INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER, UK.

HEFT – AN INTEGRATED FRAMEWORK FOR NETWORK COMPUTATION

A dissertation by

Nageswaran Keshan – 14448696 (2012034)

Supervised by

Mr. Sriyal Jayasinghe

Submitted in partial fulfillment of the requirements for the
BSc (Hons) Software Engineering Degree
Department of Computing

April 2016

Supervisor Signature: …… Date: ……..

Abstract

The importance of data has risen in the past decade due the evolvement of technological innovation for user needs. Large scale of real time, rapid changing datasets are being created due to an exceptional evolution of communication networks, social networks and internet of things. The above statement shows network analysis has become popular and important where this lead to creation of various specialized graph systems for network analysis. This paved way for many data-parallel frameworks to incorporate them. However, use of relational databases for network analysis is ignored though most data is still collected and managed in relational databases.

This situation of ignoring relational databases raises a question whether relational databases have limitation for network analytics. The relational model is inefficient for network analysis where it will take many expensive joins to do a computation. SQL query language also doesn't support network analysis operations but relational databases comes with great features, such as integrity constraints, fault tolerance, query optimization and secure transaction and so on.

This thesis presents an integrated framework for network analytics that consist of a data model that extends support for relational databases with network analytics. This also presents a query language to manipulate data for relational and network analytics or a combination of them. Along with that, this integrated framework also adds a query engine that is built on top of relational database (PostgreSQL) to process queries created with the query language. The testing results prove that query engine and model introduced were able to achieve equivalent or better performance in almost all scenarios.

Subject Descriptors

- **G.2.2 [Discrete Mathematics]: Graph Theory - Graph Algorithms, Graph Labelling, Network Problems, Path and Circuit Problems**
- **H.2.1 [Database Management]: Logical Design – Data models, Schema and Subschema**
- **H.2.3 [Database Management]: Languages – Query Languages**

Keywords

Graph Processing, Graph Algorithms, Big data, Iterative computation, Computing model, and Storage