

# Tutorial: Data Analytics with Graph Algorithms — A Hands-on Tutorial with Neo4J

Lena Wiese<sup>1</sup>

**Abstract:** This tutorial presents perspectives for advanced graph data analytics and covers the background of graph data management in modern data stores. It provides an overview of several well-established graph algorithms. The three categories covered are path-based algorithms, community detection and centrality scores. A deeper understanding of graph algorithms is a major precondition to efficiently analyze graph-structured data. The tutorial hence enables participants to achieve an informed decision about what kind of algorithm is appropriate for which use case.

## 1 Introduction

Many scientists as well as practitioners work with graph-structured data. Such data often occur in many modern data science applications. Graph theory offers different variants of graph algorithms as well as many optimizations for them. Often it is however difficult to assess which algorithms will be applicable to specific use cases.

## 2 Outline

The tutorial will proceed in the following steps.

**Background** Graph data are predominant in many applications like medicine and biology, social networks, the internet and the semantic web. We present several of these applications to highlight the importance of graph data management and analytics.

**Graph theory** This topic explains some basics of graph theory. Having presented several choices for graph data structures (from adjacency matrix to incidence list), it describes the predominant data model for graph databases: the property graph model.

---

<sup>1</sup> University of Goettingen, Institute of Computer Science , Goldschmidtstraße 7, 37077 Göttingen, wiese@cs.uni-goettingen.de

**The graph database** The Neo4J graph database is a world-leading open source graph database. Its declarative query language *Cypher* makes graph data management convenient even for novice users. The database offers built-in visualization and implementations of all covered graph algorithms. In order to gain first practical experiences with the database, Neo4J offers an online sandbox [Sa] including the graph algorithm library.

**Graph algorithms** The following categories of graph algorithms will be presented:

- **Path-based** algorithms find optimal traversals of the graph. We will cover Minimum Weight Spanning Tree as well as Shortest Path.
- **Community Detection** algorithms identify clusters of nodes in the graph and assess the quality of these clusters. We will cover Label Propagation, Louvain, Weakly and Strongly Connected Components and Triangle Count.
- **Centralities** are algorithms that assign a score to each node in the graph. These scores help identify those nodes most important for certain applications. We will cover Page Rank, Betweenness Centrality as well as Closeness Centrality.

All presented algorithms will be visualized with appropriate examples in the Neo4J database.

### 3 Related Work

Theoretical exposition of the graph algorithms will be based on Chapter 5 and Chapter 10 of [LRU14]. Hands-on examples with the Neo4J database are based on [Co] and [NH19]. Background on graph data structures and graph databases will be based on [Wi15].

### 4 Speaker's Biography

Dr. Lena Wiese is head of the research group Knowledge Engineering and lecturer at the Georg August University Göttingen. She has been teaching advanced courses on data management and database technology for several years at both graduate and undergraduate level. The Neo4J database is used in these courses as a convenient test environment. She is author of the book “Advanced Data Management for SQL, NoSQL, Cloud and Distributed Databases” (DeGruyter/Oldenbourg, 2015) [Wi15]. She holds a PhD from the University of Dortmund and worked as a postdoctoral researcher at the Japanese National Institute of Informatics in Tokyo. She acts as a reviewer for international conferences and journals on a regular basis.

## References

- [LRU14] Leskovec, Jure; Rajaraman, Anand; Ullman, Jeffrey David: Mining of massive datasets. Cambridge university press, 2014.
- [Nea] Neo4J Committers: , Neo4j Graph Algorithms User Guide. <https://neo4j.com/docs/graph-algorithms/current/>.
- [Neb] Neo4J Sandbox: . <https://neo4j.com/sandbox-v2/>.
- [NH19] Needham, Mark; Hodler, Amy E.: Graph Algorithms: Practical Examples in Apache Spark and Neo4j. O'Reilly, 2019.
- [Wi15] Wiese, Lena: Advanced Data Management for SQL, NoSQL, Cloud and Distributed Databases. DeGruyter, 2015.