SLUGGER: Lossless Hierarchical Summarization of Massive Graphs (Software User Guide)

Kyuhan Lee* (kyuhan.lee@kaist.ac.kr),
Jihoon Ko* (jihoonko@kaist.ac.kr),
Kijung Shin (kijungs@kaist.ac.kr)

1 General Information

- Version: 1.0

2 Introduction

SLUGGER (**S**calable **L**ossless S**u**mmarization of **G**raphs with Hi**er**archy) is a lossless hierarchical graph summarization algorithm that has the following superiorities:

- *Effective*: yielding up to 29.6% more concise summaries than its state-of-the-art competitors.
- *Fast*: summarizing a graph with 0.8 billion edges in a few hours.
- *Scalable*: scaling linear with the number of edges in the input graph.

3 Installation

- In order to compile all the tools, it requires OpenJDK 11 or later be installed in the system.
- For compilation (optional), type *./compile.sh*.
- For demo (optional), type *make*.

4 Input File Format of SLUGGER

SLUGGER assumes that the input graph $G = (V, E)$ is undirected without self-loops. Thus the format of an input file is as follows. Each line represents a single edge. Each edge $\{u, v\} \in E$ joins two distinct nodes $u \neq v \in V$, separated by a tab. Each node $v \in V$ is assigned to a unique integer id. The format of the example file is given below.

4.1 Example File 'input.txt'

```
1       3
2       3
2       4
```

The example 'input.txt' file consist of 4 nodes with 3 edges.

5 Output File Format of SLUGGER

The output consists of three files: `H.txt`, `Pminus.txt`, and `Pplus.txt`, which corresponds to (a) hierarchy trees of supernodes, (b) positive edges between supernodes, and (c) negative edges between supernodes, respectively. For all three files, each line represents a single edge of each kind. To distinguish leaf nodes (input node's unique id) from the internal nodes and root nodes, we prefix the non-leaf node ids with 'n' to differentiate them from the leaf node ids. The format of the example file, especially `H.txt`, is given below.

5.1 Example File 'H.txt'

```
1       n90
2       n90
3       n90
.
.
.
n90     n120
n94     n122
.
.
.
n120    n122
```

6 Running SLUGGER

6.1 How to Execute

./run.sh input_path optimize

6.2 Parameters

- *input_path*: Path to the input text file which follows the format described above.
- *optimize*: Boolean value whether to run with the SLUGGER-optimized version or not.