Densest subgraph

Maximilien Danisch

In this practical, we use the following technical vocabulary.

- The average degree density of a subgraph refers to the value of its average degree divided by 2 (that is its number of edges divided by its number of nodes).
- The size of a subgraph refers to its number of nodes.
- The densest subgraph is the subgraph with maximum average degree density and maximum size.
- The edge density of a subgraph refers to the number of edges divided by the maximum number of edges that could exist between the nodes of that subgraph.

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).
For each of the following 4 graphs, give (i) the core value of the graph, as well as (ii) the average degree density, (iii) the edge density and (iv) the size of a densest core ordering prefix[1]

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

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.

[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.