

# Current Challenges in Graph Databases

Juan L. Reutter

School of Engineering, Pontificia Universidad Catolica, Santiago, Chile  
IMFD, Santiago, Chile

---

## Abstract

---

As graph databases grow in popularity, decades of work in graph query languages and models are materialising in industry standards and in the construction of new graph database systems. However, this surge in graph systems has in turn opened up a series of new, interesting research problems related to graph databases.

Our first set of problems has to do with more efficient ways of computing the answers of graph queries, specifically graph patterns, path queries, and combinations between them. Traditionally, researchers in graph databases have pointed out that relational systems are ill-equipped to process these types of queries, and if one looks at the performance of native graph database systems, there is clearly a lot of room for improvement. The talk focuses on two possible directions for improving the state of the art in graph query processing. The first is implementing worst-case optimal algorithms for processing graph patterns that traduce in relational queries with several joins. Some advances are already in development (see e.g. Nguyen, Dung, et al. “Join processing for graph patterns: An old dog with new tricks.” GRADES’15. or Hogan, Aidan, et al. “A Worst-Case Optimal Join Algorithm for SPARQL.” ISWC’19.), but we are still far from a full fledged solution: most algorithms require complex data structures, or need further support in terms of heuristics to select an order in which joins are processed. Second, we need to understand what is the best way of evaluating path queries (that is, finding all pairs of nodes connected by a path), in such a way that these results can be further integrated with other query results in a graph system pipeline. We already have complexity results regarding path computation and enumeration for different semantics of path queries (see e.g. Martens, Wim, and Tina Trautner. “Evaluation and enumeration problems for regular path queries.” ICDT’18. or Bagan, Guillaume, Angela Bonifati, and Benoit Groz. “A trichotomy for regular simple path queries on graphs.” PODS’13.), but still very little is known in terms of optimal processing of path queries when inside a tractable fragment.

Our second set of problems is related to graph analytics, one of the current selling points of graph databases. Systems should be able to run more complex analytical queries involving tasks such as more complex path finding, centrality or clustering. It is also important to be able to run these algorithms not over native graphs, but perhaps over a certain set of nodes or edges previously selected by a graph query, and one may also want to pose further queries over the result of the analytics task. Finally, all of this should be done in an efficient way, specially in the prospect that graph databases may contain a huge amount of nodes. In this talk I will discuss possible approaches to perform these operations, covering aspects from the design of languages for graph analytics to efficient ways of processing them, and also comparing the expressive power of graph analytics solutions with other forms of graph computation.

**2012 ACM Subject Classification** Theory of computation → Database query processing and optimization (theory)

**Keywords and phrases** Graph databases, Join algorithms, path queries, graph analytics

**Digital Object Identifier** 10.4230/LIPIcs.ICDT.2020.3

**Category** Invited Talk

**Funding** This work was partially funded by Millennium Institute for Foundational Research on Data (IMFD).



© Juan L. Reutter;

licensed under Creative Commons License CC-BY

23rd International Conference on Database Theory (ICDT 2020).

Editors: Carsten Lutz and Jean Christoph Jung; Article No. 3; pp. 3:1–3:1

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany