

MoSSo: Lossless Graph Summarization in Fully Dynamic Graph Streams (Software User Guide)

1 General Information

- Version: 1.0

2 Introduction

MoSSo (**M**ove if **S**aved, **S**tay **o**therwise) is an algorithm for lossless summarization of fully dynamic graphs. MoSSo has the following advantages:

- **Fast and 'any time'**: processing each change in near-constant time, up to **7-orders of magnitude faster** than running state-of-the-art batch methods.
- **Scalable**: summarizing graphs with hundreds of millions of edges, requiring sub-linear memory during the process.
- **Effective**: achieving comparable compression ratios even to state-of-the-art batch methods.

3 Installation

- This package requires that OpenJDK 12 or greater be installed in the system and set in PATH.
- for compilation, type `./compile.sh`
- for running algorithms, type `./run.sh`
- for demo, type `./demo.sh`

4 Input File Format

The input file lists the additions and deletions in an undirected and unweighted graph in the order that they arrive. Each line corresponds to an edge addition or deletion. Each line consists of a source node id, a destination node id, and an indicator (1 for addition and -1 for deletion), which are integers separated by a tab. Additionally, we assume that there are no parallel edges. That is, if an edge has been added and has not been deleted yet, the same edge cannot be added. See *example_graph.txt* for an example input file.

5 Output File Format

The output file contains information about supernodes (S), superedges (P) and corrections (C^+ , C^-). The first line of the file contains the number of nodes $|V|$ and the number of edges $|E|$ of the input graph. Each of the next $|V|$ lines contains two integers v and S_v . v is the index of a subnode in the input graph and S_v is the index of the supernode contains v if $\deg(v) > 0$, -1 otherwise. Each of the next $|P|$ lines contains two integers S_u and S_v , which means $\{S_u, S_v\} \in P$. The next line is “-1 -1”. Each of the next $|C^+|$ lines contains two integers u and v , which means $\{u, v\} \in C^+$. The next line is “-1 -1”. Each of the next $|C^-|$ lines contains two integers u and v , which means $\{u, v\} \in C^-$. The next line is “-1 -1”.

6 Running

6.1 Running MoSSo-Greedy

```
./run.sh input_path output_path sgreedy interval
```

- *input_path*: path of the input file. See Section 4 for the detailed format of the input file.
- *output_path*: path of the output file. See Section 5 for the detailed format of the output file.
- *interval*: parameter for printing the compression ratio and accumulated execution time after processing every *interval* changes.

6.2 Running MoSSo-MCMC

```
./run.sh input_path output_path mcmc interval
```

- *input_path*, *output_path*, *interval*: see Section 6.1.

6.3 Running MoSSo-Simple

```
./run.sh input_path output_path simple escape_probability number_of_samples interval
```

- *input_path*, *output_path*, *interval*: see Section 6.1.
- *escape_probability*: parameter for the escape probability.
- *number_of_samples*: parameter for the number of samples per each update.

6.4 Running MoSSo

```
./run.sh input_path output_path mosso escape_probability number_of_samples interval
```

- *input_path*, *output_path*, *interval*, *escape_probability*, *number_of_samples*: see Section 6.3.