Graph prossesing [*]

Graph prossesing by using the SNAP library [†]

Margot Duek[‡]
Instituto Tecnologico de Estudios Superiores de Monterrey
Mexico City
margotduek@gmail.com

ABSTRACT

This paper provides a set of instructions of how to import a dataset that includes a sample of a Facebook graph from the SNAP (Stanford Network Analysis Platform) library which was created for analysis and graph mining.

This also includes the steps of how to export the graph to GraphML, GEXF, GDF and JSON Graph Format with GraphSON afterwards we export the graph with each of the formats to Gephi.

By reading these paper you will also find how to create a new document on LaTeX using the ACM format which is being used on these paper.

Keywords

ACM proceedings; LaTeX; text tagging

1. INTRODUCTION

A graph is a data structure with a finite set of vertex or nodes that get together by edges, there are directed graphs, that are the graphs wich edges point to a direction, and undirected graphs, which edges point in both ways.

Stanford Network Analysis Platform is a library being actively developed since 2004, this is a very efficient library wich can analyze hundreds of millions of nodes and billions of edges, this can scale to massive networks and is written in c++.

2. GRAPHS

We have already talked about what are graphs, what is SNAP and what can we do when them, but we haven't talked about what can we do with these tools or how to use them.

Is important to say that we will be working with the facebook graph sample, but you can do it with any graph, just remember to change the name of the file every time we need to make a route.

2.1 Importing the Facebook graph

The first thing we are going to talk about is how to import a dataset that includes a sample of a Facebook graph. You need to get into `https://snap.stanford.edu/data/index.html` and choose which graph you want to use, in these case, we will get into social networks directory and inside there we will choose Facebook, remember you can choose anything.

You will need to download the facebook_combined.txt.gz, unzip it in the folder

where you want to use it (Don't forget to add the SNAP library in the directory where you will work) and when you see the file, you will see a bunch of numbers, those numbers are the graph.

Crete your main.cpp, there you will have to import the graph and check if it was imported successfully, inside the main, you will need to put a line that should look like these, `PUNGraph g=TSnap::LoadEdgeList<PUNGraph>("../facebook_combined.txt",0,1);` where you are declaring a new PUNGraph variable and giving to that variable the Facebook graph, now we have imported the graph.

The only way to know if something really worked is by checking it, we will print the amount of nodes that we have on the graph by printing the name of your graph.GetNodes(), in this case, g->GetNodes(); and it it prints the amount of nodes that the SNAP's website says that means you have successfully imported the graph.

2.2 Exporting the graph

Now is time to talk about how to export the graph, we will export the graph in four different ways, GraphML, GEXF, GDF and JSON Graph Format with GraphSON, each of them have their own implementation but these time we will talk about the general process we should make and after that, you should be able to make the same thing with the other ones.

Basically what we need to do, is to make a new ofstream file and name it with the name you like and we will start printing what we want inside that file(in each case we have a different file to export the graph in each format), this method should have a condition or a cycle so that all the nodes can be printed of that file.

GraphML is the less efficient format of exporting from the

[*] (Produces the permission block, and copyright information). For use with SIG-ALTERNATE.CLS. Supported by ACM.

[†] A full version of this paper is available as *Author's Guide to Preparing ACM SIG Proceedings Using LaTeX2$_\epsilon$ and BibTeX* at `www.acm.org/eaddress.htm`

[‡]

© 2015 ACM. ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

four we tasted, it took 26 milliseconds with a linear temporal complexity in the worst of the cases because it have two for's inside of an if.

Executing GEXF function took us a total of 22 seconds, this also have in the worst of the cases a temporal complexity that is linear because inside of an if we have two separated for's to convert each node.

The total time of converting the facebook sample graph to GDF format was 12 milliseconds, which makes these function the one with the best time. These function have the same temporal complexity in the worst of the cases as the previous ones because is almost the same thing to do.

The last function implemented was to export the graph to a JSON Graph Format with GraphSON which also has a very good time, we have got 16 seconds. The implementation of these function is based on the same architecture than the other ones, an if condition and two for cycles to export each node.

2.3 LaTex

A very big problem for people who is not used to designing and thinking of how should things go on an investigation paper is that we have to think where to put each of the things in the document, like the date, the name, personal information and other things.

For all the people that is most of all interested in the real content, you can use LaTeX , this is a tool that allows people to just write content and don't worry about designing problems, and let the designers take care of that.

The only thing we should do is (if you are in Linux, if you are not, you should find a paper with a tutorial for your operating system) install any tool for LaTeX , in my case I used Texlive, after doing that, you should open your favorite text editor and start writing your paper, in my case I used a template from ACM and I edited it.

To see what you have created, you need to make your .tex file into a .pdf by doing ctrl + Alt + B, if when you do that nothing happens, that means you have a compilation error, you should check the syntaxis of everything that has been written in the document.

2.4 Gephi

Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.

2.5 The code

You can find the code in https://github.com/margotduek/AnalisisDisenioDeAlgoritmos/tree/master/tarea4/prueba which is a directory on a GitHub repository.

3. CONCLUSIONS

By using any library, in this case, we used SNAP but there are a lot of libraries that we can use for any subject or a project, we can make bigger and best applications in less time and in some cases, using a library could be more efficient because of the way the algorithms are programmed.

4. ACKNOWLEDGMENTS

In this investigation and while using the graph library I learned a lot of things, especially I learned how to use LaTeX for a paper with the ACM standard or for anything that I need to make an investigation paper ir report.

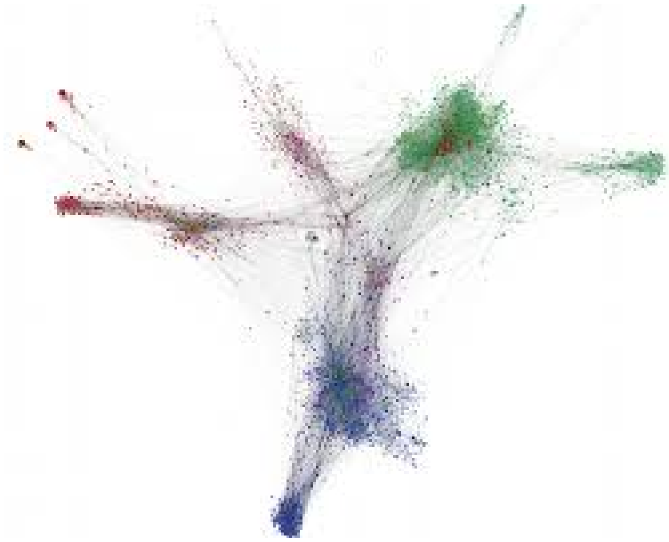

Figure 1: A sample Gephi graph

I think this is a very useful tool for all my career and all my life because for me, all my life it has been very difficult to find out where to put each information, and by using LaTeX is easier and faster for me to make my reports.

I also learned how easy is to use a library, I think in most of the cases, everyting you need is time because you need to read a lot of things but if you read all the documentation you can do almost everything.

5. REFERENCES

APPENDIX

A. HEADINGS IN APPENDICES

A.1 Introduction

A.2 Graphs

A.2.1 Importing the Facebook graph

A.2.2 Exporting the graph

A.2.3 LaTeX;

A.2.4 Gephi

A.2.5 The code

A.3 Conclusions

A.4 Acknowledgments