

Online Social Networks and Media

Πανεπιστήμιο Ιωαννίνων

Τμήμα Μηχανικών Η/Υ και Πληροφορικής

2019

Υπεύθυνοι Καθηγητές: Πιτουρά Ευαγγελία, Τσαπάρας Παναγιώτης

Φοιτητές: Ζώγος Σωτήριος, Τριανταλή Δήμητρα

Στόχοι της εργασίας:

Ο κύριος στόχος της εργασίας είναι να μετρήσουμε την ομοιοφιλία (homophily) σε ένα πραγματικό κοινωνικό δίκτυο. Για τον λόγο αυτόν επιλέξαμε το Twitter.

Η ομοιοφιλία είναι η τάση των ατόμων να συσχετίζονται και να συνδέονται με παρόμοιους άλλους. Περισσότερες από 100 μελέτες έχουν παρατηρήσει ομοιοφιλία σε κάποια μορφή και αποδεικνύουν ότι η ομοιότητα δημιουργεί σύνδεση. Αυτά περιλαμβάνουν την ηλικία, το φύλο, την τάξη και τον οργανωτικό ρόλο. Τα άτομα με ομόφιλες σχέσεις έχουν κοινά χαρακτηριστικά (πεποιθήσεις, αξίες, εκπαίδευση, κλπ.) που διευκολύνουν την επικοινωνία και τη δημιουργία σχέσεων.^[2]

Συλλογή Δεδομένων:

Πρόκειται να συλλέξουμε δεδομένα από έναν προκαθορισμένο αριθμό χρηστών του twitter. Η επιλογή των χρηστών θα γίνει επιλέγοντας δύο διάσημους εκπροσώπους από κάποιες συγκεκριμένες κατηγορίες (π.χ. αθλητικά, πολιτική κλπ.). Τα δεδομένα που θα συλλέξουμε αφορούν για κάθε χρήστη τη λίστα των ακολούθων του (followers), τη λίστα των χρηστών που ακολουθεί (following), τη λίστα με τα likes και τα retweets που έχει κάνει σε κάποιο συγκεκριμένο χρονικό διάστημα και τα

hashtags που έχει χρησιμοποιήσει στα tweets του στο ίδιο χρονικό διάστημα.

Για τον σκοπό αυτό, αρχικά δημιουργήσαμε ένα προφίλ στο Twitter και ζητήσαμε άδεια developer για να μπορέσουμε να χρησιμοποιήσουμε τα endpoints του Twitter API. Κάθε endpoint επιστρέφει τα δεδομένα που του ζητήσαμε σε μορφή json, το οποίο έπειτα θα χρησιμοποιηθεί από το πρόγραμμά μας (σε python) για τους διάφορους αναγκαίους υπολογισμούς.

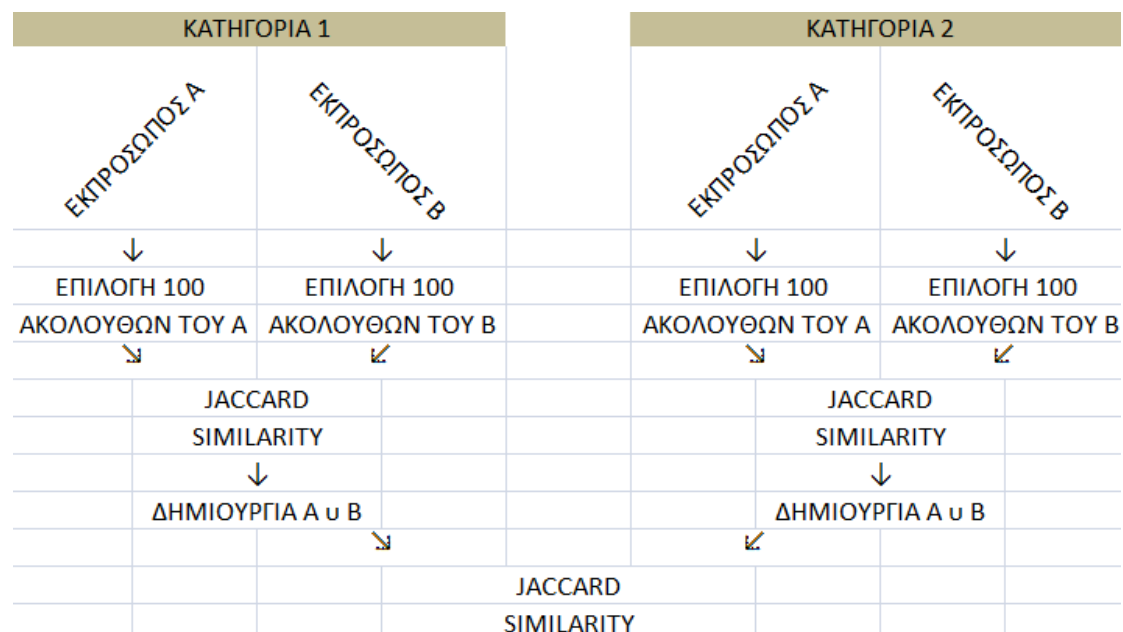
Ανάλυση:

Για την ερευνά μας θα επιλέξουμε αρχικά τέσσερις κατηγορίες. Οι κατηγορίες αυτές θα είναι “ποδόσφαιρο”, “μουσική”, “πολιτική” και “πρακτορεία ειδήσεων”. Σε κάθε μία από αυτές τις κατηγορίες θα διαλέξουμε δύο διάσημους εκπροσώπους της. Για κάθε έναν από τους παραπάνω διασήμους θα διαλέξουμε 100 ακολούθους τους. Το

μόνο κριτήριο στην επιλογή των 800 ακολούθων είναι να έχουν ενεργό προφίλ (δηλαδή να έχουν κάνει έναν συγκεκριμένο αριθμό δραστηριοτήτων με το προφίλ τους στο twitter τον τελευταίο μήνα). Ως δραστηριότητες εννοούνται τα likes, τα tweets και τα retweets.

Με τα δεδομένα αυτά θα μετρήσουμε για κάθε μία από τις τέσσερις κατηγορίες ξεχωριστά, το similarity του κάθε ζευγαριού των εκπροσώπων της ανάλογα με τα hashtags τα οποία χρησιμοποιούν οι followers τους στα tweets τους. Για παράδειγμα, έστω ότι στην κατηγορία του ποδοσφαίρου έχουμε διαλέξει τον Pique και τον Messi. Οπότε θα κατασκευάσουμε δύο ομάδες A και B. Στην ομάδα A θα υπάρχουν τα hashtags που χρησιμοποίησαν στα tweets τους οι ακόλουθοι του Pique και στην ομάδα B τα αντίστοιχα για τον Messi. Για τον υπολογισμό του similarity θα χρησιμοποιήσουμε την build in συνάρτηση υπολογισμού της Jaccard similarity σε python, η οποία θα δεχτεί σαν είσοδο τα σύνολα A και B που προαναφέραμε. Η συνάρτηση αυτή είναι η jaccard_score της βιβλιοθήκης sklearn.metrics.

Στη συνέχεια για κάθε μία από τις τέσσερις κατηγορίες θα πάρουμε την ένωση των συνόλων A και B που ορίσαμε προηγουμένως και θα την ορίσουμε ως ένα καινούριο σύνολο. Μ' αυτόν τον τρόπο θα έχουμε δημιουργήσει τέσσερα νέα σύνολα, ένα για κάθε κατηγορία, που θα περιέχουν όλα τα hashtags που έχουν χρησιμοποιήσει στα tweets τους οι χρήστες που ακολουθούν τους δύο εκπροσώπους της κάθε κατηγορίας. Σκοπός αυτής της κατασκευής είναι να βρούμε το similarity μεταξύ των διαφορετικών κατηγοριών ανά δύο.



Τέλος, θα δημιουργήσουμε μια σελίδα στο Github, στην οποία θα παρουσιάσουμε εν συντομία τα βήματα της έρευνας που θα πραγματοποιήσουμε, καθώς επίσης τον κώδικα της εργασίας και το συνοδευτικό υλικό (παρουσιάσεις, αναφορές, papers κλπ.)

Σχετικά paper:

[1] Marina Drosou, H.V. Jagadish, Evaggelia Pitoura, and Julia Stoyanovich. Diversity in Big Data: A Review. Big Data. Jun 2017

[2] McPherson, M.; Smith-Lovin, L.; Cook, J. M. (2001). "Birds of a Feather: Homophily in Social Networks". Annual Review of Sociology. 27: 415–444.