

Data Structures and Algorithms

Project Assignment

Patrick De Causmaecker¹, San Tu Pham²

1 Patrick.DeCausmaecker@kuleuven-kulak.be

2 san.pham@kuleuven-kulak.be

November 2016

1. Project

In this project, the following papers are given:

1. Morone, Flaviano, and Hernán A. Makse. "Influence maximization in complex networks through optimal percolation." *Nature* (2015).
2. Morone, Flaviano, et al. "Collective Influence Algorithm to find influencers via optimal percolation in massively large social media." *arXiv preprint arXiv:1603.08273* (2016).

The first paper presents the Collective Influence Algorithm to solve the optimal influence problem in massively large social media while the second one presents the linear-time implementation of the algorithm.

Students are expected to:

- Read and comprehend the papers.
- Implement the algorithm based on the second paper.
- Conduct experiment on big data set.

2. Tasks

The project's plan and requirements are as followed:

a. Comprehension:

In the first phase, a clear understanding of the papers is important. Students will work in groups (from 2 to 3 students per group) to read and discuss the papers together. The aim of this phase is to comprehend the algorithm and its implementation. Some background knowledge on graph theory might be required to understand the algorithm thoroughly. However, some difficult details can be ignored. The most important thing is to understand how the algorithm works and how the authors implement it. The first paper and its supplementary information are solely devoted for the algorithm while the second one presents its detailed linear-time implementation.

On 24.11.2011, 09:00 – 10:30, C615, the first meeting will take place. Each group will do a 20-minute (at most) presentation about the algorithm. The aim of this meeting is to discuss the papers and make sure everyone fully understand the algorithm.

b. Analysis and design

From this phase, students will work individually.

In this phase, students need to analyse the algorithm and design the data structure. Keep in mind that this step is very important since it decides the efficiency and flexibility of the source code.

The second (individual) meeting will take place on 1/12, 14:00-16:00, A233. Each student will have 15 minutes to talk about their design. Prior to the meeting, a report should be sent to TA's mailbox. The report should include flowcharts and/or UML model diagrams and/or pseudo code to present the design of the implementation. During the meeting, TA will discuss with students about their design. A revised report based on those discussions should be submitted before the next meeting.

c. Implementation

There are 2 individual meetings during this phase: 8/12, 15/12, all from 14:00 to 16:00 at A233. During this phase, students can send an email to the TA to ask for help in need.

d. Report and presentation

After finishing the implementation and experiment, final report and presentation should be prepared. In the final report, students will summary their work, analyze their implementation and experimental result. The final presentation should be sent to the TA the night before the final session. In the final session (19 or 20 December), each student will have 20 minutes to present their work.

After the final presentation, the following documents should be submitted before 6/1/2017:

- The source code (including a README file explaining how to compile and run your source code, how to read your output files, etc).
- A revised report and presentation based on feedback from Prof. De Causmaecker and the TA during the final presentation.

3. Evaluation

Students will be assessed based on the following criteria:

- Presentation and discussion
 - Understanding the papers' content
 - Discussion and proposals during meetings
 - Final presentation
- Implementation
 - Efficiency
 - Completeness
 - Accuracy
 - The code is well-designed, easy to read and flexible
- Experiments