

## Social Network Analysis Assignments

Στα πλαίσια των διαλέξεων του μαθήματος διδάσκονται διάφορα θεωρητικά και πρακτικά θέματα γύρω από την ανάλυση των κοινωνικών δικτύων. Σκοπός της εργασίας του μαθήματος είναι να μπορείτε να εφαρμόσετε όσα μαθαίνετε σε πραγματικά σύνολα δεδομένων. Για το σκοπό αυτό θα χρειαστεί να παραδώσετε σε τρία στάδια την εργασία σε ένα σύνολο δεδομένων της επιλογής σας.

Το σύνολο δεδομένων μπορεί να είναι κάποια από τα σύνολα που υπάρχουν στο Konect Project (<http://konect.cc/>) ή στο SNAP collection (<https://snap.stanford.edu/data/>). Τα σύνολα αυτά διαφέρουν μεταξύ τους σε πλήθος ακμών και κόμβων, στον τύπο των ακμών (κατευθυνόμενες και μη, με βάρη ή όχι, με επιπλέον γνωρίσματα) καθώς και στις εφαρμογές που χρησιμοποιούνται.

Τα στάδια (παραδοτέα) έχουν ως ακολούθως:

1ο στάδιο) Δημιουργία συνθετικού κοινωνικού δικτύου με παρόμοια χαρακτηριστικά ενός πραγματικού δικτύου.

- Επιλέγετε ένα σύνολο από τις πιο πάνω συλλογές
- Υπολογίζετε βασικές μετρικές του όπως αυτές που διδάσκονται στο πρώτο μέρος του μαθήματος
- Χρησιμοποιείτε κάποιον από τους network generators που διδάσκονται στο μάθημα προσπαθώντας να παράξετε ένα συνθετικό δίκτυο με όσο το δυνατόν πιο όμοια χαρακτηριστικά με το πραγματικό. Τουλάχιστον με παραπλήσιο αριθμό κόμβων και ακμών.
- Υπολογίζετε τις ίδιες μετρικές και στο συνθετικό δίκτυο

2ο στάδιο) Υλοποίηση αλγορίθμων πάνω στους 2 γράφους.

- Επιλέξτε έναν αλγόριθμο πρόβλεψης ακμών και υλοποιήστε ότι είναι απαραίτητο ώστε να τον εφαρμόσετε με επιτυχία στον πάνω στον συνθετικό (και στον πραγματικό) γράφο.
- Αξιολογείστε τον αλγόριθμο με χρήση μιας τεχνικής validation (k-fold, leave one out).
- Εντοπίστε κλίκες στους γράφους με εφαρμογή κατάλληλου αλγορίθμου. Παρουσιάστε χαρακτηριστικά παραδείγματα.

3ο στάδιο) Προσομοίωση διάδοσης πάνω στον γράφο

- Στο γράφο που επιλέξατε (ή σε κάποιο άλλο γράφο σε περίπτωση που ο αρχικός δεν είναι κατάλληλος) επιλέξτε τα  $k$ -seeds που θα μεγιστοποιήσουν την επιρροή στο γράφο
- Ορίστε τις παραμέτρους για την εύρεση του βέλτιστου πλήθους των  $k$ -seeds καθώς και των παραμέτρων που επιτυγχάνουν influence το 60% του δικτύου.

Σε κάθε παραδοτέο θα πρέπει να παραδώσετε κώδικα καθώς και μία μικρή αναφορά που να περιγράφει την δουλειά σας.