

## Φίλτρα

- Για την υλοποίηση των φίλτρων επιλέξαμε να φτάσουμε μια δεύτερη δομή FilterGraph που κληρονομεί απο το graph και προσθέτει mapping απο vertices τα αντίστοιχα φίλτρα, όπως και mapping απο τα φίλτρα στο πλήθος των στοιχείων που έχουν το φίλτρο αυτό.
- **Mapping Φίλτρων:** Κάθε κόμβος του γράφου αντιστοιχίζεται σε ένα std set φίλτρων (των οποίων ο τύπος είναι ορισμένος με template για πολυμορφία) το οποίο περιέχει όλα τα φίλτρα του κόμβου, σε περίπτωση δηλαδή που ένας κόμβος έχει παραπάνω απο μια τιμή στο categorical attribute. Αυτό στην πράξη συμβαίνει μόνο αν το ίδιο vector διαβαστεί 2 φορές, στην αρχικοποίηση του γράφου, με διαφορετικό φίλτρο, απλά το φίλτρο θα προστεθεί στο set.

## Διαβασμα Αρχείων

- **data.bin:** Το διαβασμα των δεδομένων γίνεται μέσω constructor του FilterGraph ο οποίος είναι εμπνευσμένος απο τον δεδομένο κωδικα για read .bin αρχείων. Διαβάζει τα δεδομένα και κάνει τις αντιστοιχες εισαγωγές στον γραφο (αρχικοποίηση χωρίς ακμές).
- **queries.bin:** Το διάβασμα των queries γίνεται μέσω της συνάρτησης read\_queries (και αυτή εμπνευσμένη απο τον δεδομένο κώδικα) η οποία επιστρέφει ένα vector<vector<float>> που περιέχει τα query vectors, και ένα vector<float> που περιέχεται αντίστοιχα φίλτρα. Σε περίπτωση που το query δεν έχει φίλτρο (categorical attribute) απλά περνάμε κατευθείαν την τιμή -1 και οι συναρτήσεις που καλούν την read\_queries δρουν αναλόγως ώστε η τιμή αυτή να αντιμετωπιστεί σαν query χωρίς φίλτρο.

## Groundtruth

- Ο υπολογισμός των groundtruth αρχείων για τα filtered queries γίνεται με brute force μέσω του groundtruth.cpp και αποθηκεύει σε αντίστοιχο .bin αρχείο τα indices των αντίστοιχων groundtruth γειτόνων για κάθε query.
- Επειδη η main περιμένει να βρεί τα συγκεκριμένα αρχεία έχουμε ανεβάσει το groundtruth.bin που αφορά τα dummy αρχεία και αναμενουμε να ανέβει συντομα και togroundtruth1m.bin που αντιστοιχει στα contest1m αρχεία.

## Main

Η main λαμβάνει τα ορίσματα dataset\_name(-f), k (-k), L (-L), R (-R), a (-a), vamaana\_type (-v) και Rstitched (-Rst) με τα αντίστοιχα command line arguments with flags, μέσω της συνάρτησης get\_arguments. Το πεδίο dataset\_name παίρνει τιμές dummy, small, large, contest1m, contest10m. Το πεδίο vamaana type είναι υποχρεωτικά είτε simple, είτε filtered είτε stitched, ενώ

τα πεδία Rstitched και a δεν είναι υποχρεωτικά (το a έχει default τιμή  $a = 1.2$ ). Έχοντας λάβει τα παραπάνω ορίσματα και έχοντας κάνει τους κατάλληλους ελέγχους, η main διαβάζει (ή δημιουργεί) τον αντίστοιχο γράφο (ο οποίος αποθηκεύεται στον φάκελο bin με postfix \_graph.bin και το αντίστοιχο medoidMap αν υπάρχει με postfix \_medoid\_map.bin). Σε όλες τις περιπτώσεις θεωρούμε ότι το αντίστοιχο αρχείο groundtruth υπάρχει στον φάκελο sift/ , όπου υπάρχουν και τα αντίστοιχα datasets.

## Medoid Map

- Το medoid map στις περιπτώσεις των Filtered και Stitched Vamana υπολογίζεται και αποθηκεύεται (και διαβάζεται αντιστοίχως) σε ένα αρχείο filename\_medoid\_map.bin.