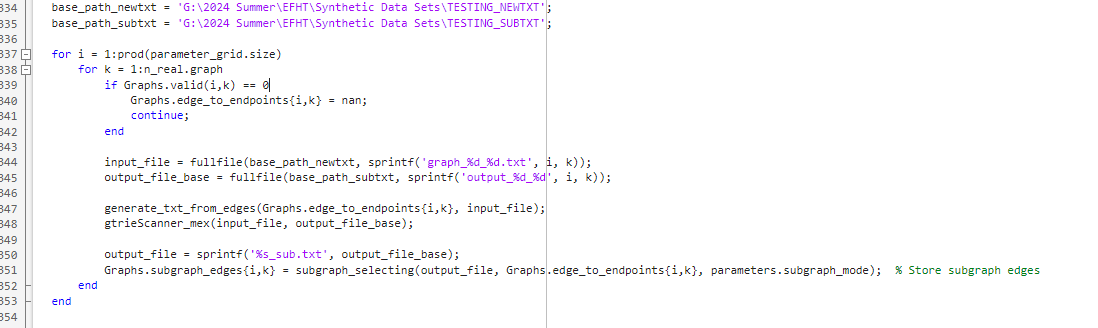
Egde Flow Hypothesis Test (EFHT) Project, Datahub Requirement

Overview of the workflow:

Graph generation -> output edge txt file-> use the external software to classify the subgraph we need by using the output edge txt file -> it will output two files for each output edge txt file used by the software -> Flow generation -> perform tests when each flow is generated

The problem that stops my lab members from running the whole program by ourselves is this part in the Random\_Flow\_Sampler.m



This is the process after the graph generated, we are running the loop with parameter i and j to output the graph edge txt file (will based on graph1.1,graph1.2,graph1.3 and so on) which means it needs a place to store. And in the loop after one txt file is generated, calling the external software to classify the subgraph we need by using the output graph edge txt. Then use the subgraph\_selecting funciton to store the subgraph edges into parameter Graph.subgraph\_edges{i,k}.

For the base\_path\_newtxt, and bath\_path\_subtxt is the path( folder) where we store the files. As we ran the program before, each time these two folder will take 2 GB storage space.

The external software part:

gtrieScanner: It is the software used for looking for specific subgraph in the graph. It needs to be run in the Linux system. And the command we use is “./gtrieScanner -s 4 -g {input\_file} -u -o {output\_file\_base}.txt -oc {output\_file\_base}\_sub.txt -f simple.”

The software description link is :http://www.dcc.fc.up.pt/gtries/

gtrieScanner\_src\_01.zip: It’s the software zip file

And we may need to change the number 4 in this command when we need to look for different subgraph type. [but it’s not our priority now]

MATLAB functionality uesd in this wrapper process is

Flow\_Examples.m: This is the frame function which contains all the steps we need to do for the workflow. So all the function files are designed for supporting for this file to do the workflow. I’ve put the part that we need to do the wrapper between the graph generation and flow generation.(which is the part in the picture shown above)

generate\_txt\_from\_edges.m :This is function file that do the step: generate txt file from the generated graph.

subgraph\_selecting.m: A key part of the subgraph extraction logic, this script selects subgraphs based on a given set of input graphs and stores the selected subgraphs for further processing.

And we may need to add more subgraph type selecting logic in this function file when we need to look for different subgraph type. [but it’s not our priority now]

gtrieScanner\_mex.cpp, gtrieScanner\_mex.mexw64 : The mexw64 file is generated by complieing the cpp file in matlab. The software needs to be run in the Linux system. And the software is put in Linux Ubuntu in my laptop. So the file is designed for calling the software and run the command I mentioned above in the windows system by calling it in Ubuntu. If you can help us do the wrapper, so it should be a similar function m file with (input\_file, output\_file\_base) as input to run the software in this wrapper.

Our priority now is to do the part in the picture as a wrapper like a function file so we can use the wrapper instead. So we can all run the entire program by ourselves.

If there is any information needs to be added or explained, please let us know!

Thank you!