



ΕυΦυησ αλληλεπιδραση με κοινΩνικα δικτυα

ΕΡΓΑΣΙΑ 2018 - 2019



|  |  |
| --- | --- |
| **Π16036** | ΙΩΑΝΝΙΔΗΣ ΠΑΝΑΓΙΩΤΗΣ |
| **Π16097** | ΝΙΚΑΣ ΔΙΟΝΥΣΙΟΣ |
| **Π16112** | ΠΑΡΑΒΑΝΤΗΣ ΑΘΑΝΑΣΙΟΣ |

Περιεχόμενα

[II. Επεξήγηση 2](#_Toc13173597)

[III. Εκτέλεση 4](#_Toc13173598)

[IV. Γραφήματα 7](#_Toc13173599)

[Εικόνα 1 Αποτελέσματα ερωτημάτων 1, 2 και 3 με N=3. 4](#_Toc13173559)

[Εικόνα 2 Αποτελέσματα ερωτημάτων 4, 5 και 6 με param\_GD=0.3, param\_CN=0.5, param\_JC=0.6, param\_A=0.7, param\_PA=0.3 για το 1ο γράφημα που προκύπτει. 5](#_Toc13173560)

[Εικόνα 3 Αποτελέσματα ερωτημάτων 6, 7 με param\_GD=0.9, param\_CN=0.6, param\_JC=0.4, param\_A=0.5, param\_PA=0.3 για το 2ο γράφημα που προκύπτει. 6](#_Toc13173561)

# Επεξήγηση

Η εργασία του μαθήματος που παραδίδουμε αναλύει σε βάθος το πακέτο δεδομένων StackOverflow temporal network, αναδεικνύοντας ουσιαστικές πληροφορίες και σχέσεις μεταξύ προγραμματιστών. Για την επίλυση της εργασίας χρησιμοποιήθηκαν οι εξής τεχνολογίες:

* Γλώσσα προγραμματισμού Python 3.7
* Βιβλιοθήκη ανάλυσης δεδομένων Pandas  
  <https://pandas.pydata.org/>
* Βιβλιοθήκη γραφημάτων NetworkX  
  <https://networkx.github.io/>
* Βιβλιοθήκη γραφικών παραστάσεων Matplotlib  
  <https://matplotlib.org/>

Επειδή τα δεδομένα που μας δίνονται περιέχουν εκατομμύρια γραμμές, αποφασίσαμε να δουλέψουμε πάνω στις πρώτες 1.000 καταχωρήσεις, δηλαδή σε ένα υποσύνολο όλων των συνεργασιών μεταξύ προγραμματιστών.

Αφού φορτώσουμε τα δεδομένα με τη βιβλιοθήκη Pandas υπολογίζουμε την ελάχιστη και μέγιστη χρονική στιγμή για το 1ο ερώτημα. Προχωρώντας στο 2ο ερώτημα, χωρίζουμε αυτή τη χρονική περίοδο σε N διαστήματα που καθορίζονται από τον χρήστη. Έχοντας υπ’ όψη όλες τις χρονικές περιόδους βασιζόμενοι πάντα στους τύπους της εκφώνησης, στο 3ο ερώτημα σχεδιάζουμε με τη βοήθεια της βιβλιοθήκης NetworkX όλα τα γραφήματα, όπως εξελίσσονται στη διάρκεια του χρόνου.

Τα γραφήματα του 3ου ερωτήματος έρχονται να συμπληρώσουν το 4ο ερώτημα, το οποίο υπολογίζει τα εξής για κάθε γράφημα: degree centrality, in-degree centrality, out-degree centrality, closeness centrality, betweenness centrality, eigenvector centrality και katz centrality. Τα ευρήματα των υπολογισμών αποθηκεύονται στον ίδιο φάκελο σε μορφή γραφημάτων χρησιμοποιώντας τη βιβλιοθήκη Matplotlib. Ο αριθμός των γραφημάτων έχει άμεση σχέση με τον αριθμό των χρονικών περιόδων και με τις συνεργασίες των προγραμματιστών.

Για το 5ο ερώτημα, βρίσκουμε τις κοινές κορυφές και ακμές μεταξύ όλων των περιόδων. Για την εύρεση των κοινών κορυφών βρίσκουμε την τομή μεταξύ κάθε χρονικής περιόδου και αντίστοιχα για την εύρεση των ακμών, υπολογίζουμε τα ζευγάρια κόμβων που παραμένουν σταθερά στην εξέλιξη του χρόνου. Χρησιμοποιώντας τις πράξεις συνόλων της Python, οι υπολογισμοί αυτοί γίνονται εύκολα και γρήγορα.

Στο 6ο ερώτημα, χρησιμοποιούμε ξανά τις χρήσιμες μεθόδους τους NetworkX για τα ζεύγη κόμβων στην εξέλιξη των γραφημάτων. Υπολογίζουμε τα εξής: graph distance, common neighbors, jaccard’s coefficient, adamic / adar και preferential attachment. Αφού γίνουν οι υπολογισμοί των μετρικών αυτών, προχωράμε αμέσως στο 7ο ερώτημα που καλούμαστε να μετρήσουμε την αποτελεσματικότητα της πρόγνωσης μελλοντικών ακμών για κάθε μετρική. Οι τιμές των παραμέτρων καθορίζονται από τον χρήστη και οι τιμές τους κυμαίνονται μεταξύ του 0 και 1.

Η πηγαίος κώδικας που απαντάει στα ερωτήματα της εργασίας βρίσκεται στο αρχείο main.py. Για την εκτέλεση: python main.py αφού εγκαταστήσουμε τις βιβλιοθήκες που χρησιμοποιούνται.

# Εκτέλεση

Παρακάτω ακολουθούν στιγμιότυπα από την εκτέλεση του προγράμματος σε περιβάλλον Ubuntu με τη Python 3.7.

A screenshot of a cell phone

Description automatically generated

Εικόνα 1 Αποτελέσματα ερωτημάτων 1, 2 και 3 με N=3.

A screenshot of a cell phone

Description automatically generated

Εικόνα 2 Αποτελέσματα ερωτημάτων 4, 5 και 6 με param\_GD=0.3, param\_CN=0.5, param\_JC=0.6, param\_A=0.7, param\_PA=0.3 για το 1ο γράφημα που προκύπτει.

A screenshot of a cell phone

Description automatically generated

Εικόνα 3 Αποτελέσματα ερωτημάτων 6, 7 με param\_GD=0.9, param\_CN=0.6, param\_JC=0.4, param\_A=0.5, param\_PA=0.3 για το 2ο γράφημα που προκύπτει.

# Γραφήματα