Counting Triangles In Large Graphs

Panagiotas Ioannis

August 2016

Dissertation submitted in partial fulfillment for the degree of
Master of Science in *(Big Data)*

Computing Science and Mathematics
University of Stirling

Abstract

In this thesis we dealt with the problem of counting triangles in graphs. Though this is a well studied problem in Computer Science there currently is a need to develop new algorithms that can scale well as the graphs grow in size. Standard algorithms are not very efficient when they need to run for large graphs that contain many millions or even billions of edges. Such large graphs are becoming more and more common nowadays and in fact there are many situations where we might be required to find the number of triangles they contain.

In the largest part of the thesis we present different methods and approaches to solve this problem that are well accustomed for real works. We begin by describing distributed algorithms that run in the Map-Reduce environment. Afterwards we describe approximation algorithms that are applied mostly in the streaming model of computation. These algorithms do not return an exact algorithm but are still usually able to return a number that is very close to the real answer. It is quite probable that a fast approximation that is very close to the real answer would be preferred over an exact yet slow algorithm in most real life situations.

In the last part of the thesis we conduct experimental analysis on the algorithms to try and evaluate their performance. This is done in the spirit of verifying the author's results as the source code used in the original papers for most of the algorithms is not available online. Due to technical limitations we had to focus solely on approximation algorithms.

In the majority of tests we have been able to verify the claims that the authors make. This apart from validating the author's findings also verifies that our codes have been correctly developed.

In the end, we have developed a collection of algorithms in java that are able to analyse large graphs in a very fast amount of time and provide quite accurate results in most cases.