SubMatch

Overview

This tool aims to match any given query metagraph (i.e., to compute the instances of the metagraph) over a large input graph. For the definition and examples of metagraphs, refer to the citation below. Currently, the tool only works on undirected graphs, where nodes are typed (but edges are not typed).

Citation

Y. Fang, W. Lin, V. W. Zheng, M. Wu, K. C.-C. Chang and X. Li. Semantic Proximity Search on Graphs with Metagraph-based Learning. In ICDE 2016, pp. 277--288. [PDF]

Code Download

Module	Requirement	Comment	Link
Operating System	Windows	nil	nil
Runtime DLL	MinGW	Only the following DLLs are needed: **libgcc-4.5.2-1-mingw32-dll-1** **libstdc++-4.5.2-1-mingw32-dll-6** Downloading newer versions may not work. Extract the two DLLs into the same directory as SubMatch.exe or any directory in PATH environment.	Download
Main Program	SubMatch	Code author: Wenqing Lin. Sample data are also included in the download.	Download

Usage

```
SubMatch.exe mode=2 data=<String> query=<String> maxfreq=
<Integer> subgraph=<String> stats=<String>
```

data=<String>
The input graph filename. The file is in the Labeled Graph Format. The graph is treated as undirected, and edge types are not considered at the moment.

query=<String>
The input filename for a list of query metagraphs, in the Metagraph Query Format. These query metagraphs can be mined from the input graph using a modified version

of GRAMI.

maxfreq=<Integer>
The maximum number of instances to match, for each query metagraph. The program immediately moves on to the next query after the specified maximum number of instances are found.

subgraph=<String>
The filename to output the metagraph database, which contains a list of processed metagraphs. The file is in the Metagraph Database Format.

stats=<String>
The directory name to output matched instances of each metagraph.

- One instance file per metagraph.
- Instance filenames are named according to the ID of each processed metagraph (see **subgraph**).
- Each line in an instance file representing one instance, containing tab delimited **NodeID**'s of the input graph, in the order according to the order of nodes in the processed metagraph (see **subgraph**).

Sample Data

Sample data of two input graphs and their correponding metagraph queries are included, which are also used in our citation above. They are derived from SNAP's Facebook data and Forward's LinkedIn data.

Disclaimer

We provide any code and/or data on an as-is basis. Use at your own risk.