

3rd Graph-based Technologies and Applications

18th March 2015

Program

9:00

Registration

9:30 – 10:45

Welcome by Fernando Orejas Vice-rector of research UPC

Presentation session I

1

RDF Graph Data Management in Oracle Database and NoSQL Platforms

Speaker: Xavier Lopez

Abstract extract: This presentation highlights support for RDF graph data management on Oracle relational and NoSQL databases. It will illustrate how customers are applying this technology to develop enterprise-class linked data and knowledge management solutions. The presentation will describe the comprehensive support for the W3C RDF/OWL/SPARQL specifications.

2

Consensus Strategies applied to deduct a Mean of two Graph Correspondences

Speaker: Francesc Serratosa

Abstract extract: Consensus strategies have been recently studied to help machine learning ensure better results. Likewise, optimisation in graph matching has been explored to accelerate and improve pattern recognition systems. We present a fast and simple consensus method which, given two correspondences of graphs generated by separate entities, enounces a final consensus correspondence. It is based on an optimisation method that minimises the cost of the correspondence while forcing it (to the most) to be a weighted mean.

3

GRAPHITE: An Extensible Graph Traversal Framework for Relational Database Management Systems

Speaker: Marcus Paradies

Abstract extract: We introduce GRAPHITE, an extensible graph traversal framework as the central component of a graph engine for columnar main-memory database management systems. In the presentation we describe a top-down approach from the query layer down to the physical operator layer and outline how we integrated GRAPHITE into a relational database system.

4

On the Discovery of Novel Drug-Target Interactions from Dense SubGraphs

Speaker: Maria-Esther Vidal

Abstract extract: Knowledge graphs are networks of semantically related concepts, described in terms of attributes, semantic types, annotations, and relationships. For example, knowledge graphs in the biomedical domain represent relationships among drugs, diseases, targets, and side effects. Exploring these semantically rich networks can lead to novel discoveries, e.g., interactions between drugs and targets, or drug side effects. We tackle the problem of suggesting associations between concepts from a knowledge graph where similarity measures allow to determine related-ness among concepts of the same type.

5

Graphalytics: A big data benchmark for graph-processing platforms

Speaker: Mihai Capotă

Abstract extract: Graph data is increasingly used in many sectors of the economy, such as commerce, health, or transport. However, selecting the right platform for a particular application is a difficult process, because performance depends not only on the processing platform, but also on the workload, i.e., the algorithm being executed and the graph data itself. In this presentation, we introduce Graphalytics, a big data benchmark for graph-processing platforms, which includes a comprehensive selection of real-world graphs and representative algorithms that stress the choke points of processing platforms.

6

Autograph: an evolving lightweight graph tool

Speaker: Joan Vilaltella

Abstract extract: There are many powerful software libraries for the calculation of graph properties and for the visualization of large networks, and there are many excellent tools to draw and edit diagrams or, in particular, graphs. However, if you have ever felt somewhat frustrated looking for a lightweight tool that combines usual graph theoretic operations with easy, stable, graph editing (including undo/redo), you may have a look at the Autograph application.

7

Understanding Graph Structure in Knowledge Bases

Speaker: Joan Guisado

Abstract extract: Knowledge bases are very good sources for Knowledge Extraction, the ability to create knowledge from structured and unstructured sources. However, extracting knowledge from unstructured source is challenging. In this respect, understanding the structure of knowledge bases can provide significant benefits for the effectiveness if such purpose. We analyze the structure of Wikipedia and understand how different categories of data within relate to each other and how those relationships can be used in the benefit of a particular aspect of knowledge extraction, which is query expansion.

8

Finding patterns of chronic disease and medication prescriptions from a large set of electronic health records

Speaker: Ricard Gavaldà

Abstract extract: We describe a prototype of a system for discovering, analyzing, and visualizing the co-occurrence of diagnostics, interventions, and medication prescriptions in a large patient database. The final tool is intended to be used both by health managers and planners and for primary attention clinicians in direct contact with patients.

10:45 – 11:45

11:45 – 13:00

Poster session I (in conjunction with coffee break)

Presentation session II

1

Towards Temporal
Graph Management
and Analytics

Speaker: Yinglong Xia

Abstract extract: Many real graphs are highly dynamic, such as the Twitter interaction graph and the Call Detailed Record network, which can be reviewed as a series of graph slices over time axis with possible connections across adjacent slices. For a wide range of applications, studying the graph morphing in terms of both topology and properties is as important, if not more, than exploring the interaction among graph vertices/edges within a certain time stamp. In this work, we present our undergoing work on natively supporting temporal property graph query and analytics, where we extend the IBM System G graph store for representing the cross-time edges as special types of connections; the time cursor mechanism is introduced for retrieving graph data across time stamps.

2

Lighthouse: large-scale
graph pattern matching
on Giraph

Speaker: Claudio Martella

Abstract extract: Lighthouse is a system that leverages the Pregel model, and its open-source implementation Giraph, to perform graph pattern matching on large graphs across hundreds of commodity machines. Graph queries are expressed through a subset of Neo4j's Cypher query language, and Lighthouse computes a query plan and executes it as a Giraph computation on a Hadoop cluster. This talk gives an overview of the project and the system architecture.

3

A Cypher extension for
efficient shortest
weighted path queries

Speaker: Peter Rutgers

Abstract extract: Finding shortest paths based on edge weights has many applications in data analysis. We propose an extension to the Cypher language to easily express shortest weighted path queries. This language extension is demonstrated in the Lighthouse system, which can execute queries in (a subset of) Cypher on the Giraph system using BSP processing, in a scalable fashion.

4

MobiCS: Mobile Crowd
Sensing Platform

Speaker: Petar Mrazovic

Abstract extract: Current participatory sensing approaches usually do not consider device carriers as intelligent participants in sensing processes. However, modern mobile communication devices can allow users to express their opinions and judgements which can complement to captured sensor data. We present MobiCS platform which offers real-time services to users by collecting contributed information which is not limited to passively-generated sensor readings from the device, but also includes proactively-generated opinions and perspectives from voluntary participants.

5

Analysing the degree distribution of real graphs by means of several probabilistic models.

Speaker: Ariel Duarte

Abstract extract: Given that, usually there exists a lot of nodes with few connections, and just very few nodes with a huge amount of connections, it is proposed in the literature the Zipf's model as an appropriate one. As it is well known, the Zipf's model requires that the logarithm of the probabilities be a linear function of the logarithm of the values, a property not always verified in real graphs. We have considered several alternative models that include: the Marshall-Olkin Extended Zipf (MOEZipf) and the Altmann, which relax the log-log linearity requirement for high frequency observations.

6

SPIMBENCH: A Scalable, Schema-Aware, Instance Matching Benchmark for the Semantic Publishing

Speaker: Tzanina Saveta

Abstract extract: Choosing the right framework for Data Web remains tedious, as current instance matching benchmarks fail to provide end users and developers with the necessary insights pertaining to how current frameworks behave when dealing with real data. We present the Semantic Publishing Instance Matching Benchmark (SPIMBENCH) which allows the benchmarking of instance matching systems against not only structure based and value-based test cases, but also against semantics aware test cases based on OWL axioms. SPIMBENCH features a scalable data generator and a weighted gold standard that can be used for debugging instance matching systems and for reporting how well they perform in various matching tasks.

7

Generating synthetic online social network graph data and topologies

Speaker: David F. Nettleton

Abstract extract: One of the difficulties for data analysts of online social networks is the public availability of data, while respecting the privacy of the users. One alternative is to use synthetically generated data. However, this presents a series of challenges related to generating a realistic dataset in terms of topologies, attribute values, communities, data distributions, and so on. In the following we present an approach for generating a graph topology and populating it with synthetic data for an online social network.

8

Model Synchronization with Triple Graph Grammars

Speaker: Elvira Pino

Abstract extract: Nowadays, software systems or communication networks are becoming more and more interconnected, and, we expect to access services and applications anywhere and anytime. In that context, the model synchronization is the problem of restoring the consistency of the synchronization between components of systems after an update. Based on the Triple Graph Grammar approach, we study the problem of restoring synchronization after concurrent updates.

13:00 – 14:45

Poster session II (in conjunction with lunch break)

14:45 – 16:10

Presentation session III

1

Deriving an Emergent Relational Schema from RDF Data

Speaker: Minh-Duc

Abstract extract: We motivate and describe techniques that allow to detect an “emergent” relational schema from RDF data. We show that on a wide variety of datasets, the found structure explains well over 90% of the RDF triples. Further, we also describe technical solutions to the semantic challenge to give short names that humans find logical to these emergent tables, columns and relationships between tables.

2

Managing RDF data with graph databases

Speaker: Renzo Angles

Abstract extract: Considering that RDF defines a graph data model and that its query language SPARQL is based on graph pattern matching, the idea of developing an RDF Triple Store on top of a graph database results intuitive. We concentrate our research on the study of methods for managing RDF data by using graph databases. Currently, we have identified three methods for storing RDF data over databases following the property graph data model and implemented a prototype by using the Jena RDF API and the Sparksee graph database. After some empirical experiments, we identified the advantages and disadvantages of the methods and showed the feasibility of implementing a graph-based RDF Triple Store.

3

On the prime cordial labeling of Möbius Ladder M_n

Speaker: Mohammed Mominul

Abstract extract: A graph with vertex set V is said to have a prime cordial labeling if there is a bijection f from V to $\{1, 2, \dots, |V|\}$ such that if each edge uv is assigned the label 1 for the greatest common divisor $\gcd(f(u), f(v)) = 1$ and 0 for $\gcd(f(u), f(v)) > 1$ then the number of edges labeled with 0 and the number of edges labeled with 1 differ by at most 1. In this paper, we show that Möbius Ladder M_n is prime cordial for all n except M_4 .

4

Graph based word spotting approach for large document collections

Speaker: Pau Riba Fiérrez

Abstract extract: Given a handwritten text, graphemes are extracted from the shape convexities and they are used as stable units of handwriting. Afterwards, an attributed graph is constructed using a part-based approach where the graphemes are associated to graph nodes. Moreover, the spatial relations between them determine graph edges. Afterwards the spotting is defined in terms of an error-tolerant graph matching using bipartite-graph matching algorithm. Although this approach has good performance in word-to-word comparison, we need to find a word graph inside the document collection. In order to minimize the high computational complexity of subgraph matching, we had proposed a fast indexation formalism for graph retrieval.

5 Use of graph technologies for political analysis

Speaker: Francisco Rodriguez, Diego Olano
Abstract extract: The aim of this project is to discover and analyze relationships between politicians and relevant third party actors from news articles and political data. To do this, we detect patterns of co-occurrence of entities in 1) individual news articles and 2) clusters of articles related to topics modeled through bills. Based on these patterns, we extract relevant links to build and study graphs which can be used to summarize the media coverage of politicians and to characterize their lobbying process. For this project, we work with data related to the Catalan Parliament and the Texas State Legislature.

6 Graphium Chrysalis: Exploiting Graph Database Engines to Analyze RDF Graphs

Speaker: Maria-Esther Vidal
Abstract extract: We present Graphium Chrysalis, a visualization tool that exploits different graphical representations to report on the results of evaluating a variety of RDF graphs, and graph invariants implemented on top of Neo4j and Sparksee. Graph invariants implemented in Graphium Chrysalis facilitate the understanding of the structure of RDF graphs, and graph connectivity properties that support the discovery of novel patterns and interactions.

7 Graph analytics for IT Management

Speaker: Jaume Ferrarons
Abstract extract: IT Management helps to improve companies' performance by making them more efficient and competitive. It eases the execution of complex operations such as resource planning or defining the topology of a computer network. Consequently, IT Management has to deal with dynamic complex systems. Graph-based models turn out to suit many of the requirements posed by IT Management problems. There are many success cases in industry and a representative set of examples are summarized in this poster to open a discussion about typical graph operations executed in the IT domain.

8 Langford Sequences through a product of labeled digraphs

Speaker: Susana Clara Lopez Masip
Abstract extract: Skolem, Langford sequences and their many generalizations have applications in numerous areas. In this paper, we study Skolem and Langford sequences through (extended) Skolem and Langford labelings. We will show that our procedure can also be applied to obtain Langford sequences from existing ones, and that in this way, we can obtain a lower bound for the number of Langford sequences, for particular values of the defect.

16:10 – 17:00

Poster session III (in conjunction with coffee break)

17:00 – 17:10

Closing