Εργασία στο μάθημα Ανάλυση Κοινωνικών Δικτύων για το ακαδημαϊκό έτος 2021-2022

Ραυτόπουλος Μάριος με αριθμό μητρώου 3180163 (φοιτητής τμήματος Πληροφορικής)

The Quaker Social Network

A picture containing person, indoor, group

Description automatically generated

ΠΊΝΑΚΑΣ ΠΕΡΙΕΧΟΜΈΝΩΝ:

1. ΣΎΝΟΨΗ

2. ΠΡΟΕΤΟΙΜΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

3. ΑΝΑΛΥΣΗ ΔΙΚΤΥΟΥ

4. ΕΠΊΛΟΓΟΣ

5. ΒΙΒΛΙΟΓΡΑΦΙΑ

ΚΕΦΑΛΑΙΟ 1 : ΣΥΝΟΨΗ

Ερευνούμε τη δομή του δικτύου των Κουάκερων.

Οι σχέσεις μεταξύ των Κουακέρων του 17ου αιώνα προσφέρονται για ανάλυση κοινωνικών δικτύων, διότι οι μελετητές έχουν συνδέσει εδώ και καιρό την ανάπτυξη και την αντοχή των Κουακέρων με την αποτελεσματικότητα των δικτύων τους:

Πριν υπάρξουν οι φίλοι στο Facebook, υπήρχε η Εταιρεία των Φίλων, γνωστή ως Κουάκεροι. Ιδρυμένοι στην Αγγλία στα μέσα του 17ου αιώνα, οι Κουάκεροι ήταν προτεστάντες χριστιανοί που διαφωνούσαν με την επίσημη Εκκλησία της Αγγλίας και προωθούσαν την ευρεία θρησκευτική ανεξιθρησκεία, προτιμώντας το υποτιθέμενο "εσωτερικό φως" και τη συνείδηση των χριστιανών από την επιβεβλημένη από το κράτος ορθοδοξία. Ο αριθμός των Κουάκερων αυξήθηκε ραγδαία στα μέσα και στα τέλη του 17ου αιώνα και τα μέλη τους εξαπλώθηκαν στις Βρετανικές Νήσους, στην Ευρώπη και στις αποικίες του Νέου Κόσμου.

ΚΕΦΑΛΑΙΟ 2: ΠΡΟΕΤΟΙΜΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Εργαλεία και μέθοδοι

1.

Σύνολα δεδομένων(datasets) από το GitHub: Ανοιχτού κώδικα πλατφόρμα για το κοινό.

Συγκεκριμένα πάρθηκαν δύο αρχεία: το quaker-nodes.csv και το quaker-edges.csv

Στη συνέχεια, αυτά τα δύο αρχεία γίνονται import στο project που θα δημιουργήσουμε στο Gephi, για να προχωρήσουμε μετέπειτα στην ανάλυση του κοινωνικού δικτύου.

2.

Gephi 0.9.2: Λογισμικό ανοικτού κώδικα για την οπτικοποίηση δικτύων και την ανάλυση των χαρακτηριστικών του δικτύου.

Το csv των κόμβων (quaker-nodes.csv) περιέχει 96 διαφορετικούς Κουάκερους, δηλαδή, oι κόμβοι που χρησιμοποιούνται είναι unimodal και αναφέρονται σε ένα τύπο κόμβου( αναπαριστούν Κουάκερους).

Το quaker-nodes.csv περιέχει επίσης πληροφορίες χαρακτηριστικών όπως "ιστορική σημασία", "φύλο", "ημερομηνία γέννησης" και "ημερομηνία θανάτου".

Το csv των ακμών (quaker-edges.csv) περιέχει 162 μη σταθμισμένες (unweighted) σχέσεις μεταξύ αυτών των χαρακτήρων, οι οποίες προέκυψαν υπολογιστικά μέσω του \*Oxford Dictionary of National Biography\* και επιβεβαιώθηκαν και επεκτάθηκαν μέσω των συνεισφορών πλήθους μελετητών. Προφανώς, οι ακμές είναι αμφίδρομες (undirected) διότι μας ενδιαφέρει απλά η σχέση των συγκεκριμένων χαρακτήρων.

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΔΙΚΤΥΟΥ

ΜΕΡΟΣ 1: Αρχική γραφική αναπαράσταση του δικτύου μας:

Chart

Description automatically generated

Το layout που επιλέξαμε είναι το Force Atlas με τα αντίστοιχα settings:

Table

Description automatically generated

ΜΕΡΟΣ 2: Βασικές τοπολογικές ιδιότητες

Ο Γράφος μας έχει 96 κόμβους και 162 ακμές. Οι κόμβοι αναπαριστούν τους Κουάκερους και οι ακμές τις μεταξύ τους σχέσεις .

Η διάμετρος ( το μεγαλύτερο shortest-path ανάμεσα σε όλα τα ζεύγη κόμβων) είναι ίση με 8

Η ακτίνα(προσδιορίζει την ελάχιστη εκκεντρότητα) είναι ίση με 4

Και το μέσο μήκος μονοπατιού ίσο με 3.37

Οι παραπάνω τιμές βρίσκονται μέσω της ενότητας statistics που παρέχει το Gephi και συγκεκριμένα κάνοντας run το network diameter που δίνει τα αποτελέσματα που αναφέρθηκαν. Ενδεικτικά παρέχεται το παρακάτω screenshot:

Graphical user interface, text, application, email

Description automatically generated

ΜΕΡΟΣ 3: Επιμέρους συνιστώσες

Μέσω της ενότητας filters που παρέχει το Gephi μπορούμε να εφαρμόσουμε Giant Component αλλά δε θα διαπιστώσουμε κάποια αλλαγή. Αυτό γίνεται διότι όλες οι οντότητες μας μπορούν να αλληλοεπιδράσουν μεταξύ τους και δεν εξαρτώνται από κάτι.

ΜΕΡΟΣ 4: Βαθμοί κόμβων

Chart, scatter chart

Description automatically generated

Από το διάγραμμα εξάγουμε τα εξής:

Μέσος βαθμός κόμβου = 3.375 , δηλαδή κάθε κόμβος επικοινωνεί κατά μέσο όρο με 3 άλλους κόμβους. Επίσης διαπιστώνουμε πως οι περισσότεροι κόμβοι επικοινωνούν με τουλάχιστον 1-2 κόμβους ενώ μόνο ένας επικοινωνεί με 22 που είναι και το μέγιστο

A picture containing necklet

Description automatically generated

Γράφημα με βάσει τον βαθμό του κόμβου

\*\* Παρατήρηση για τα γραφήματα που δημιουργήθηκαν από το Partition/Ranking: Σε όλα ισχύουν τα εξής ( εκτός του Modularity, Homophily):

Όσο πιο έντονο γίνεται το κόκκινο τόσο μεγαλύτερη αντίστοιχη τιμή έχει ο κόμβος

\*\*

ΜΕΡΟΣ 5: Μέτρα κεντρικότητας

Chart

Description automatically generated

Η κεντρικότητα betweenness μετράει όλες τις συντομότερες διαδρομές μεταξύ κάθε ζεύγους κόμβων του δικτύου και στη συνέχεια μετράει πόσες φορές ένας κόμβος βρίσκεται σε μια συντομότερη διαδρομή μεταξύ δύο άλλων. Στο διάγραμμα παρατηρούμε χαμηλές τιμές διότι δεν αλληλοεπιδρούν όλοι οι κομβόι μεταξύ τους εκτός από μερικούς με υψηλότερη τιμή όπως και ο κόμβος με το μέγιστο βαθμό(που αλληλοεπιδρά με τους περισσότερους).

Chart, radar chart

Description automatically generated

Γράφημα βάσει του Betweenness Centrality

Table

Description automatically generated

Η κεντρικότητα εγγύτητας είναι ένας τρόπος εντοπισμού των κόμβων που είναι σε θέση να διαδίδουν πληροφορίες πολύ αποτελεσματικά σε έναν γράφο. Η κεντρικότητα εγγύτητας ενός κόμβου μετρά τη μέση εγγύτητά του (αντίστροφη απόσταση) προς όλους τους άλλους κόμβους. Οι κόμβοι με υψηλό σκορ εγγύτητας έχουν τις μικρότερες αποστάσεις από όλους τους άλλους κόμβους. Η κατανομή του διαγράμματος ακολουθεί την κανονική κατανομή καθώς βλέπουμε τιμές στο διάστημα [0, 1]. Προφανώς δεν λαμβάνονται οι ακραίες τιμές καθώς όλοι οι κόμβοι ανήκουν σε κάποιο shortest path , όχι όμως σε όλα .

A picture containing necklet, accessory

Description automatically generated

Γράφημα με βάση το closeness centrality

A picture containing chart

Description automatically generated

Το μέτρο της εκκεντρότητας αποτυπώνει την απόσταση μεταξύ ενός κόμβου και του κόμβου που απέχει περισσότερο από αυτόν. Άρα, μια υψηλή εκκεντρότητα σημαίνει ότι ο πιο απομακρυσμένος κόμβος στο δίκτυο είναι πολύ μακριά, ενώ μια χαμηλή εκκεντρότητα σημαίνει ότι ο πιο απομακρυσμένος κόμβος είναι στην πραγματικότητα αρκετά κοντά.

Όπως φαίνεται στο διάγραμμα, επιβεβαιώνεται και η τιμή της διαμέτρου αφού δεν υπάρχει απόσταση μεγαλύτερη από 8 μεταξύ όλων των κόμβων.

Chart, radar chart

Description automatically generated

Γράφημα βάσει Eccentricity

Chart

Description automatically generated with medium confidence

Η κεντρικότητα ιδιοδιανύσματος είναι ένας δείκτης κεντρικότητας που υπολογίζει την κεντρικότητα ενός φορέα όχι μόνο με βάση τις συνδέσεις του, αλλά και με βάση την κεντρικότητα των συνδέσεων του φορέα αυτού. Η σημαντικότητα αυτού του ελέγχου δε φαίνεται καλά στο δίκτυο μας διότι το άτομο με το μεγαλύτερο σκορ στο διάγραμμα είναι και αυτό με τις περισσότερες συνδέσεις, πράγμα το οποίο θα μπορούσε να μην ισχύει (πχ αν είχαμε υποομάδες στο δίκτυο μας που επικοινωνούσαν όλες με ένα κόμβο αλλά πρακτικά δεν επηρεάζουν το σύνολο οπότε και ο υποτιθέμενος κόμβος του παραδείγματος δεν θα είχε μεγάλο σκορ ακόμα και να επικοινωνούσε με όλους αυτούς τους κόμβους επειδή αυτοί δε θα επικοινωνούσαν με πολλούς ή/και καθόλου άλλους κόμβους.)

A picture containing necklet, accessory

Description automatically generated

Γράφημα σύμφωνα με το Eigenvector Centrality

ΜΕΡΟΣ 6: Φαινόμενα ομαδοποίησης στο δίκτυο

Chart

Description automatically generated with medium confidence

Ο συντελεστής ομαδοποίησης δείχνει πόσο καλά συνδεδεμένη είναι η γειτονιά του κόμβου. Εάν η γειτονιά είναι πλήρως συνδεδεμένη, ο συντελεστής ομαδοποίησης είναι 1, ενώ μια τιμή κοντά στο 0 σημαίνει ότι δεν υπάρχουν σχεδόν καθόλου συνδέσεις στη γειτονιά. Παρατηρούμε πως οι ‘γειτονιές’ των κόμβων δεν είναι ισχυρά συνδεδεμένες αφού οι τιμές της clustering coefficient συγκεντρώνονται κάτω από 0.5, κάτω από το μισό του διαστήματος δηλαδή , το οποίο φαίνεται και από το μέσο όρο που είναι 0.403.

Παρατηρούνται 59 τρίγωνα , τα οποία σίγουρα δεν είναι πολλά σύμφωνα με τις ακμές μας που είναι 162, καθώς και αφορούν ελάχιστους κόμβους σε σχέση με το υπόλοιπο σύνολο.(τρίγωνα εμφανίζονται κυρίως γύρω από το κόμβο με το μέγιστο βαθμό , κάτι που είναι λογικό και διαμορφώνει ισχυρή γειτονιά πράγμα όμως που δε συμβαίνει στο υπόλοιπο ποσοστό των κόμβων)

Παρακάτω δείχνουμε τους κόμβους με μηδενικό clustering coefficient που δεν σχηματίζουν τρίγωνα χωρίς απαραίτητα να έχουν μικρό βαθμό

Table

Description automatically generated

Παρακάτω δείχνουμε τους κόμβους με clustering coefficient = 1 που σχηματίζουν τρίγωνα χωρίς απαραίτητα να έχουν μεγάλο βαθμό

Table

Description automatically generated

Παρατήρηση:

Μας παρέχετε επίσης και στήλη που δείχνει ακριβώς πόσα τρίγωνα σχηματίζονται ανά κόμβο που τα περισσότερα είναι στους κόμβους που εμφανίζουν τιμή σχεδόν ίση με το average clustering coefficient .

A picture containing necklet, accessory

Description automatically generated

Γράφημα σύμφωνα με το clustering coefficient

ΜΕΡΟΣ 7: Γέφυρες

Αφού εγκαταστήσουμε το κατάλληλο plugin(Bridging Centrality (BriCe)) ‘τρέχουμε’ τον έλεγχο Bridging Centrality

Table

Description automatically generated

Παρατηρούμε χαμηλές τιμές, κάτι που είναι λογικό αφού στο γράφο μας δεν έχουμε υποκοινότητες που να ενώνονται δηλαδή μεγάλες ομάδες κόμβων. Το γράφημα μας αποτελείται από μια μόνο συνιστώσα, το οποίο διαπιστώθηκε και με την εφαρμογή του φίλτρου Giant Component.

A picture containing chart

Description automatically generated

Γράφημα με βάσει το Bridging Centrality.

Παρατηρούμε κόμβους με εντονότερο κόκκινο, αλλά οι υποομάδες που ‘ενώνονται’ έχουν μικρό αριθμό κόμβων γεγονός που δικαιολογεί και τις μικρές τιμές του ελέγχου.

ΜΕΡΟΣ 8: Φύλο και ομοιογένεια

Αφού εγκαταστήσουμε το κατάλληλο plugin (Circular Layout) επιλέγουμε Radial Axis Layout και στο group nodes by επιλέγουμε το Modularity Class και δημιουργούμε το παρακάτω γράφημα.

A picture containing chart

Description automatically generated

Όπως φαίνεται και στο διάγραμμα οι κόμβοι μας έχουν την τάση να δημιουργούν σχέσεις σχεδόν αποκλειστικά με την κοινότητά τους με ελάχιστες εξαιρέσεις.

ΜΕΡΟΣ 9: Πυκνότητα Γράφου

A picture containing text

Description automatically generated

Η πυκνότητα γραφήματος αντιπροσωπεύει την αναλογία μεταξύ των ακμών που υπάρχουν σε ένα γράφημα και του μέγιστου αριθμού ακμών που μπορεί να περιέχει το γράφημα. Εννοιολογικά, παρέχει μια ιδέα για το πόσο πυκνός είναι ένας γράφος όσον αφορά τη συνδεσιμότητα ακμών.

Στο γράφο μας έχουμε πυκνότητα 0.036 , αρκετά μικρός αριθμός δηλαδή που δείχνει ότι οι κόμβοι μας δεν έχουν ισχυρή σύνδεση . Θα προτιμούσαμε πυκνότητα κοντά στην τιμή ‘1’ που είναι και η μέγιστη δυνατή αλλά αυτό συμβαίνει μόνο αν επικοινωνούσαν όλοι οι κόμβοι με όλους .

ΜΕΡΟΣ 10: Δομή της κοινότητας

Table

Description automatically generated

Παρατηρούμε 7 κοινότητες με τους περισσότερους κόμβους να ανήκουν στην 4η κοινότητα. Επιβεβαιώνεται και από το παρακάτω γράφημα

Chart, radar chart

Description automatically generated

Γράφημα βάσει Modularity Class

ΜΕΡΟΣ 11: PageRank

Table

Description automatically generatedΤο PageRank μετράει τη σημασία των ιστοσελίδων χρησιμοποιώντας τη δομή του δικτύου υπερσυνδέσμων του διαδικτύου. Και η βασική ιδέα, είναι ότι το PageRank θα αποδίδει μια βαθμολογία σπουδαιότητας σε κάθε κόμβο. Και η υπόθεση που κάνει, είναι ότι σημαντικοί κόμβοι είναι αυτοί που έχουν πολλούς εσωτερικούς συνδέσμους από σημαντικές σελίδες ή σημαντικούς άλλους κόμβους.

Λειτουργεί καλύτερα για δίκτυα που έχουν κατευθυνόμενες ακμές. Στην πραγματικότητα, οι σημαντικές σελίδες είναι εκείνες που έχουν πολλούς εσωτερικούς συνδέσμους από πιο σημαντικές σελίδες.

Στην περίπτωσή μας δεν έχουμε κατευθυνόμενες ακμές . Οι παράμετροι είναι οι default που δίνει το Gephi. Το epsilon είναι κριτήριο για τερματισμό του αλγορίθμου και η πιθανότητα αντιπροσωπεύει την πιθανότητα ένας χρήστης που θα κλικάρει έναν τυχαίο σύνδεσμο να συνεχίσει να κλικάρει συνδέσμους πράγμα που σημαίνει πως οι κόμβοι αν ήταν σελίδες στο διαδίκτυο δεν έχουν μεγάλη πιθανότητα να κλικαριστούν.

Chart, radar chart

Description automatically generated

Γράφημα βάσει PageRank

ΚΕΦΑΛΑΙΟ 4: ΕΠΙΛΟΓΟΣ

Η ανάλυση δικτύου είναι χρήσιμη σε πολλές εργασίες εφαρμογών που καθορίζουν την σημερινή κοινωνία . Μας βοηθά στη βαθιά κατανόηση της δομής μιας σχέσης στα κοινωνικά δίκτυα, μιας δομής ή μιας διαδικασίας αλλαγής στα φυσικά φαινόμενα ή ακόμη και στην ανάλυση των βιολογικών συστημάτων των οργανισμών.

Με την ανάλυση του παραπάνω δικτύου μαθαίνουμε σε βαθύτερο επίπεδο τον τρόπο οργάνωσης των actors του δικτύου , των ρόλων τους και των σχέσεων τους. Επίσης φαίνεται ότι η ανάλυση κοινωνικών δικτύων είναι ικανή να μας δώσει πληροφορίες όχι μόνο για fictional δίκτυα αλλά και πραγματικά ειδικά και σε μια εποχή που δεν υπήρχαν τα social media.

ΚΕΦΑΛΑΙΟ 5: ΒΙΒΛΙΟΓΡΑΦΙΑ

[1]: Σημειώσεις διδάσκοντα

[2]: <https://gephi.org/users/> Εγκατάσταση του εργαλείου και χρήση του manual

[3]: <https://towardsdatascience.com/network-analysis-d734cd7270f8> και

<https://en.wikipedia.org/wiki/Centrality> για εξακρίβωση εννοιών.

[4]: Χρήση online-streaming πλατφόρμας για μια πιο πρακτική ανάλυση του εργαλείου Gephi.

[5]:Χρήση GitHub για άντληση dataset <https://github.com/melaniewalsh/sample-social-network-datasets/tree/master/sample-datasets/quakers>

[6]: Ο σύνδεσμος της εικόνας βρίσκεται με ένα απλό γκουγκλάρισμα Quakers/images