Ανάλυση Κοινωνικού Δικτύου Twitter με NodeXL και Kumu

Φράνκο Καρόλι

A.M 3119178

Τμήμα Ψηφιακών Συστημάτων, Πανεπιστήμιο Θεσσαλίας,
Γαιόπολις, Λάρισα, Ελλάδα

fkaroli@uth. gr

www.ds.uth.gr

1. Εισαγωγή

Στις μέρες μας, τα μέσα κοινωνικής δικτύωσης είναι ένα αναπόσπαστο κομμάτι της ζωής μας. Μεγάλες εταιρίες όπως η Google, το Facebook, η Microsoft και το Amazon, παράγουν καθημερινά πληροφορία όγκου 44 zettabytes(10^{13} gigabytes), καθώς οι χρήστες δεν καταναλώνουν απλά τα δεδομένα, αλλά παράγουν κι εκείνοι εξίσου. Ένα μεγάλο μέρος όλων αυτόν των πληροφοριών παραμένει ανεκμετάλλευτο. Για αυτόν τον λόγο, έχουν δημιουργηθεί εργαλεία με το οποία μπορούμε να αναλύσουμε αυτά τα δεδομένα και να τα οπτικοποιήσουμε ώστε να έχουμε μια καλύτερη, λιγότερο χαοτική, εικόνα. Όπως θα δούμε και στη συνέχεια, έχοντας στην κατοχή μας κάποια δεδομένα μπορούμε να σχηματίσουμε δίκτυα και να εξάγουμε συμπεράσματα.

Γενικά ο όρος δίκτυο μπορεί να αναφέρεται σε οποιαδήποτε αλληλοσυνδεόμενη ομάδα η σύστημα. Υπάρχουν διαφορετικοί τύποι δικτύων, μεταξύ των οποίων, τα δίκτυα επιχειρήσεων, τα νομικά, τα δίκτυα ξενοδοχείων και φυσικά τα κοινωνικά δίκτυα. Ένα κοινωνικό δίκτυο είναι μια κοινωνική δομή αποτελούμενη από κόμβους (άτομα ή επιχειρήσεις) οι οποίοι συνδέονται μεταξύ τους με κάποιου τύπου αλληλεξάρτηση (φιλία, συναλλαγή, συγγένεια κ.α). Είναι δηλαδή μια τεχνική συσχέτισης των κόμβων για τη μέτρηση και την απεικόνιση των σχέσεων και των ροών ανάμεσα σε ανθρώπους, ομάδες, επιχειρήσεις ή οποιαδήποτε άλλες μονάδες επεξεργασίας γνώσης και πληροφορίας.

2. Μεθοδολογία

Αρχικά, τα δεδομένα για την ανάλυση του κοινωνικού δικτύου συγκεντρώθηκαν από τον ιστοχώρο κοινωνικής δικτύωσης Twitter, που ιδρύθηκε το 2006 και επιτρέπει στους χρήστες του να στέλνουν και να διαβάζουν σύντομα μηνύματα, τα οποία ονομάζονται “tweets”. Όλοι μπορούν να διαβάσουν tweets στην πλατφόρμα αλλά μόνο οι εγγεγραμμένοι χρήστες μπορούν να πατήσουν “like” ή να κοινοποιήσουν.

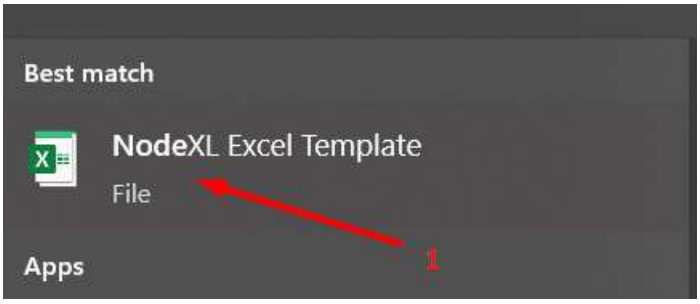
Για να «τραβήξουμε» τα δεδομένα χρειαζόμαστε ένα twitter λογαριασμό.

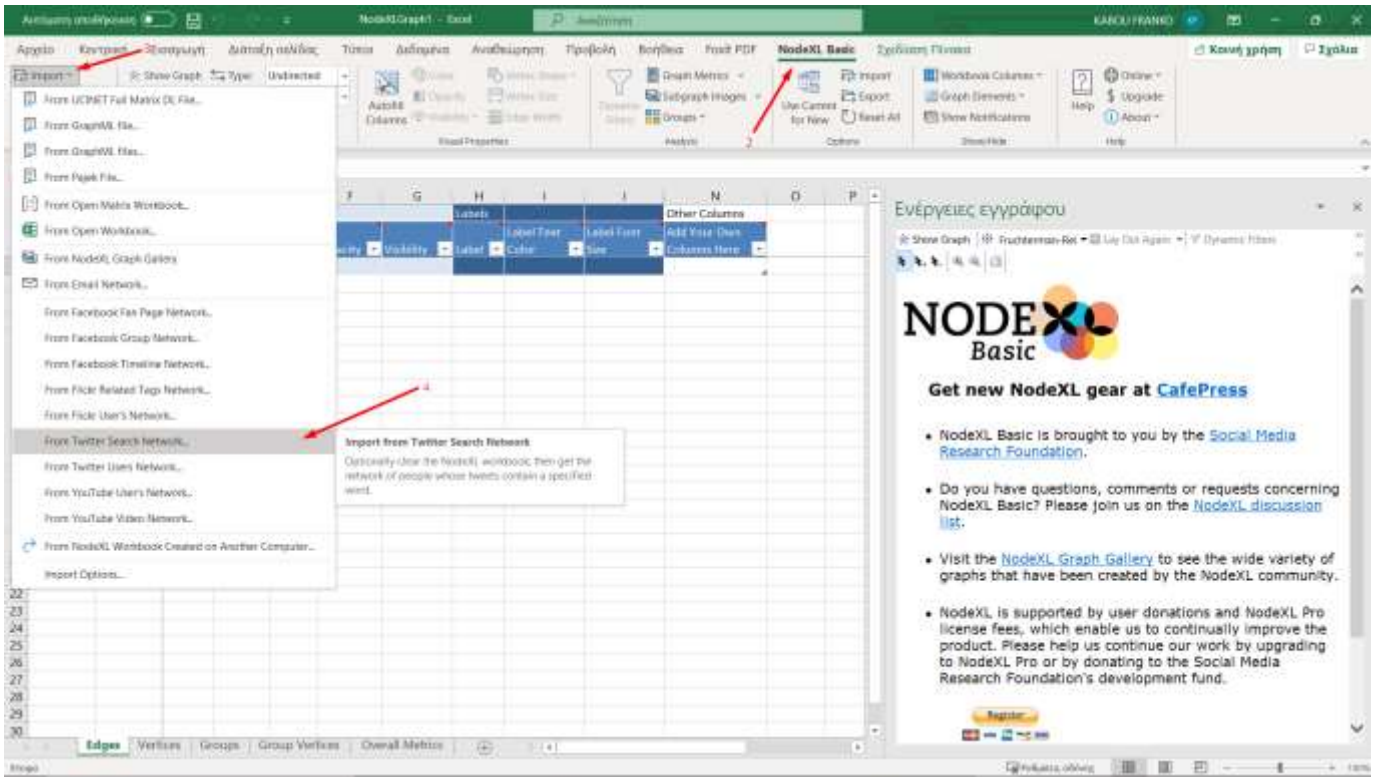
Οπότε φτιάχνουμε λογαριασμό στο https://twitter.com/i/flow/signup.

Τα δεδομένα μπορεί να τα παρέχει το twitter αλλά χρειαζόμαστε και κάποιο εργαλείο ώστε να μπορέσει να ομαδοποιήσει αυτά τα δεδομένα. Αυτή την διαδικασία θα την κάνουμε με το εργαλείο Node XL το οποίο είναι ένα plugin για το Excel της Microsoft.

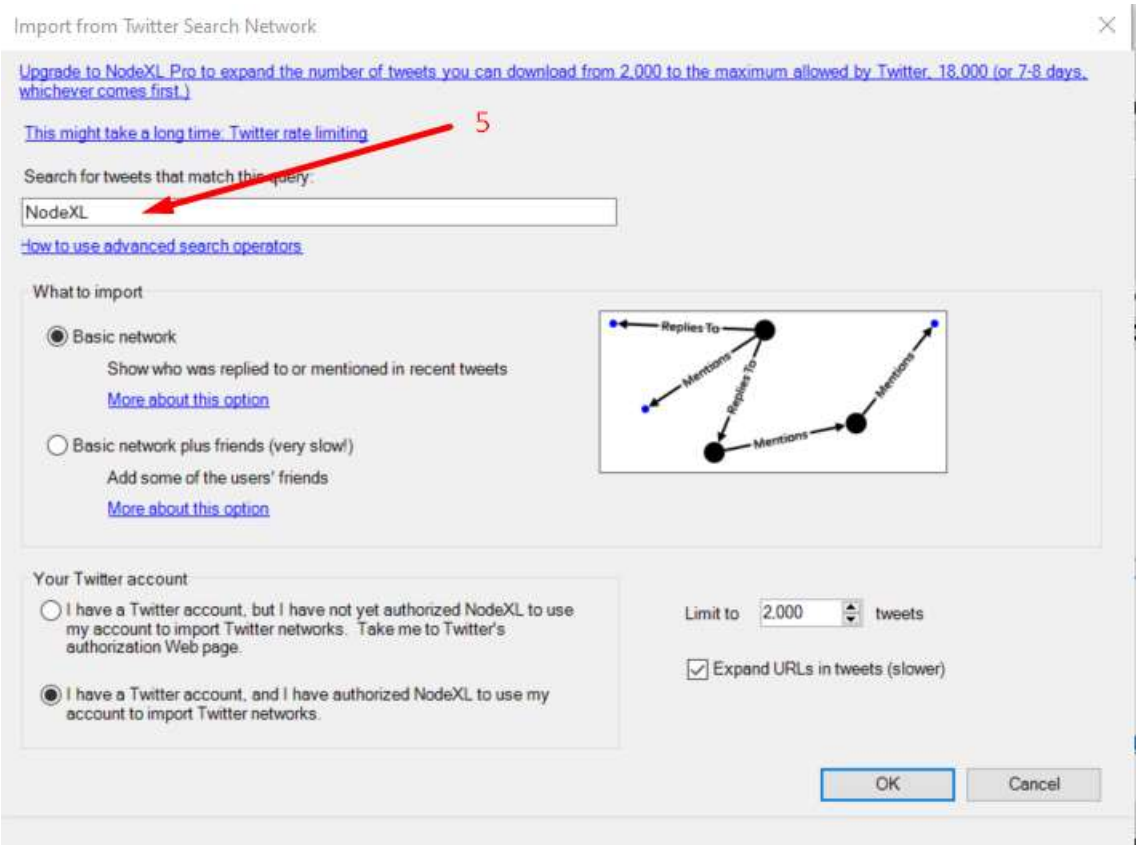
Μπαίνουμε λοιπόν στην ιστοσελίδα του NodeXL https://nodexl.com/ και κατεβάζουμε την δωρεάν έκδοση. Θα μας ζητήσει κάποια στοιχεία και ύστερα θα μας στείλει έναν σύνδεσμο μέσω email για να κατεβάσουμε το εργαλείο.

Αφού ανοίξουμε την εφαρμογή(1), πατάμε στην ετικέτα NodeXL Basic(2) > import(3) > from Twitter Search Network(4).

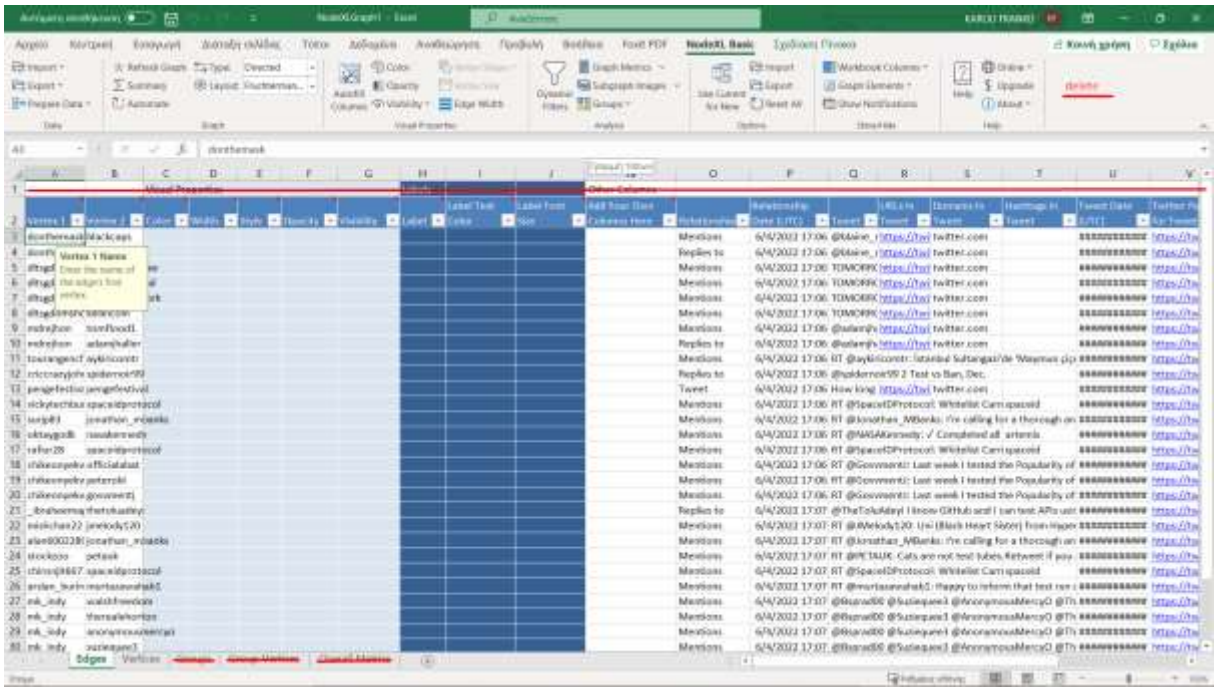




Αφού αποδεχτούμε τους όρους που εμφανίζει ο λογαριασμός μας στο twitter θα εμφανιστεί το παρακάτω παράθυρο όπου εκεί αποφασίζουμε το hashtag που μας ενδιαφέρει(5) καθώς και το πλήθος των tweets (περισσότερα tweets = περισσότερος χρόνος αναμονής) ,με μέγιστο όριο τα 2000, για την δωρεάν έκδοση, όπου και πατάμε «Ok».

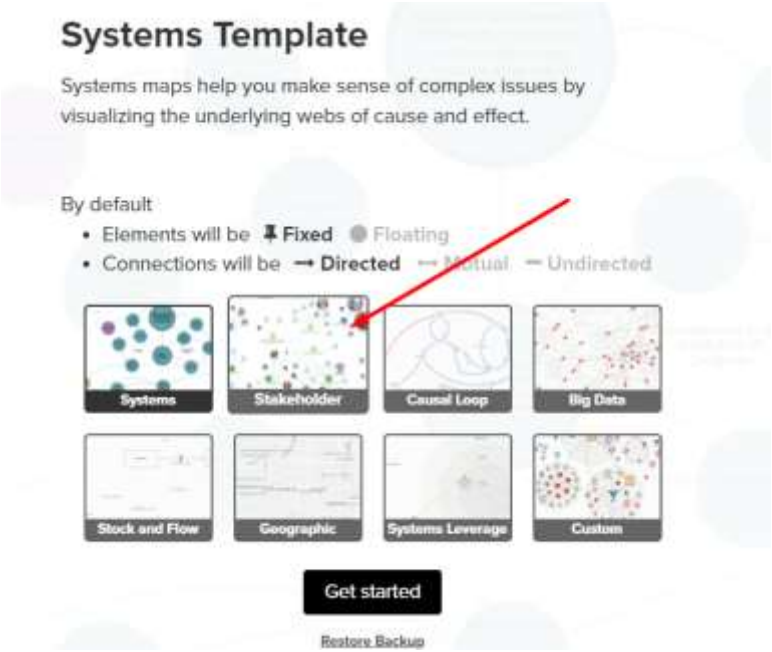


Αφού πατήσουμε «Ok», σύντομα θα εμφανιστούν κάποια δεδομένα με την παρακάτω μορφή, όπου θα διαγράψουμε(Δεξί κλίκ > Διαγραφή) τα σημεία που είναι με κόκκινη γραμμή καθώς είναι περιττά για την ανάλυσή μας. Αποθηκεύουμε το έγγραφο.

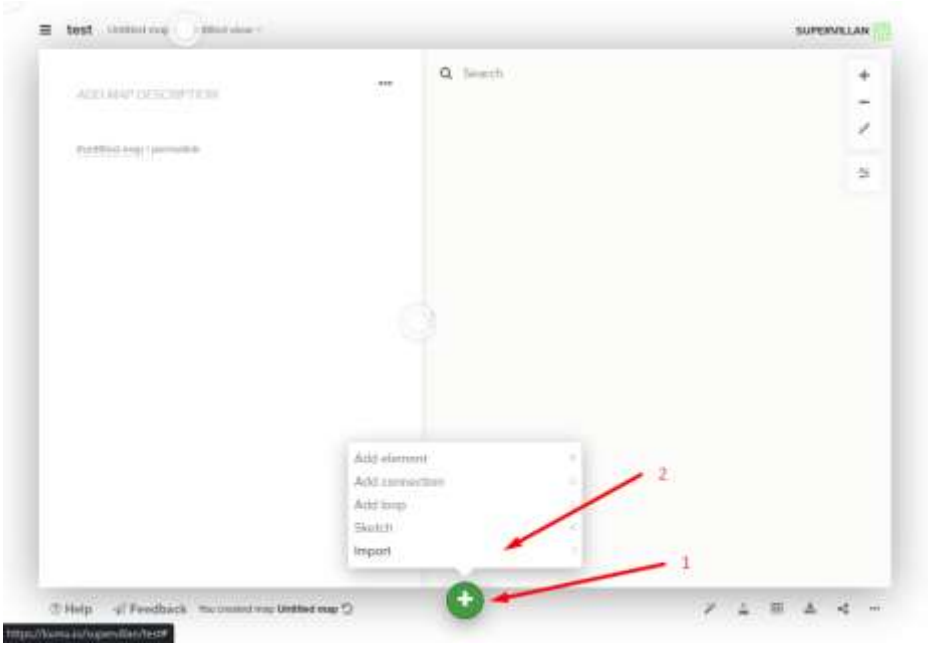


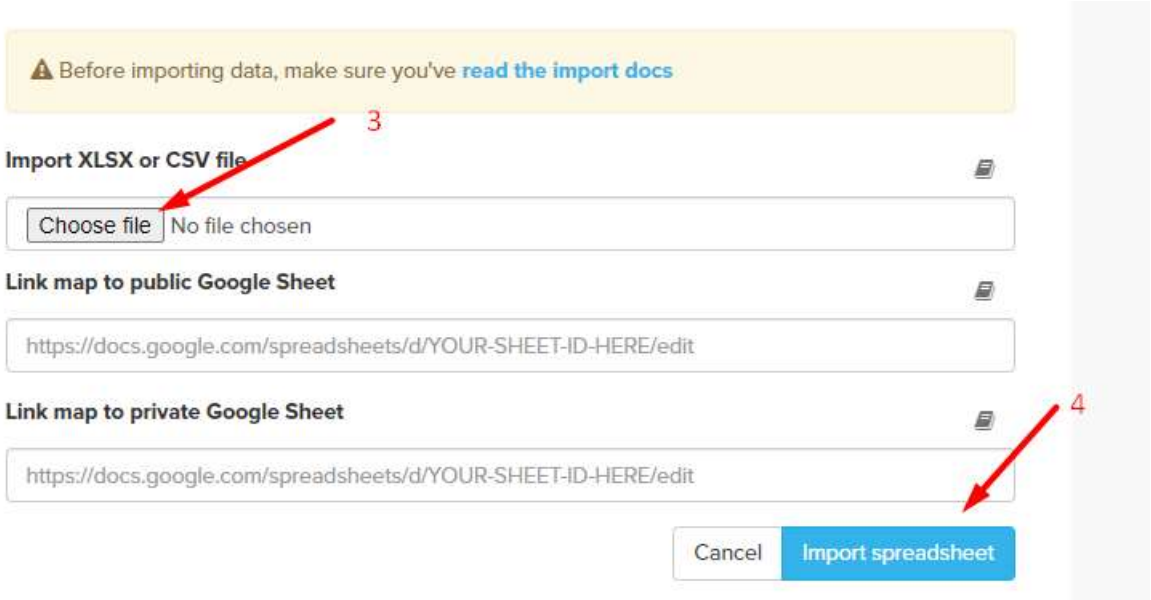
Τώρα που έχουμε τα δεδομένα με την μορφή που τα θέλουμε, θα χρησιμοποιήσουμε μια web εφαρμογή ανοιχτού κώδικα, το Kumu. Η εφαρμογή αυτή μας δίνει έτοιμο ένα template στο οποίο μπορούμε να βάλουμε τα δεδομένα μας για να ξεκινήσουμε την ανάλυση.

Έχοντας ανοίξει λοιπόν την web εφαρμογή Kumu (https://kumu.io/), δημιουργούμε λογαριασμό και ξεκινάμε ένα νέο project. Επιλέγουμε όνομα και περιγραφή και πατάμε δημιουργία. Επιλέγουμε το template Stakeholder και μετά Get Started.



Επιλέγουμε το πράσινο «+»(1) > import(2) > “Choose File”(3) όπου και επιλέγουμε το αρχείο μας, και “Import Spreadsheet”(4).

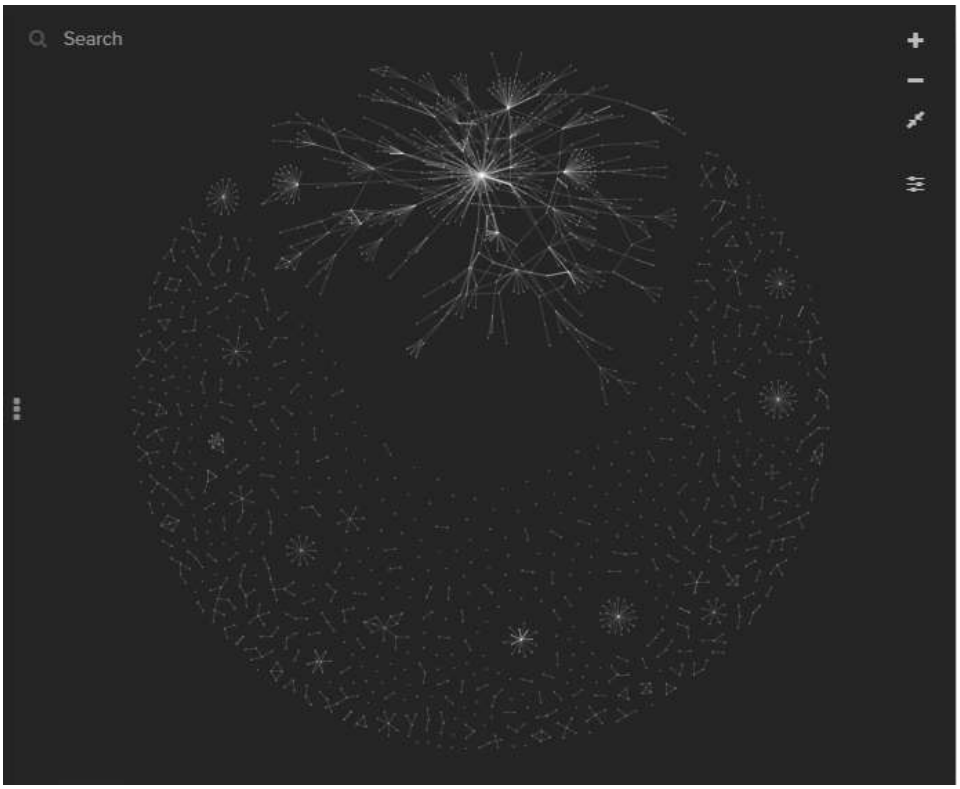




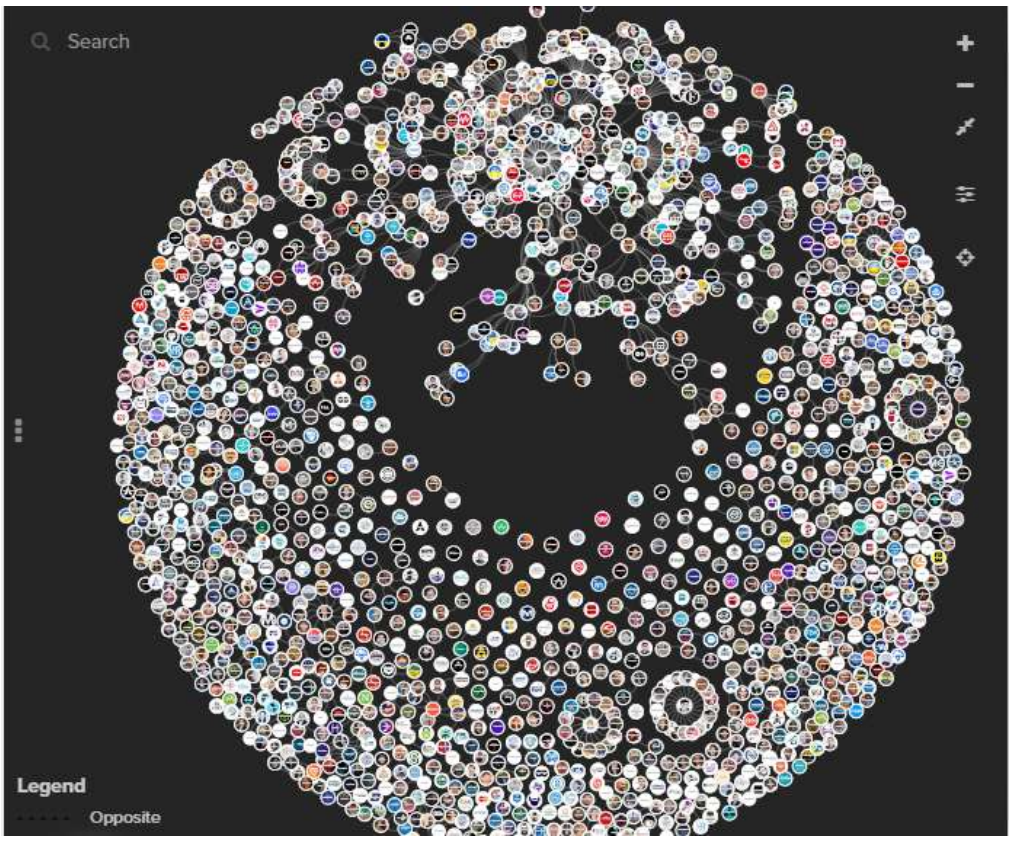
Αμέσως θα γίνει ένα review όπου θα πατήσουμε “save import” και θα εμφανιστεί το γράφημα που θέλουμε.

3. Αποτελέσματα

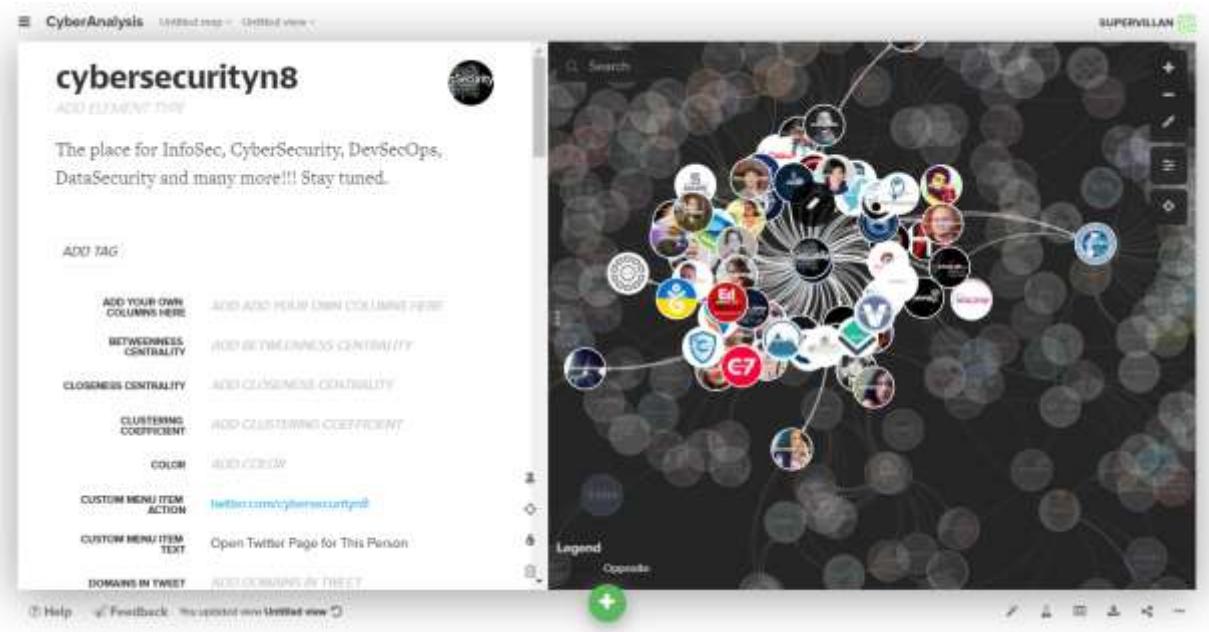
Έχοντας δεδομένα από το twitter τα οποία «τραβήχτηκαν» με το NodeXL και οπτικοποιήθηκαν με την web εφαρμογή Kumu, έχουμε το παρακάτω γράφημα σαν αποτέλεσμα.



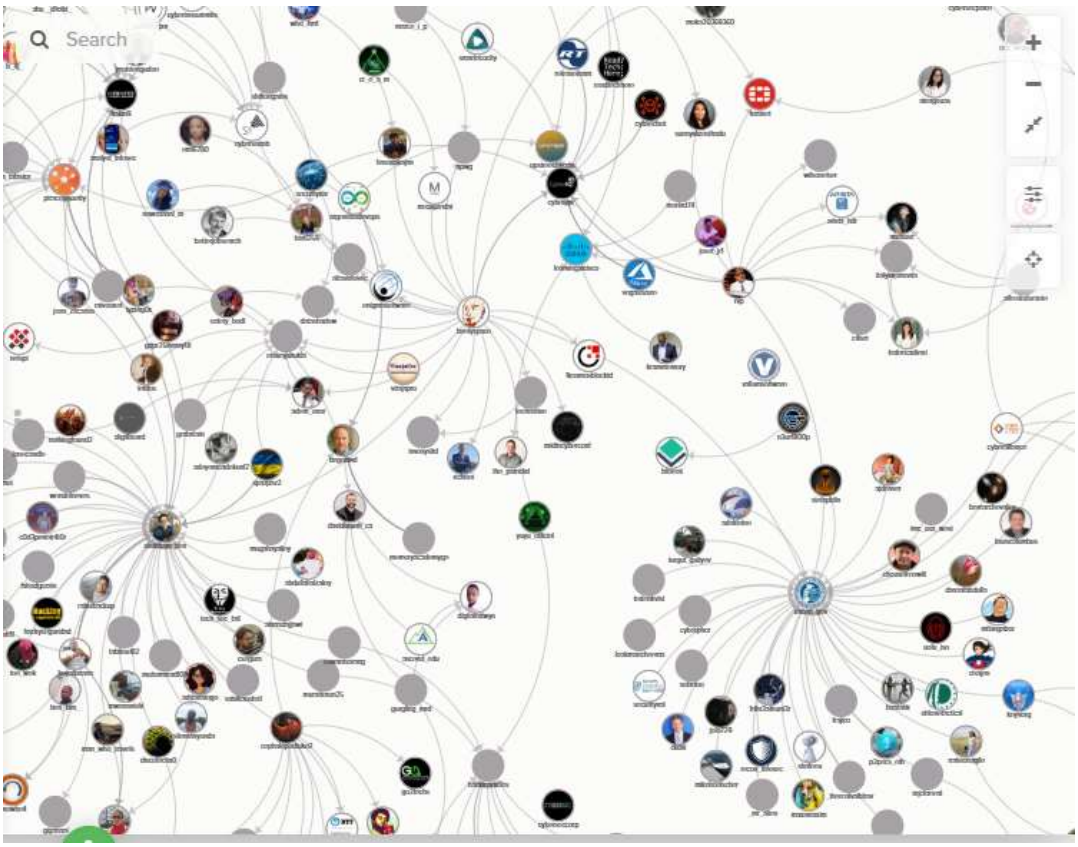
Πατώντας Settings > General Settings και αλλάζοντας την τιμή Template από “Social Network Analysis” σε “Systems” έχουμε το λίγο διαφοροποιημένο ως προς την εμφάνιση γράφημα.



Κάνοντας zoom και βάζοντας τον κέρσορα σε έναν κεντρικό κόμβο, βλέπουμε τις συνδέσεις που έχει με τους γύρω κόμβους, το οποίο μας δείχνει ότι ο κεντρικός αυτός κόμβος συσχετίζεται με τους γύρω, μέσω του hashtag που έχουμε χρησιμοποιήσει. Οι σχέσεις αυτές είναι κατευθυνόμενες. Οι ακμές έχουν ένα βέλος προς την φορά τους.

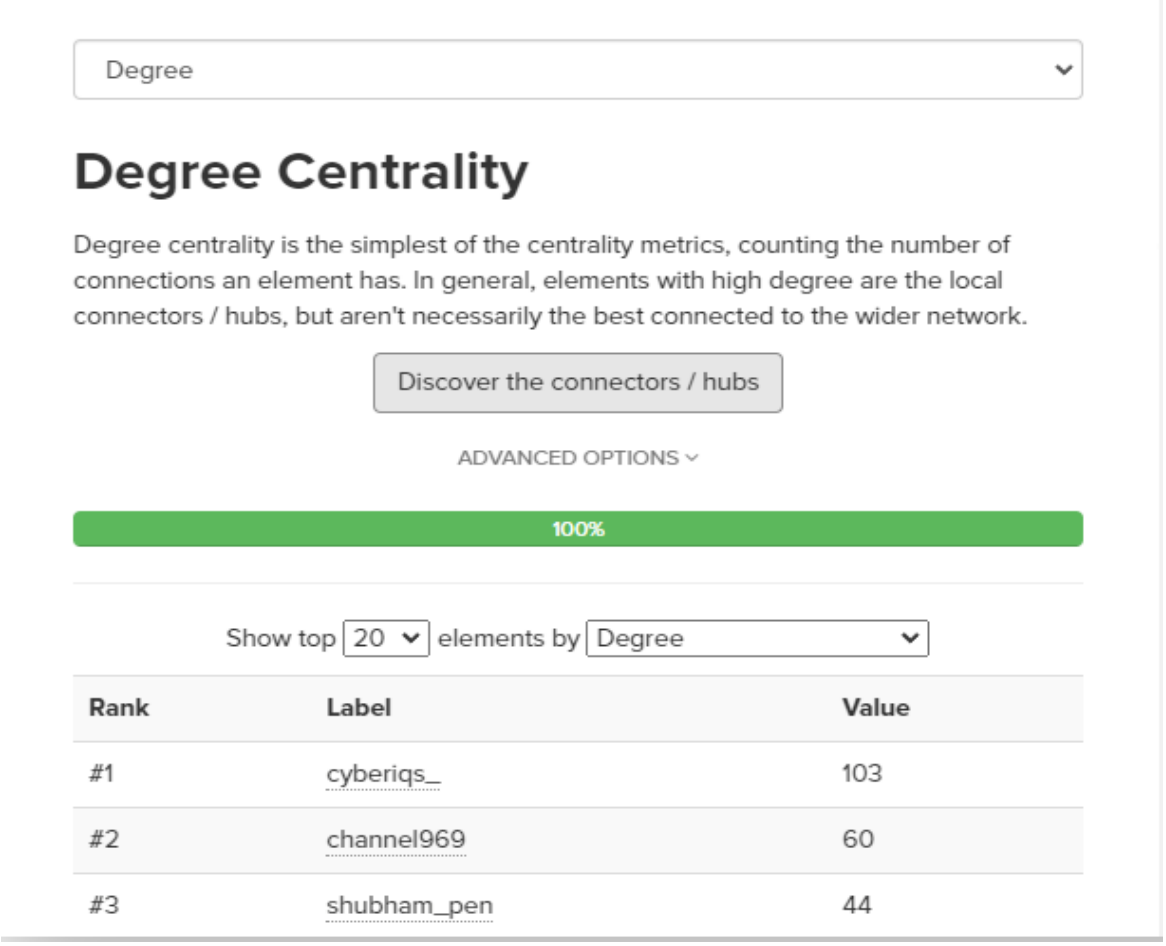


Αν επιλέξουμε έναν κεντρικό κόμβο και πατήσουμε δεξί κλικ > Hide τότε το γράφημα αναδιαμορφώνεται και οι κόμβοι που συνδεόντουσαν με άλλους μέσω του κεντρικού που κρύψαμε, εξαφανίζονται, αφού αποτελούν έναν μοναδικό χωρίς ακμές κόμβο.

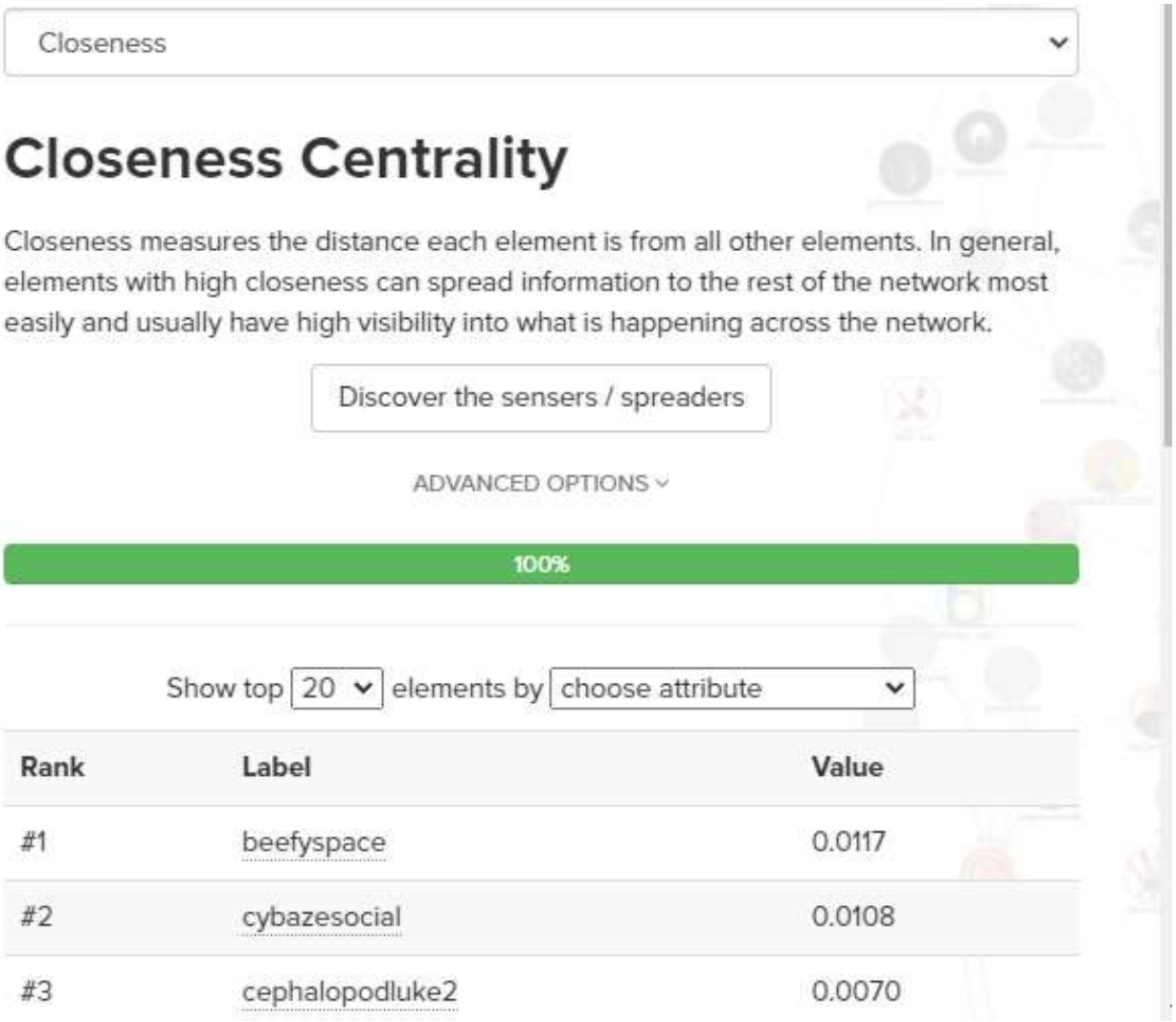


Στο τελικό μας γράφημα μπορούμε να υλοποιήσουμε κάποιες συναρτήσεις μέσω του Kumu Metrics. Στην επιλογή Metrics > Social Media Analysis, επιλέγουμε αρχικά Degree Centrality μετά Closeness Centrality και Betweenness Centrality.

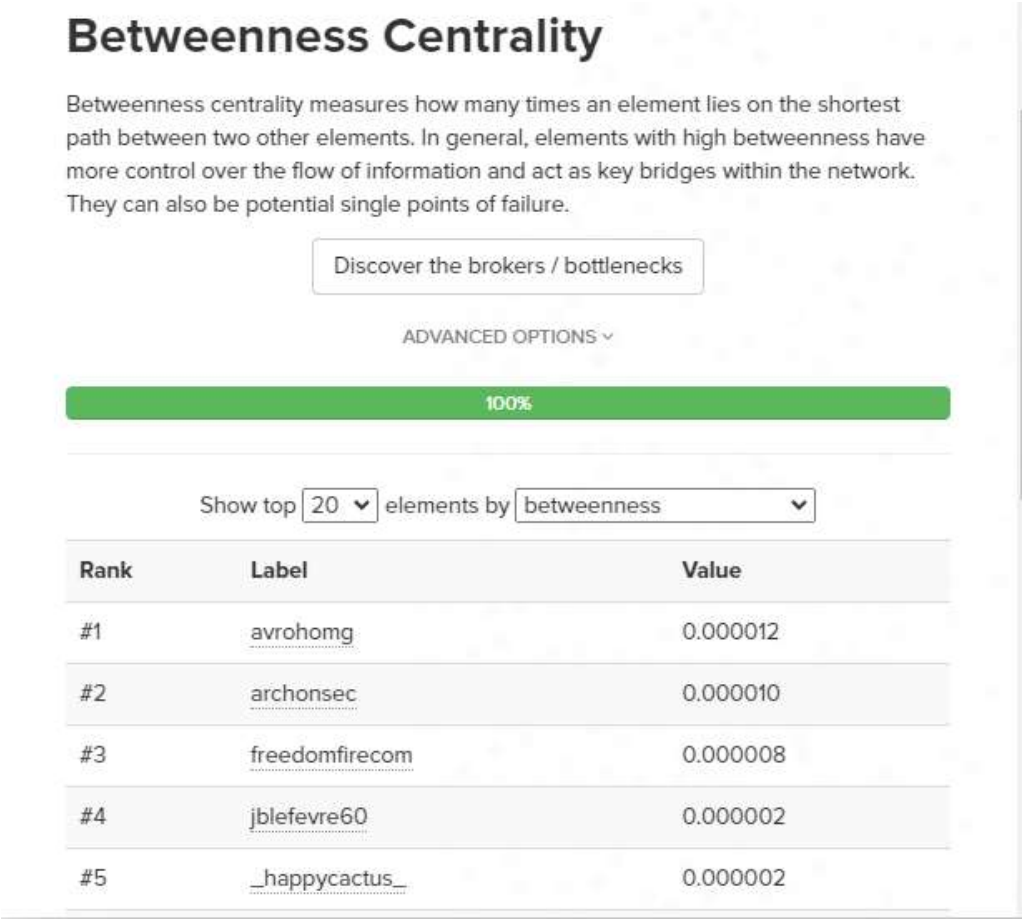
Το Degree Centrality μας δίνει πληροφορίες για το ποιος είναι ο κόμβος με τις περισσότερες συνδέσεις. Οι πολλές συνδέσεις δεν σημαίνουν απαραίτητα ότι ο κόμβος αυτός είναι ο πιο γνωστός αλλά είναι κόμβος που έχει αναφέρει πολλούς άλλους.



Το closeness Centrality δείχνει πόσα βήματα (hops) μακριά είναι οι υπόλοιποι κόμβοι. Αυτό μπορεί να μας φανεί εξαιρετικά χρήσιμο αφού κόμβοι με “high closeness”, μπορούν να κοινοποιήσουν γρηγορότερα πληροφορίες στους υπόλοιπους κόμβους, και είναι αυτοί με την μεγαλύτερα δημοσιότητα (Higher visibility).



Με την επιλογή Betweenness Centrality βλέπουμε πόσες φορές ένας κόμβος έχει απαντήσει σε κάποιον άλλον. Μας βοηθάει δηλαδή να παρακολουθήσουμε κάποιες συζητήσεις που έχουν γίνει (retweets). Αρκετές φορές όμως, ο πρώτος κόμβος που θα πάρουμε ως αποτέλεσμα, μπορεί να είναι ένας μοναδικός κόμβος που αποτυγχάνει.



Rank	Label	Value
#1	avrohomg	0.000012
#2	archonsec	0.000010
#3	freedomfirecom	0.000008
#4	jblefevre60	0.000002
#5	_happycactus_	0.000002

4. Συζήτηση

Όπως είδαμε, ένας μεγάλος όγκος δεδομένων χωρίς κάποια επεξεργασία μπορεί να φανεί χαοτικός. Με την ανάλυση δικτύου όμως, χωρίς να βλέπουμε νούμερα ή να υλοποιούμε περίπλοκες μαθηματικές συναρτήσεις μπορούμε να βγάλουμε συμπεράσματα για το γράφημα που αντικρίζουμε. Για παράδειγμα στην ανάλυση κοινωνικού δικτύου twitter, είδαμε ότι κόμβοι που βρίσκονται στο κέντρο είναι πιο σημαντικοί από τους άλλους που βρίσκονται στην άκρη, αφού έχουν περισσότερες συσχετίσεις. Στην άκρη βρίσκονται οι κόμβοι με λίγες συνδέσεις, και τα tweets τους είναι λιγότερο ορατά στους άλλους κόμβους.

Η πληροφορία που παράγεται μέσω της ανάλυσης μπορεί επίσης να χρησιμοποιηθεί για να διακρίνουμε τις ομάδες χρηστών (groups) και τις αλληλεπιδράσεις που έχουν μεταξύ τους. Επιπλέον, υπάρχουν αμέτρητα εργαλεία που μπορούν να συνδυαστούν για να φέρουν μια ακόμα πιο σημαντική πληροφορία, όπως για παράδειγμα η χρήση συγκεκριμένου λεξικού για το «φιλτράρισμα» των κοινοποιήσεων (tweet) , για συγκεκριμένες λέξεις κλειδιά που επιστρέφουν τιμή για το αν η δημοσίευση υποστηρίζει κάτι θετικό ή κάτι αρνητικό σε σχέση με το συγκεκριμένο hashtag.

5. Βιβλιογραφία

http://imu.ntua.gr/static/courses/strategicISmanagement/lectures/11-Social%20Networks.pdf
https://en.wikipedia.org/wiki/Twitter
https://www.youtube.com/watch?v=THdrju-UWjo&t=550s&ab_channel=DerekCaelin
https://seedscientific.com/how-much-data-is-created-every-day/