Processing Large Graphs: Representations, Storage, Systems and Algorithms

[Tutorial Overview Abstract]

Deepak Ajwani
Bell Laboratories Ireland
deepak.ajwani@alcatel-lucent.com

Marcel Karnstedt
Bell Laboratories Ireland
marcel.karnstedt@alcatel-lucent.com

Alessandra Sala
Bell Laboratories Ireland
alessandra.sala@alcatel-lucent.com

Description

Analyzing and processing large graphs is of fundamental importance for an ever-growing number of applications. Significant advancements in the last few years at both, systems and algorithmic side, let graph processing become increasingly scalable and efficient. Often, these advances are still not well-known and well-understood outside the systems and algorithms communities. In particular, there is very little understanding of the various trade-offs involved in the usage of particular combinations of algorithms, data structures, and systems. This tutorial will have a particular focus on this aspect, imparting theoretical knowledge intertwined with hands-on experience.

Since there is no clearly winning system/algorithm combination that performs best on all the different metrics, it is of utmost importance to understand the pros and cons of the various alternatives. The tutorial will enable application developers in industry and academics, students as well as researchers to make corresponding decisions in an informed way. The participants do neither require any particular a-priori knowledge apart from a basic understanding of core computer science concepts, nor any special equipment apart from their laptop.

After a general introduction, we will describe the critical dimensions that need to be tackled together to effectively and efficiently overcome problems in large graph processing: data representation, data storage, acceleration via multi-core programming, and horizontally scalable graph-processing infrastructures. Thereafter, we will provide an overview of existing graph-processing systems and graph databases. This will be followed by hands-on experiences with popular representatives of such systems. Finally, we will provide a detailed description of algorithms used in these systems for fundamental problems like shortest paths and Pagerank, how they are implemented, and how this affects the overall performance. We will also cover basic data structures such as distance oracles that can be built on these systems to efficiently answer distance queries for real-world graphs.

Presenters

Deepak Ajwani is a research scientist at Bell Labs Ireland. He has many years of postdoctoral research experience working at Aarhus University, University College Cork (UCC) and Bell Labs Ireland. During his postdoctoral research at UCC, he has been the principal investigator of the IRCSET and IBM funded project on "Designing Graph Partitioning and Repartitioning Techniques for Robust Resource Allocation in Exascale Stream Computing Systems." His research focuses on large graph algorithms, particularly on I/O-efficient graph traversal algorithms, algorithms for multicores and graph partitioning and routing solutions for Exascale systems.

Marcel Karnstedt is a research scientist at Bell Labs Ireland. He has many years of postdoctoral research experience working at Digital Enterprise Research Institute (DERI), National University of Ireland, Galway (NUIG). During his appointment, he worked with the Unit for Information Mining and Retrieval (UIMR) in different national and European projects on analyzing and querying large graphs and networks, specifically social networks, sensor networks, and semantic graphs. Since December 2009, he also holds an adjunct lectureship at NUIG. His research focuses on large-scale data processing and data management in elastically scalable distributed systems, with a particular focus on the support of massive streaming and graph-based analytics.

Alessandra Sala is the technical manager of the "Data Analytics and Operations Research" group in Bell Labs Ireland. In her prior appointment, she held a research associate position in the Department of Computer Science at University of California Santa Barbara. During this appointment, she was a key contributor of several funded proposals from National Science Foundation in USA and her research was awarded with the Cisco Research Award in 2011. She focused her research on modeling massive graphs with an emphasis on mitigating privacy threats for Online Social Network users. Before that, she worked for two years as post-doctoral fellow with the CurrentLab research group led by Prof. Ben Y.Zhao. Her research focus lies on distributed algorithms and complexity analysis with an emphasis on graph algorithms and privacy issues in large-scale networks.