Accepted Manuscript

Profiling distributed graph processing systems through visual analytics

Alessio Arleo, Walter Didimo, Giuseppe Liotta, Fabrizio Montecchiani

PII: S0167-739X(17)32382-8
DOI: https://doi.org/10.1016/j.future.2018.04.067
Reference: FUTURE 4145

To appear in: *Future Generation Computer Systems*

Received date : 19 October 2017
Revised date : 19 January 2018
Accepted date : 22 April 2018

Please cite this article as: A. Arleo, W. Didimo, G. Liotta, F. Montecchiani, Profiling distributed graph processing systems through visual analytics, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.04.067

Profiling Distributed Graph Processing Systems through Visual Analytics ☆

Alessio Arleo, Walter Didimo, Giuseppe Liotta, Fabrizio Montecchiani

Dipartimento di Ingegneria
Università degli Studi di Perugia, Italy

Abstract

Analyzing large-scale graphs provides valuable insights in different application scenarios, including social networking, crime detection, content ranking, and recommendations. While many graph processing systems working on top of distributed infrastructures have been proposed to deal with big graphs, the task of profiling their massive computations remains time consuming and error-prone. This paper presents GiViP, a visual profiler for distributed graph processing systems based on a Pregel-like computation model. GiViP captures the huge amount of messages exchanged throughout a computation and provides a powerful user interface for the visual analysis of the collected data. We discuss the effectiveness of our approach and show how to take advantage of GiViP to detect anomalies related to the computation and to the infrastructure, such as slow computing units, anomalous message patterns, unbalanced graph partitions, and links with high latency.

Keywords: Distributed Platforms, Apache Giraph, Vertex-centric Frameworks, Profiling, Anomaly Detection, Visual Analytics

1. Introduction

The analysis of large-scale graphs provides valuable insights in different application scenarios, including social networking [38, 48, 53], crime detection [27], content ranking [21, 56], and recommendations [55]. On the other hand, graph computations are often difficult to scale and parallelize, due to the inherent interdependencies within graph data. Furthermore, graph algorithms are usually iterative and hence poorly suited for popular Big Data processing systems such

☆A preliminary version of this research appeared at the 25th International Symposium on Graph Drawing & Network Visualization. Research supported in part by the project: "Algoritmi e sistemi di analisi visuale di reti complesse e di grandi dimensioni" - Ricerca di Base 2017, Dipartimento di Ingegneria dell'Università degli Studi di Perugia.

Email addresses: ingarleo@icloud.com (Alessio Arleo), [walter.didimo,giuseppe.liotta,fabrizio.montecchiani]@unipg.it (Walter Didimo, Giuseppe Liotta, Fabrizio Montecchiani)