

## Editorial

Ulrich Meyer\* and Ziawasch Abedjan

# Algorithms for Big Data

<https://doi.org/10.1515/itit-2020-0008>

**Keywords:** Algorithms, Big Data Technologies

The vast amount of existing data in various fields of industry, such as health, finance, and automotives, and its fast growth through social networks, sensors, and smart devices makes continuous research on the impact, opportunities, and boundaries of Big Data necessary and inevitable. At the same time, distributed processing systems, such as Hadoop, Flink, and Spark allow engineers to create data processing software that can handle large volumes of data and fast paced streams. In order to achieve the best possible speedups and scalability, however, new algorithmic insights and their efficient implementation are crucial, too. Furthermore, current research still tries to overcome challenging dimensions, such as variety and veracity of data. Also data privacy is becoming of significant importance by the day. In Germany, several Big Data projects and initiatives try to tackle Big Data problems in a focused manner. For example, the priority programme DFG-SPP 1736 on Algorithms for Big Data has been funding various projects in the targeting technological challenges, fundamental algorithmic techniques, and applications. The Federal Ministry of Education and Research (BMBF) is expanding its funding for Big Data research from two competence centers for Big Data: the Berlin Big Data Center (BBDC) and the Competence Center on Scalable Data Solutions and Services (ScaDS) Dresden/Leipzig to several AI competence centers throughout Germany now also in Tübingen, Darmstadt, and Munich.

For this special issue we have invited contributions from German researchers who conduct research on theoretical boundaries of big data as well the realization of end-to-end data processing systems. After careful reviewing by several experts and revision of the papers, we have finally accepted the following seven contributions for this special issue on “Algorithms for Big Data”.

- “Dictionary learning for transcriptomics data reveals type-specific gene modules in a multi-class setting”

---

\*Corresponding author: **Ulrich Meyer**, Goethe Universität Frankfurt, Frankfurt, Germany, e-mail: [umeyer@ae.cs.uni-frankfurt.de](mailto:umeyer@ae.cs.uni-frankfurt.de)  
**Ziawasch Abedjan**, Technische Universität Berlin, Berlin, Germany, e-mail: [abedjan@tu-berlin.de](mailto:abedjan@tu-berlin.de)

by Mona Rams and Tim O. F. Conrad from FU Berlin describes the application of the RNA gene sequencing algorithm and a corresponding benchmark for dictionary learning.

- The article “Large-scale graph generation: Recent results of the SPP 1736 – Part II” by Ulrich Meyer and Manuel Penschuck from the Goethe University in Frankfurt describes continuous work on SPP projects for large-scale graph generation that enable large-scale research and experiments on graph data.
- Abdulrahman Kaitoua, Tilmann Rabl, and Volker Markl from TU Berlin address the data movement problem in polystores with their paper “A distributed data exchange engine for polystores”. The presented system Muses is a distributed data migration engine that is able to interconnect distributed data stores by forwarding, transforming, or broadcasting data among distributed engines’ instances.
- Claudio Hartmann, Lars Kegel, and Wolfgang Lehner from TU Dresden present in their paper “Feature-aware forecasting of large-scale time series data sets” a technique for forecasting a set of time series with one single model, and a feature-aware partitioning approach.
- “Optimization frameworks for machine learning: Examples and case study” is the topic of the article by Joachim Giesen, Soeren Laue, and Matthias Mitterreiter from Friedrich Schiller University Jena. The authors provide an introduction into the area with an exemplary treatment of some frameworks including their own GENO (GENeric Optimization) tool.
- In his paper “Solving subset sum with small space – Handling cryptanalytic Big Data”, Alexander May from Ruhr University Bochum reviews recent progress on memory-less combinatorial algorithms for data that appears in the context of cryptographic protocols.
- The article “Scaling up network centrality computations – A brief overview” by Alexander van der Grinten, Eugenio Angriman, and Henning Meyerhenke (Humboldt University Berlin) reviews several common and some recent (and not necessarily exact) performance-oriented algorithmic techniques that enable significantly faster processing of network centralities than the previous state of the art.

## Bionotes



**Ulrich Meyer**  
Goethe Universität Frankfurt, Frankfurt,  
Germany  
[umeyer@ae.cs.uni-frankfurt.de](mailto:umeyer@ae.cs.uni-frankfurt.de)

Prof. Dr. Ulrich Meyer is a full professor at Goethe University Frankfurt, Germany. He received his Ph.D. in computer science from Saarland University in 2002. Subsequently he was a postdoc and eventually senior researcher at Max-Planck Institute for Computer Science in Saarbrücken. Ulrich Meyer is working in algorithm engineering with a focus on graph algorithms and advanced models of computation. He is leading the priority programme Algorithms for Big Data funded by the German Research Foundation.



**Ziawasch Abedjan**  
Technische Universität Berlin, Berlin,  
Germany  
[abedjan@tu-berlin.de](mailto:abedjan@tu-berlin.de)

Prof. Dr. Ziawasch Abedjan is Juniorprofessor and head of the “Big Data Management” (BigDaMa) Group at TU Berlin. Prior to that, Ziawasch was a postdoc at MIT working on various data integration problems. Ziawasch received his PhD on from the Hasso Plattner Institute in Potsdam, Germany. He is a PI in the Berlin Institute for the Foundations of Learning and Data (BIFOLD) and his research is funded by the DFG and the Federal Ministry for Research and Education.