CUTNPEEL (Software User Guide)

1 General Information

- Version: 1.0

2 Introduction

CUTNPEEL is a fast search algorithm for a high-quality set of near bi-cliques. CUTNPEEL has the following advantages :

- *High Quality*: provides near bi-cliques o up to 51.2% better quality than the second best one.

- *Speed*: up to $68.8\times$ faster than the competitors that is second best in terms of quality.

- *Scalability*: empirically scales near linearly with the size of the input graph.

- *Applicability*: successfully applicable to lossless graph compression and interesting pattern

3 Installation

- This package requires java 1.8 or greater be installed in the system.

- For compilation (optional), type *./compile.sh*.

- For demo (optional), type *make*.

4 Input File Format

CUTNPEEL assumes that the input graph G is dynamic graph, whose edge has a source node, a destination node, and time. G can contain self-loops and to make undirected graph, insert symmetric edge by changing destination and source of each edge. Thus the format of an input file is as follows. Each line represents a single edge. Each edge $\{s, d, t\} \in E$ joins

source node $s \in S$, destination node $d \in D$, time $t \in T$, and integer 1 as existence sign, separated by comma.

example_data.txt is an example of the input file. The entire dataset used in the experiments can be found at `https://www.dropbox.com/sh/ag4ghglt04g7cg8/AADrC5OD7zQiPWdhDFpJWsuC` dl=0.

5   Output Files Format

The output consists of three files: `bicliques.txt`, `missingE.txt`, and `remainingE.txt`. The details are following:

- `bicliques.txt`: this file lists the set of exact bi-cliques. Each bi-cliques are described within 4 lines, 1) the size and the actual number of edges in corresponding near bi-clique, 2) source nodes, 3) destination nodes, and 4) time.

- `missingE.txt`: this file lists the set of missing edges, which are included in exact bi-cliques B but not in original graph G. Each line consists of an edge, (src,dst,time).

- `remainingE.txt`: this file lists the set remaining of edges in G that do not belong to any bi-cliques. Each line consists of an edge, (src,dst,time).

6   Running CutNPeel

6.1   How to Execute

```
./run_cutnpeel.sh input_path threshold_decrement iteration output_path
```

6.2   Parameters

- *input_path*: Path to the input text file which follows the format described above.

- *threshold_decrement*: Double value, the rate of threshold decrement.

- *iteration*: Integer value, which is the number of iteration.

- *output_path*: Name of directory to write output files.