

PROJECT HY-383

2013-2014

Στα πλαίσια της εργασίας για το μάθημα **HY-383 Δυναμική Πολύπλοκων Δικτύων** θα επεξεργαστείτε και θα αναλύσετε δεδομένα που προέρχονται από τον πραγματικό κόσμο. Με αυτό τον τρόπο, αφενός, θα εμπεδώσετε καλύτερα όσα έχετε διδαχτεί και αφετέρου θα αντιληφθείτε πληρέστερα το εύρος των πρακτικών εφαρμογών που προκύπτουν από τις εν λόγω θεωρίες.

Τα δεδομένα βρίσκονται στον παρακάτω διαδικτυακό τόπο:

<http://snap.stanford.edu/data/#socnets>

και είναι ταξινομημένα σε διάφορες θεματικές κατηγορίες. Όπως θα διαπιστώστε καλύπτεται ένα μεγάλο φάσμα εφαρμογών και προβλημάτων. Από το συγκεκριμένο κατάλογο θα επιλέξετε ένα σύνολο δεδομένων (data set) με το οποίο και θα εργαστείτε. Συνίσταται να δουλέψετε πάνω σε υλικό που να σχετίζεται με ένα θέμα που βρίσκετε ενδιαφέρον ούτως ώστε να έχετε ένα επιπλέον κίνητρο για την ποιοτικότερη υλοποίηση της εργασίας .

Η εργασία αυτή καθαυτή, μετά την επιλογή των δεδομένων, χωρίζεται σε δύο διακριτά στάδια:

(A) Κατά το **πρώτο** θα πρέπει να εφαρμόσετε όσες από τις τεχνικές έχετε διδαχτεί στο μάθημα, και κρίνετε πλέον κατάλληλες για διεξοδικότερη επεξεργασία των δεδομένων.

(B) Ακολουθώντας, κατά το **δεύτερο** στάδιο, στηριζόμενοι στα αποτελέσματα που εξάχθηκαν προηγουμένως θα πραγματοποιήσετε μια ποιοτική ανάλυση. Με άλλα λόγια, θα πρέπει να ανακαλύψετε ποια είναι τα κύρια χαρακτηριστικά και οι σημαντικότερες ιδιότητες του υλικού που μελετάτε . Επιπλέον, θα πρέπει να αποφανθείτε κατά πόσο συμφωνούν τα συμπεράσματα σας με τα αναμενόμενα αποτελέσματα βάση της θεωρίας.

Τονίζεται ότι και τα δύο στάδια είναι εξίσου σημαντικά:

Δεν αρκεί να εφαρμόσετε σωστά τις τεχνικές.

Είναι απαραίτητο να μπορείτε να κατανοείτε σε τι αντιστοιχούν τα εκάστοτε αποτελέσματα που παράγονται κατά τα διαδοχικά βήματα της εργασίας.

Για παράδειγμα, έστω ότι μελετάτε ένα δίκτυο που αφορά την εξάπλωση ενός νοσήματος . Αρχικά θα πρέπει να υπολογίσετε κάποια ποσοτικά χαρακτηριστικά του δικτύου όπως το clustering coefficient, το μέσο μήκος του συντομότερου μονοπατιού ,τον μέσο βαθμό των κόμβων και ότι άλλο κρίνετε απαραίτητο. Η επεξεργασία των προηγούμενων αποτελεσμάτων οδηγεί σε διάφορα συμπεράσματα σε σχέση με το δίκτυο : π.χ. το clustering coefficient και το μέσο μήκος του συντομότερου μονοπατιού ενδέχεται να συνηγορούν στο ότι το δίκτυο διαθέτει τα χαρακτηριστικά του

small-world. Με τον ίδιο τρόπο η ύπαρξη ή μη λίγων hubs και πολλών κόμβων μικρού βαθμού συσχετίζεται σε κάποιο άλλο ποιοτικό χαρακτηριστικό και ούτω καθεξής. Αφού ολοκληρώσετε και την ποιοτική ανάλυση θα συγκρίνεται τα συμπεράσματά σας με ότι προβλέπεται από τη θεωρία για τα δίκτυα του συγκεκριμένου τύπου -στην περίπτωση του παραδείγματος δηλαδή με τα δίκτυα που αφορούν την εξάπλωση ενός νοσήματος παρεμφερούς με αυτό που αναλύετε. Έτσι αν πρόκειται για νόσημα βραδείας εξάπλωσης και υψηλού ποσοστού θνησιμότητας ο γράφος του δικτύου αναμένεται να μην διαθέτει μεγάλη διακύμανση ανάμεσα στους βαθμούς των κόμβων. Πάντως, το γεγονός ότι τα συμπεράσματά δεν συμφωνούν με τη θεωρία δεν συνεπάγεται απαραίτητα ότι η ανάλυσή σας είναι λανθασμένη - ίσως το δίκτυο σας διαθέτει κάποια μοναδικά χαρακτηριστικά. Σε κάθε περίπτωση θα πρέπει μπορείτε να αιτιολογήσετε την συμφωνία ή ασυμφωνία μεταξύ του δικτύου που μελετάτε και του τυπικού μοντέλου της θεωρίας.

Για την υλοποίηση της εργασίας είστε ελεύθεροι να χρησιμοποιήσετε όποια γλώσσα προγραμματισμού θέλετε (Java, C++) ή όποιο εργαλείο (π.χ. Matlab) πιστεύετε ότι θα σας βοηθήσει. Επίσης τυχόν πρωτοβουλίες που θα αναλάβετε θα προσμετρηθούν θετικά. Για παράδειγμα, ενδέχεται να επιθυμείτε να εφαρμόσετε κάποιον αλγόριθμο ή κάποια τεχνική που δεν περιλαμβάνεται στην ύλη του μαθήματος αλλά κρίνετε ότι είναι συναφή με αυτό.

Αναφορικά με τα διαδικαστικά της εργασίας: μπορείτε είτε να δουλέψετε σε ομάδες των δύο ατόμων είτε ατομικά. (Και στις δύο περιπτώσεις ο φόρτος που αναμένεται από πλευράς σας είναι ο ίδιος). Θα πρέπει, συνεπώς, να έχετε αποστείλει μέχρι τις 29/11 mail με θέμα HY383_Project_Team στο gouidis@csd.uoc.gr με τα στοιχεία (login και AM) τα δικά σας και του συνεργάτη σας (αν έχετε) καθώς και του data set που θα αναλύσετε. Τα παραδοτέα θα είναι τα εξής: **α)** ο πλήρης κώδικας της εργασίας μαζί με το data set, ο οποίος θα πρέπει να εκτελείται δίχως πρόβλημα και **β)** μια μικρή αναφορά στην οποία θα παρουσιάζετε συνοπτικά τα αποτελέσματα που προέκυψαν από την ανάλυσή σας, δηλ., μια σύντομη περιγραφή της διαδικασίας και του σκεπτικού που ακολουθήσατε καθώς και τα συμπεράσματά που αντλήσατε από την ανάλυσή σας. Τέλος, μετά την παράδοση των εργασιών, κάθε ομάδα θα παρουσιάσει την δουλειά της ενώπιον του καθηγητή, των βοηθών και των υπόλοιπων φοιτητών.

Σημαντικές ημερομηνίες:

29/11/2013: Φροντιστήριο αφιερωμένο στην εργασία.

09/01/2013: Παράδοση όλων των παραδοτέων.

14, 16, 17, η/και 21/01/21013: Παρουσιάσεις (10 λεπτών) εργασιών από ομάδες φοιτητών.