Graph TA

Program

4th Graph-based Technologies and Applications

4th March 2016

8:50	Registration
9:15 – 10:35	Welcome & Presentation session I

1 Using Evolutionary Computing for Feature-driven Graph Generation

Speaker: Merijn Verstraaten (UVA)

Abstract extract: Parallel processing is one of the desirable ways to handle the increase in scale and complexity of graph analytics. However, the increasing heterogeneity of hardware and software solutions poses significant challenges for the systematic performance analysis of different types of applications, datasets, and systems. In this context, our work focuses on the empirical analysis of the impact of graph structural features on the performance of parallel graph processing algorithms. In this talk, we discuss how we address this challenge with synthetic graphs,

2 HOBBIT: Holistic Benchmarking of Big Linked Data

Speaker: Irini Fundulaki (FORTH)

Abstract extract: A key step towards abolishing the barriers to the adoption and deployment of Big Data is to provide open benchmarking reports that allow users and developers to assess the fitness of existing solutions for their purposes. However, achieving this goal demands: (a) the deployment of benchmarks on data that reflects reality within realistic settings (b) the provision of corresponding industry-relevant key performance indicators (KPIs) and (c) the computation of comparable results on standardized hardware. Although some efforts have already been undertaken to benchmark particular steps of the Big Linked Data (BLD) Processing chain, there was no significant and coordinated effort so far that aimed to measure the fitness of current frameworks based on real data along complete BLD processing pipelines.

3 Polyglot graph databases using OCL as pivot

Speaker: Raquel Pau (Sparsity Technologies)

Abstract extract: Conceptual schemas, which are the functional specifications of information systems, are composed by a structural schema (i.e data model) usually expressed using the Unified Modeling Language(UML) and a behavioral schema (i.e the set of supported operations), which can be declaratively expressed with the Object Constraint Language (OCL). UML and OCL support the graph-property model and have enough expressive power to express hypergraphs.

4 Reactive Databases for BigData Applications

Speaker: Humberto Rodríguez Ávila (Vrije Universiteit Brussel)

Abstract extract: Software development is currently one of most active and changing fields in computer science. Especially in the modern I/O hardware, systems need to be highly responsive and support an effective way for real-time interactions with their users. In this context, Reactive Programming (RP) has emerged as a promising avenue to solve the increasing requirements of interactive and real-time systems.

5 Using WordNet to study the evolution of the polysemy

Speaker: Bernardino Casas (LARCA & MACDA)

Abstract extract: WordNet is a lexical database that contains the relationship between word meanings, also called synsets. The vocabulary and meanings that someone knows is ideally a subgraf of WordNet. Children when acquire vocabulary follow some criteria to learn some word instead of another. We hypothesize that the polysemy is a factor that influence the vocabulary acquisition process of the children. We analyze massive electronic corpora of transcriptions of conversations between adults and children in English, to obtain the evolution of the polysemy along the children age.

6 Live Graph - Towards Memory-Driven Analytics

Speaker: Tomer Sagi (HPE Labs)

Abstract extract: Graph database workloads pose a unique challenge for traditional software. Data integrity requires persistence, precluding usage of in-memory only solutions. The variable out-degrees and non-locality of query results require databases to retrieve a small amount of results from multiple disparate physical storage locations to answer even simple queries. Furthermore, the extent of non-local retrieval a specific query induces is not easily predictable. This goes against everything that traditional hardware and software is designed around, which is to maximize sequential, predictable data retrieval. We present memory-driven analytics, new computing paradigm in which more and faster shared memory, rather than faster CPU, drives analytics. We share our experiences and discuss the potential of this new paradigm for analytics and its unique properties.

7 Synthetic Data Generation.Using Exponential Random Graph Modelling

Speaker: Burcu Kolbay (DAMA-UPC & MACDA)

Abstract extract: Statistical network analysis based on the family of exponential random graph models gives us huge advantage to discover the dependencies in the network. We use these dependencies to explore the network and based on the model, we check the goodness of fit and then we simulate new networks with different number of nodes.

8 The scarcity of crossing dependencies: a direct outcome of a specific constraint?

Speaker: Ramon Ferrer i Cancho (LARCA & MACDA)

Abstract extract: Crossing syntactic dependencies have been observed to be infrequent in natural language, to the point that some syntactic theories and formalisms disregard them entirely. This leads to the question of whether the scarcity of crossings in languages arises from an independent and specific constraint on crossings. We provide statistical evidence suggesting that this is not the case, as the proportion of dependency crossings in a wide range of natural language treebanks can be accurately estimated by a simple predictor based on the local probability that two dependencies cross given their lengths the network and based on the model.

10:35 – 11:20	Poster session I (in conjunction with coffee break)
11:20 – 12:20	Keynote - Peter Boncz (Vrije Universeit Amsterdam)
12:20 – 13:40	Presentation session II

1 DASL: A Scala-based DSL for Graph Analytics on GPUs

Speaker: Olaf Hartig (Hasso Plattner Institute)

Abstract extract: For data intensive analytic challenges, memory bandwidth, not processor speed, is the primary performance limitation. Graphics processing units (GPUs) provide superior bandwidth to main memory and can deliver significant speedups over CPUs. In contrast, to scale applications onto multicore, parallel architectures it requires significant expertise, including intimate knowledge of the CPU and GPU memory systems, and detailed knowledge of a GPU programming framework such as OpenCL or CUDA. To enable analytic experts to implement complex graph applications that efficiently run on GPUs we have developed a domain-specific language called DASL and a corresponding execution system processing algorithms.

2 Modelling the Clustering Coefficient of a Random Graph

Speaker: Ariel Duarte (DAMA-UPC & MACDA)

Abstract extract: Acquiring graph-like datasets with realistic degree distributions or other structural properties such as clustering coecient is not always feasible, either due to privacy preserving concerns or technical issues. In this work in progress, we present a graph generation algorithm that is able to generate graphs following a given degree distribution and a target clustering coefficient.

3 Computing on Event-sourced Graphs

Speaker: Benjamin Erb (Institute of Distributed Systems University of Ulm)

Abstract extract: While traditional graph computing usually employs batch processing, near-realtime computations on streaming data are often accomplished using event processing technologies. However, an increasing number of applications requires both computing capabilities for scenarios with highly dynamic and highly interconnected data fed by a stream of events. We suggest a novel platform architecture for graph computing that enables event-driven graph dynamicity while also supporting complex, long-running graph computations within the same system.

4 Use of graphs for cloud service selection in Multi-Cloud Environments

Speaker: Jaume Ferrarons (CA Technologies)

Abstract extract: The need for providing online resilient services is driving companies to adopt multi-cloud architectures in order to sustain their applications online. The selection process to choose the right cloud services for each application turns out to be a complex problem. The outcome from this selection process is conditioned by several dimensions as the costs, flexibility, adaptability and quality of the final setup. We explore the use of weighted multigraphs to depict the information involved in this kind of decisions and then discuss the applicability of graph algorithms on top of it in order to find the best deployment solution.

5 Benchmarking Versioning systems for Big Linked Data

Speaker: Irini Fundulaki (FORTH)

Abstract extract: The evolution of datasets and the management of their links would often require storing different versions of the same dataset, so that interlinked datasets can refer to older versions of an evolving dataset and upgrade at their own pace, if at all. Supporting the functionality of accessing and querying past and multiple versions of an evolving dataset is the main challenge for versioning systems. In this presentation we are going to provide some ideas on how in the HOBBIT H2020 project we intend to develop benchmarks for versioning systems that will challenge the capability of versioning systems to address multiple versions of Linked Data and complex cross snapshot queries that span across multiple versions.

6 Identifiability in Dynamic Causal Networks

Speaker: Gilles Blondel (LARCA & MACDA)

Abstract extract: Causal networks are a variant of the well-known Bayesian networks that intend to model causality instead of simple association between a fixed, finite set of variables. Pearl (1995) introduced the do-calculus, a set of rules for reasoning about Causal Networks, and the ID algorithm which uses do-calculus rules for deducing whether the effect of an experiment can be calculated from observational (instead of interventional) data only. On the other hand, Dynamic Bayesian Networks model the case in which a set of variables evolve. In this work we put together the Dynamic and Causal aspects for the first time: We present Dynamic Causal Networks (DCN), we show how to apply do-calculus in this setting, and we present a dynamic variant of Pearl's ID algorithm for DCN.

7 Improving Mobility in Smart Cities with Intelligent Tourist Trip Planning

Speaker: Petar Mrazovic (KTH)

Abstract extract: Selecting the most interesting attractions and planning optimal sightseeing tours can be a difficult task for individuals visiting unfamiliar tourist destinations. On the other hand, the massive amounts of tourists in big cities can collapse certain areas causing transport inefficiency, unbalanced economic growth and nuisance among tourists and citizens. Therefore, the tourist trip planning problem can also arise as a means for the city government to manage the urban environment and achieve a balanced and sustainable growth. In this work we introduce the tourist trip planning problem which serves both individual (tourist) and global (city) needs. The planning problem is modelled as an extension of the mixed orienteering problem and can be controlled by deployment of mobility policies which put restrictions on points of interest and routes between them.

8 LDBC Graphalytics: Benchmarking Platforms for Large-Scale Graph Processing

Speaker: Alexandru Iosup (TU Delft)

Abstract extract: Graphs model social networks, human knowledge, and other vital information for business, governance, and academic practice. Although both industry and academia are developing and tuning many graph-processing algorithms and platforms, the performance of graph-processing platforms has never been explored or compared in-depth. Moreover, graph processing exposes new bottlenecks in traditional and new systems (see, for example, the large differences between Top500 and Graph500 rankings). LDBC Graphalytics is a benchmark that explores for batch full-graph analytics the performance dependency Platform-Algorithm-Dataset.

13:40 Closing & Poster session II (in conjunction with lunch)

GRAPH-TA is an event organized by

DAMA-UPC. DATA MANAGEMENT
UNIVERSITAT POLITÈCNICA DE CATALUNYA

LARCA. Laboratori d'Algorísmia Relacional, Complexitat i Aprenentatge
UNIVERSITAT POLITÈCNICA DE CATALUNYA