

Θέματα Πτυχιακών Εργασιών για το Ακαδημαϊκό Έτος 2012-2013 Εαρινό Εξάμηνο

Ευαγγελία Πιτουρά

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 1 : Οπτικοποίηση ποικιλομορφίας (diversity) με δυνατότητα zooming

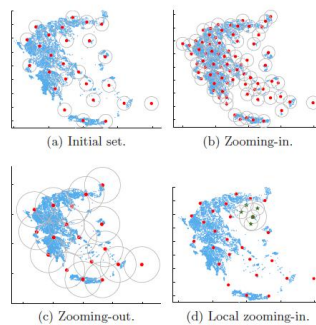
Λόγω του μεγάλου όγκου των δεδομένων, πρόσφατα έχει δοθεί ιδιαίτερο ενδιαφέρον στη διαφορετικότητα ή ποικιλομορφία. Η ιδέα είναι να επιλέξουμε να δείχνουμε στους χρήστες ως αποτέλεσμα των ερωτήσεων τους δεδομένα που καλύπτουν διαφορετικές πτυχές του αποτελέσματος.

Σε προηγούμενη δουλειά μας προτεινάμε ένα νέο τρόπο για να επιλέγεις αυτά τα διαφορετικά δεδομένα που σου επιτρέπει να κάνεις zoom-in και zoom-out, δηλαδή, να μπορείς να βλέπεις λιγότερα και πιο διαφορετικά σημεία ή περισσότερα και όχι τόσο διαφορετικά σημεία μέσω του προσδιορισμού μιας ακτίνας (βλέπε Εικόνα παρακάτω).

Πιο αναλυτικά:

Έστω ένα σύνολο δεδομένων S . Δοσμένης μιας ακτίνας r , διαλέγεις ένα υποσύνολο S' του S έτσι ώστε (α) να καλύπτεται όλο το S , που σημαίνει ότι για κάθε σημείο p του S , πρέπει να υπάρχει ένα σημείο στο S' κοντινό του p , δηλαδή σε απόσταση $\leq r$ από το p και (β) να μην υπάρχουν στο S' κοντινά σημεία (δηλαδή, όλα τα σημεία του S' να έχουν απόσταση μεταξύ τους $> r$). Μπορείς να κάνεις zoom in (μικραίνοντας το r) και zoom out (μεγαλώνοντας το r).

Πως οπτικοποιούμε το παραπάνω, δηλαδή, πως δείχνουμε οπτικά το αποτέλεσμα; Για 2-διαστάσεις και αριθμητικά δεδομένα είναι σχετικά εύκολο (δες Εικόνα), αλλά πως θα δείχνουμε το αποτέλεσμα όταν έχουμε πολλές διαστάσεις ή μη αριθμητικά δεδομένα;



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 2: Ευρετήρια για Μεγάλα Χρονικά Μεταβαλλόμενα Γραφήματα

Τα κοινωνικά δίκτυα (π.χ., Twitter, Facebook) συνήθως αναπαρίστανται ως γράφοι όπου οι κόμβοι αντιστοιχούν στους χρήστες και οι ακμές στις μεταξύ τους σχέσεις. Ένα χαρακτηριστικό αυτών των κοινωνικών γράφων είναι ότι το περιεχόμενο και η δομή τους αλλάζει συνεχώς, π.χ. με την προσθήκη νέων χρηστών, τη προσθήκη και τη διαγραφή συσχετίσεων μεταξύ τους κλπ.

Στη βιβλιογραφία έχουν προταθεί ευρετήρια έτσι ώστε να απαντώνται γρήγορα ερωτήσεις τόσο προσπέλασης (reachability) όσο και κοντινότερης διαδρομής (shortest path).

Θα εξετάσουμε μερικά από τα παρακάτω:

- Τη δυνατότητα να επεκτείνουμε αυτά τα ευρετήρια με δεδομένα (ως ετικέτες στους κόμβους ή στις ακμές)
- Πως συμπεριφέρονται σε ενημερώσεις
- Πως απαντάμε top-k ερωτήσεις (για παράδειγμα: ποιοι είναι οι k πιο κοντινοί κόμβοι;)

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 3: Σύγκριση Διαφορετικών Μεθόδων Αποθήκευσης Γραφημάτων

Η εργασία θα εξετάσει διαφορετικούς μεθόδους για την αποθήκευση μεγάλων γραφημάτων. Η επιθυμητή υλοποίηση θα πρέπει να είναι κατανομημένη σε cluster. Θα εξεταστεί η χρήση μη σχεσιακών βάσεων δεδομένων, όπως

Couchdb <http://couchdb.apache.org/> και

HBase <http://hbase.apache.org/>

και θα γίνει σύγκριση των εναλλακτικών σχεδιασμών και υλοποιήσεων, π.χ., πως θα γίνει η κατανομή του γράφου (partition), ο σχεδιασμός του σχήματος της βάσης, κλπ

Για όλες τις εργασίες απαιτείται καλή γνώση αλγορίθμων, δομών δεδομένων και βάσεων δεδομένων. Επίσης, θα χρειαστεί να διαβάσετε ερευνητικές εργασίες στα Αγγλικά.

ΓΙΑ ΕΡΩΤΗΣΕΙΣ ΣΤΕΙΛΤΕ ΜΟΥ Email ΠΕΡΑΣΤΕ ΑΠΟ ΤΟ ΓΡΑΦΕΙΟ ΜΟΥ