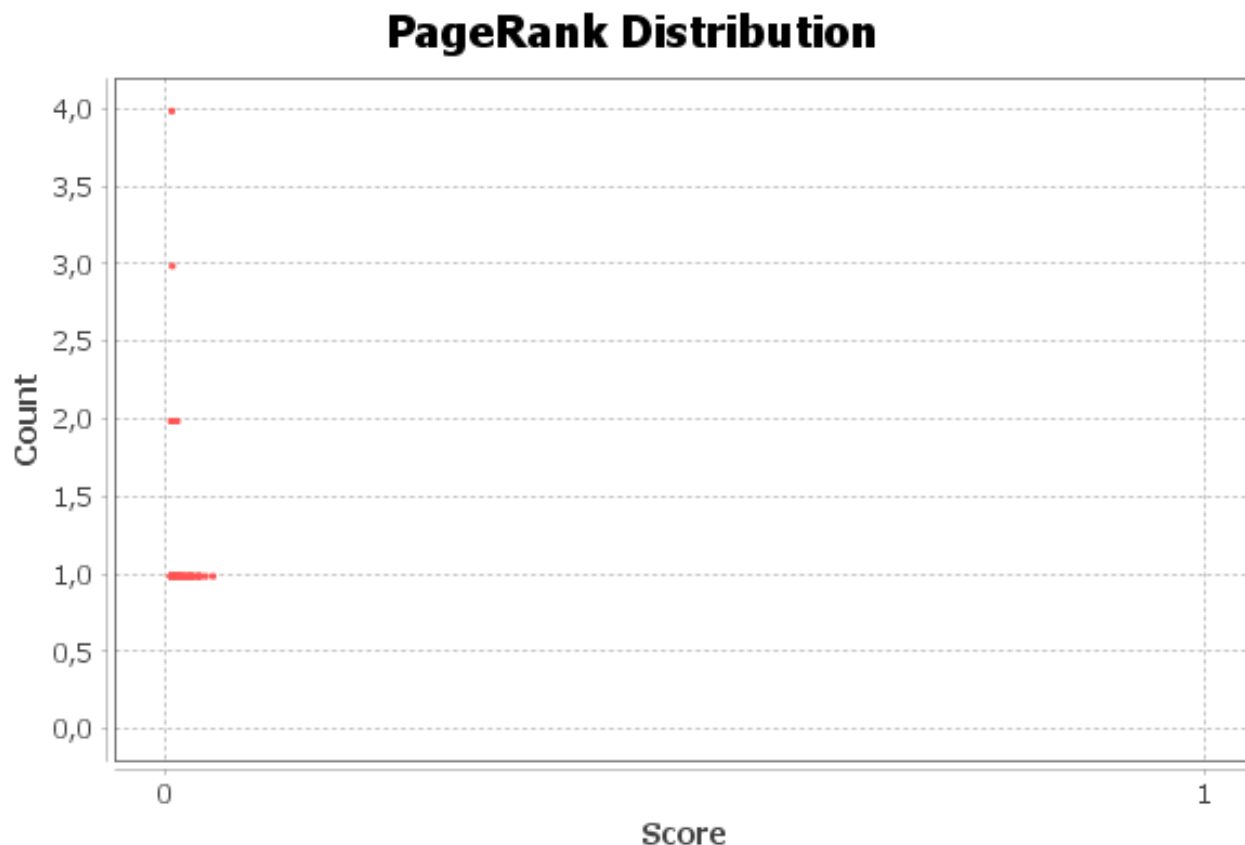
2.7 Pagerank

Parameters:

Epsilon = 0.001

Probability = 0.85

Results:



Σχήμα 2.24: Pagerank distribution.

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. Αξίζει να σημειωθεί ότι σε αντίθεση με την eigenvector centrality, η τιμή pagerank μιας κορυφής δεν εξαρτάται από την τιμή pagerank των γειτόνων της, αλλά από την σημαντικότητά τους. Για τις παραμέτρους, έχουμε διαλέξει $\epsilon = 0.001$, $p = 0.85$, οι οποίες ήταν οι προκαθορισμένες τιμές από το Gephi. Η τιμή $p = 0.85$ αφορά την πιθανότητα ενός χρήστη που κάνει κλικ σε τυχαίους συνδέσμους να συνεχίσει να κάνει κλικ και σε άλλους συνδέσμους. Η τιμή $\epsilon = 0.001$ είναι το κριτήριο τερματισμού του αλγορίθμου, οπότε όσο μικρότερη τιμή έχει, τόσο πιο πολύ θα αργήσει ο υπολογισμός του Pagerank. Δεν έχουν χρησιμοποιηθεί τα βάρη των ακμών. Παρατηρούμε ότι τα σκορ αυτής της μεθόδου είναι ιδιαίτερα χαμηλά, δηλαδή δεν υπάρχουν κορυφές οι οποίες να είναι σημαντικές για τη μέθοδο Pagerank. Συνεπώς, αν υποθέσουμε ότι οι χαρακτήρες του βιβλίου ήταν ιστοσελίδες, η πιθανότητα κάποιος τυχαίος χρήστης στο διαδίκτυο να βρεθεί τυχαία σε μια από τις ιστοσελίδες αυτές είναι πολύ μικρή. Μια άλλη παρατήρηση, η οποία προκύπτει από το παρακάτω γράφημα, είναι ότι οι κορυφές με μεγάλο Pagerank είναι κορυφές που έχουν μεγάλο βαθμό (είναι συνδεδεμένες με κορυφές μέσω πυκνών ακμών).



Σχήμα 2.25: Γράφημα μεθόδου Pagerank.

2.8 Modularity

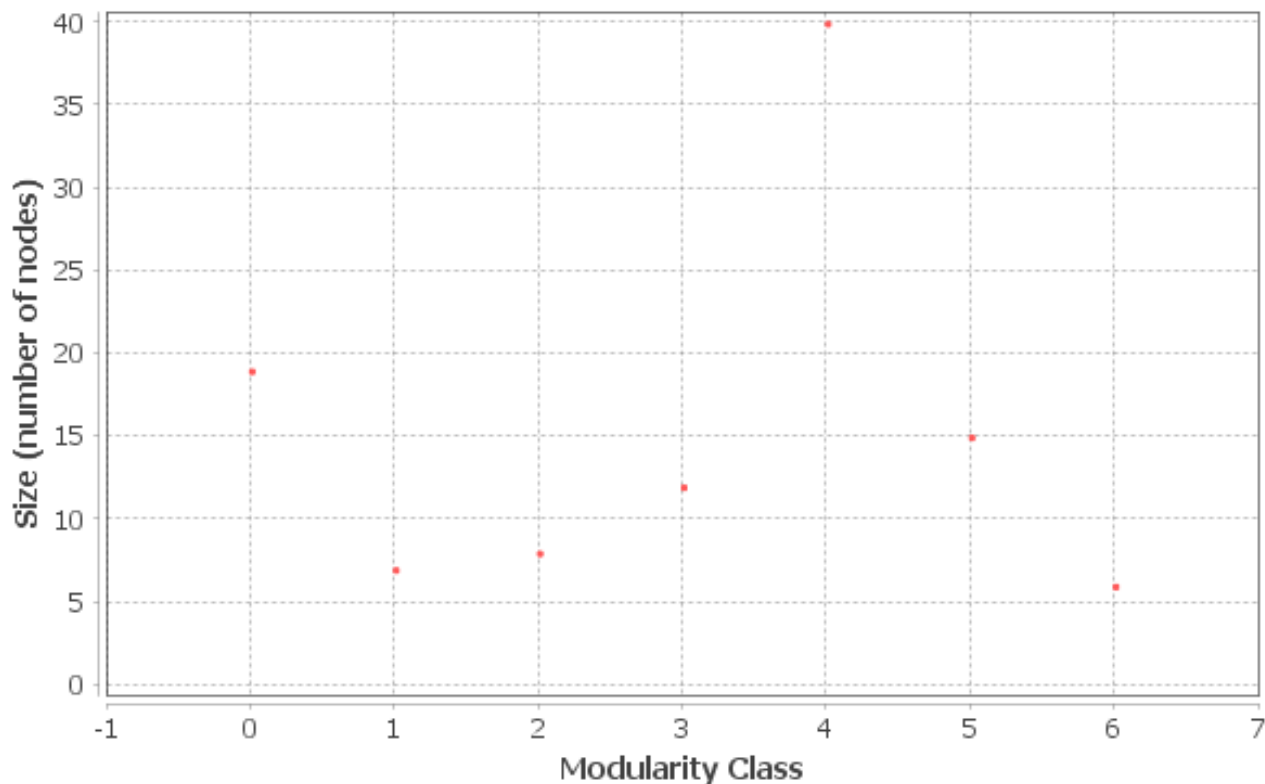
Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

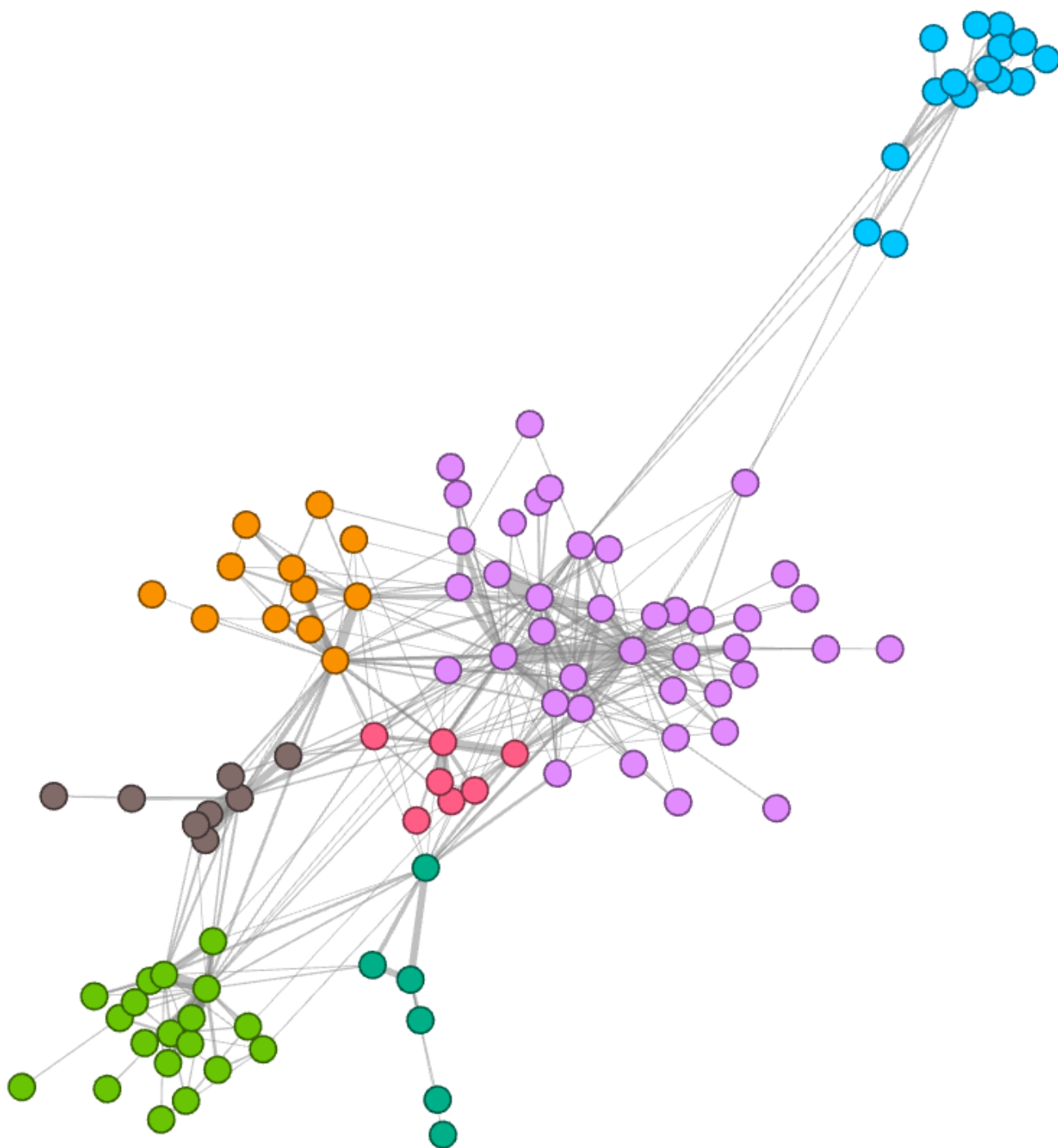
Modularity: 0,600
Modularity with resolution: 0,600
Number of Communities: 7

Size Distribution



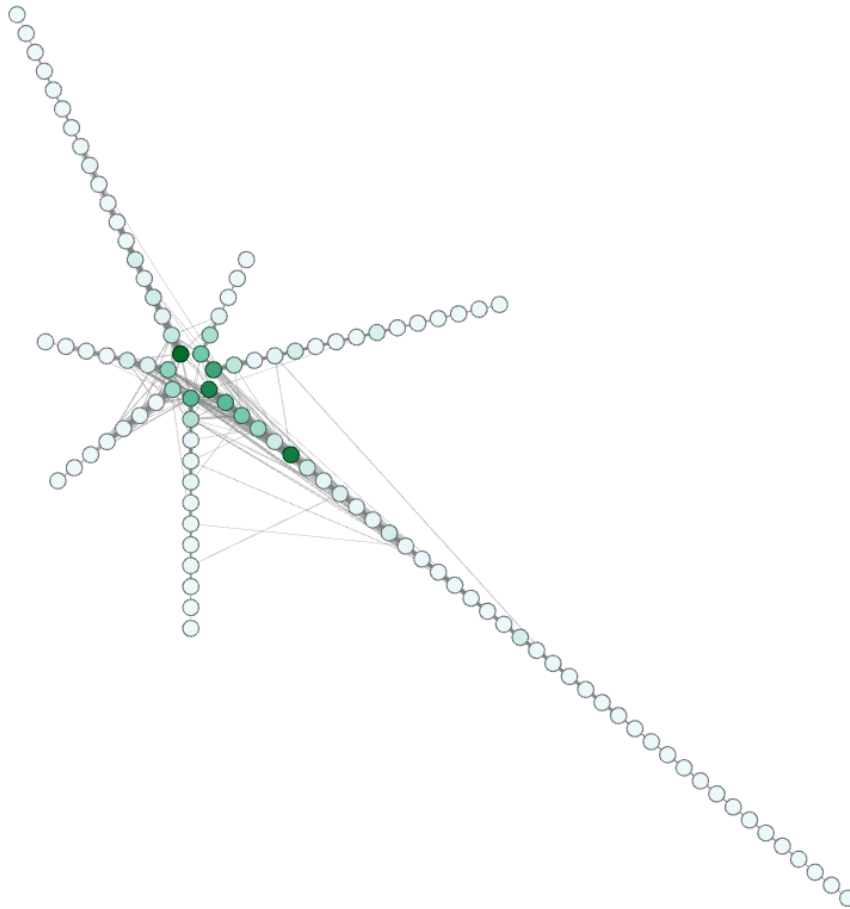
Σχήμα 2.26: Αριθμός ατόμων ανά κοινότητα (modularity).

Παρατηρούμε ότι στο γράφημά μας υπάρχουν 7 κοινότητες, συνεπώς ο αλγόριθμος εύρεσης του modularity του γραφήματος βρήκε το σωστό αποτέλεσμα. Επιπλέον, παρατηρούμε ότι οι κορυφές που ανήκουν σε μια κοινότητα αλληλεπιδρούν μεταξύ τους περισσότερο από ότι με τις υπόλοιπες κοινότητες. Σε αρκετές περιπτώσεις υπάρχουν και κορυφές που αλληλεπιδρούν αποκλειστικά με κορυφές της κοινότητας της οποίας ανήκουν. Στο γράφημα παρατηρούμε ότι η κοινότητα με ετικέτα 4 περιλαμβάνει τους περισσότερους χαρακτήρες (40).



Σχήμα 2.27: Οι 7 κοινότητες στο γράφημά μας, χωρισμένες βάσει χρωμάτων.

2.9 Homophily



Σχήμα 2.28: Αναπαράσταση του homophily.

Στο παραπάνω γράφημα, έχουμε χρωματίσει τις κορυφές βάσει του betweenness centrality. Οι ανοιχτόχρωμες κορυφές, έχουν $\text{betweenness centrality}=0$, και όπως παρατηρούμε, κατά πλειοψηφία, δεν επικοινωνούν με κορυφές από άλλες κοινότητες. Οι πιο σκουρόχρωμες κορυφές έχουν μεγαλύτερη τιμή betweenness centrality, και όσο πιο σκούρο χρώμα έχουν τόσο πιο πολλές οι συνδέσεις τους με κορυφές από άλλες κοινότητες. Στα layouts του Gephi, διαλέξαμε το Radial Axis Layout, και ομαδοποιήσαμε τις κορυφές βάσει του modularity τους. Με το homophily εξετάζουμε κατά πόσο κορυφές με κοινά χαρακτηριστικά (δηλαδή χαρακτήρες που ανήκουν στην ίδια κοινότητα για την περίπτωσή μας) τείνουν να δημιουργούν φιλίες με κορυφές από άλλες κοινότητες. Όπως βλέπουμε και από το γράφημα, είναι μικρός ο αριθμός των ατόμων ανά κοινότητα που κάνει φιλίες με άτομα άλλων κοινοτήτων, κάτι που συνάδει με την μικρή πυκνότητα του γραφήματος.

Κεφάλαιο 3

Συμπεράσματα

Τα συμπεράσματα του συντάκτη της εργασίας είναι προφανώς υποκειμενικά, αφού κάποιος άλλος θα μπορούσε να τα ερμηνεύσει αλλιώς. Ωστόσο, αρκετά από τα συμπεράσματα της παρούσας εργασίας συνάδουν με την πλοκή του έργου, συνεπώς βλέπουμε ότι η ανάλυση κοινωνικών δικτύων μπορεί να εξάγει συμπεράσματα ακόμα και για την πλοκή ενός βιβλίου. Πιο συγκεκριμένα, βρέθηκαν οι κοινότητες του γραφήματος μας, βρέθηκαν και αναλύθηκαν τα μέτρα κεντρικότητας του γραφήματος από τα οποία συμπεράναμε πράγματα όπως ότι υπάρχουν αρκετοί χαρακτήρες που έχουν μικρό ποσοστό αλληλεπίδρασης με τους υπόλοιπους, ενώ υπάρχουν και χαρακτήρες που αλληλεπιδρούν πολύ καθ' όλη τη διάρκεια του βιβλίου. Επιπλέον, είδαμε ότι λόγω του μεγάλου αριθμού τριγώνων στο γράφημα υπάρχει ισχυρή τάση για δημιουργία συστάδων (οι οποίες είναι οι 7 κοινότητες που υπάρχουν στο βιβλίο). Τέλος, εφαρμόστηκε η μέθοδος Pagerank στους χαρακτήρες μας για να δούμε πως συμπεριφέρεται σε ένα γράφημα που δεν αφορά ιστοσελίδες και συνδέσμους, ενώ αναλύθηκε το μέτρο Homophily που μας έδειξε ότι γενικά, οι χαρακτήρες του βιβλίου δεν έχουν μεγάλη τάση να αλληλεπιδρούν με χαρακτήρες από άλλες κοινότητες.

Βιβλιογραφία

- [1] Ι. Μανωλόπουλος, Α. Παπαδόπουλος, Κ. Τσίχλας, Θεωρία και Αλγόριθμοι Γράφων, Εκδόσεις Νέων Τεχνολογιών, Αθήνα, 2014.
- [2] Κεντρικότητες: <https://en.wikipedia.org/wiki/Centrality>
- [3] A. Beveridge, J. Shan, Network of thrones, διαθέσιμο στον σύνδεσμο: [https://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20\(1\).pdf](https://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20(1).pdf)
- [4] Gephi manual, διαθέσιμο στον σύνδεσμο: <https://gephi.org/users/>
- [5] Σημειώσεις μαθήματος Ανάλυση Κοινωνικών Δικτύων, Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας, ΟΠΑ.
- [6] Σύνολο δεδομένων, διαθέσιμο στον σύνδεσμο: <https://github.com/melaniewalsh/sample-social-network-datasets/tree/master/sample-datasets>