



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

**SOCIAL NETWORK ANALYSIS OF POLITICO EUROPE ON
GERPI**

ΜΑΘΗΜΑ: ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

ΔΙΔΑΣΚΟΝ:

ΔΗΜΗΤΡΙΟΣ ΠΟΥΡΝΑΡΑΚΗΣ

ΦΟΙΤΗΤΡΙΑ:

ΖΩΗ ΣΠΥΡΙΔΗ – 8210139

Αθήνα, 2025

Περιεχόμενα

1. Εισαγωγή.....	3
2. Γραφική αναπαράσταση του δικτύου	4
3. Basic Topological Properties	6
4. Component Measures.....	8
5. Degree measures	10
5.1. In-Degree :	10
5.2. Out-Degree :	12
5.3. Total Degree	15
5.4. Average Degree.....	17
5.5. Degree Distribution.	17
6. Centrality measures.....	20
6.1. Degree Centrality	20
6.2. Betweenness Centrality	23
6.3. Closeness Centrality	25
6.4. Eigenvector Centrality	27
7. Clustering effects in the network	30
8. Bridges and local bridges.....	33
9. Gender and homophily	35
10. Graph density	37
11. Community structure (modularity)	38
12. PageRank.....	40
13. Βιβλιογραφία.....	44

1. Εισαγωγή

Η παρούσα εργασία επικεντρώνεται στην ανάλυση ενός δικτύου από την πλατφόρμα **BlueSky**, με τη χρήση του λογισμικού Gephí, ενός από τα πιο δημοφιλή εργαλεία για την οπτικοποίηση και ανάλυση δικτύων. Το δίκτυο που επιλέχθηκε σχετίζεται με τον λογαριασμό **POLITICO Europe**. Ο συγκεκριμένος λογαριασμός ανήκει σε έναν από τους κορυφαίους φορείς ειδήσεων, γνωστό για την παροχή αξιόπιστης ενημέρωσης γύρω από πολιτικά και οικονομικά ζητήματα που επηρεάζουν την Ευρώπη.

Ο **POLITICO Europe** επιλέχθηκε για αυτήν την ανάλυση λόγω της σημασίας του ως πηγή πληροφοριών και της δυνατότητάς του να λειτουργεί ως "κόμβος" στο δίκτυο. Αποτελείται από 168,8 χιλιάδες ακόλουθους και ακολουθεί έναν περιορισμένο αριθμό χρηστών, δηλαδή 70 λογαριασμούς, γεγονός που τον καθιστά ιδανικό για τη διερεύνηση της επιρροής που ασκούν οι λογαριασμοί που ακολουθεί.

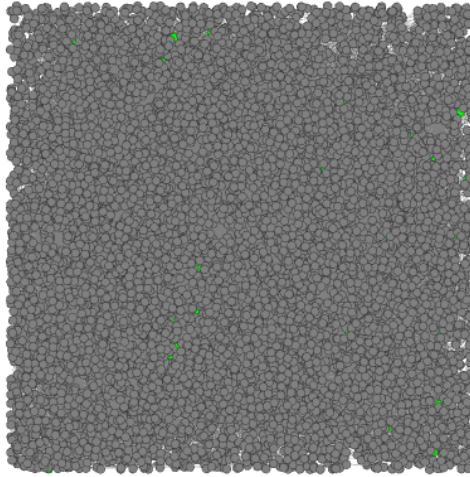
Ο στόχος αυτής της ανάλυσης είναι να κατανοηθεί η δομή του δικτύου που δημιουργείται από τους λογαριασμούς που ακολουθεί το **POLITICO Europe** και από τους λογαριασμούς που αυτοί με τη σειρά τους ακολουθούν. Συγκεκριμένα, την δομή του δικτύου, δηλαδή το δίκτυο που δημιουργείται από τους χρήστες που ακολουθεί αυτός ο λογαριασμός και στην συνέχεια αυτούς που οι χρήστες ακολουθούν. Δηλαδή κινούμαστε από following σε following, χωρίς να μας απασχολούν οι followers. Αυτή η ανάλυση παρουσιάζει ενδιαφέρον από την άποψη της ιεραρχικής δομής και της επιρροής που δέχεται ο κεντρικός λογαριασμός/λογαριασμός στόχος Politico Europe. Το συγκεκριμένο δίκτυο δίνει τη δυνατότητα να μελετηθούν ενδιαφέροντα μοτίβα, όπως κοινότητες χρηστών, η επιρροή συγκεκριμένων κόμβων, και η εξάπλωση πληροφοριών.

Συνεπώς, η μελέτη αυτή φιλοδοξεί να προσφέρει μια ολοκληρωμένη ανάλυση των ιδιοτήτων του δικτύου, εντοπίζοντας χρήστες με αυξημένη επιρροή και αξιολογώντας τη συνολική του δομή. Παράλληλα, θα παραχθούν γραφικές αναπαραστάσεις και στατιστικά χαρακτηριστικά, χρησιμοποιώντας το εργαλείο Gephí, για να κατανοηθούν καλύτερα τα δεδομένα και να εξαχθούν ενδιαφέροντα συμπεράσματα.

Η διαδικασία θα ξεκινήσει με της εισαγωγή των δεδομένων/δικτύου στο Gephí. Συγκεκριμένα το account από το BlueSky είναι το **@politico.eu** πάνω σε αυτό θα βασιστούμε στην δυνατότητα "fetch follows" και "fetch n+1" στο εργαλείο συλλογής δεδομένων. Το "fetch follows" προσθέτει στο δίκτυο τους λογαριασμούς που ακολουθεί το **POLITICO Europe** ως κόμβους και δημιουργεί ακμές (edges) που αντιπροσωπεύουν τη σχέση "following". Το "fetch n+1" διευρύνει το δίκτυο με έναν επιπλέον βαθμό, προσθέτοντας τους λογαριασμούς που ακολουθούνται από τους κόμβους του αρχικού επιπέδου.

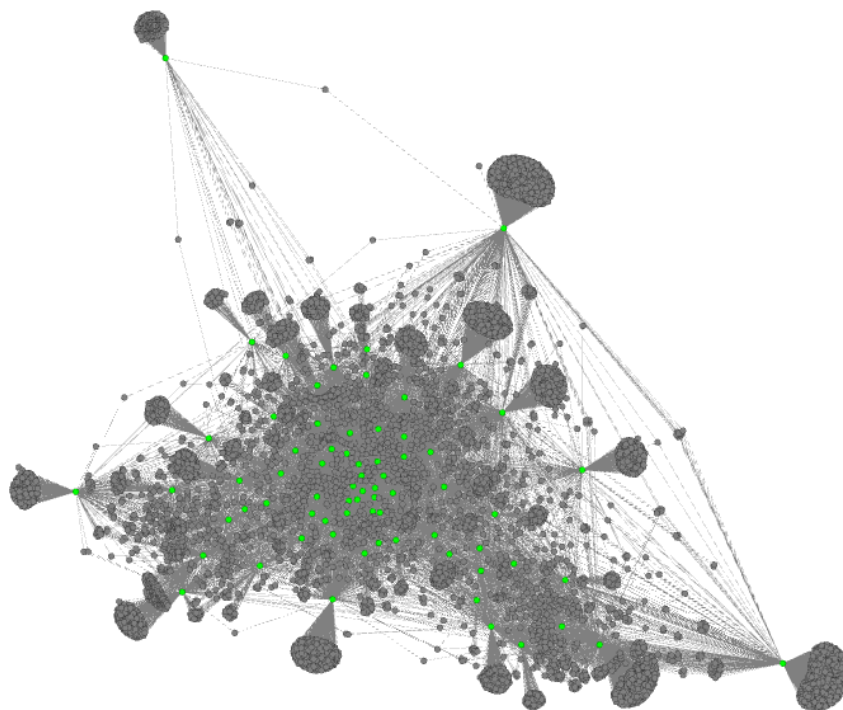
2. Γραφική αναπαράσταση του δικτύου

Όπως είναι λογικό στην αρχή το δίκτυο μας είναι χωρίς οργάνωση με αποτέλεσμα να μην βγάζουμε κάποιο λογικό συμπέρασμα.



Για την καλύτερη οπτικοποίησή του λοιπόν χρησιμοποιήθηκαν δύο αλγόριθμοι διάταξης. Αρχικά, εφαρμόστηκε ο αλγόριθμος **ForceAtlas 2**, ο οποίος είναι ιδιαίτερα κατάλληλος για την ομαδοποίηση κόμβων και την ανάδειξη της συνολικής δομής ενός δικτύου. Μέσα από την εφαρμογή του, του αναγνωρίστηκαν οι πρώτες ομάδες κόμβων.

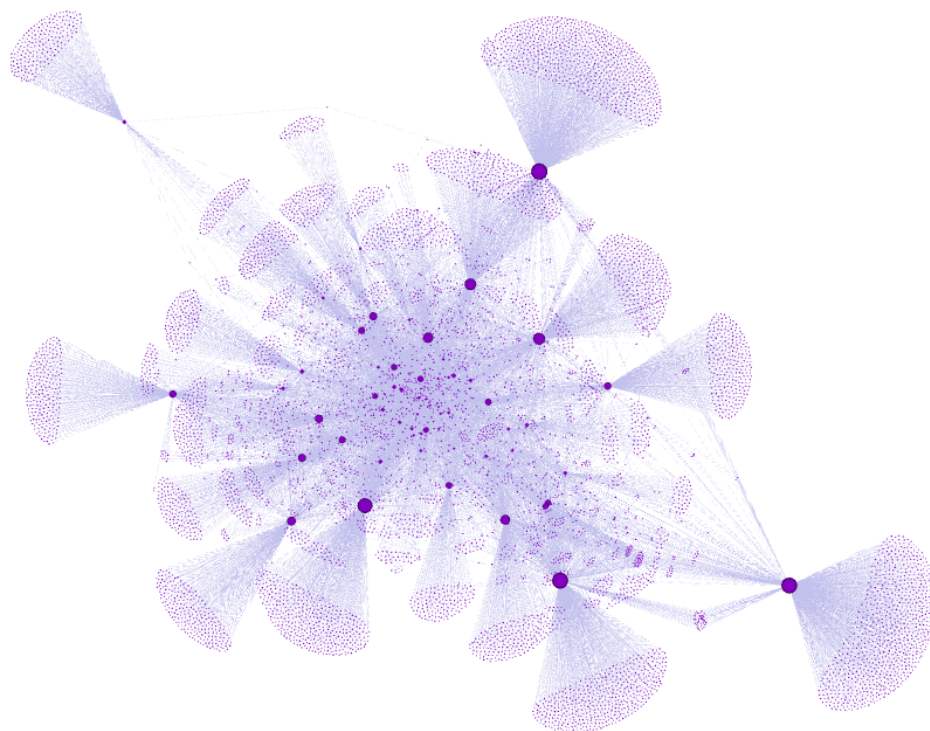
Εδώ είναι το αποτέλεσμα του **Force Atlas 2** :



Επόμενο βήμα ήταν η ακόμα καλύτερη οπτικοποίηση του δικτύου τόσο σε επίπεδο χρωμάτων όσο και σε επίπεδο μεγέθους σημαντικότητας. Συγκεκριμένα το μέγεθος των κόμβων ρυθμίστηκε βάσει του **Degree**, επιτρέποντας την ανάδειξη των πιο “κεντρικών” και συνδεδεμένων κόμβων. Αυτή η προσέγγιση προσφέρει μια πιο σαφή κατανόηση του ρόλου κάθε κόμβου στο δίκτυο.



Για τη βελτίωση της αισθητικής και της κατανόησης του διαγράμματος, χρησιμοποιήθηκε ο αλγόριθμος **Yifan Hu**, ο οποίος προσφέρει πιο καθαρό διαχωρισμό μεταξύ των κόμβων. Με τη χρήση αυτού του αλγορίθμου, οι σχετιζόμενοι κόμβοι διαχωρίστηκαν καλύτερα, ενώ οι ομάδες έγιναν πιο ευδιάκριτες, καθιστώντας την ανάλυση του δικτύου πιο προσιτή.



3. Basic Topological Properties

Το δίκτυο που εξετάζεται είναι κατευθυνόμενο (**directed graph**) και περιλαμβάνει **8822** κόμβους(**Nodes**) και **17120** ακμές(**Edges**). Κάθε ακμή αναπαριστά μία σχέση "**ακολουθεί(follows)**" μεταξύ δύο χρηστών.

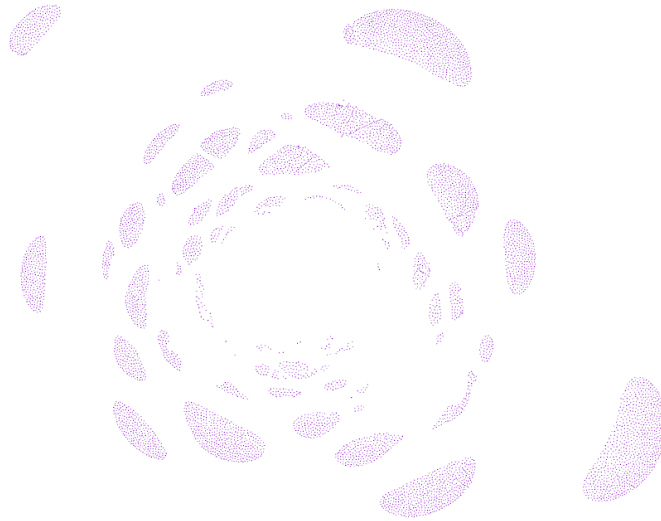
Οι τιμές που υπολογίστηκαν για το μέσο μήκος διαδρομής(Average Path Length), τη διάμετρο (Diameter) και την ακτίνα (Radius) του δικτύου είναι οι εξής:

- Average Path Length: 2.37957
- Diameter: 4
- Radius: 0

Για την κατανόηση ωστόσο των τιμών αυτών πρέπει να γνωρίζουμε τι υποδηλώνει κάθε μετρική. Αρχικά το μέσο μήκος διαδρομής(Average Path Length) αναφέρεται στον μέσο όρο των συντομότερων διαδρομών (shortest paths) μεταξύ όλων των ζευγών κόμβων στο δίκτυο. Στη συγκεκριμένη περίπτωση, η τιμή περίπου 2.379 υποδεικνύει ότι, κατά μέσο όρο, ένας κόμβος μπορεί να συνδεθεί με οποιονδήποτε άλλον στο δίκτυο μέσω περίπου 2.38 βημάτων. Αυτή η σχετικά χαμηλή τιμή επιβεβαιώνει την ύπαρξη του φαινομένου του "μικρού κόσμου" (small-world effect), δηλαδή ότι οι περισσότεροι κόμβοι συνδέονται μεταξύ τους με λίγα μόνο βήματα, χαρακτηριστικό που παρατηρείται συχνά σε κοινωνικά δίκτυα. Η διάμετρος του δικτύου(Diameter) είναι η μεγαλύτερη συντομότερη διαδρομή (longest shortest path) μεταξύ δύο κόμβων. Η τιμή που έχουμε υποδεικνύει ότι ακόμη και οι πιο απομακρυσμένοι κόμβοι στο δίκτυο μπορούν να συνδεθούν μέσω μόλις 4 βημάτων. Τέλος, η ακτίνα(radius) του δικτύου είναι η μικρότερη απόσταση από τον πιο κεντρικό κόμβο στον πιο απομακρυσμένο κόμβο που μπορεί να προσεγγίσει. Η τιμή 0

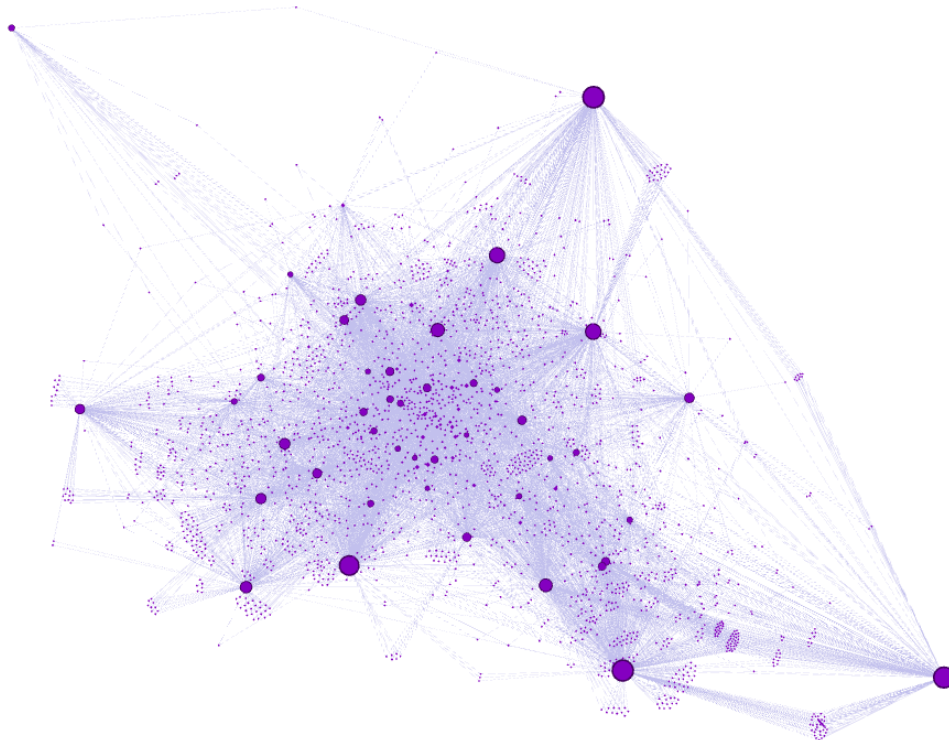
υποδηλώνει την παρουσία Peripheral or Lead Nodes που δεν συνδέονται άμεσα με άλλους στο δίκτυο.

Στην παρακάτω εικόνα απεικονίζονται οι Peripheral or Lead Nodes:



Αυτοί προέκυψαν φιλτράροντας με βάση τον βαθμό(Degree) ίσο με 1. Αυτοί οι κόμβοι είναι 6506 και αποτελούν το 73.75% του δικτύου. Το χαρακτηριστικό τους είναι πως συνδέονται μόνο μέσω κεντρικών κόμβων (Hubs) και δεν έχουν συνδέσεις μεταξύ τους.

Φιλτράροντας λοιπόν με $\text{degree} \geq 2$ το δίκτυο που προέκυψε είναι:



Αν ξανά υπολογίσουμε τις μετρικές για το νέο φιλτραρισμένο δίκτυο παρατηρούμε ότι η διάμετρος και η ακτίνα παραμένουν ίδιες το μόνο που αλλάζει είναι το average path length σε 2.1917. Αυτό που παρατηρούμε λοιπόν είναι ότι η ακτίνα παραμένει μηδενική ακόμα και αν αφαιρεθούν οι Peripheral or Lead Nodes, άρα το πρόβλημα δεν προκύπτει αποκλειστικά από την ύπαρξη αυτών των κόμβων. Αντίθετα, η αιτία της μηδενικής ακτίνας φαίνεται να σχετίζεται με τη γενικότερη δομή του δικτύου. Πιο συγκεκριμένα το δίκτυο φαίνεται να έχει χαρακτηριστικά **δενδροειδούς δομής** (tree-like structure), όπου πολλοί κόμβοι είναι συνδεδεμένοι μέσω κεντρικών κόμβων (**hubs**), αλλά δεν υπάρχουν πολλές άμεσες συνδέσεις μεταξύ των υπόλοιπων κόμβων.

4. Component Measures

Τα component measures αναφέρονται σε μεθόδους που χρησιμοποιούνται για τον εντοπισμό και την ανάλυση των connected components ενός δικτύου. Η ανάλυση αυτή είναι χρήσιμη για την κατανόηση της συνολικής δομής και συνδεσιμότητας του δικτύου. Συγκεκριμένα για τον εντοπισμό κατακερματισμού ή απομονωμένων ομάδων. Με τον όρο connected component εννοούμε ένα υποσύνολο κόμβων του γραφήματος, όπου οποιοιδήποτε δύο κόμβοι του υποσυνόλου συνδέονται με μία διαδρομή, και κανένα από αυτούς δεν συνδέεται με κόμβους εκτός του υποσυνόλου.

Σε αυτή την ενότητα, θα υπολογίσουμε τον αριθμό των connected components, την ύπαρξη giant component και την component size distribution.

Για τον υπολογισμό του αριθμού των connected components, χρησιμοποιούμε την επιλογή **Connected Components** στο Gephi. Υπάρχουν δύο είδη που μπορούν να υπολογιστούν :

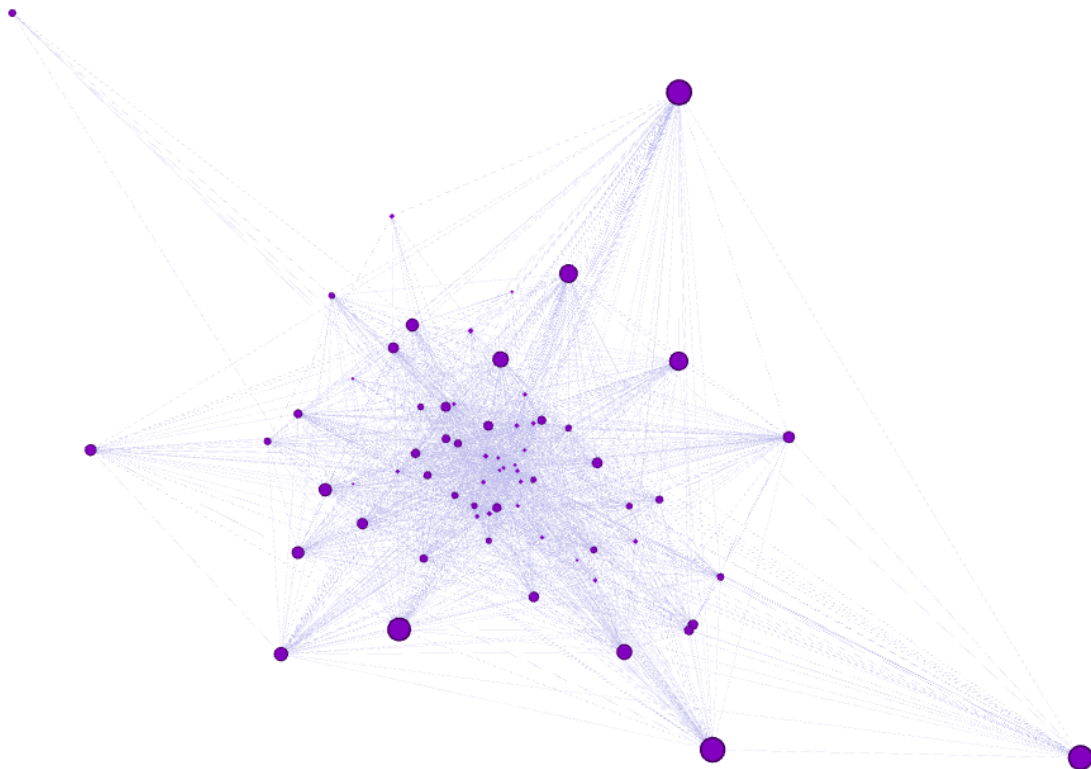
- a) **Weakly Connected Components:** Αυτή η μετρική υπολογίζει connected components αγνοώντας την κατεύθυνση των ακμών. Στο δίκτυό μας, η τιμή είναι 1, που σημαίνει ότι αν παραβλέψουμε την κατεύθυνση των ακμών, όλοι οι κόμβοι στο δίκτυο είναι προσβάσιμοι ο ένας από τον άλλον.
- b) **Strongly Connected Components :** Αυτή η μετρική λαμβάνει υπόψη την κατεύθυνση των ακμών και υπολογίζει connected components στα οποία όλοι οι κόμβοι είναι συνδεδεμένοι μεταξύ τους και στις δύο κατευθύνσεις. Στο δίκτυό μας, υπάρχουν 8754 strongly connected components. Αυτό σημαίνει ότι το δίκτυο είναι εξαιρετικά κατακερματισμένο όταν λαμβάνεται υπόψη η κατεύθυνση των ακμών. Η ύπαρξη τόσο μεγάλου αριθμού strongly connected components δείχνει ότι οι περισσότεροι κόμβοι ή ομάδες κόμβων συνδέονται μόνο σε μία κατεύθυνση, με αποτέλεσμα τη δημιουργία πολλών μικρών ισχυρών συνδεδεμένων συστατικών. Το γεγονός αυτό είναι τυπικό για κατευθυνόμενα κοινωνικά δίκτυα, όπου η "κατεύθυνση" περιορίζει την αλληλεπίδραση. Παράλληλα, επιβεβαιώνεται ότι στο δίκτυό μας, οι χρήστες που ακολουθούν άλλους χωρίς να ακολουθούνται πίσω δημιουργούν μικρά strongly connected components.

Για να εξετάσουμε την ύπαρξη Giant Component πρέπει πρώτα να καταλάβουμε τι είναι. Το Giant Component σε ένα γράφημα αναφέρεται στο μεγαλύτερο component

που καταλαμβάνει ένα σημαντικό μέρος των κόμβων, όπου κάθε κόμβος μπορεί να προσεγγιστεί από κάθε άλλο κόμβο σε αυτό το στοιχείο. Μπορούμε να υπολογίσουμε δύο Giant Component, ένα για τα Weakly Connected Components και ένα για τα Strongly Connected Components.

Λαμβάνοντας υπόψη το Weakly Connected Component και εφόσον υπάρχει μόνο 1, ολόκληρο το γράφημα είναι giant weakly connected component. Αυτό μπορούμε να το επιβεβαιώσουμε και στο Data Laboratory, μιας και στη στήλη Component Id όλες οι τιμές είναι ίσες με 0. Αυτό δείχνει ότι ολόκληρο το δίκτυο θεωρείται ένα giant connected component.

Όσον αφορά το Giant Component of Strongly Connected Component, αν φιλτράρουμε το Strongly Connected ID θα παρατηρήσουμε όλες τις πιθανές τιμές και τα ποσοστά τους. Το μεγαλύτερο είναι το Giant Component περιλαμβάνει 69 nodes (εκ των οποίων 0.78% ορατοί) και 2344 edges (εκ των οποίων 13.69% ορατές). Αυτό το Giant Component απεικονίζεται παρακάτω:



Τέλος, η component size distribution αναφέρεται στη συχνότητα και το μέγεθος των connected components σε ένα δίκτυο, δηλαδή δείχνει πόσα connected components υπάρχουν στο γράφημα και πόσο μεγάλο είναι το κάθε στοιχείο. Θα το χωρίσουμε σε δύο ενότητες, ένα για τα Weakly Connected Components και ένα για τα Strongly Connected Components.

Σχετικά με τα Weakly Connected Components ο υπολογισμός είναι απλός. Αφού γνωρίζουμε πως έχουμε μόνο 1 τέτοιο component και ο αριθμός των κόμβων(nodes) στο δίκτυο είναι 8822, το size of Weakly Connected Components είναι το ίδιο δηλαδή 8822.

Σχετικά τώρα με το Strongly Connected Components όπως είδαμε και παραπάνω υπάρχουν 8754 τέτοια components. Η μεγαλύτερη Strongly Connected Component distribution/κατανομή αποτελεί το 0,78% των κόμβων και όλα τα υπόλοιπα μικρότερα connected components έχουν κατανομή 0,01% και αποτελούνται από 1 κόμβο/node το καθένα.

5. Degree measures

Στην ενότητα αυτή, υπολογίζουμε μετρικές βαθμού του δικτύου για να αποκτήσουμε πληροφορίες σχετικά με τη συνδεσιμότητα και τη δομή του.

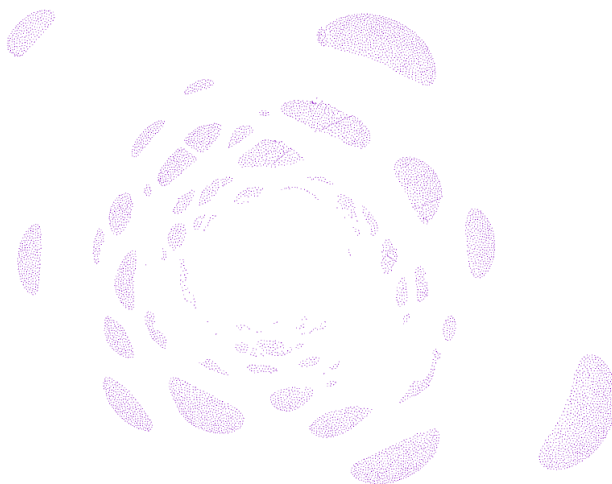
Αναλυτικότερα, θα αναφερθούμε στα:

5.1. In-Degree :

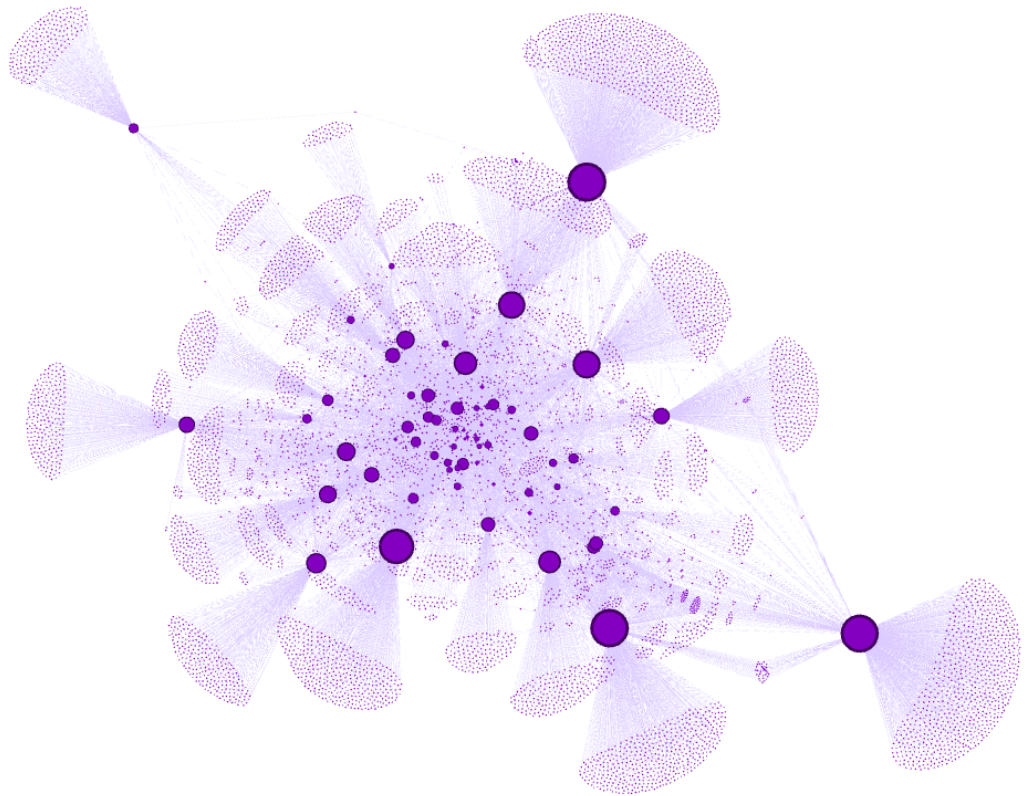
Το In-Degree ενός κόμβου αντιπροσωπεύει τον αριθμό των εισερχόμενων συνδέσεων του, δηλαδή τον αριθμό των "followers" στο συγκεκριμένο δίκτυο.

Πάμε όμως να δούμε και οπτικά τα in-degree. Αρχικά στο Data Laboratory αφού έχουμε εκτελέσει πρώτα τη λειτουργία Degree, μπορούμε να παρατηρήσουμε ότι η ελάχιστη(minimum) τιμή για το in-degree είναι 1 και η μέγιστη(maximum) 64.

Σε αυτό το σχήμα παρατηρούμε τους κόμβους(nodes) με In-degree 1, καθώς αποτελούν και το μεγαλύτερο μέρος του δικτύου.



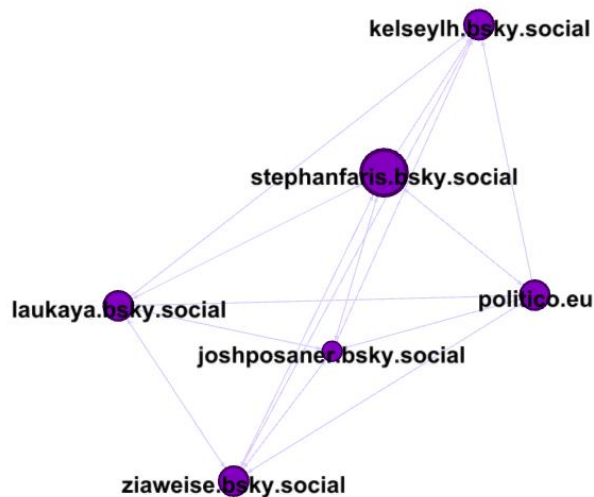
Για να οπτικοποιήσουμε ολόκληρο το δίκτυο μας με το μέγεθος των nodes να αλλάζει βάση του in-degree, είναι μια καλή πρακτική να χρησιμοποιήσουμε για το μέγεθος του κόμβου ως min size το 5 ώστε οι κόμβοι με χαμηλό In-Degree να είναι μικροί αλλά ορατοί, αφού όπως είδαμε αποτελούν σημαντικό κομμάτι του δικτύου, και για max size 50 ώστε οι hubs να είναι αισθητά μεγαλύτεροι.



Μια ενδιαφέρουσα παρατήρηση από τα παραπάνω είναι ότι ο κόμβος με το μέγιστο In-Degree είναι ο **politico.eu**, ο οποίος είναι και ο κόμβος από τον οποίο ξεκινήσαμε τη δημιουργία του δικτύου.

Αυτό είναι ιδιαίτερα ενδιαφέρον, καθώς επιβεβαιώνει την υπόθεση ότι ο κόμβος αυτός λειτουργεί ως σημαντικός-hub στο δίκτυο. Παρόλο που το δίκτυο δεν περιλάμβανε εξ αρχής τους followers του politico.eu, παρατηρούμε ότι μέσω των αλληλοσυνδέσεων, ο κόμβος αυτός επανεμφανίζεται δυναμικά στο δίκτυο, προσελκύοντας μεγάλο αριθμό εισερχόμενων συνδέσεων.

Ως ένας από τους κύριους φορείς ειδησεογραφίας στην Ευρώπη, το γεγονός ότι έχει τόσες εισερχόμενες συνδέσεις σημαίνει ότι οι ειδήσεις και το περιεχόμενό του αναφέρονται συχνά και πιθανώς επηρεάζουν τον διάλογο στο δίκτυο.



Εκτός από τον κόμβο politico.eu, που κατέχει την κορυφαία θέση, ας αναλύσουμε και τους υπόλοιπους κόμβους με υψηλές τιμές In-Degree για να εξετάσουμε την επιρροή τους στο δίκτυο.

Οι 5 κορυφαίοι κόμβοι με βάση το In-Degree είναι:

1. **politico.eu**: 64
2. **ziawaise.bsky.social**: 55
3. **laukaya.bsky.social**: 53
4. **kelseylh.bsky.social**: 49
5. **stephanfaris.bsky.social** & **joshposaner.bsky.social**: 47

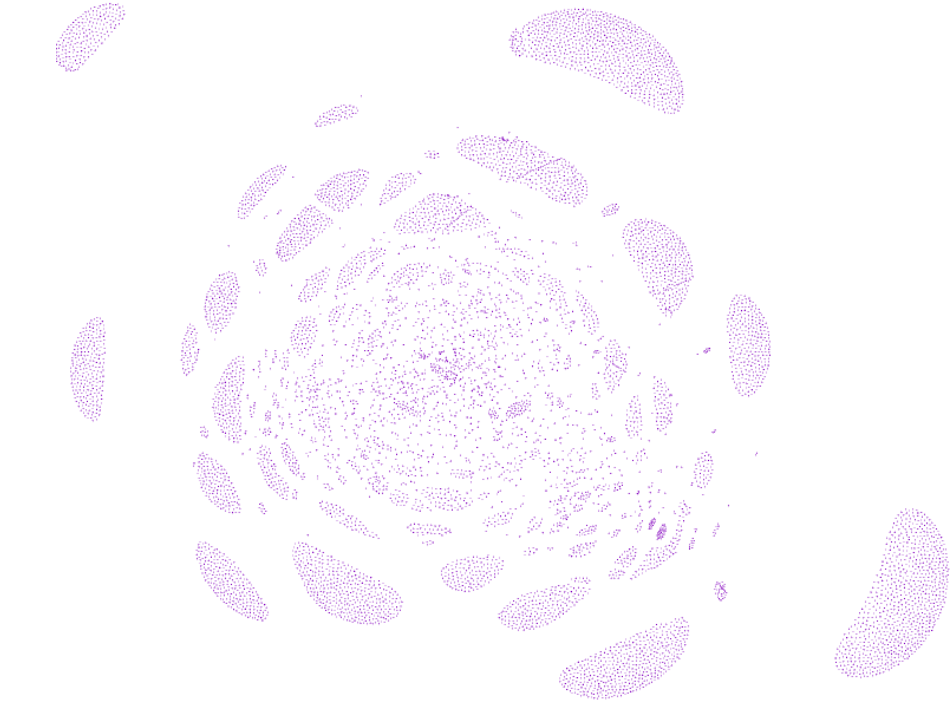
Οι κόμβοι αυτοί είναι από τους πιο συνδεδεμένους και πιθανότατα παίζουν κεντρικό ρόλο στο δίκτυο. Οι υψηλές τιμές In-Degree δείχνουν ότι αυτοί οι λογαριασμοί δέχονται μεγάλο αριθμό εισερχόμενων συνδέσεων (followers), γεγονός που υποδηλώνει την επιρροή τους ή τη δημοφιλία τους. Αναλυτικότερα, πρόκειται για δημοσιογράφους και συντάκτες του Politico Europe, γεγονός που ενισχύει την άποψη ότι το κοινό ενδιαφέρεται έντονα για το περιεχόμενό του. Ο λογαριασμός ziawaise.bsky.social πρόκειται για μια Senior EU climate reporter at Politico στην Κωνσταντινούπολη με 6,2 χιλ followers Και 923 follows. Ο λογαριασμός laukaya.bsky.social είναι εξίσου μια γυναίκα που εργάζεται ως defense correspondent at POLITICO Europe στο Παρίσι. Έχει και αυτή μεγάλο αριθμό followers και follows, 3,6χιλ. και 259 αντίστοιχα. Η kelseylh.bsky.social είναι chief subeditor Ex-POLITICO με 2,1χιλ. followers και 1,5χιλ follows. Τέλος, τόσο ο stephanfaris.bsky.social όσο και ο joshposaner.bsky.social εργάζονται και αυτοί για το Politico Europe ως Associate editor και ως Senior Defense and Space Reporter αντίστοιχα.

5.2. Out-Degree :

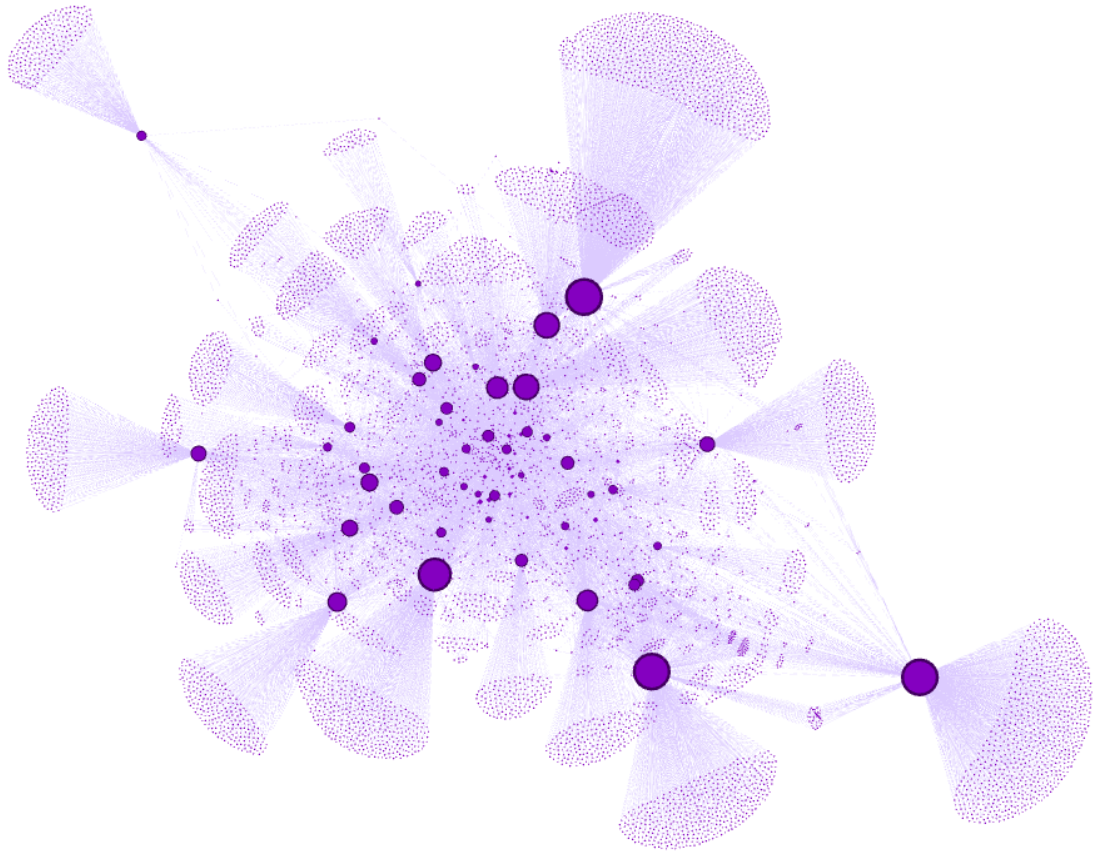
Δηλαδή τον αριθμό των εξερχόμενων συνδέσεων. Στο συγκεκριμένο δίκτυο οι εξερχόμενες συνδέσεις είναι ο αριθμός των “follows”.

Πάμε όμως να δούμε και οπτικά τα out-degree. Αρχικά στο Data Laboratory μπορούμε να παρατηρήσουμε ότι η ελάχιστη(minimum) τιμή για το out-degree είναι 0 και η μέγιστη(maximum) 997.

Σε αυτό το σχήμα παρατηρούμε τους κόμβους(nodes) με out-degree 0 καθώς αποτελούν και το μεγαλύτερο μέρος του δικτύου, συγκεκριμένα το 99.21%.

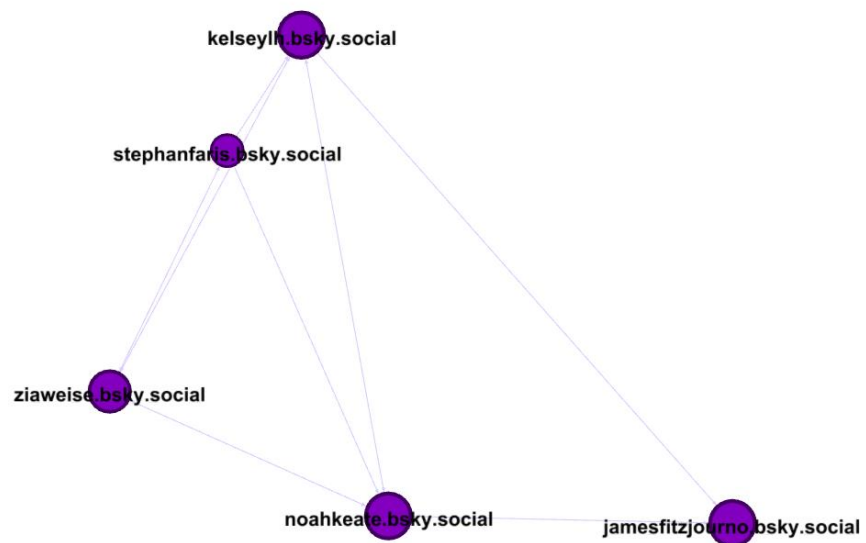


Για να οπτικοποιήσουμε ολόκληρο το δίκτυο μας με το μέγεθος των nodes να αλλάζει βάση του out-degree, είναι μια καλή πρακτική να χρησιμοποιήσουμε για το μέγεθος του κόμβου ως min size το 5 ώστε οι κόμβοι με χαμηλό out-Degree να είναι μικροί αλλά ορατοί, αφού όπως είδαμε αποτελούν σημαντικό κομμάτι του δικτύου, και για max size 100 ώστε οι hubs να είναι αισθητά μεγαλύτεροι.



Οι 5 κορυφαίοι κόμβοι με βάση το Out-Degree είναι:

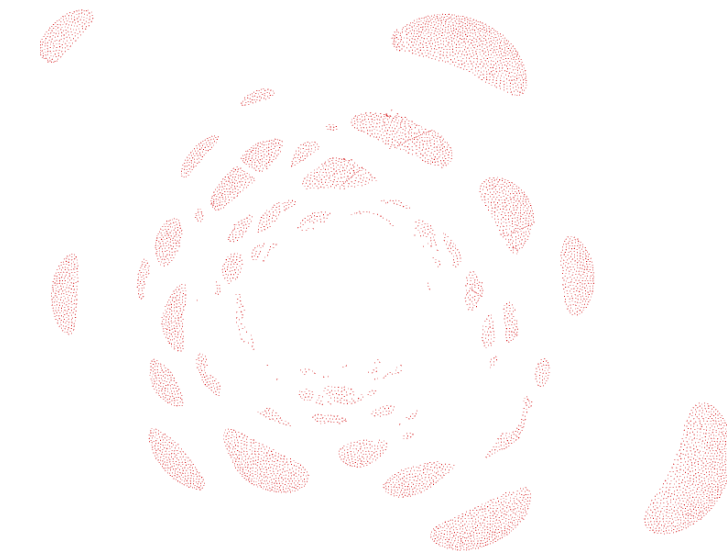
1. **kelseylh.bsky.social**: 997
2. **noahkeate.bsky.social**: 991
3. **jamesfitzjournno.bsky.social**: 988
4. **ziawaise.bsky.social**: 892
5. **stephanfarris.bsky.social**: 691



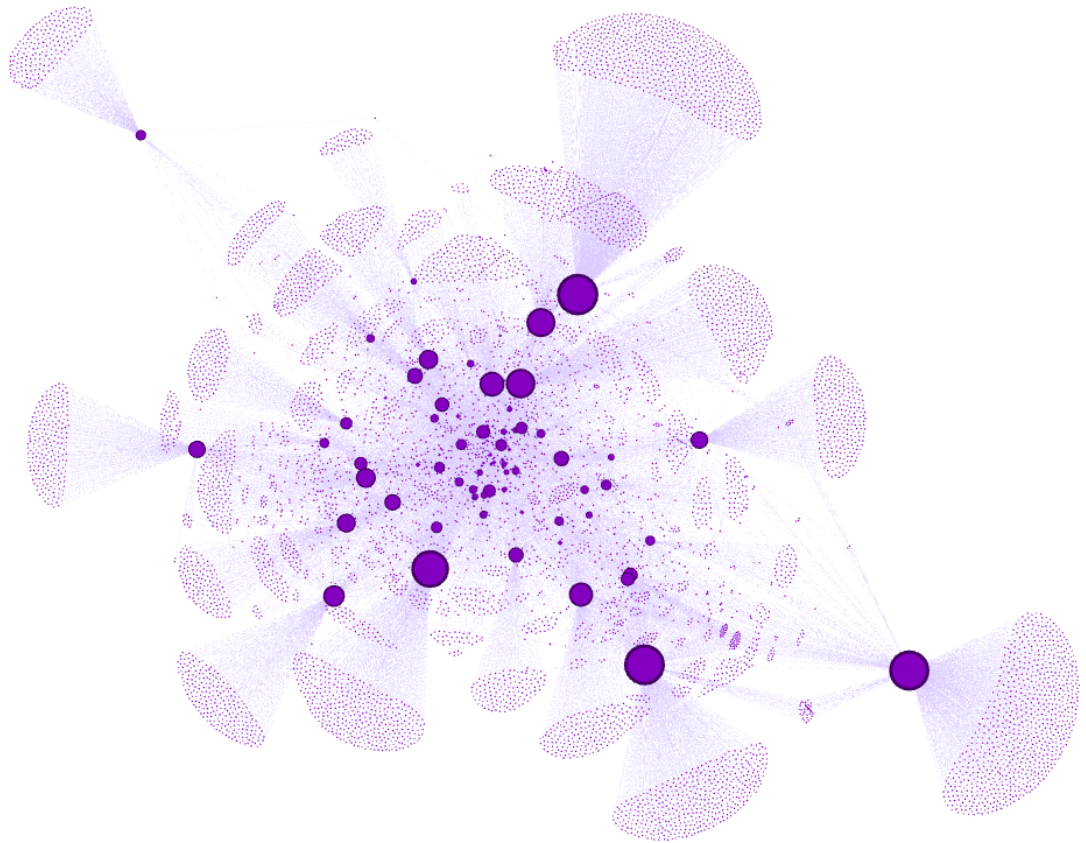
Οι κόμβοι αυτοί εμφανίζουν υψηλή εξερχόμενη συνδεσιμότητα, γεγονός που υποδεικνύει ότι είναι ιδιαίτερα ενεργοί στο να ακολουθούν άλλους κόμβους. Παρατηρούμε ότι οι λογαριασμοί `kelseyjh.bsky.social`, `ziawaise.bsky.social` και `stephanfaris.bsky.social` είναι σημαντικοί και στις 2 κατευθύνσεις, κάτι που δείχνει ότι συμμετέχουν ενεργά στη διάδραση στο δίκτυο. Αν αναλογιστούμε κιόλας πως οι δύο από αυτούς είναι δημοσιογράφοι του Politico και είναι λογικό να έχουν ισχυρή παρουσία στο πλαίσιο της ενημέρωσης. Όσον αφορά τους άλλους δυο λογαριασμούς πρόκειται εξίσου για άλλους δύο δημοσιογράφους στο Politico Europe.

5.3. Total Degree

Αρχικά όπως μπορούμε να παρατηρήσουμε και στο Data Laboratory η ελάχιστη(minimum) τιμή για το συνολικό Degree είναι 1 και η μέγιστη(maximum) 1046. Οι κόμβοι με Degree 1 αποτελούν το μεγαλύτερο μέρος του δικτύου και σε αυτή την μετρική.

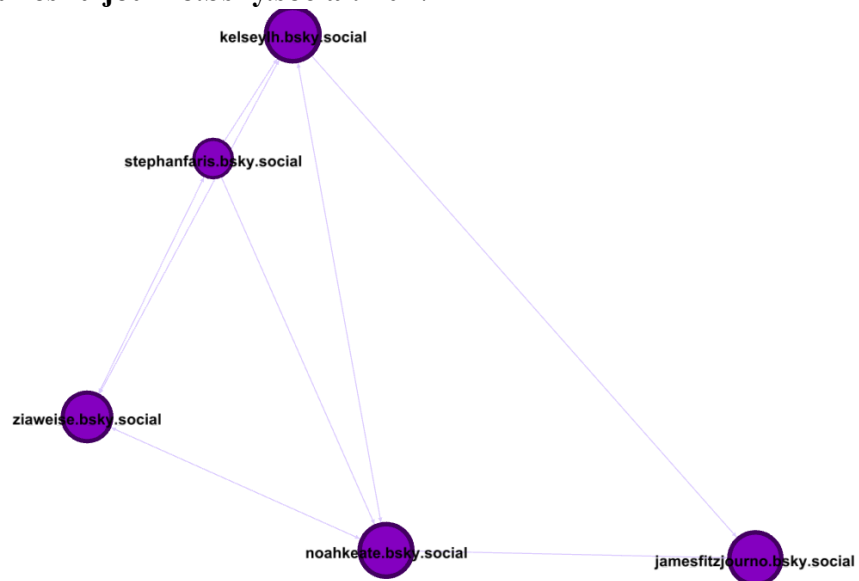


Για να οπτικοποιήσουμε ολόκληρο το δίκτυο μας με το μέγεθος των nodes να αλλάζει βάση του Degree, είναι μια καλή πρακτική να χρησιμοποιήσουμε για το μέγεθος του κόμβου ως min size το 5 ώστε οι κόμβοι με χαμηλό Degree να είναι μικροί αλλά ορατοί, αφού όπως είδαμε αποτελούν σημαντικό κομμάτι του δικτύου, και για max size 110 ώστε οι hubs να είναι αισθητά μεγαλύτεροι.



Αυτό που επίσης παρατηρούμε για το total degree είναι πως 3 κορυφαίοι κόμβοι/nodes του είναι :

1. **kelseylh.bsky.social**: 1046
2. **noahkeate.bsky.social**: 1029
3. **jamesfitzjournno.bsky.social**: 1017



Αυτοί οι τρεις συμπίπτουν με τους κορυφαίους του Out-Degree, πράγμα που σημαίνει ότι δεν είναι απλοί influencers που λαμβάνουν προσοχή, αλλά και ενεργοί χρήστες που αλληλεπιδρούν με πολλούς άλλους. Όλοι εργάζονται για το POLITICO Europe, γεγονός που ενισχύει τη θέση του συγκεκριμένου μέσου ως κυρίαρχη δύναμη στο δίκτυο και μέσου διάδοσης πληροφοριών.

Συμπερασματικά, το δίκτυο είναι ισχυρά επηρεασμένο από το POLITICO Europe, καθώς οι κορυφαίοι κόμβοι (In-Degree, Out-Degree και Total Degree) είναι σχεδόν όλοι δημοσιογράφοι ή μέλη του POLITICO. Άρα, οι ειδήσεις και οι αναρτήσεις του POLITICO έχουν υψηλή διάχυση στο δίκτυο και πιθανότατα το δίκτυο αποτελείται από άτομα που παρακολουθούν πολιτικές, οικονομικές και διεθνείς εξελίξεις.

5.4. Average Degree

Αφού εξετάσαμε τις maximum degree και minimum degree values, είναι σημαντικό να υπολογίσουμε και να αναλύσουμε το Average Degree του δικτύου.

Το average degree σε ένα δίκτυο αντιπροσωπεύει τον μέσο αριθμό συνδέσεων ανά κόμβο. Στην περίπτωση ενός κατευθυνόμενου δικτύου, όπως το δικό μας, η μέτρηση αυτή λαμβάνει υπόψη τόσο τις εισερχόμενες (In-Degree) όσο και τις εξερχόμενες (Out-Degree) συνδέσεις.

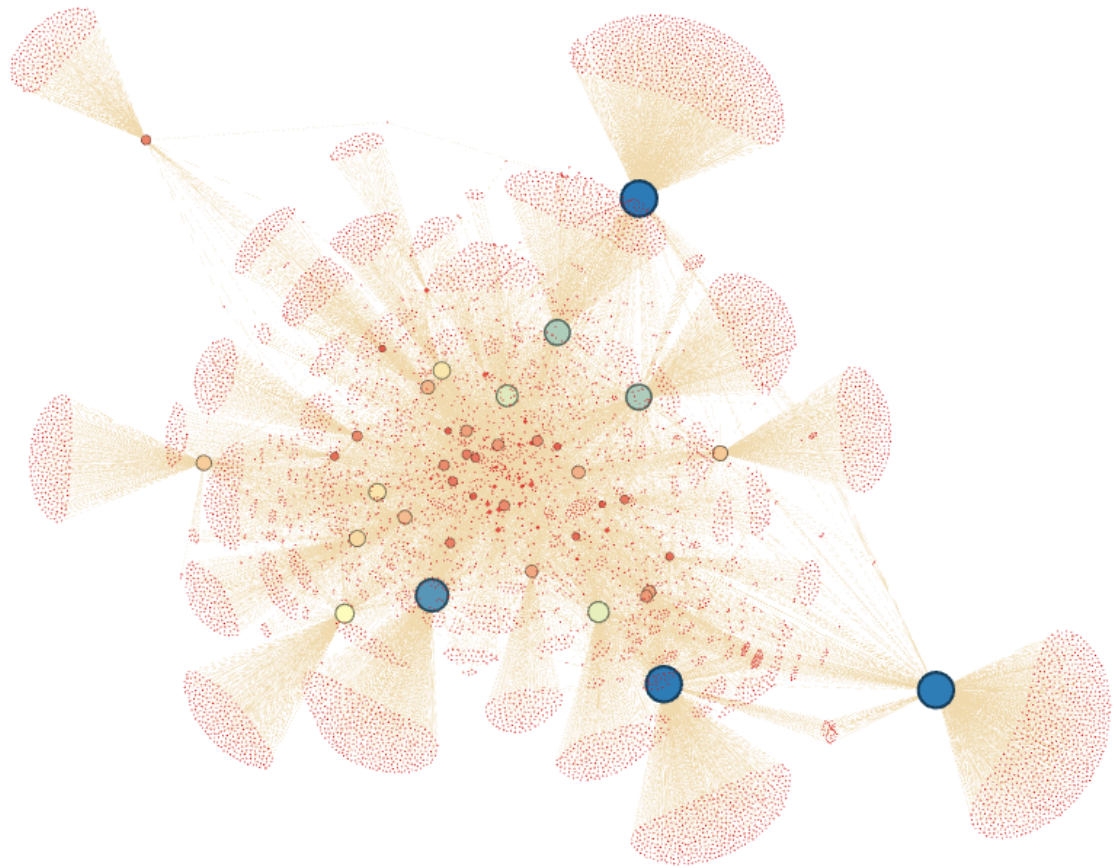
Ο average degree του δικτύου είναι 1.941. Αυτό σημαίνει ότι, κατά μέσο όρο, κάθε κόμβος συνδέεται με περίπου 1.94 άλλους κόμβους. Η τιμή αυτή είναι πολύ κοντά στο 2, γεγονός που υποδηλώνει ότι οι περισσότεροι κόμβοι έχουν μικρό αριθμό συνδέσεων, είτε εισερχόμενων είτε εξερχόμενων. Η χαμηλή τιμή του Average Degree υποδηλώνει ότι το δίκτυο είναι αραιό (sparse) και όχι ιδιαίτερα διασυνδεδεμένο. Ωστόσο, αυτή είναι μια συνηθισμένη συμπεριφορά σε παγκόσμια κοινωνικά δίκτυα ιστού.

5.5. Degree Distribution.

Άλλη μια ενδιαφέρουσα μετρική στην ανάλυση δικτύων είναι η Degree Distribution, η οποία περιγράφει τη συχνότητα των κόμβων με έναν δεδομένο αριθμό συνδέσεων σε ένα γράφημα. Στην ανάλυση αυτή, θα υπολογίσουμε την κατανομή για το Out-Degree, το In-Degree και το Total Degree.

5.5.1 Out-Degree Distribution

Παρακάτω παρατίθεται η οπτικοποίηση του δικτύου, όπου οι κόμβοι έχουν χρωματιστεί ανάλογα με την κατανομή του Out-Degree και το μέγεθός τους έχει οριστεί βάσει των τιμών του Out-Degree.



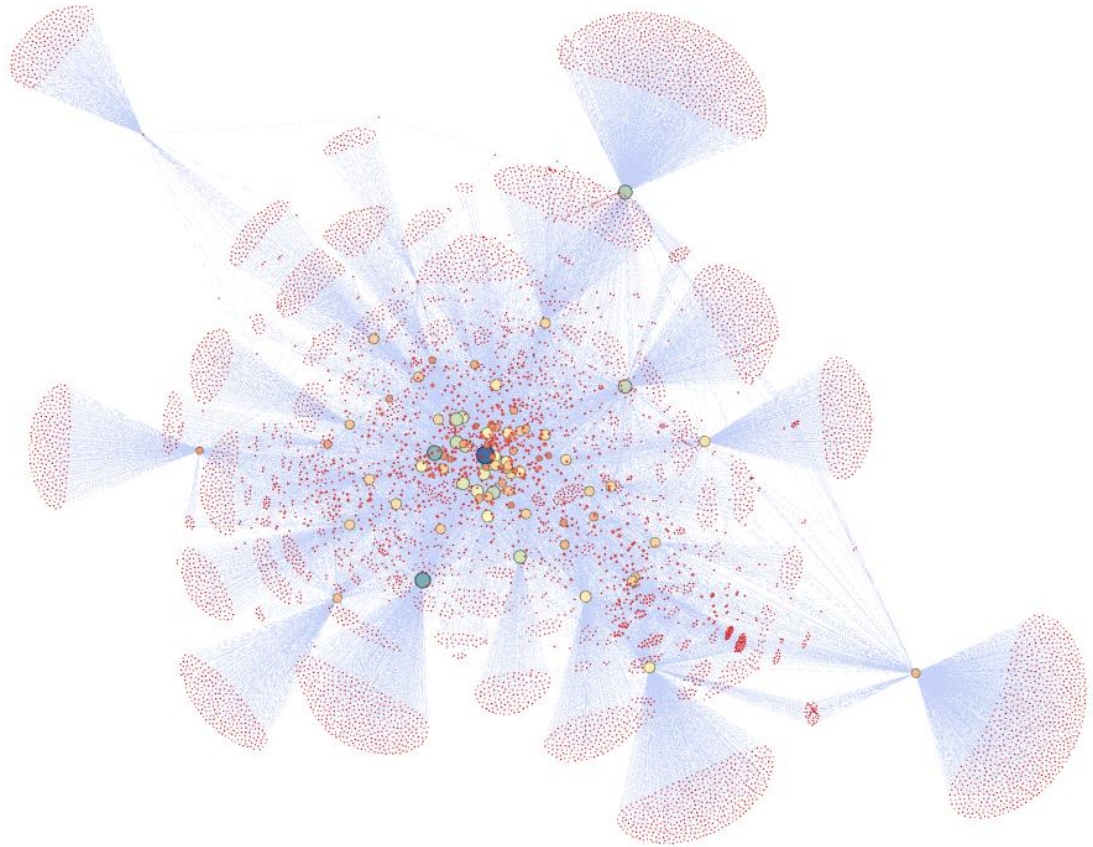
Η κατανομή του Out-Degree δείχνει ότι:

- Το 99.21% του δικτύου, δηλαδή 8752 κόμβοι, έχει Out-Degree = 0, πράγμα που σημαίνει ότι αυτοί οι κόμβοι δεν έχουν καμία εξερχόμενη σύνδεση.
- Όλοι οι άλλοι κόμβοι αποτελούν το 0.2%-0.1% του δικτύου με το out-degree να μεταβάλλεται από 32 μέχρι και 997.

Η πλειονότητα των κόμβων με Out-Degree = 0 υποδηλώνει ότι οι περισσότεροι κόμβοι δεν δημιουργούν ενεργά συνδέσεις με άλλους.

5.5.2 In-Degree Distribution

Παρακάτω παρατίθεται η οπτικοποίηση του δικτύου, όπου οι κόμβοι έχουν χρωματιστεί ανάλογα με την κατανομή του In-Degree και το μέγεθός τους έχει οριστεί βάσει των τιμών του In-Degree.



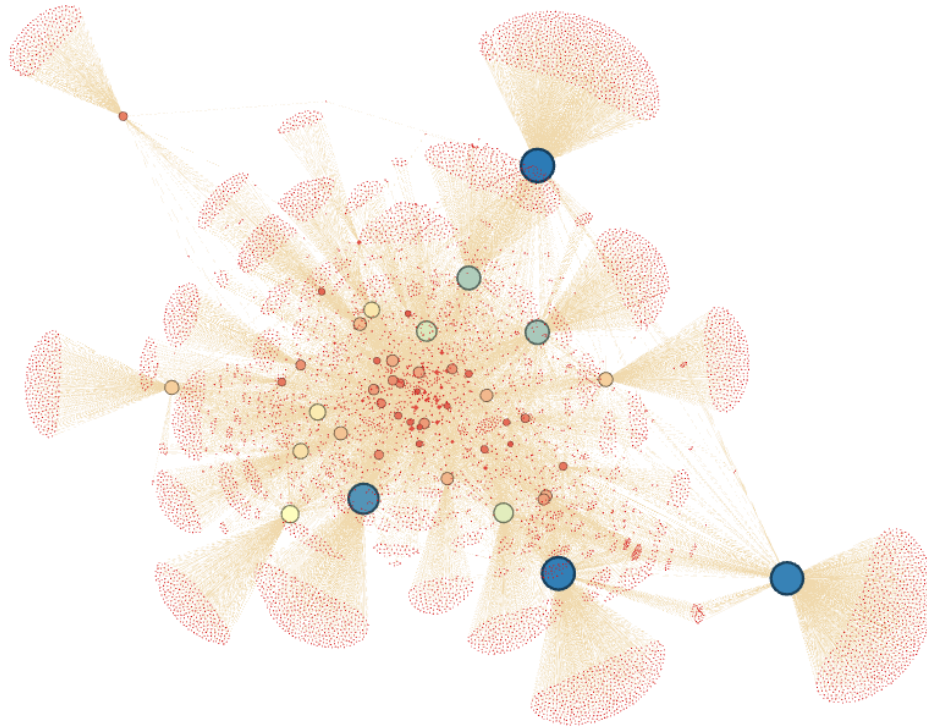
Η κατανομή του In-Degree δείχνει ότι:

- Το 73.75% του δικτύου, δηλαδή 6506 κόμβοι, έχουν In-Degree = 1.
- Το 12.04%, δηλαδή 1067 κόμβοι, έχουν In-Degree = 2.
- Οι κόμβοι με In-Degree ίσο με 3 καταλαμβάνουν 4.95% ,ενώ εκείνοι με τιμές από 5 και 6 αποτελούν αντίστοιχα το 1.69% και το 1.05%.
- Τέλος, για In-Degree μεγαλύτερο από 6, οι κόμβοι αποτελούν λιγότερο από το 1% του δικτύου. Ο μέγιστος βαθμός In-Degree είναι 64, ο οποίος ανήκει στον κόμβο politico.eu.

Αυτή η κατανομή υποδεικνύει ότι οι περισσότεροι κόμβοι έχουν χαμηλές τιμές In-Degree, γεγονός που επιβεβαιώνει την ύπαρξη ενός αραιού και αποκεντρωμένου δικτύου.

5.5.3. Total Degree Distribution

Παρακάτω παρατίθεται η οπτικοποίηση του δικτύου, όπου οι κόμβοι έχουν χρωματιστεί ανάλογα με την κατανομή του Degree και το μέγεθός τους έχει οριστεί βάσει των τιμών του Degree.



Η κατανομή του Total Degree έχει ως εξής:

- Το 73.75% των κόμβων (7568 κόμβοι) έχουν Total Degree = 1.
- Το 12.04% (1062 κόμβοι) έχει Total Degree = 2, ενώ οι κόμβοι με Total Degree = 3 αποτελούν 4.95% (437 κόμβοι).
- Οι κόμβοι με Total Degree από 4 έως 6 κατανέμονται από 3.15% έως 1.05%.
- Για τιμές Total Degree > 6, οι κόμβοι αποτελούν λιγότερο από 1%, με την υψηλότερη τιμή Total Degree να φτάνει το 1,046.

Αυτή η κατανομή ενισχύει την παρατήρηση ότι το δίκτυο είναι αραιό, με την πλειονότητα των κόμβων να συνδέονται μόνο με 1-2 κόμβους, ενώ μόνο ένα πολύ μικρό ποσοστό είναι εξαιρετικά συνδεδεμένοι κόμβοι (hubs).

6. Centrality measures

Οι μετρικές κεντρικότητας (Centrality Measures) είναι κρίσιμες για την ανάλυση δικτύων, καθώς μας βοηθούν να κατανοήσουμε τη σημασία και την επιρροή κάθε κόμβου στο δίκτυο. Μέσα από αυτές, μπορούμε να αναγνωρίσουμε κόμβους που διαδραματίζουν κεντρικό ρόλο, είτε μέσω των συνδέσεών τους (Degree), είτε μέσω της ικανότητάς τους να συνδέουν άλλους κόμβους (Betweenness), είτε με βάση την εγγύτητά τους (Closeness) ή την επιρροή τους σε συνάρτηση με τους συνδεδεμένους γείτονές τους (Eigenvector).

6.1. Degree Centrality

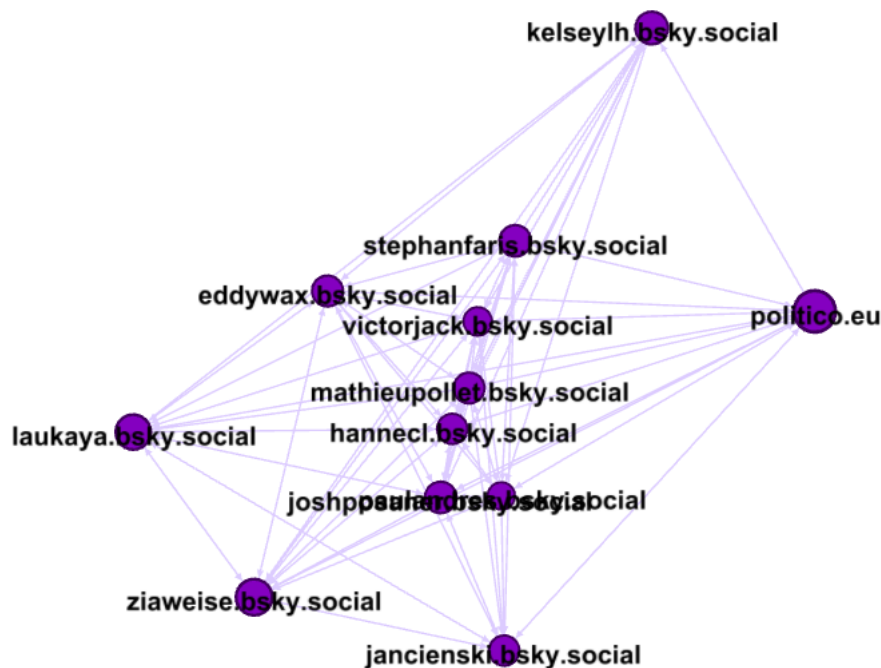
Η Degree Centrality μετρά την άμεση επιρροή ενός κόμβου υπολογίζοντας τον αριθμό των ακμών που συνδέονται με αυτόν. Στο συγκεκριμένο δίκτυο,

που είναι κατευθυνόμενο, υπολογίζουμε ξεχωριστά τις In-Degree και Out-Degree Centrality, ενώ παράλληλα εξετάζουμε και την Total Degree Centrality.

Στην προηγούμενη ενότητα, αναλύσαμε λεπτομερώς την κατανομή των In-Degree, Out-Degree και Total Degree, και εντοπίσαμε τους κορυφαίους κόμβους σε κάθε κατηγορία.

6.1.1. In-Degree Centrality

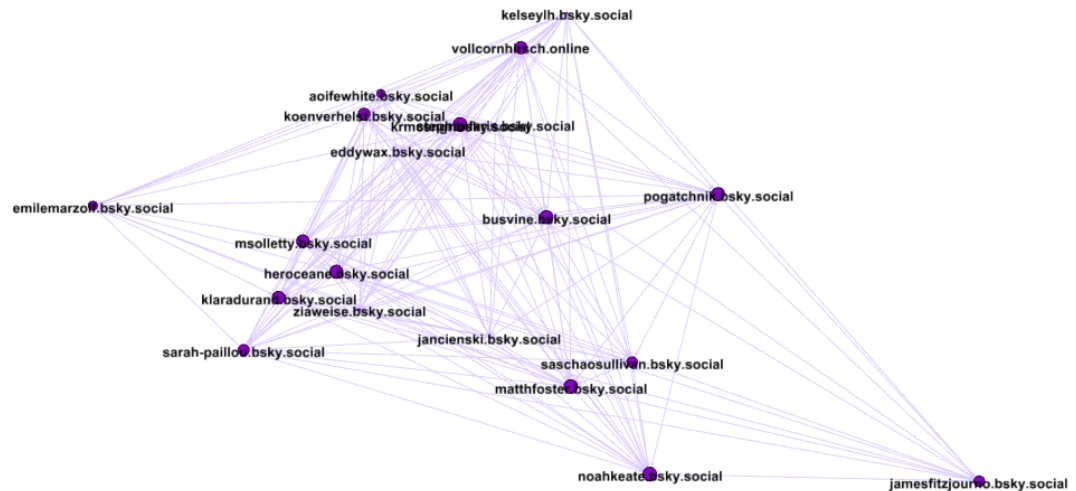
Στην προηγούμενη ενότητα, αναλύσαμε λεπτομερώς την κατανομή των In-Degree και είδαμε ποιοι είναι οι 5 κορυφαίοι κόμβοι/nodes. Αν δεχτούμε την παραδοχή ότι οι κόμβοι με $\text{in-degree} \geq 45$ θεωρούνται influencers, τότε προκύπτει ότι είναι 12 και συγκεκριμένα αυτοί που παρουσιάζονται στο παρακάτω σχήμα. Οι κορυφαίοι από αυτούς έχουν ήδη αναφερθεί αναλυτικά στην προηγούμενη ενότητα, με τον κόμβο **politico.eu** να ξεχωρίζει ως ο πιο επιδραστικός με In-Degree = 64.



Γενικά υπάρχουν διάφορες θεωρίες σχετικά με την σημαντικότητα και την επιρροή ενός κόμβου σε ένα δίκτυο. Μια από αυτές περιγράφει πώς δίκτυα, όπως τα κοινωνικά δίκτυα, είναι συχνά scale-free, με λίγους κόμβους (hubs/influencers) που συγκεντρώνουν την πλειοψηφία των συνδέσεων. Σε τέτοια δίκτυα, οι κόμβοι που βρίσκονται στο άνω 5% ή 1% της κατανομής χαρακτηρίζονται ως influencers. Εμείς εδώ δεχτήκαμε ένα αρκετά αυστηρό όριο για την ανάδειξη ενός hub/influencer καθώς τα $\text{in-degree} \geq 45$ αποτελούν το 0.02%-0.01%.

6.1.2. Out-Degree Centrality

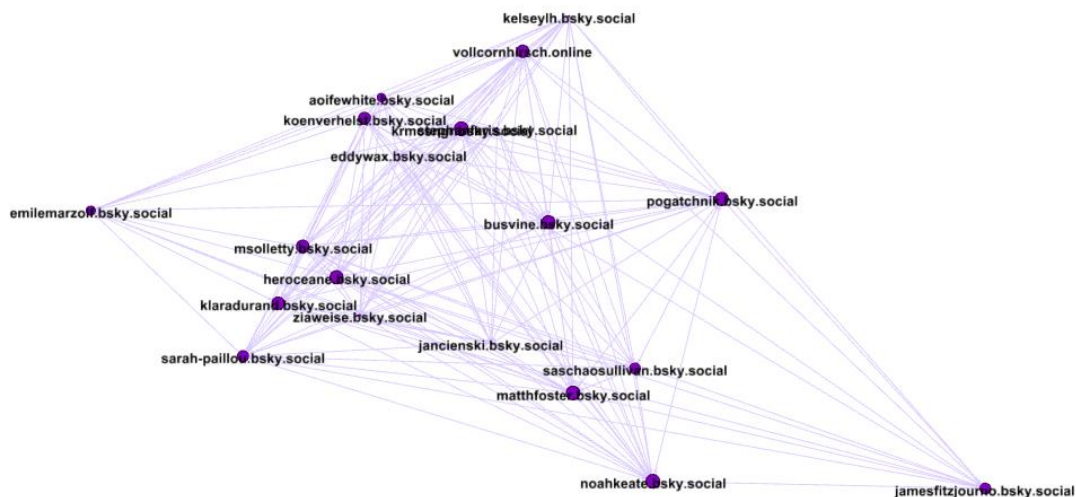
Στην προηγούμενη ενότητα αναλύσαμε και της κατανομή του out-degree, καθώς και τους 5 κορυφαίους κόμβους. Τώρα θα δεχτούμε την παραδοχή ότι οι κόμβοι με $\text{out-degree} > 250$ θεωρούνται followers, καθώς δημιουργούν πολλές εξερχόμενες συνδέσεις. Η υπόθεσή μας και πάλι είναι αρκετά αυστηρή καθώς συνολικά όλα τα out-degree που πληρούν τις προϋποθέσεις από απαρτίζουν το 0.2% του συνολικού δικτύου.



Ωστόσο και εδώ οι πιο σημαντικοί κόμβοι είναι οι πέντε κορυφαίοι που αναλύθηκαν στην ενότητα 5.2.

6.1.3. Total-Degree Centrality

Την ίδια ανάλυση και διαδικασία με τα προηγούμενα ακολουθήσαμε και για το Total-Degree Centrality. Σε αυτή την περίπτωση θα θεωρήσουμε actives τους λογαριασμούς-nodes με $\text{Degree} > 300$, λόγω της συνολικής τους δραστηριότητας στο δίκτυο. Πρόκειται για ένα εξίσου «αυστηρό» όριο που αναδεικνύει μόνο το 0.2% του δικτύου αλλά έτσι και αλλιώς το ενδιαφέρον βρίσκεται στους 3 κορυφαίους που συμπίπτουν με αυτούς του out-degree, δηλαδή προφανώς ένας λογαριασμός που αναδείχθηκε ως follower θεωρείται και αρκετά active/ενεργός.

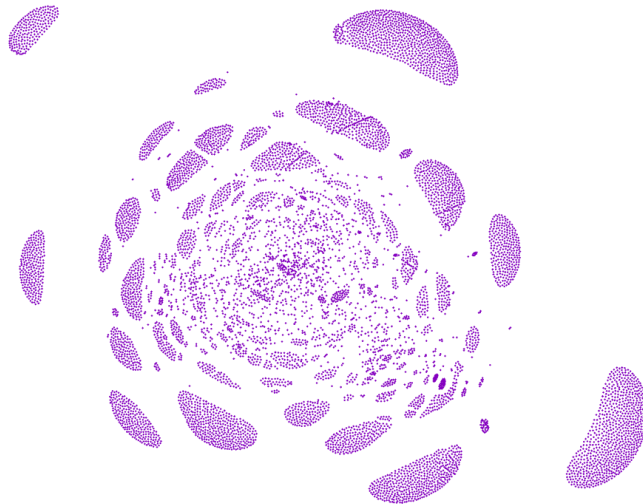


6.2. Betweenness Centrality

Το Betweenness Centrality είναι μία από τις πιο σημαντικές μετρικές σε ανάλυση δικτύων, καθώς μετρά πόσο συχνά ένας κόμβος εμφανίζεται στις συντομότερες διαδρομές (shortest paths) μεταξύ δύο άλλων κόμβων. Οι κόμβοι με υψηλή τιμή betweenness centrality συχνά λειτουργούν ως γέφυρες (bridges) που συνδέουν διαφορετικά μέρη του δικτύου. Αυτοί οι κόμβοι ελέγχουν τη ροή πληροφοριών ή αλληλεπιδράσεων, και η απώλειά τους μπορεί να διασπάσει το δίκτυο σε ξεχωριστά υποσύνολα.

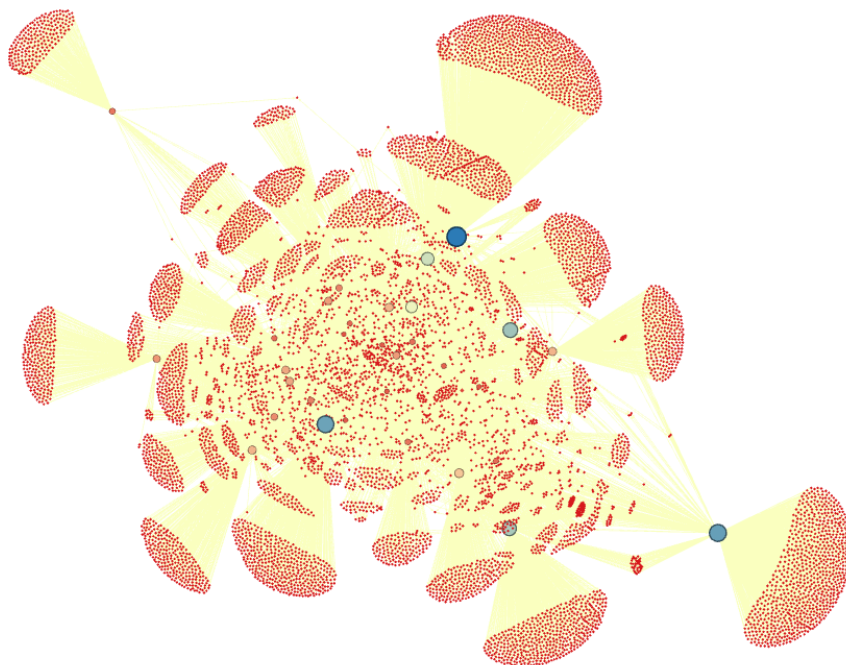
Το Betweenness Centrality, μας βοηθάει να εντοπίσουμε τους σημαντικούς κόμβους. Οι κόμβοι με υψηλή τιμή betweenness centrality είναι συνήθως στρατηγικά σημαντικοί, καθώς διευκολύνουν τη σύνδεση διαφορετικών κοινοτήτων ή περιοχών του δικτύου. Επιπλέον, συμβάλλει στην ανάλυση της επιρροής και την ανίχνευση ευπάθειας. Στα κοινωνικά δίκτυα, οι κόμβοι με υψηλή τιμή μπορεί να είναι χρήστες με ρόλο "διαμεσολαβητή", όπως δημοσιογράφοι ή οργανισμοί που διαδίδουν ειδήσεις και είναι παράλληλα κρίσιμοι για την διατήρηση της συνδεσιμότητας.

Στο παρόν δίκτυο, οι τιμές betweenness centrality κυμαίνονται από 0 έως 66694.32, με τη συντριπτική πλειοψηφία των κόμβων να παρουσιάζει μηδενικές τιμές. Αυτή η παρατήρηση φανερώνει μία έντονα ασύμμετρη κατανομή, όπου μόνο ένα μικρό ποσοστό κόμβων λειτουργεί ως κρίσιμοι διαμεσολαβητές.



Συγκεκριμένα, όπως φαίνεται και στην παραπάνω εικόνα το 99.21% των κόμβων, δηλαδή 8752 κόμβοι, έχουν τιμή *betweenness centrality* ίση με 0, κάτι που σημαίνει ότι αυτοί οι κόμβοι δεν διαδραματίζουν ρόλο διαμεσολάβησης στο δίκτυο. Όλες οι άλλες τιμές αποτελούν μόλις το 0.01% η κάθε μια, δηλαδή έχουν αρκετούς κόμβους ο καθένας από αυτούς με διαφορετική τιμή *betweenness centrality*. Οι τιμές *betweenness centrality* των κόμβων αυτών ξεκινάνε από το 100 περίπου και φτάνουν τις 66694.

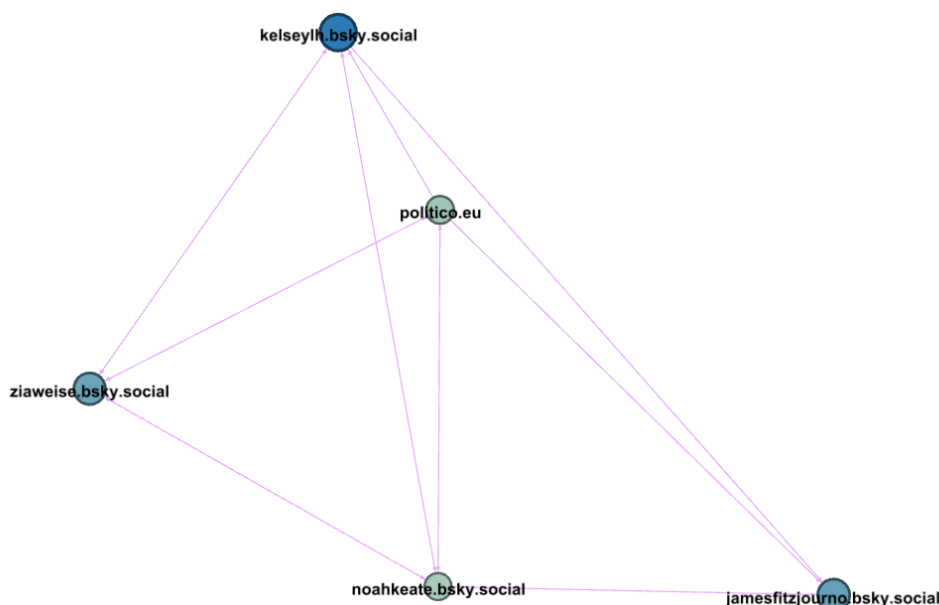
Η παρακάτω οπτικοποίηση απεικονίζει το δίκτυο, όπου οι κόμβοι έχουν μέγεθος και χρώμα ανάλογα με την τιμή της *betweenness centrality*.



Όπως αντιλαμβανόμαστε η αφαίρεση κόμβων με υψηλό *betweenness centrality* οδηγεί σε πλήρη αποσύνδεση του δικτύου, με αποτέλεσμα τη

δημιουργία απομονωμένων τμημάτων χωρίς συνδέσεις. Αυτό υπογραμμίζει τον κρίσιμο ρόλο αυτών των κόμβων στη διατήρηση της δομικής ακεραιότητας και στη ροή πληροφοριών μέσα στο δίκτυο.

Ενδιαφέρον είναι να δούμε ποιοι είναι οι 5 κορυφαίοι κόμβοι βάσει betweenness centrality:



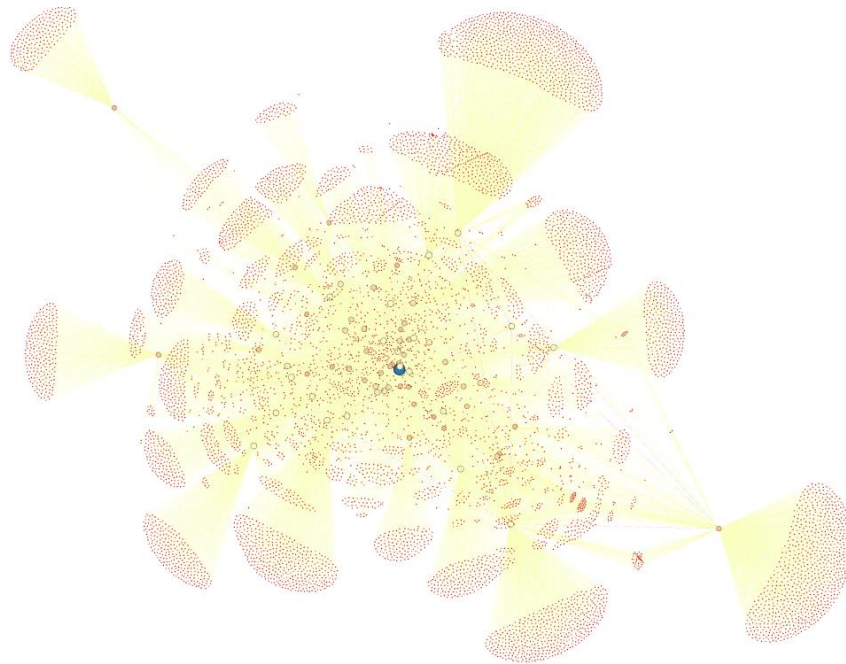
Αυτοί οι κόμβοι έχουν κεντρικό ρόλο, συνδέοντας διαφορετικές υπο-κοινότητες μέσα στο δίκτυο. Όπως παρατηρούμε ο λογαριασμός **politico.eu** ανήκει μέσα σε αυτούς, αναδεικνύοντας έτσι ακόμα περισσότερο το καθοριστικό ρόλο που παίζει μέσα στο δίκτυο.

Η ανάλυση της μετρικής Betweenness Centrality καταδεικνύει τη σημασία ενός μικρού αριθμού κόμβων στη διατήρηση της συνδεσιμότητας και της λειτουργικότητας του δικτύου. Αυτοί οι κόμβοι είναι ζωτικής σημασίας όχι μόνο για τη ροή πληροφοριών, αλλά και για τη γενικότερη σταθερότητα της δομής. Η αφαίρεσή τους θα οδηγούσε σε κατακερματισμό του δικτύου.

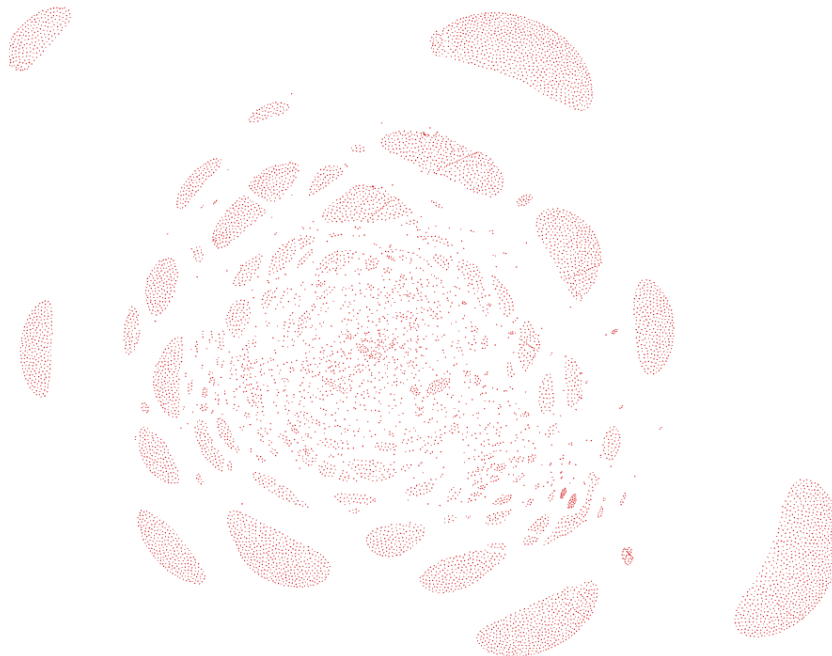
6.3. Closeness Centrality

Η μετρική Closeness Centrality μετράει πόσο κοντά βρίσκεται ένας κόμβος σε όλους τους άλλους κόμβους του δικτύου. Συγκεκριμένα, επικεντρώνεται στο πόσο αποδοτικά μπορεί ένας κόμβος να φτάσει σε όλους τους άλλους. Κόμβοι με υψηλή τιμή closeness centrality βρίσκονται σε κεντρικές θέσεις του δικτύου και είναι ικανοί να φτάσουν γρήγορα και αποδοτικά σε άλλους κόμβους. Αυτοί οι κόμβοι είναι εξαιρετικά χρήσιμοι για τη διανομή πληροφοριών ή την επιρροή σε άλλους. Η μέτρηση αυτή κυμαίνεται συνήθως μεταξύ 0 και 1, όπου τιμές κοντά στο 1 υποδηλώνουν υψηλή σημασία.

Η παρακάτω οπτικοποίηση απεικονίζει το δίκτυο, όπου οι κόμβοι έχουν μέγεθος και χρώμα ανάλογα με την τιμή Closeness Centrality.

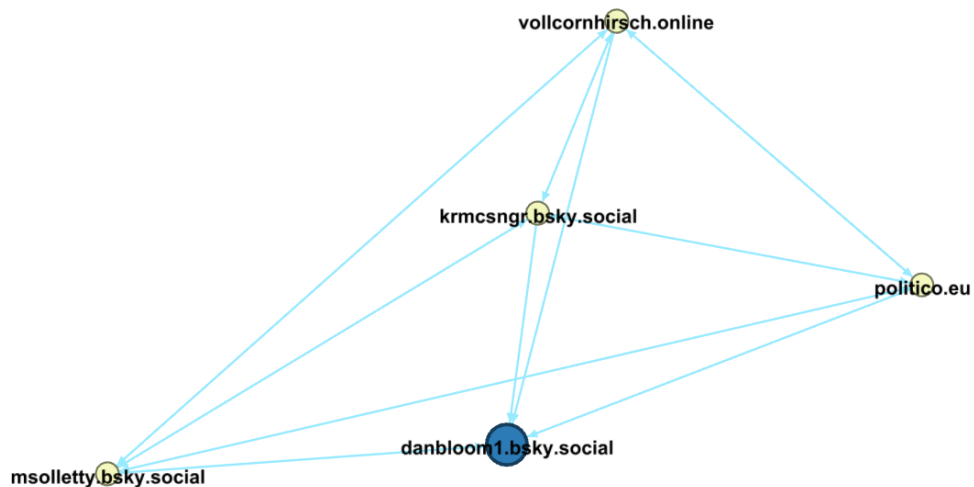


Η κατανομή του Closeness Centrality στο δίκτυό μας είναι εξαιρετικά ασύμμετρη. Συγκεκριμένα, το 99.21% των κόμβων έχει closeness centrality ίσο με 0, γεγονός που υποδηλώνει ότι αυτοί οι κόμβοι είναι είτε απομονωμένοι είτε πολύ απομακρυσμένοι από άλλους. Όπως φαίνονται και στην παρακάτω εικόνα:



Οι υπόλοιποι κόμβοι έχουν τιμές που κυμαίνονται από 0.33 έως 1, με καθένα από αυτά τα επίπεδα να αντιπροσωπεύει μόλις το 0.01% των κόμβων. Η μεγάλη πλειοψηφία των κόμβων δεν έχει άμεση ή έμμεση πρόσβαση στο υπόλοιπο δίκτυο. Ωστόσο, ένας μικρός αριθμός κόμβων με υψηλότερες τιμές

closeness centrality φαίνεται να λειτουργούν ως «κεντρικοί» κόμβοι, επιτρέποντας γρήγορη διάχυση πληροφοριών. Συγκεκριμένα οι 5 κορυφαίοι κόμβοι είναι οι παρακάτω:



Όπως παρατηρούμε από το γράφημα, ο κόμβος **danbloom1.bsky.social** έχει την υψηλότερη τιμή closeness centrality, η οποία είναι ίση με 1. Ωστόσο, αναλύοντας τις πληροφορίες που διαθέτουμε στο Data Laboratory, διαπιστώνουμε μια ανακρίβεια. Συγκεκριμένα, ο συγκεκριμένος κόμβος έχει In-Degree = 29, Out-Degree = 1 και συνολικό Degree = 30. Η τιμή Closeness Centrality = 1 συνεπάγεται ότι ο κόμβος είναι άμεσα συνδεδεμένος με όλους τους υπόλοιπους κόμβους του δικτύου (δηλαδή, μπορεί να φτάσει οποιονδήποτε άλλο κόμβο μέσω μίας μόνο σύνδεσης). Ωστόσο, με συνολικό Degree = 30, ο κόμβος συνδέεται μόνο με 30 άλλους κόμβους, κάτι που έρχεται σε πλήρη αντίθεση με το συνολικό μέγεθος του δικτύου μας, το οποίο περιλαμβάνει 8822 κόμβους.

Αυτή η ασυμβατότητα πιθανώς οφείλεται σε σφάλμα στον τρόπο που το Gephi υπολογίζει την τιμή του Closeness Centrality για κόμβους που βρίσκονται σε απομονωμένα ή λιγότερο συνδεδεμένα μέρη του δικτύου. Υποστηρίζοντας αυτή την υπόθεση, έχει αναφερθεί ένα σχετικό πρόβλημα στο GitHub της Gephi ([Gephi Issue #2951](#)), όπου περιγράφεται ότι σε δίκτυα με πολλαπλά ασύνδετα υποδίκτυα ή σε δίκτυα με μεγάλα επίπεδα αραιότητας, οι υπολογισμοί της μετρικής μπορεί να οδηγούν σε λανθασμένες τιμές. Βάσει αυτών των δεδομένων, η τιμή Closeness Centrality = 1 που παρατηρείται στον κόμβο danbloom1.bsky.social δεν είναι αξιόπιστη και ενδέχεται να επηρεάζει την ανάλυση μας.

6.4. Eigenvector Centrality

Το Eigenvector Centrality είναι μια μετρική που αξιολογεί τη σημασία ενός κόμβου όχι μόνο με βάση τον αριθμό των συνδέσεών του αλλά και τη σημασία των κόμβων στους οποίους συνδέεται. Με άλλα λόγια, ένας κόμβος που συνδέεται με άλλους ιδιαίτερα κεντρικούς κόμβους αποκτά υψηλό Eigenvector Centrality, ακόμα κι αν ο ίδιος δεν έχει πολλές συνδέσεις. Αυτή η

μετρική παρέχει μια πιο «ποιοτική» προσέγγιση στην κεντρικότητα, καθώς αξιολογεί όχι μόνο τις άμεσες συνδέσεις, αλλά και την ευρύτερη επιρροή ενός κόμβου στο δίκτυο. Οι τιμές κυμαίνονται στο διάστημα από 0 έως 1.

Η μετρική αυτή είναι ιδιαίτερα χρήσιμη σε κοινωνικά δίκτυα για τον εντοπισμό κόμβων που έχουν ευρεία επιρροή και δύναμη στη διάδοση της πληροφορίας, ακόμα και αν αυτοί δεν έχουν πολλές άμεσες συνδέσεις.

Για το δίκτυό μας ο αλγόριθμος του Eigenvector Centrality εκτελέστηκε με τα παρακάτω χαρακτηριστικά:

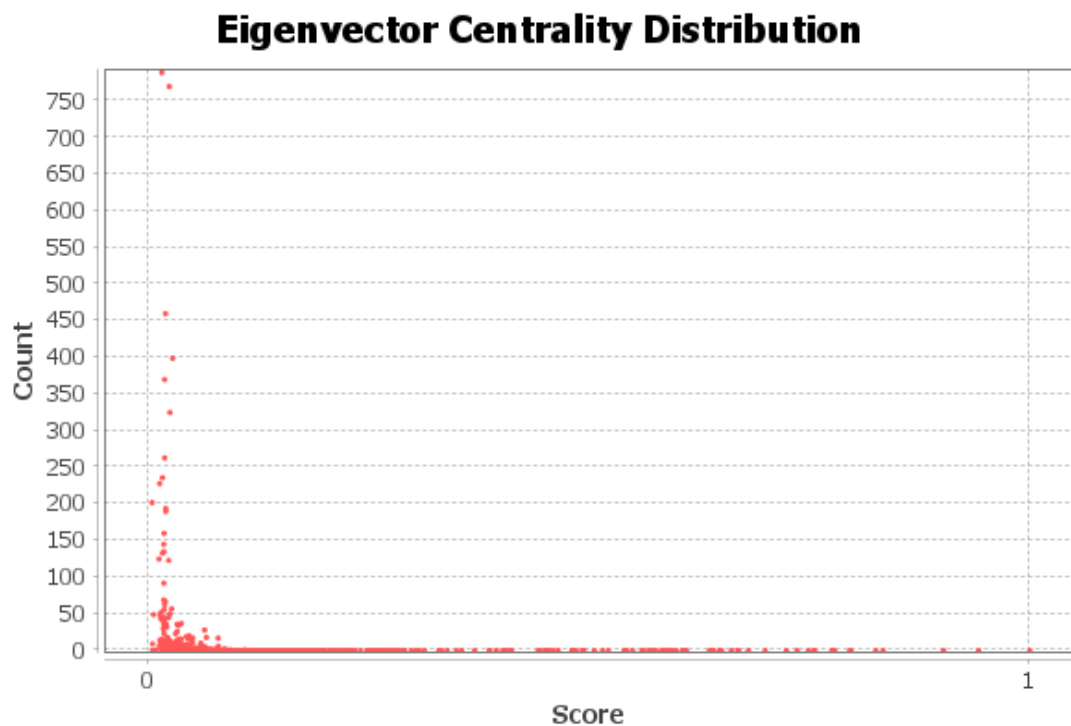
Parameters:

Network Interpretation: directed

Number of iterations: 100

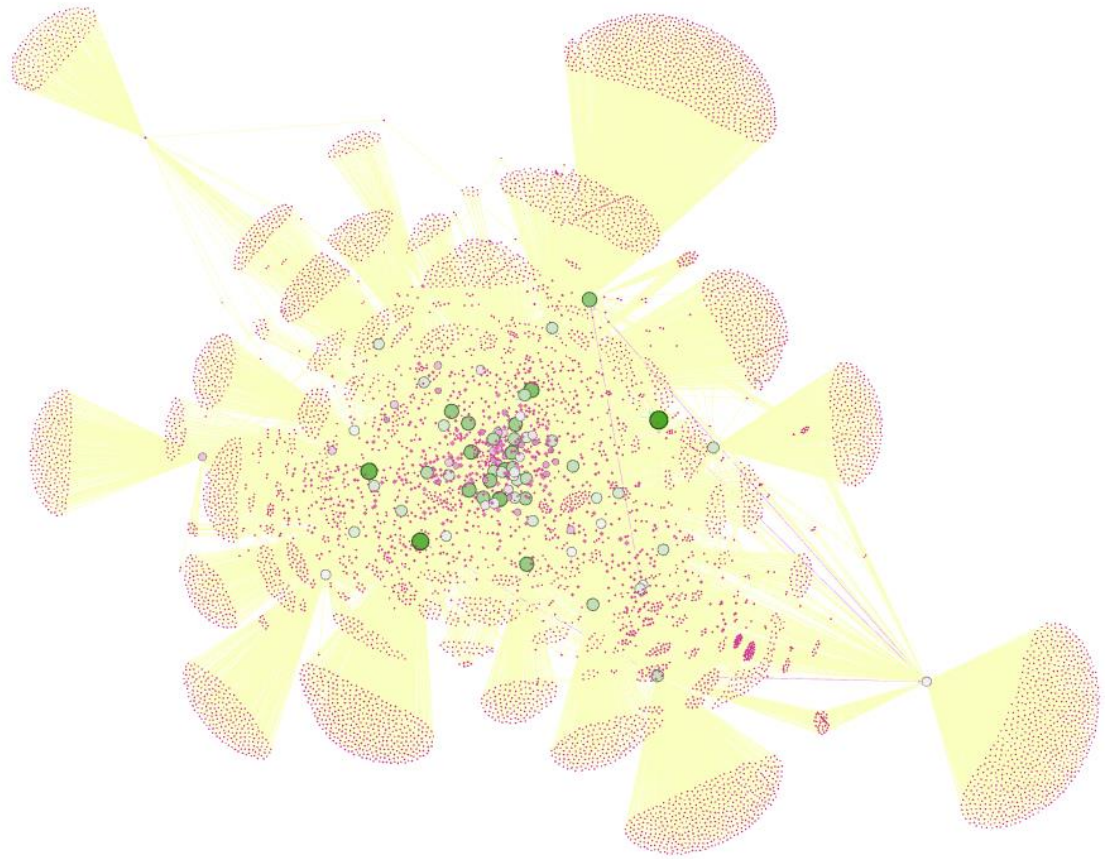
Sum change: 0.009823151333318354

Results:

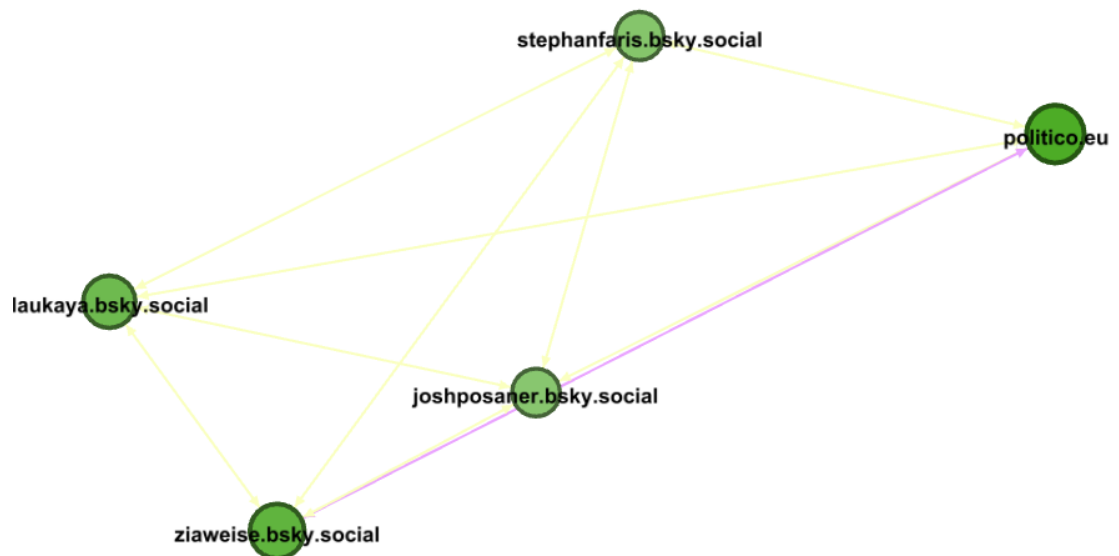


Η διαδικασία ολοκληρώθηκε σε 100 επαναλήψεις και η πολύ μικρή τιμή του Sum change δείχνει ότι ο αλγόριθμος συγκλίνει ικανοποιητικά, με ελάχιστες αλλαγές στις centrality values.

Στο παρακάτω σχήμα, οπτικοποιείται το δίκτυο με τους κόμβους να έχουν μέγεθος και χρώμα που μεταβάλλεται ανάλογα με τις τιμές του Eigenvector Centrality:



Οι τιμές του Eigenvector Centrality στο δίκτυό μας κυμαίνονται από 0.0139552 (8.95%) έως 1 (0.01%). Παρατηρούμε ότι οι τιμές είναι εξαιρετικά άνισα κατανομημένες, με την πλειονότητα των κόμβων να παρουσιάζει χαμηλές τιμές, ενώ οι υψηλότερες τιμές περιορίζονται σε έναν πολύ μικρό αριθμό κόμβων. Ειδικότερα, οι πέντε μεγαλύτερες τιμές ανήκουν σε πέντε κόμβους που καλύπτουν μόλις 0.01% του συνολικού δικτύου.



Παρατηρούμε πως πέρα από το **politico.eu** που εμφανίζεται σταθερά στις κορυφαίες πεντάδες σε κάθε μετρική, όλοι οι άλλοι κόμβοι είναι διαφορετικοί, γεγονός που δείχνει ότι διαφορετικοί κόμβοι διαδραματίζουν κεντρικούς ρόλους ανάλογα με τη μέτρηση που εξετάζουμε. Αυτό ενισχύει την ιδέα ότι κάθε μέτρο κεντρικότητας αναδεικνύει διαφορετικές πτυχές της δικτύωσης. Όσον αναφορά το κόμβο politico.eu η συνεχής εμφάνισή του τον καθιστά έναν από τους πιο σημαντικούς κόμβους στο δίκτυο.

7. Clustering effects in the network

Σε αυτή την ενότητα θα αναλύσουμε τα φαινόμενα clustering στο δίκτυό μας, εστιάζοντας σε τέσσερις βασικές μετρήσεις:

1. **Average Clustering Coefficient:** Ένα μέτρο που περιγράφει πόσο πιθανό είναι οι γείτονες ενός κόμβου να είναι συνδεδεμένοι μεταξύ τους.
2. **Number of Triangles:** Ο αριθμός των πλήρως συνδεδεμένων υποδικτύων τριών κόμβων στο δίκτυο.
3. **Clustering Coefficient Distribution:** Η κατανομή των τιμών του συντελεστή συνοχής μεταξύ των κόμβων.
4. **Existence of the Triadic Closure Phenomenon:** Η τάση δημιουργίας τριγώνων μεταξύ συνδεδεμένων κόμβων, που συχνά παρατηρείται στα κοινωνικά δίκτυα.

Ο συντελεστής clustering coefficient ενός κόμβου μετρά πόσοι από τους γείτονές του είναι επίσης γείτονες μεταξύ τους. Οι τιμές κυμαίνονται από 0 έως 1. Αν clustering coefficient σε έναν κόμβο είναι 0 τότε οι γείτονές του δεν είναι συνδεδεμένοι μεταξύ τους, αν είναι 1 σημαίνει ότι όλοι οι γείτονές του συνδέονται απευθείας μεταξύ τους.

7.1. Average Clustering Coefficient

Στο δίκτυο μας, χρησιμοποιήσαμε δύο μεθόδους για να υπολογίσουμε το Clustering Coefficient, από την μια την μέθοδο Triangle η οποία επικεντρώνεται σε πιο τοπικά φαινόμενα, δίνοντας έμφαση στη δημιουργία τριγώνων και από την άλλη την Basic μέθοδος που παρέχει μια γενική εικόνα της συνοχής του δικτύου.

Με την Triangle method, ο Clustering Coefficient υπολογίστηκε κοντά στην τιμή 0.0284, αυτή η μέθοδος βασίζεται στον αριθμό των τριγώνων και των διαδρομών μήκους 2 στο δίκτυο. Ενώ, με την Basic Method, ο Average Clustering Coefficient υπολογίστηκε 0.2383. Αυτό υποδεικνύει ότι περίπου το 23.83% των πιθανών τριγώνων στο δίκτυο έχουν ολοκληρωθεί.

Η διαφορά μεταξύ των δύο μεθόδων αποτυπώνει τις διαφορετικές προσεγγίσεις τους για τον υπολογισμό της συνοχής. Η Triangle Method εστιάζει στο τοπικό επίπεδο, εξετάζοντας τις διαδρομές μήκους 2 που μπορεί να καταλήξουν σε τρίγωνα. Αντίθετα, η Basic Method παρέχει μια πιο γενική εικόνα της συνοχής του δικτύου, εξετάζοντας το συνολικό ποσοστό συνδεδεμένων γειτόνων. Η Triangle Method υποδεικνύει ότι το δίκτυο έχει

μια πιο αραιή τοπική δομή, ενώ η Basic Method δείχνει μια συνολική εικόνα πιο υψηλής συνοχής.

7.2. *Number of Triangles*

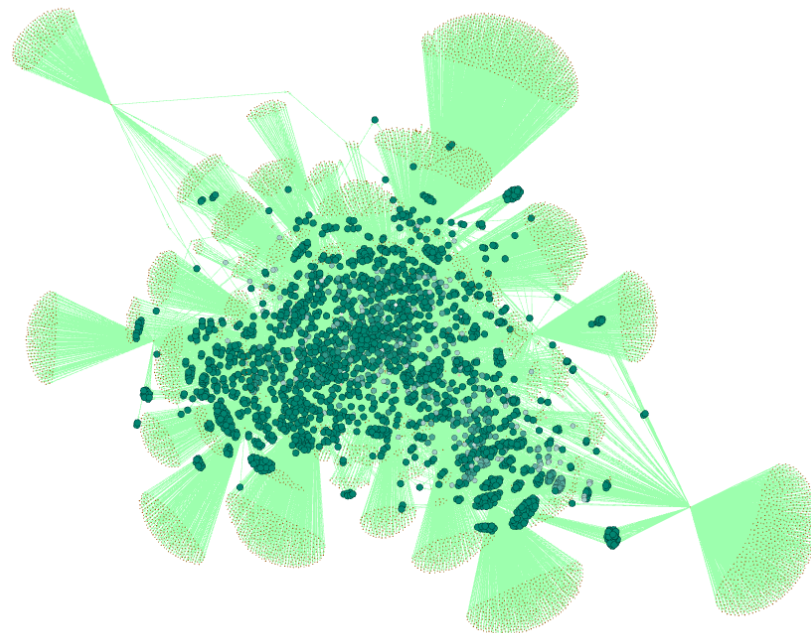
Σύμφωνα με την Triangle μέθοδο το δίκτυο περιέχει συνολικά 40375 τρίγωνα, αριθμός που υποδηλώνει την ύπαρξη τοπικών πυκνών περιοχών, παρά την αραιότητα που παρατηρείται γενικά στο δίκτυο. Αυτές οι πυκνές περιοχές αποτελούν συχνά σημαντικά υποδίκτυα, όπου οι κόμβοι είναι ισχυρά συνδεδεμένοι.

Επιπλέον, ο συνολικός αριθμός των διαδρομών μήκους 2 (paths of length 2) στο δίκτυο υπολογίστηκε σε 4268312. Από αυτές τις δυνητικές συνδέσεις, μόνο ένα μικρό ποσοστό σχηματίζει τρίγωνα, κάτι που επιβεβαιώνει τη γενική αραιότητα του δικτύου.

Οι διαδρομές μήκους 2 είναι αλυσίδες της μορφής $A \rightarrow B \rightarrow C$, που αντιπροσωπεύουν δυνητικά τρίγωνα. Εάν ο κόμβος A συνδεθεί με τον κόμβο C, τότε η διαδρομή μήκους 2 μετατρέπεται σε τρίγωνο.

Η παρουσία σημαντικού αριθμού διαδρομών μήκους 2, αλλά σχετικά λίγων τριγώνων, υποδηλώνει ότι το δίκτυο έχει περιορισμένο φαινόμενο triadic closure. Στην επόμενη ενότητα, θα εξετάσουμε το φαινόμενο αυτό λεπτομερέστερα.

Στο παρακάτω σχήμα, οπτικοποιείται το δίκτυο με τους κόμβους να έχουν μέγεθος και χρώμα που μεταβάλλεται ανάλογα με τις τιμές του Clustering Coefficient:



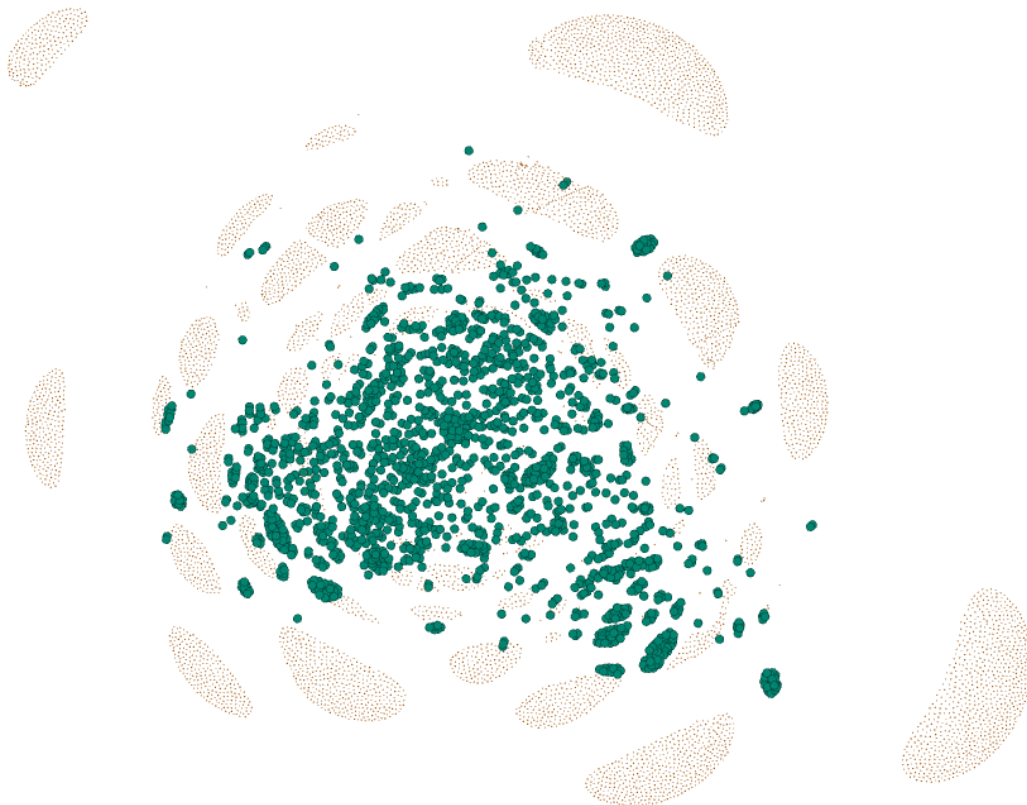
7.3. *Clustering Coefficient Distribution*

Η Clustering Coefficient Distribution στο δίκτυο αποκαλύπτει ενδιαφέροντα χαρακτηριστικά σχετικά με τη δομή του. Σύμφωνα με την Ttriangle μέθοδο παρατηρείται ότι:

- Το 74.53% των κόμβων έχουν συντελεστή συνοχής 0.0, γεγονός που δείχνει ότι οι γείτονές τους δεν είναι συνδεδεμένοι μεταξύ τους.
- Το 19.37% των κόμβων έχουν συντελεστή συνοχής 1.0, υποδεικνύοντας πλήρη σύνδεση μεταξύ των γειτόνων τους.
- Ένα μικρό ποσοστό κόμβων έχει ενδιάμεσες τιμές, όπως 0.833 (0.85%), 0.66 (0.83%), 0.89 (0.53%) και όλες οι υπόλοιπες τιμές έχουν ποσοστό εμφάνισης από 0.26% - 0.01%.

Η άνιση κατανομή του συντελεστή συνοχής δείχνει ότι το δίκτυο περιλαμβάνει κυρίως αραιά συνδεδεμένες περιοχές, αλλά ταυτόχρονα περιέχει και πυκνές τοπικές κοινότητες.

Στην παρακάτω απεικόνιση φαίνονται οι κόμβοι με Clustering Coefficient ίσο με 0 (οι πιο μικροί κόκκινοι κόμβοι) και 1 (μεγαλύτεροι πράσινοι κόμβοι):



7.4. *Existence of the Triadic Closure Phenomenon*

Το φαινόμενο Triadic Closure αναφέρεται στην τάση δύο κόμβων που μοιράζονται έναν κοινό γείτονα να συνδεθούν απευθείας μεταξύ τους, σχηματίζοντας ένα τρίγωνο.

Για να αξιολογήσουμε την ύπαρξη αυτού του φαινομένου, υπολογίσαμε τη σχέση μεταξύ των υπαρχόντων και των δυνητικών τριγώνων (Triadic Closure Ratio):

$$\text{Triadic Closure Ratio} = \frac{\text{Number of Triangles}}{\text{Number of Paths of Length 2}} = \frac{40375}{4268312} \approx 0.00946$$

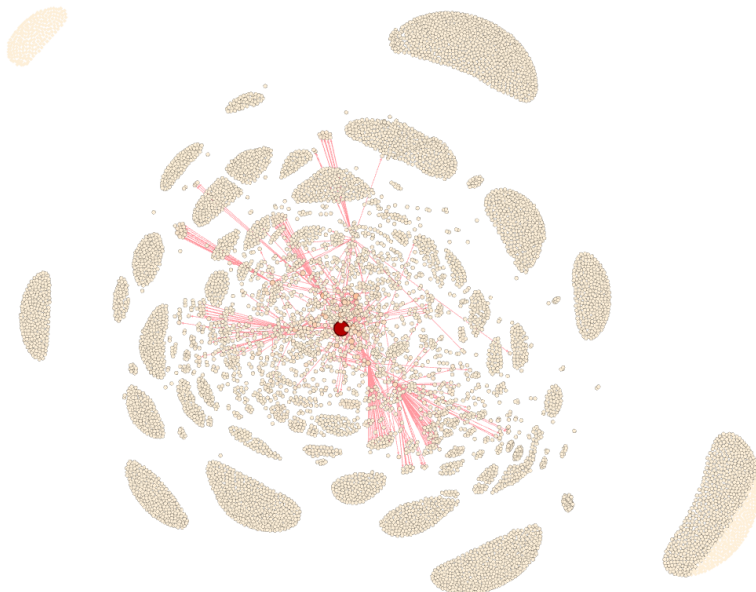
Η αναλογία αυτή δείχνει ότι μόνο το 0.946% των δυνητικών τριγώνων έχουν πραγματικά σχηματιστεί στο δίκτυο. Αυτό υποδηλώνει ότι, παρόλο που το φαινόμενο του Triadic Closure παρατηρείται, είναι αρκετά περιορισμένο στο δίκτυο.

8. Bridges and local bridges

Σε αυτή την ενότητα, θα εξετάσουμε το ρόλο των γέφυρων (bridges) και των τοπικών γεφυρών (local bridges) στο δίκτυο, οι οποίες παίζουν σημαντικό ρόλο στην κατανόηση της δομής και της συνοχής του δικτύου.

Μια bridge είναι μια σύνδεση μεταξύ δύο κόμβων που, αν αφαιρεθεί, θα διαχωρίσει το δίκτυο σε δύο ή περισσότερα απομονωμένα μέρη. Αυτές οι γέφυρες είναι κρίσιμες για τη συνολική συνδεσιμότητα του δικτύου, καθώς η αφαίρεσή τους μπορεί να προκαλέσει απομόνωση μεγάλων τμημάτων του δικτύου. Γενικά, οι bridge συνδέουν διαφορετικές περιοχές του δικτύου και αν αφαιρεθούν, το δίκτυο θα σπάσει σε πιο μικρά και ανεξάρτητα υποδίκτυα.

Στο δίκτυό μας, οι bridges μπορούν να εντοπιστούν μέσω διαφόρων τρόπων. Για αρχή μέσω των Connected Components και φιλτράροντας για 1-100 κόμβους μπορούμε να δούμε απομονωμένα μικρά components, δηλαδή να καταλάβουμε ποιοι κόμβοι δεν συνδέονται καλά με το υπόλοιπο δίκτυο. Στο παρακάτω σχήμα παρατηρούμε την παρουσία πολλών μικρών υποδικτύων που υποδηλώνουν ότι υπάρχουν ακμές που λειτουργούν ως bridges, καθώς η αφαίρεσή τους έχει οδηγήσει στη δημιουργία αποσυνδεδεμένων περιοχών.



Επιπλέον, γνωρίζουμε ότι αν ένας κόμβος έχει $\text{in-degree} = 1$ και $\text{out-degree} = 0$, τότε η απομάκρυνση της εισερχόμενης ακμής του αυξάνει τα disconnected components, καθιστώντας την ακμή bridge.

Οι local bridges είναι ακμές που συνδέουν δύο κόμβους, οι οποίοι δεν έχουν άλλες εναλλακτικές διαδρομές σύνδεσης μέσω του δικτύου. Αν αφαιρεθούν, οι συγκεκριμένοι κόμβοι δεν θα αποκοπούν εντελώς από το δίκτυο, αλλά η μεταξύ τους σύνδεση θα γίνει πολύ πιο δύσκολη, καθώς η μικρότερη διαδρομή (shortest path) θα αυξηθεί. Παρόλο που η αφαίρεσή τους δεν διασπά το δίκτυο σε μικρότερα μέρη, οι local bridges είναι σημαντικές για τη σύνδεση διαφορετικών περιοχών ή κοινοτήτων μέσα στο δίκτυο, που διαφορετικά δεν θα επικοινωνούσαν μεταξύ τους. Η απομάκρυνση αυτών των ακμών μπορεί να οδηγήσει στην απομόνωση μικρότερων τμημάτων του δικτύου ή να περιορίσει την επικοινωνία μεταξύ διαφορετικών περιοχών.

Ο κανόνας για τον εντοπισμό των local bridges είναι ο εξής: μια τοπική γέφυρα είναι μια ακμή που, αν αφαιρεθεί, αυξάνει το μήκος της μικρότερης διαδρομής (shortest path length) μεταξύ των δύο συνδεδεμένων κόμβων σε πάνω από 2. Η αναγνώριση αυτών των ακμών είναι κρίσιμη για τη μελέτη της ανθεκτικότητας του δικτύου και της ροής πληροφορίας, καθώς υποδεικνύει τις συνδέσεις που διαδραματίζουν σημαντικό ρόλο στη διασφάλιση της σύνδεσης μεταξύ διαφορετικών περιοχών.

Καθώς δεν υπάρχει ένας άμεσος θεωρητικός τρόπος να εντοπίσουμε local bridges, για να εντοπιστούν, θα γράψουμε ένα script χρησιμοποιώντας τη βιβλιοθήκη NetworkX της Python. Θα εξάγουμε πρώτα το αρχείο CSV του δικτύου μας από το Gephi και στη συνέχεια μέσω της συνάρτησης `find_local_bridges()` θα υπολογίσουμε τη μικρότερη διαδρομή μεταξύ των κόμβων πριν και μετά την αφαίρεση μιας ακμής και θα καταγράψουμε τις ακμές που πληρούν τον ορισμό της τοπικής γέφυρας. Παρακάτω βρίσκεται ο σχετικός κώδικας που χρησιμοποιήθηκε:

```

import pandas as pd
import networkx as nx # type: ignore

# Load the CSV file
df = pd.read_csv("politico_graph.csv")

# Function to find local bridges
def find_local_bridges(G):
    local_bridges = []
    edges = list(G.edges())

    for u, v in edges:
        # Compute the shortest path before removal
        try:
            before_removal = nx.shortest_path_length(G, source=u, target=v)
        except nx.NetworkXNoPath:
            continue # Skip if there is no path initially

        # Remove edge temporarily
        G.remove_edge(u, v)

        # Compute shortest path after removal
        try:
            after_removal = nx.shortest_path_length(G, source=u, target=v)
        except nx.NetworkXNoPath:
            after_removal = float("inf") # No path exists

        # Restore the edge
        G.add_edge(u, v)

        # Restore the edge
        G.add_edge(u, v)

        # If a path exists and removing the edge increases path length beyond '2', it's a local bridge
        if after_removal != float("inf") and after_removal - before_removal > 2:
            local_bridges.append((u, v, after_removal))

    return local_bridges

# Create a directed graph
G = nx.DiGraph()
G.add_edges_from(zip(df['Source'], df['Target']))

# Find local bridges
local_bridges = find_local_bridges(G)

# Print results
print("Local Bridges (edges whose removal increases shortest path):")
for edge in local_bridges:
    print(edge)

```

Το αποτέλεσμα της παραπάνω διαδικασίας είναι:

```

Local Bridges (edges whose removal increases shortest path):
('did:plc:ca5r7drvift7w46fyzrt2ga', 'did:plc:4k6dakv7cskxttdvfpzadq7e', 4)
('did:plc:ca5r7drvift7w46fyzrt2ga', 'did:plc:qc6xzgctorfsm35w6i3vdebx', 4)

```

Κάθε γραμμή δείχνει μια τοπική γέφυρα, δηλαδή μια ακμή που συνδέει δύο κόμβους και η αφαίρεσή της αυξάνει τη μικρότερη διαδρομή μεταξύ αυτών των κόμβων σε 4, στην προκειμένη περίπτωση.

9. Gender and homophily

Στο συγκεκριμένο δίκτυο, δεν υπάρχουν διαθέσιμες πληροφορίες σχετικά με το gender των κόμβων/λογαριασμών. Αυτό οφείλεται στο γεγονός ότι πρόκειται για κοινότητα που σχετίζεται με το Politico.eu, το οποίο δεν συνδέεται άμεσα με συγκεκριμένες πληροφορίες φύλου και επιπλέον πολλά από τα προφίλ που περιλαμβάνονται στο δίκτυο δεν παρέχουν στοιχεία ή ενδείξεις για το φύλο των χρηστών. Συνεπώς, η ανάλυση βασισμένη στο gender δεν είναι εφικτή στο πλαίσιο

αυτού του δικτύου και κατά συνέπεια οι παραδοσιακές προσεγγίσεις για ανάλυση gender-based homophily δεν μπορούν να εφαρμοστούν.

Ωστόσο, η ομοφιλία (homophily) σε κοινωνικά δίκτυα αναφέρεται στη τάση των ατόμων να συνδέονται με άλλους που έχουν παρόμοια χαρακτηριστικά ή συμπεριφορές. Σε αυτό το δίκτυο, μπορούμε να παρατηρήσουμε την degree-based homophily, η οποία υποδεικνύει ότι οι κόμβοι με υψηλό degree, δηλαδή οι πιο συνδεδεμένοι κόμβοι του δικτύου, είναι πιθανότερο να συνδέονται με άλλους υψηλού degree. Αντίστοιχα, οι κόμβοι με χαμηλό degree τείνουν να συνδέονται με άλλους με μικρό αριθμό συνδέσεων.

Αυτή η μορφή homophily, παρά το γεγονός ότι δεν αφορά το φύλο, έχει σημαντική επίδραση στην δομή του δικτύου, επηρεάζοντας την εμφάνιση clusters ή κοινοτήτων και τη ροή των πληροφοριών. Παραδοσιακά, αυτή η μορφή homophily οδηγεί στη δημιουργία μιας core-periphery δομής, όπου οι κόμβοι υψηλού degree συγκεντρώνονται στον πυρήνα του δικτύου και συνδέονται πυκνά μεταξύ τους, ενώ οι κόμβοι χαμηλού degree βρίσκονται στην περιφέρεια και είναι αραιά συνδεδεμένοι. Ωστόσο, η παρακάτω οπτικοποίηση του δικτύου υποδεικνύει ότι η κατανομή των κόμβων δεν είναι απόλυτα συγκεντρωμένη στον πυρήνα. Αντίθετα, παρατηρούνται ομάδες κόμβων που σχηματίζουν πιο διάσπαρτες μικρο-κοινότητες, οι οποίες δεν συνδέονται αποκλειστικά με έναν κεντρικό πυρήνα.

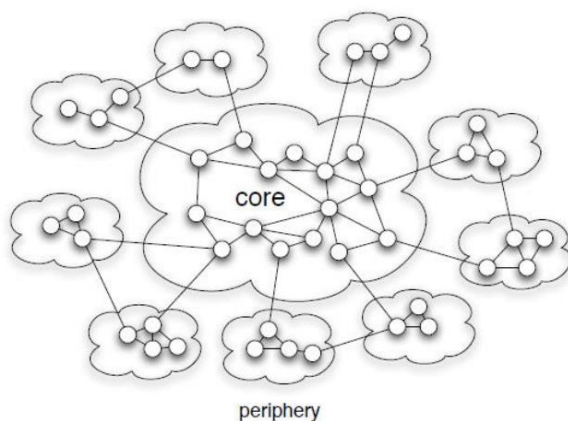


Figure 20.13: The core-periphery structure of social networks.

Ένα μεγάλο ποσοστό των κόμβων(73.75%) έχει μικρό degree, με τιμή ίση με 1. Αυτοί οι κόμβοι βρίσκονται στις πιο αραιές περιοχές του δικτύου. Οι κόμβοι υψηλού degree, όπως οι λογαριασμοί noahkeate.bsky.social, jamesfitzjournno.bsky.social, καιkelsesylh.bsky.social, λειτουργούν ως κεντρικοί κόμβοι ή hubs, παρότι δεν σχηματίζουν έναν απόλυτα πυκνό πυρήνα. Ετσι, το γράφημα εμφανίζει μια διασπορά των κόμβων σε clusters που συνδέονται χαλαρά μεταξύ τους, αντί για μια παραδοσιακή πυρήνα-περίμετρο.

Η παρακάτω οπτικοποίηση του δικτύου αντικατοπτρίζει αυτή τη δομή. Τα μεγαλύτερα labels αντιστοιχούν σε κόμβους με υψηλότερο degree, ενώ οι μικρότεροι και λιγότερο συνδεδεμένοι κόμβοι απλώνονται στις περιφερειακές περιοχές. Παρά την "χαοτική" εμφάνιση της δομής, διακρίνονται ορισμένες ομάδες.



10. Graph density

Σε αυτή την ενότητα, θα αναφερθούμε στην πυκνότητα-Graph Density, η οποία υπολογίζει την αναλογία των υπαρχόντων ακμών του δικτύου σε σχέση με τον μέγιστο δυνατό αριθμό ακμών που θα μπορούσαν να υπάρχουν, αν κάθε κόμβος συνδεόταν με όλους τους άλλους κόμβους. Ο υπολογισμός της πυκνότητας μας βοηθά να κατανοήσουμε πόσο συνδεδεμένο είναι το δίκτυο, δηλαδή πόσο "πυκνό" ή "αραιό" είναι.

Αρχικά, εκτελώντας τη λειτουργία Graph Density του Gephi για να υπολογίσουμε την πυκνότητα του δικτύου παρατηρούμε πως το αποτέλεσμα που προκύπτει είναι 0.000. Τεχνικά, αυτό σημαίνει ότι το δίκτυο δεν έχει ακμές. Ωστόσο, στην περίπτωση μας η πυκνότητα είναι τόσο χαμηλή που, όταν στρογγυλοποιείται στις 4 δεκαδικές θέσεις, το αποτέλεσμα εμφανίζεται λανθασμένα ως 0.

Για να ξεπεράσουμε αυτό το θέμα και να υπολογίσουμε την πραγματική πυκνότητα, χρησιμοποιούμε το μαθηματικό πρότυπο. Η Graph Density υπολογίζεται με τον εξής τύπο:

$$\text{Graph Density} = \frac{\text{Number of Edges}}{\text{Number of Nodes} * (\text{Number of Nodes} - 1)}$$

Όπου για το δίκτυό μας, με 8822 κόμβους και 17120 ακμές, η πυκνότητα υπολογίζεται ως 0.00022. Αυτός ο υπολογισμός μας δίνει μια πιο ακριβή εικόνα της πραγματικής πυκνότητας του δικτύου, υποδεικνύοντας ότι το δίκτυο είναι εξαιρετικά αραιό, γεγονός που αντανακλά στην τάση που ακολουθούν συνήθως τα κοινωνικά δίκτυα.

11. Community structure (modularity)

Σε αυτή την ενότητα, θα αναλύσουμε τη δομή των κοινοτήτων (Community structure) στο δίκτυό μας, όπως αυτή καθορίζεται μέσω της μετρικής Modularity.

Η Modularity μετράει τη δύναμη της διάκρισης του δικτύου σε ομάδες/κοιότητες. Κάθε ομάδα αντιπροσωπεύει μία ξεχωριστή κοινότητα. Η μετρική αυτή αξιολογεί πόσο καλά οι κόμβοι είναι ομαδοποιημένοι σε κοινότητες σε σύγκριση με μία τυχαία κατανομή των ακμών. Η τιμή της Modularity κυμαίνεται από -1 έως 1. Όταν η τιμή είναι 0 ή αρνητική, η δομή του δικτύου πλησιάζει την τυχαία κατανομή και δεν υπάρχουν ουσιαστικές κοινότητες, ενώ θετικές τιμές υποδεικνύουν την ύπαρξη ισχυρών κοινοτήτων.

Ο αλγόριθμος της Modularity χρησιμοποιεί ένα παγκόσμιο μοντέλο για να αποφασίσει πώς πρέπει να συνδέονται οι κόμβοι, αλλά τα πραγματικά δίκτυα συχνά ακολουθούν τοπικά μοτίβα, όπου οι κόμβοι αλληλεπιδρούν μόνο μέσα σε περιορισμένο "ορίζοντα". Αυτή η διαφορά προκαλεί προβλήματα όταν αναλύουμε πολύ μεγάλα δίκτυα. Συγκεκριμένα, ο αλγόριθμος Modularity υποθέτει ότι κάθε κόμβος θα μπορούσε να συνδεθεί με οποιονδήποτε άλλο, κάτι που δεν είναι ρεαλιστικό για μεγάλα δίκτυα. Καθώς το δίκτυο μεγαλώνει, ο αναμενόμενος αριθμός συνδέσεων μεταξύ δύο ομάδων κόμβων γίνεται πολύ μικρός. Αν απλώς μία ακμή συνδέει δύο ομάδες, η Modularity μπορεί να θεωρήσει ότι οι δύο ομάδες είναι ισχυρά συνδεδεμένες και να τις συγχωνεύσει σε μία κοινότητα, ακόμη και αν δεν είναι παρόμοιες. Αυτό οδηγεί στην απώλεια μικρών κοινοτήτων, καθώς αυτές συγχωνεύονται σε μεγαλύτερες.

Αυτή η προκατάληψη της μεθόδου είναι ο λόγος ύπαρξης μιας άλλης μετρικής, της Modularity with resolution. Σύμφωνα με τη λειτουργία του Gephi, τιμές ανάλυσης μεγαλύτερες από 1 οδηγούν στη συγχώνευση κοινοτήτων σε μεγαλύτερους τομείς, ενώ τιμές μικρότερες από 1 επιτρέπουν στον αλγόριθμο να εντοπίσει περισσότερες, μικρότερες κοινότητες.

Για να εντοπίσουμε τις κοινότητες στο δίκτυό μας, εκτελούμαι τη λειτουργία Modularity στο Gephi χρησιμοποιώντας Resolution = 0.8, ώστε να αποφύγουμε τη συγχώνευση μικρών κοινοτήτων που δεν πρέπει να συγχωνευθούν. Ο αλγόριθμος υπολόγισε την Modularity και τις κοινότητες με τις εξής τιμές:

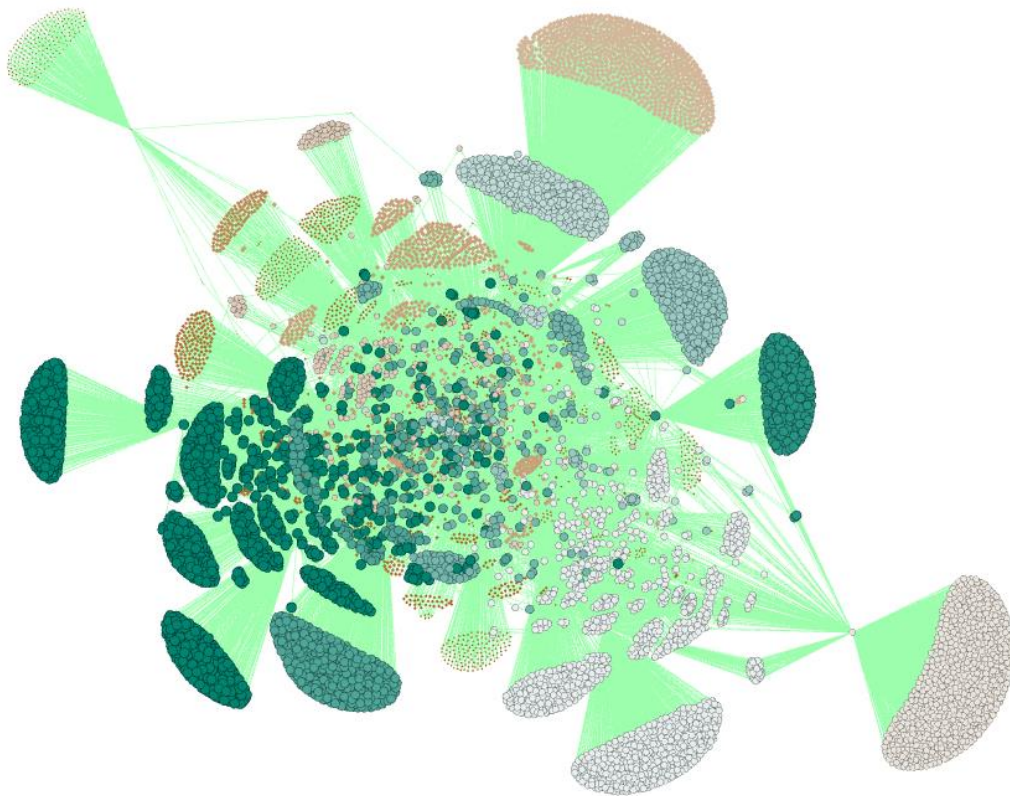
- Modularity: 0.590
- Modularity με Ανάλυση Ανάλυσης: 0.454
- Αριθμός Κοινοτήτων: 17

Αυτές οι τιμές μας δείχνουν ότι το δίκτυο περιλαμβάνει 17 διακριτές κοινότητες, με ισχυρές εσωτερικές συνδέσεις και αδύναμες συνδέσεις μεταξύ τους. Το αποτέλεσμα της Modularity (0.590) υποδηλώνει ότι οι κόμβοι εντός των κοινοτήτων αλληλεπιδρούν περισσότερο μεταξύ τους παρά με τους κόμβους άλλων κοινοτήτων, γεγονός που αποτελεί ένδειξη μέτριας συνοχής στις κοινότητες.

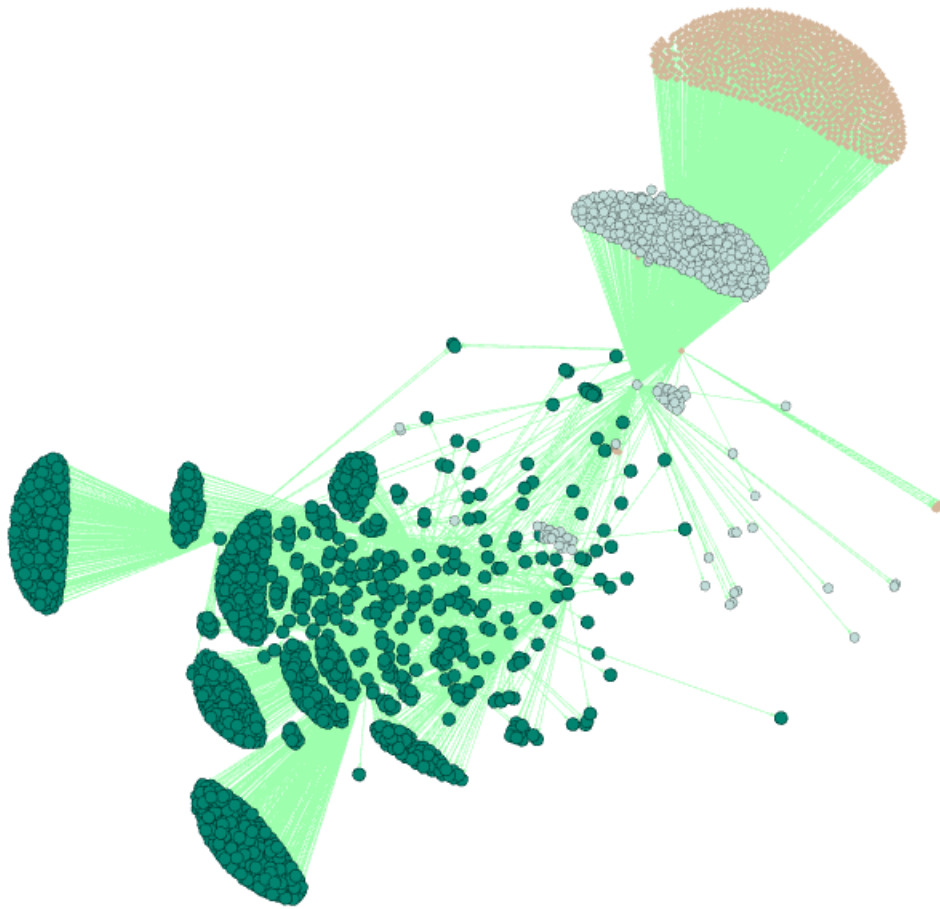
Οι 17 κοινότητες που εντοπίστηκαν κατανέμονται ως εξής, με τα αντίστοιχα ποσοστά των κόμβων που περιλαμβάνει η καθεμία:

1. Κοινότητα 16: 15.95%
2. Κοινότητα 10: 15.17%
3. Κοινότητα 9: 9%
4. Κοινότητα 7: 8.9%
5. Κοινότητα 1: 6.95%
6. Κοινότητα 14: 6.64%
7. Κοινότητα 6: 6.54%
8. Κοινότητα 11: 4.92%
9. Κοινότητα 12: 4.59%
10. Κοινότητα 2: 4.01%
11. Κοινότητα 13: 3.12%
12. Κοινότητα 15: 3.07%
13. Κοινότητα 8: 3.02%
14. Κοινότητα 4: 2.63%
15. Κοινότητα 0: 2.45%
16. Κοινότητα 3: 1.75%
17. Κοινότητα 5: 1.3%

Η παρακάτω απεικόνιση απεικονίζει το δίκτυο με τις 17 κοινότητες που εντοπίστηκαν, και το κάθε χρώμα αναπαριστά μια διαφορετική κοινότητα.



Μπορούμε να φιλτράρουμε κιάλας για να δούμε κάθε κοινότητα χωριστά. Για παράδειγμα στην παρακάτω εικόνα παρουσιάζονται 3 από τις 17 κοινότητες:



12. PageRank

Στην παρούσα ενότητα, θα εξετάσουμε τη μετρική PageRank για το δίκτυό μας, με στόχο να αξιολογήσουμε τη σημαντικότητα των κόμβων. Το PageRank είναι ένας αλγόριθμος που υπολογίζει τη σχετική επιρροή κάθε κόμβου σε ένα δίκτυο με βάση τις συνδέσεις του.

Σύμφωνα με το επίσημη τεκμηρίωση του Gephi, το PageRank είναι ένας επαναληπτικός αλγόριθμος που μετρά τη σημασία κάθε κόμβου μέσα στο δίκτυο. Ο αλγόριθμος υπολογίζει την πιθανότητα να βρίσκεται κάποιος σε έναν συγκεκριμένο κόμβο μετά από πολλές κινήσεις. Οι τιμές του PageRank αντιστοιχούν στα στοιχεία του ιδιοδιανύσματος που σχετίζεται με τη μέγιστη ιδιοτιμή ενός κανονικοποιημένου πίνακα γειτνίασης A' , όπου το άθροισμα των στοιχείων κάθε στήλης ισούται με 1.

Η τιμή PageRank κυμαίνεται από 0 έως 1, με υψηλότερες τιμές να υποδηλώνουν μεγαλύτερη σημασία.

Για τον υπολογισμό του PageRank στο δίκτυό μας, χρησιμοποιήσαμε την ενσωματωμένη λειτουργία του Gephi με τις ακόλουθες παραμέτρους:

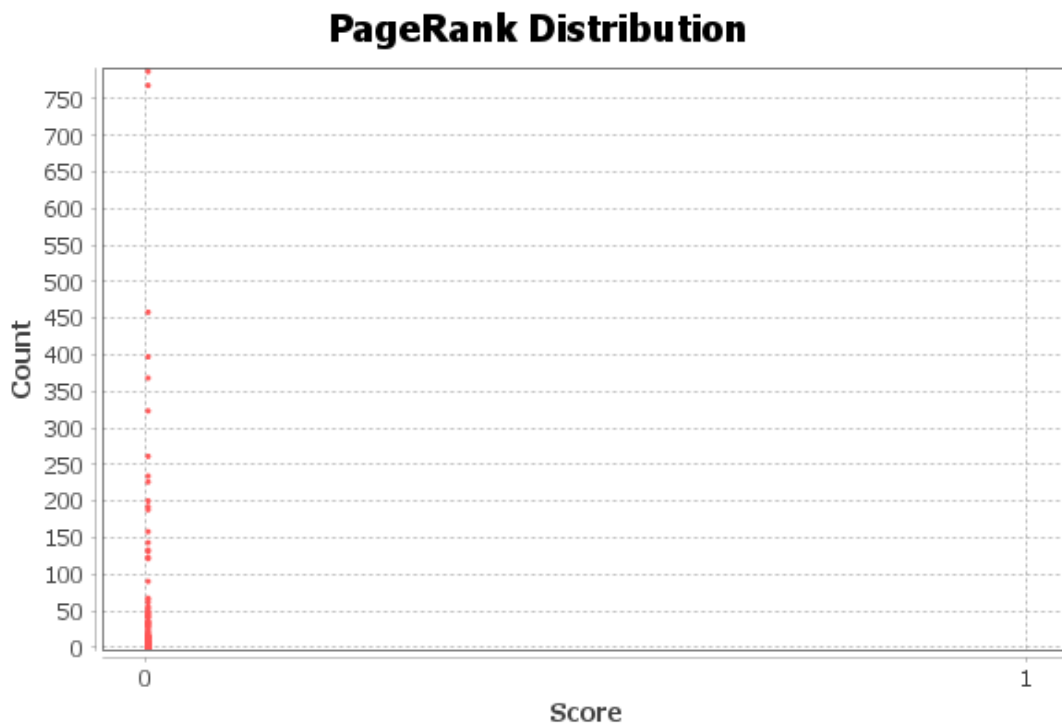
➤ **Probability (p): 0.85**

Αυτή η παράμετρος ελέγχει την πιθανότητα ο "τυχαίος περιηγητής" στον αλγόριθμο να συνεχίσει να ακολουθεί τις ακμές αντί να ξεκινήσει ξανά από έναν τυχαίο κόμβο. Η προεπιλεγμένη τιμή είναι 0.85.

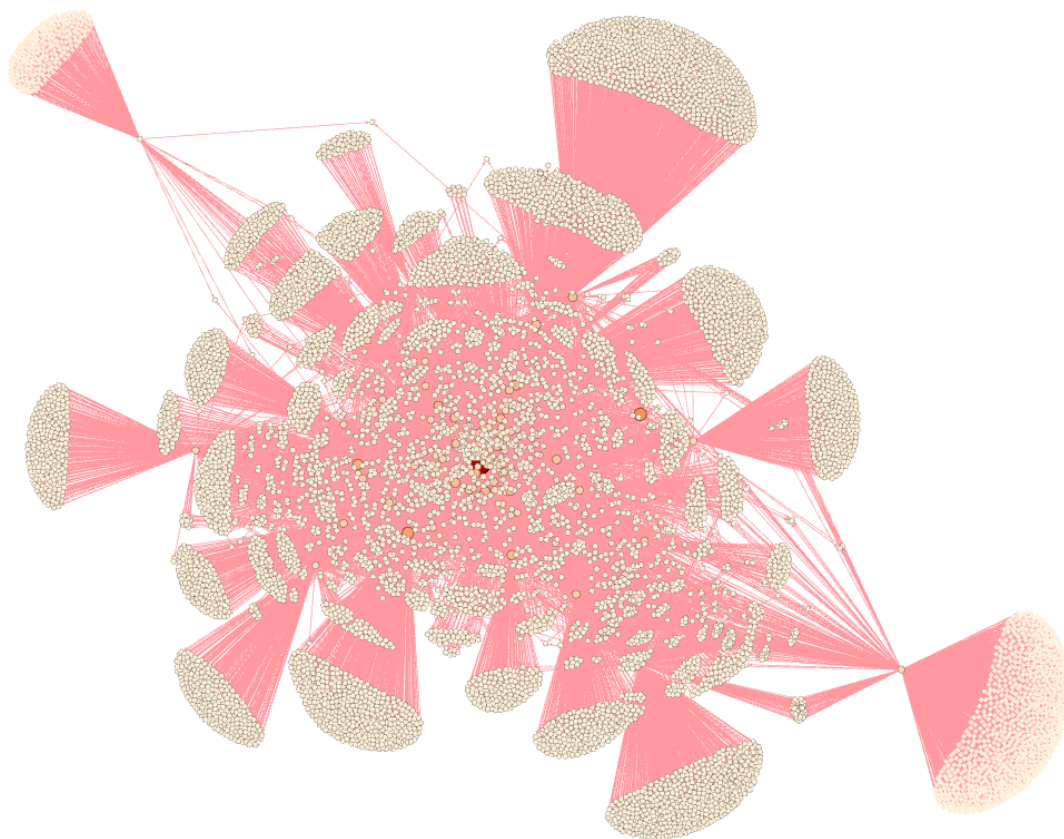
➤ **Epsilon: 0.001**

Αυτή η παράμετρος καθορίζει πότε θα σταματήσει ο αλγόριθμος PageRank. Όσο μικρότερη είναι η τιμή, τόσο περισσότερος χρόνος απαιτείται για τη σύγκλιση, καθώς ο αλγόριθμος προσπαθεί να προσεγγίσει όσο το δυνατόν περισσότερο τις πραγματικές τιμές του PageRank.

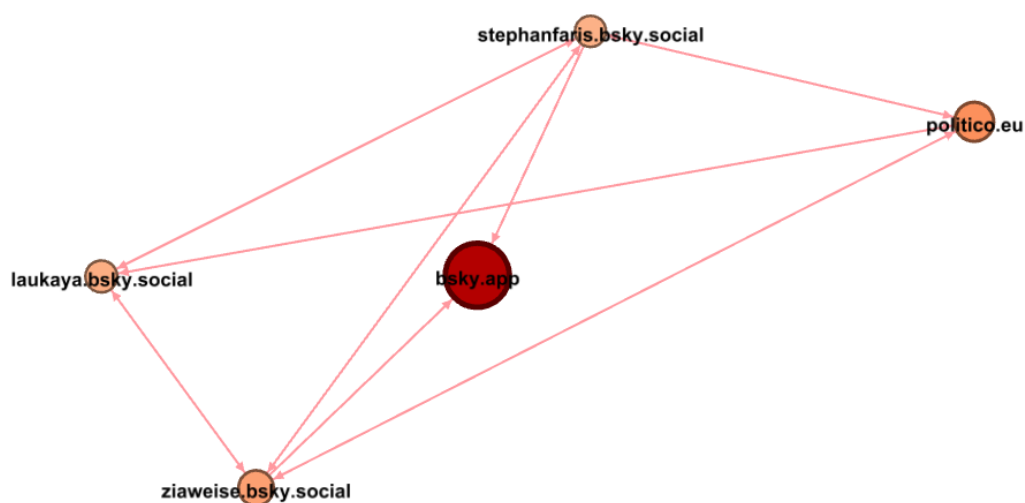
Εκτελώντας, λοιπόν, το PageRank έχουμε τα παρακάτω αποτελέσματα:



Για την καλύτερη κατανόηση θα δούμε και οπτικά την μορφή του δικτύου μας με το χρώμα και το μέγεθος των κόμβων να μεταβάλλεται βάση των τιμών του PageRank.



Οι τιμές του PageRank είναι όλες πάρα πολύ κοντά στο 0. Αυτό το αντιλαμβανόμαστε από τις παραπάνω απεικονίσεις αλλά και από το ακόλουθο διάγραμμα που περιλαμβάνει τους 5 κορυφαίους κόμβους με τη μεγαλύτερη τιμή PageRank. Συγκεκριμένα, οι τιμές τους κυμαίνονται από **1.6189E-4(0.00016189)** έως **2.6828E-4(0.00026828)**.



Οι τιμές του PageRank επηρεάζονται από διάφορους παράγοντες, οι οποίοι εξηγούν γιατί είναι τόσο χαμηλές στο δίκτυό μας. Για αρχή το δίκτυό μας είναι αραιό, με σχετικά λίγες ακμές ανάμεσα στους κόμβους. Σε τέτοιου είδους δίκτυα, οι πιθανότητες που υπολογίζει ο αλγόριθμος PageRank για κάθε κόμβο είναι πολύ μικρές, καθώς οι "διαδρομές" που μπορεί να ακολουθήσει ένας τυχαίος περιηγητής είναι περιορισμένες. Αυτό έχει ως αποτέλεσμα μικρές τιμές PageRank. Επιπλέον, η κανονικοποίηση του πίνακα γειτνίασης (adjacency matrix), που εξασφαλίζει ότι το άθροισμα κάθε στήλης ισούται με 1, διανέμει την πιθανότητα ισομερώς σε όλο το δίκτυο. Καθώς το δίκτυό μας έχει μεγάλο αριθμό κόμβων, η συνολική "σημασία" κατανέμεται σε πολλούς κόμβους, γεγονός που οδηγεί σε ακόμη μικρότερες τιμές. Η ύπαρξη μεγάλου αριθμού κόμβων με περιορισμένες συνδέσεις μειώνει τη συνολική δυνατότητα μεταφοράς "σημασίας" μέσω του δικτύου. Ακόμα και οι πιο σημαντικοί κόμβοι λαμβάνουν χαμηλές τιμές PageRank, αφού το συνολικό "βάρος" είναι περιορισμένο. Παρά τις χαμηλές απόλυτες τιμές, παρατηρούμε ότι ορισμένοι κόμβοι ξεχωρίζουν. Αυτοί είναι οι πιο "σημαντικοί" κόμβοι του δικτύου, καθώς δέχονται περισσότερες συνδέσεις από άλλους ή είναι συνδεδεμένοι με κόμβους υψηλής σημασίας. Αυτή η διαφοροποίηση είναι κρίσιμη για την κατανόηση της δυναμικής του δικτύου.

Οι κόμβοι με τις υψηλότερες τιμές PageRank παίζουν σημαντικό ρόλο στη ροή της πληροφορίας μέσα στο δίκτυο, ακόμα κι αν οι απόλυτες τιμές είναι χαμηλές. Αυτό υποδηλώνει ότι το PageRank δεν μετράει μόνο την ποσότητα των συνδέσεων, αλλά και την ποιότητά τους. Ένας κόμβος μπορεί να είναι σημαντικός επειδή συνδέεται με άλλους σημαντικούς κόμβους, ακόμα κι αν οι ίδιοι δεν έχουν πολλές άμεσες συνδέσεις.

Στο δίκτυό μας, όπως είδαμε παραπάνω, ο κόμβος **Politico.eu**, ο οποίος αποτελεί τον κύριο κόμβο του δικτύου μας, συγκαταλέγεται στην κορυφαία πεντάδα με τις μεγαλύτερες τιμές PageRank. Συγκεκριμένα, ο κόμβος με τη μεγαλύτερη τιμή PageRank είναι ο **bsky.app**, ενώ αμέσως μετά ακολουθεί ο **Politico.eu**. Αυτό δείχνει ότι, παρά την πιθανή διαφορά στον αριθμό συνδέσεων, ο Politico.eu διατηρεί μια ισχυρή επιρροή στο δίκτυο λόγω των σημαντικών του συνδέσεων.

13. Βιβλιογραφία

- Baeldung. (n.d.). Graph density. Retrieved January 31, 2025, from <https://www.baeldung.com/cs/graph-density>
- Bloom, D. (n.d.). Dan Bloom on BlueSky. <https://bsky.app/profile/danbloom1.bsky.social>
- BlueSky. (n.d.). BlueSky official profile. <https://bsky.app/profile/bsky.app>
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2), 163-177. <https://doi.org/10.1080/0022250X.2001.9990249>
- Faris, S. (n.d.). Stephan Faris on BlueSky. <https://bsky.app/profile/stephanfaris.bsky.social>
- Fitzgerald, J. (n.d.). James Fitzgerald on BlueSky. <https://bsky.app/profile/jamesfitzjournio.bsky.social>
- Gephi. (2015). PageRank. GitHub. <https://github.com/gephi/gephi/wiki/PageRank>
- Gephi. (2021, October 28). [Issue #2951: Closeness centrality produces wrong results]. GitHub. <https://github.com/gephi/gephi/issues/2951>
- Gephi. (n.d.). Average clustering coefficient. Retrieved January 31, 2025, from <https://github.com/gephi/gephi/wiki/Average-Clustering-Coefficient>
- Keate, N. (n.d.). Noah Keate on BlueSky. <https://bsky.app/profile/noahkeate.bsky.social>
- Kelsey, L. H. (n.d.). Kelsey L. H. on BlueSky. <https://bsky.app/profile/kelseylh.bsky.social>
- Kleinberg, J. (n.d.). Networks, crowds, and markets: Reasoning about a highly connected world (Chapter 3). Retrieved from <https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch03.pdf>
- Lambiotte, R., Blondel, V. D., de Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., & Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. arXiv. <https://arxiv.org/abs/0803.0476>
- Laukaya, B. (n.d.). Laukaya on BlueSky. <https://bsky.app/profile/laukaya.bsky.social>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- NetworkX. (n.d.). NetworkX documentation. Retrieved January 29, 2025, from <https://networkx.org/>
- Politico.eu. (n.d.). Politico Europe. <https://www.politico.eu/>

Pournarakis, D. (2025). Homophily and small worlds [PowerPoint slides]. Athens University of Economics and Business.
https://edu.dmst.aueb.gr/pluginfile.php/53704/mod_resource/content/0/Homophily%20and%20Small%20Worlds.pdf

ScienceDirect. (n.d.). Betweenness centrality. Retrieved January 31, 2025, from <https://www.sciencedirect.com/topics/computer-science/betweenness-centrality#:~:text=Betweenness%20centrality%20is%20defined%20as,their%20position%20in%20these%20paths>

ScienceDirect. (n.d.). Closeness centrality. Retrieved January 31, 2025, from <https://www.sciencedirect.com/topics/computer-science/closeness-centrality>

ScienceDirect. (n.d.). Degree centrality. Retrieved January 31, 2025, from <https://www.sciencedirect.com/topics/computer-science/degree-centrality>

ScienceDirect. (n.d.). Eigenvector centrality. In *Encyclopedia of Computer Science*. Retrieved January 29, 2025, from <https://www.sciencedirect.com/topics/computer-science/eigenvector-centrality#:~:text=Eigenvector%20centrality%20refers%20to%20the,other%20nodes%20and%20their%20centrality>

ScienceDirect. (n.d.). Giant component. In *Encyclopedia of Computer Science*. Retrieved January 29, 2025, from <https://www.sciencedirect.com/topics/computer-science/giant-component>

Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2), 146–160. <https://doi.org/10.1137/0201010>

Weise, Z. (n.d.). Zia Weise on BlueSky. <https://bsky.app/profile/ziaweise.bsky.social>

Wikipedia contributors. (n.d.). Politico Europe. Wikipedia. Retrieved January 31, 2025, from https://en.wikipedia.org/wiki/Politico_Europe