M2 – NSD (Practical Work 5)

Densest subgraph

Maximilien Danisch, Marwan Ghanem, Lionel Tabourier

Exercise 1 — *k-core decomposition*

Implement an efficient algorithm to compute the k-core decomposition (that is to compute a k-core ordering and the core value of each node in the graph).
What are the core values of the following graphs:

- `http://snap.stanford.edu/data/com-Youtube.html`
- `http://snap.stanford.edu/data/com-Orkut.html`
- `http://snap.stanford.edu/data/com-LiveJournal.html`

For the tree graphs, give (i) the average degree density, (ii) the edge density and (iii) the size of "a densest core ordering prefix"[1].

Exercise 2 — *Graph mining with k-core*

Download the google scholar dataset at:
`https://drive.google.com/open?id=0B6cGK503Ibt0dXA3Z2lJcHlLX28`.
Download two files: (i) the list of undirected co-authorship links and (ii) the names of the authors (corresponding to each node ID).
Using the google scholar dataset, make a plot similar to the ones shown on slide 11 of the course. Try to find some "anomalous" authors.

Exercise 3 — *Densest subgraph*

Make an efficient implementation of the algorithm given in slide 17.
Fix the number of iterations t to 100. For the three graphs given in exercise 1, give (i) the average degree density, (ii) the edge density and (iii) the size of a densest prefix for a non-increasing density score ordering.
Compare these values to the ones obtained in exercise 1.

Exercise 4 — *(Optional) Graphs not fitting in main memory*

Implement the algorithm without storing all edges in main memory, but reading them from disk several times.
Fix the number of iterations t to 10. Download the graph at `http://snap.stanford.edu/data/com-Friendster.html` and give (i) the average degree density, (ii) the edge density and (iii) the size of a densest prefix for a non-increasing density score ordering.

Exercise 5 — *(Optional) Triangle densest subgraph*

Make an efficient implementation of the algorithm you guessed in slide 20 leading to "a triangle density score".
Fix the number of iterations t to 100. For the three graphs given in exercise 1, give (i) the average degree density, (ii) the edge density and (ii) the size of a triangle densest prefix[2] following a non-increasing triangle density score ordering.
Compare these values to the ones obtained in exercise 3.

---

[1] Meaning a subgraph with the highest average degree density among the subgraphs induced on the p first nodes of a core ordering for any p

[2] Meaning the prefix maximising the fraction between the number of triangles and the number of nodes