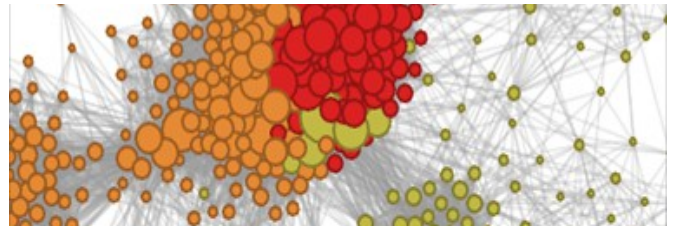


## Bachelor's/Master's Thesis

# Network Analysis on Distributed Systems

**Description.** When the term „big data“ is mentioned, it often refers to data sets with a highly irregular structure. Such data sets may come in the form of large graphs, such as social networks, web graphs, and biological networks. One way to deal with the enormous size such datasets can have is to use distributed systems for processing and analytics. In recent years there has been a variety of projects targeting distributed computation, especially in the context of graph algorithms:



- **Apache Hadoop** is a open-source framework for distributed data analysis, supporting the map-reduce programming model and written in Java; while Hadoop itself is not tailored to graph computations, several other frameworks built on top of Hadoop are
- **Apache Giraph** is a system for large scale graph processing, based on Hadoop and supporting the Bulk Synchronous Parallel programming model.
- **GraphLab** is an open-source project for distributed computation on graphs, written in C++
- **GraphChi** is a spin-off of GraphLab, meant to run on a single machine and using external memory
- ...



**Tasks.** In the course of your thesis you are going to explore such current technologies for distributed graph processing. This includes implementation and experimental comparison of graph algorithms for network analysis on different platforms. The breadth and depth of the required work will be adapted to the thesis type (Bachelor or Master).

**Requirements.** You will need the ability to adapt to new frameworks and programming models as well as enough skill at the command line to set up and troubleshoot their installation. Some previous experience with graph algorithms is of course beneficial.

**Contact.** *Please do not hesitate to contact us via e-mail or face-to-face if you are interested.*

Forschungsgruppe Paralleles Rechnen (<http://parco.iti.kit.edu>)  
Juniorprof. Dr. Henning Meyerhenke  
**Kontakt: Dipl.-Inform. Christian Staudt**  
[christian.staudt@kit.edu](mailto:christian.staudt@kit.edu) | Gebäude 50.34 Raum 249

