

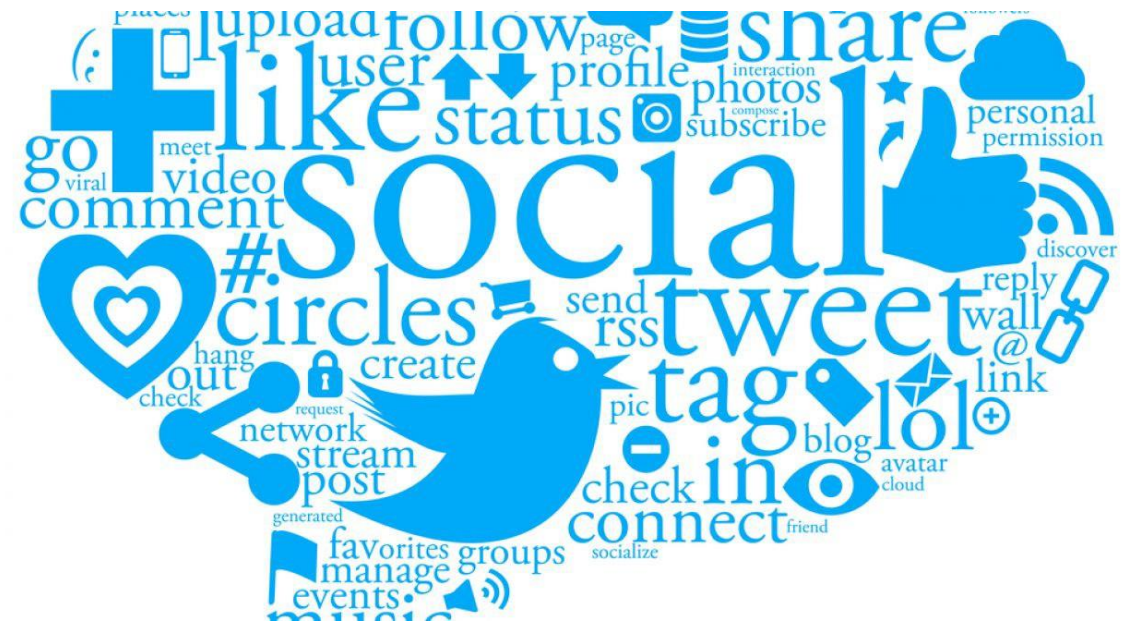


Online Social Networks and Media

ΖΩΓΟΣ ΣΩΤΗΡΙΟΣ, ΤΡΙΑΝΤΑΛΗ ΔΗΜΗΤΡΑ

Στόχος της εργασίας

- ▶ Ο τίτλος του topic που επιλέξαμε είναι "Content homophily in a real social network".
- ▶ Ο κύριος στόχος της εργασίας είναι να μετρήσουμε την ομοιοφιλία (homophily) σε ένα πραγματικό κοινωνικό δίκτυο.
- ▶ Το κοινωνικό δίκτυο που επιλέξαμε είναι το Twitter.



Τι είναι η ομοιοφιλία?

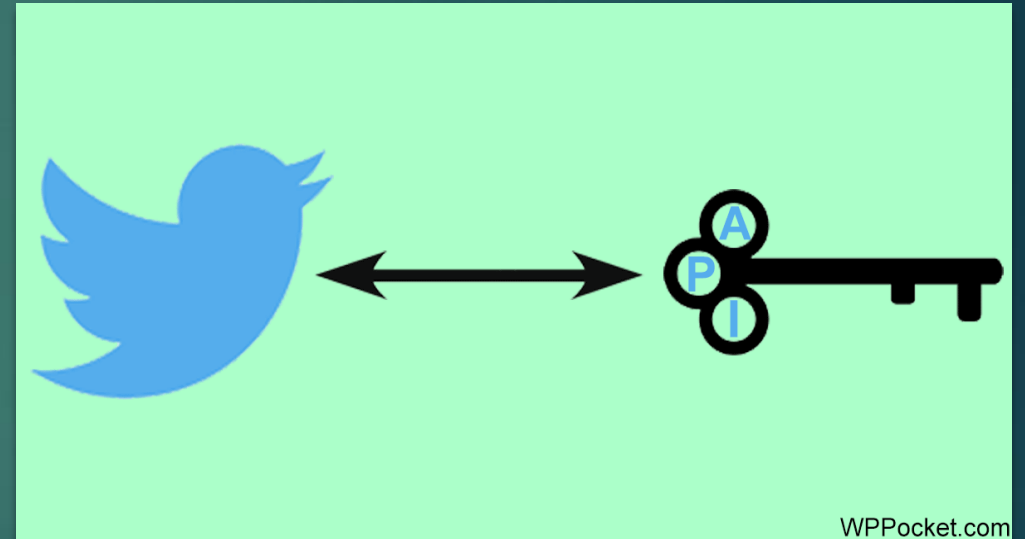
- ▶ Η ομοιοφιλία είναι η τάση των ατόμων να συσχετίζονται και να συνδέονται με παρόμοιους άλλους. Περισσότερες από 100 μελέτες έχουν παρατηρήσει ομοιοφιλία σε κάποια μορφή και αποδεικνύουν ότι η ομοιότητα δημιουργεί σύνδεση. Αυτά περιλαμβάνουν την ηλικία, το φύλο, την τάξη και τον οργανωτικό ρόλο. Τα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

Συλλογή Δεδομένων

- ▶ Πρόκειται να συλλέξουμε δεδομένα από έναν προκαθορισμένο αριθμό χρηστών του twitter. Η επιλογή των χρηστών θα γίνει πιθανότατα βάσει της διασημότητάς τους, δηλαδή τον αριθμό των ακολούθων τους. 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.

Πώς ξεκινάμε?

- ▶ 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 (Πρώτο βήμα)

- ▶ Στην έρευνά μας θα διακρίνουμε τους χρήστες που επιλέξαμε σε τρεις κατηγορίες.
 - Στην Κατηγορία 1 για οποιουσδήποτε δύο χρήστες A και B ισχύει ότι ο χρήστης A ακολουθεί τον χρήστη B και αντίστροφα.
 - Στην Κατηγορία 2 για οποιουσδήποτε δύο χρήστες A και B ισχύει ότι ο χρήστης A ακολουθεί τον χρήστη B, ενώ ο χρήστης B δεν ακολουθεί τον χρήστη A.
 - Τέλος, στην Κατηγορία 3 για οποιουσδήποτε δύο χρήστες A και B, κανένας δεν ακολουθεί κανέναν.

Ανάλυση (Δεύτερο βήμα)

- ▶ Με τα δεδομένα αυτά θα μετρήσουμε για κάθε κατηγορία ξεχωριστά, το similarity του κάθε ζευγαριού ανάλογα με τα hashtags τα οποία χρησιμοποιούν στα tweets τους, καθώς και τα likes και τα retweets που κάνουν σε άλλους χρήστες. Για τον υπολογισμό του similarity θα χρησιμοποιήσουμε την build in συνάρτηση υπολογισμού της Jaccard similarity σε python. Η συνάρτηση αυτή είναι η jaccard_score της βιβλιοθήκης sklearn.metrics.

Ανάλυση (Τρίτο βήμα)

- ▶ 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.)

Βιβλιογραφικές αναφορές

- ▶ Marina Drosou, H.V. Jagadish, Evaggelia Pitoura, and Julia Stoyanovich. Diversity in Big Data: A Review. Big Data. Jun 2017
- ▶ McPherson, M.; Smith-Lovin, L.; Cook, J. M. (2001). "Birds of a Feather: Homophily in Social Networks". Annual Review of Sociology. 27: 415–444.