

Title of your report

S. Tudent (12345678)

In these guidelines we provide a more detailed idea of what is expected of your experiments and of the report. For each section there is a rough indication of how much text is expected and which figures. Depending on your results certain parts may require more or less text or more or fewer figures. The main goal of your experiments is to study the performance, running time, and number of rounds of one of the MapReduce algorithms that is presented in the class. We have learned two algorithms for the MST problem. The first algorithm is for dense graphs and needs $O(1/\epsilon)$ communication rounds. The second algorithm is for general graphs and needs $O(\log(n))$ rounds. You can choose either one that you think you can implement better and do your experiments.

1 Introduction

[1-2 paragraphs] A brief introduction of what the goal of your experiments is.

2 Data sets and experiments

[3-4 paragraphs, with images of data types] A description of your data sets and experimental setup (which algorithms are you running and with which parameters). Also briefly explain how these help you answer your research question.

3 Experimental evaluation

[2-3 paragraphs with figures] A presentation of the results of your experiments as well as a brief discussion of them. The presentation would generally be in the form of graphs, plots or tables. Be sure to focus on the interesting parts. In the discussion you consider what the results imply and discuss any potential irregularities or unexpected results.

Depending on your research question, you can focus on one or several of the following measures depending on your research question.

- Convergence rate of the algorithm: How many communication rounds your algorithm needs before finishing?
- Cluster quality: What is the output of your algorithm for each data set that you use with respect to the true clustering of that data set. In case you have time, you can compare the output of your algorithm against the true clustering of the datasets using measures such as *precision*, *recall*, and *F-score*.
- Actual running time: How long does your algorithm take?

Note that for algorithms that use randomness, you may need to run the same experiment multiple times and use an average value to obtain a reliable value for the convergence rate or the score. You should of course then also take this into account if you are comparing efficiency or runtime.

4 Concluding remarks

[1-2 paragraph] A summary of the main conclusions of the experiments. Do your results help answer (part of) your initial research question?