

GRAPH PROCESSING Using SNAP

Alejandra Tubilla Castellanos

Instituto Tecnológico de Estudios Superiores de Monterrey

Abstract

This paper provides instructions on how to use and import a dataset from the SNAP (Stanford Network Analysis Platform) library.

It also includes directions on how to use the imported dataset to export the graph into different formats, such as GraphML, GEXF, GDF and JSON Graph Format with GraphSON in *C++*.

And once exported how to visualize them using Gephi.

2.Introduction

In computer science a Graph is a type of data structure that consists of a finite set of points known as nodes and a set of lines known as edges which connect the nodes in the graph.

Stanford Network Analysis Platform (SNAP) is a library which is being actively developed since 2004. This library can analyze hundreds of millions of nodes and billions of edges, which can be scale to massive networks.

3.Importing Dataset

For the purpose of this paper, it is relevant to mention the Facebook graph sample dataset from the library mentioned before, is being used to fulfill all needs of development. If needed,

visit <https://snap.stanford.edu/data/egonets-Facebook.html> to obtain this dataset. You may also use the one you see fit, just remember to make the necessary changes.

To import the Facebook data set you will need to download the

facebook_combined.txt.gz, and unzip it in the folder where you will be using it. It is important to remember you will be using SNAP library, for this reason it is important to download the complete library if you have not done that already.

Create your main file (main.cpp), in it you will be importing the graph. You will need to include the following line:

```
TSnap::LoadEdgeList<DGraph>("facebook  
_combined.txt",0,1);
```

Beforehand you will need to declare a new PN-Graph variable. After you have done this, now you have imported the graph.

4.Exporting files

At this point you have now imported the dataset into our file, so you can continue to export the graph into the different formats, to later on visualize them using the GEPHI tool.

To export the graph, you will need to create a function for each of the different formats. Inside each function, crate a new **ofstream** file giving it a name

and extension, for example ***facebook.graphml***. After you have created the file, with the help of a cycle, print all the information contained in the file.

Important: each format has a different file

All four algorithms to export the file in different formats, have a temporal complexity of $O(N+E)$. (N = nodes and E = edges), and a Spacial complexity of $S(1)$.

The execution time of each of the different functions are as follows:

- GraphML = 23.219 miliseconds
- GEXF = 23.328 miliseconds
- GDF = 12.964 miliseconds
- JSON = 21.092 miliseconds

The function that had the best execution performance, was the one exporing into a GDF file format, while the GEXF had the slowest execution time.

5.Graph visualization

After all files are exported, inicialize Gephi. Once it starts, you will see a welcome window, select ***“Open graph file”***. Go to your directory where you have all the exported files and select one. Once you have selected the file, a new window will appear with all the graph information. After you select ***“Ok”*** your graph will be created.

6.Advantages and Disadvantages of Graph viewers

The main advantage of using a graph viewer, is the easy manipulation and visualization of graphs. This type of tools allows you to see not only the graph, but also all the data stored in them. It also allows you to aply filters to visualize the information you need in each case.

Figure 1: Facebook dataset graph



At the moment, most tools as it is the case of Gephi, do not support hipergraphs, making it its main disadvantage.

7.Code

If you wish to see the code, it is available in a github repository, just go to <https://github.com/alejandraturub/Algoritmos/tarea4>. There you will find all exporting functions and files created after the export.

References

- [1] *GEFX Working Group, GEFX File Format, Gephi Community Project, 2009.*
- [2] *GraphML team, The GraphML File Format, Graph Drawing, 2016.*
- [3] *Jbmusso, GraphSON Reader and Writer Library, Github, 2016.*
- [4] *Gephi, GDF Format, Gephi, 2017.*
- [5] *Lyx, Lyx the Document Processor, Lyx, 1995.*
- [6] José Federico Medrano, José Luis Alonso Berrocal and Carlos G. Figuerola, Visualización de Grafos Web, Universidad de Salamanca, 2011.
- [7] Doxygen, SNAP Library 4.0, User Reference, Stanford University, 2017.
- [8] Jure Leskovec, Stanford Large Network Dataset Collection, Stanford University, 2004.
- [9] *Javier Bezos, LaTeX, Wikipedia, 2017.*
- [10] *. LaTeX Project, LaTeX – A document preparation system, The LaTeX Project, 2017.*
- [11] *Ulrik Brandes and Christian Pich, GraphML Transformation, University of Konstanz, 2010.*
- [12] *yWorks GmbH, Graph Exchange and Graph Export Formats, yWorks, 2016.*