JOURNAL OF INNOVATION IN DIGITAL ECOSYSTEMS 3 (2016) 83–97

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jides

An enhanced Graph Analytics Platform (GAP) providing insight in Big Network Data

Anastasios Drosou[a,*], *Ilias Kalamaras*[b], *Stavros Papadopoulos*[b], *Dimitrios Tzovaras*[a]

[a] *Centre for Research & Technology Hellas - Information Technologies Institute, Greece*
[b] *Imperial College London - Department of Electrical and Electronic Engineering, United Kingdom*

HIGHLIGHTS

- The paper presents the enhanced Graph Analytics Platform (GAP).
- GAP allows for data mining promoting a top-down approach for Big Data investigation.
- GAP supports a wide range of key-features, incl. clutter minimization, HR clustering.
- GAP is demonstrated on both a mobile and a social network real-world use case.

ARTICLE INFO

Article history:
Published online 15 November 2016

Keywords:
Graph analytics
Network graph
Big data
Behavioural clustering
Information visualization
Hypothesis formulation

ABSTRACT

Being a widely adapted and acknowledged practice for the representation of inter- and intra-dependent information streams, network graphs are nowadays growing vast in both size and complexity, due to the rapid expansion of sources, types, and amounts of produced data. In this context, the efficient processing of the big amounts of information, also known as Big Data forms a major challenge for both the research community and a wide variety of industrial sectors, involving security, health and financial applications. Serving these emerging needs, the current paper presents a Graph Analytics based Platform (GAP) that implements a top-down approach for the facilitation of Data Mining processes through the incorporation of state-of-the-art techniques, like behavioural clustering, interactive visualizations, multi-objective optimization, etc. The applicability of this platform is validated on 2 istinct real-world use cases, which can be considered as characteristic examples of modern Big Data problems, due to the vast amount of information they deal with. In particular, (i) the root cause analysis of a Denial of Service attack in the network of a mobile operator and (ii) the early detection of an emerging event or a hot topic in social media communities. In order to address the large volume of the data, the proposed application starts with an aggregated overview of the whole network and allows the operator to gradually focus on smaller sets of data, using different levels of abstraction. The proposed platform offers differentiation between different user behaviors that enable the analyst to obtain insight on the network's operation and to extract the meaningful information in an effortless manner. Dynamic hypothesis formulation techniques exploited

Peer review under responsibility of Qassim University.
* *Corresponding author.*
E-mail address: drosou@iti.gr (A. Drosou).

http://dx.doi.org/10.1016/j.jides.2016.10.005

by graph traversing and pattern mining, enable the analyst to set concrete network-related hypotheses, and validate or reject them accordingly.

1. Introduction

Graphs are widely used structures that serve to represent networks consisting nodes and their inter-connections called edges. They could be exploited to describe paths in a city, circuit networks like telephone and computer ones or even social networks, whereby each node is a structure and contains information like person id, name, gender, locale, etc. It becomes apparent thus, that data used in a wide range of applications can be intuitively formulated into a graph, providing a holistic view of the correlations that an entity participates and extending to both visualization and analytics fields [1]. In this respect, the term *Graph Analytics* was introduced and refers exactly to the study and analysis of data that can be transformed into such graph representation.

Graph analytics is a fast growing field in both the big data mining and the visualization community [2] that is utilized on numerous multi-disciplinary and high-impact applications, such as network security, finance, health care, etc. [3] providing users with adequate knowledge across various patterns within a given system. Despite the fact that the analysis of unstructured collections of multi-dimensional points has already been addressed in the past by several methods, graph analytic technologies form a rather recent trend and they pose many challenges concerning not only the performance of the data mining algorithms that promotes knowledge discovery through algorithmic computation, but also producing effective graph visualizations in order to enhance human perception.

The twofold nature of graph analytics can be described by the following uses-cases:

- The task of graph analytics is known beforehand. Graph mining utilizes the vast computational resources that exist today, in order to analyze large graphs fast and efficiently. Specific tasks for which graph mining has been used, include the detection of anomalous nodes or subgraphs, community detection, and object recognition.
- The task analysis is not known beforehand. Graph visualization includes the human cognition in the analysis loop, and enables the interactive exploration of graph related datasets. It is used in many cases where the human may simply examine the data to learn more about it, gain insight about it, or make new discoveries.

In the modern digital ecosystem, the need for different advanced graph analytics to make something of that data becomes more than essential as the sources, types, and amounts of data continue to expand [4]. In this respect, the contextual impact of data and the impact of graph analytics technology on organizations seeking to discover the cause, effect, interrelate and influence of events on business outcomes needs to be defined. This way, graph analytics are not only able of identifying the key individuals in the graph and visualize them, but they also classify (via clustering) them into behavioural groups [5]. Moreover, they can detect correlations and determine their nature and their significance within the given environment.

Thus, graph analytics can be used to model all sorts of relationships and processes in a wide range of systems [6]. For instance, they can reveal patterns across varied data sets that signal the onset of cyber attacks. With a steadily increasing amount of user devices, the problem of mobile network security, i.e. monitoring a mobile phone network and identifying abnormal and malicious behavior, is nowadays becoming even more challenging. The number of connected mobile devices is expected to increase even more in the next years, including diverse types of data, such as those originating from Internet of Things (IoT) devices. The vast number of mobile phone subscribers, communicating every day, results in a huge amount of signaling and billing records, containing multiple and diverse types of information. This constant flow of information from multiple sources renders the problem of mobile network security as a *Big Data* problem [7], posing the challenge of how to reduce the amount of information and focus on the useful aspects. For instance, malicious individuals may be tempted to launch Denial of Service (DoS) attacks and affect network security and performance. Since the efficient detection and attribution of these anomalies are of major importance to the mobile network operators, graph analytics techniques can significantly assist in this direction. They can assist the mobile network operator to have an overview of various aspects of the whole network, while allowing her/him to explore and focus on gradually smaller subsets of the data, until the desired information is reached and a decision is made.

In the same context, graph techniques could also identify the root (cause) of surrounding or bigger events, i.e. this could help in finding the most influencing people in social media. Alternatively, graph analytics could help in the identification of communities that revolve around a certain theme, i.e. detect patterns in communication that might indicate a threat to national defense by identifying groups of people who have been communicating about terroristic events, something security agencies and/or authorities might be interested in.

Building upon all aforementioned issues, this paper presents a Graph Analytics Platform (GAP) that offers a rich toolkit for and is built upon the concept of an interactive data mining framework [8] that follows a top-down approach, aiming at the detection and attribution of abnormal/suspicious events or patterns in a wide range of network structures. Last but not least, its value and applicability is demonstrated in two different real world reference use-cases.