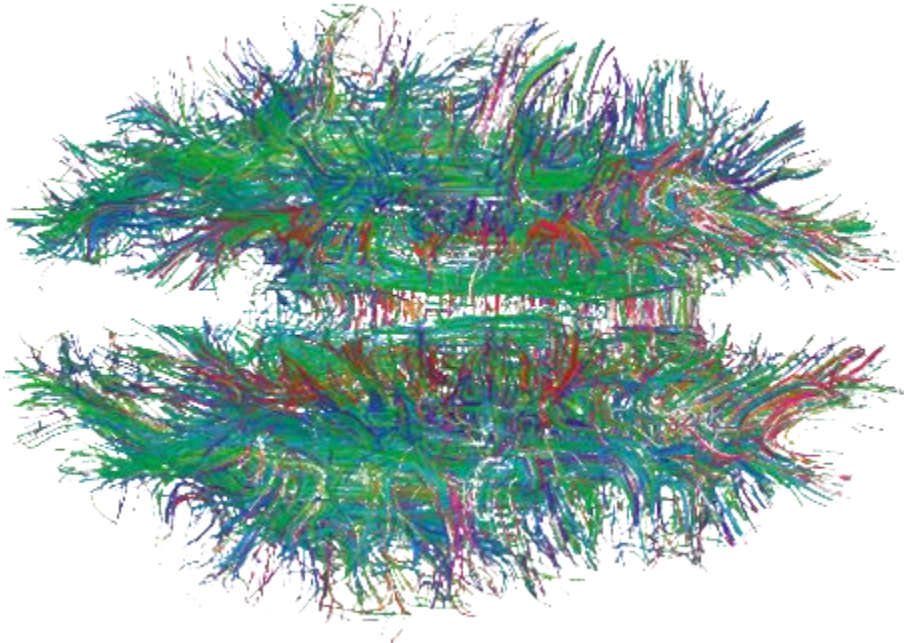
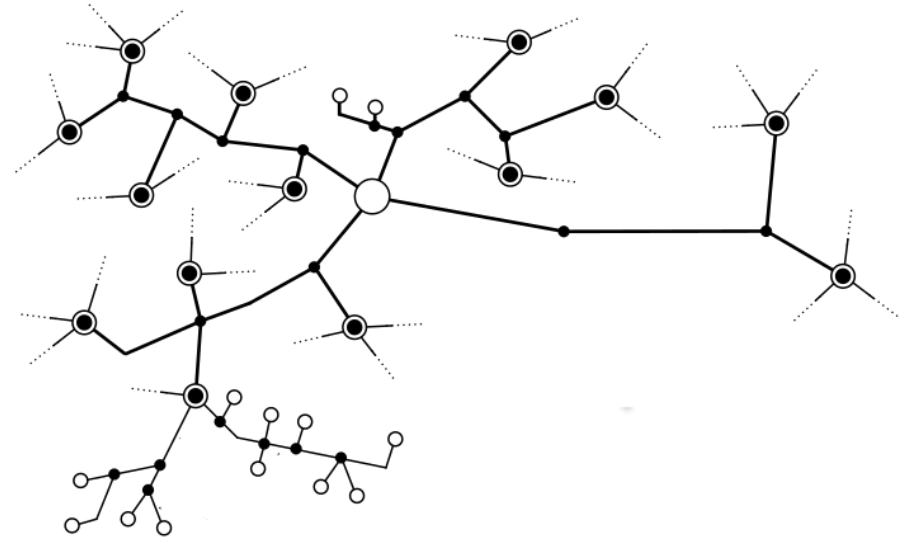


Benchmarking von Graphdatenbanken

Matthias Jordan

Dresden, 21.03.2016



GDB vs. RDB

Relationale Datenbanken für Graphen ungeeignet

Emil Eifrem, NeoTechnology

Vergleich von GDB und RDB

- 1. Messung: →GDB: 2 ms, RDB: 2000 ms
- 2. Messung: →GDB: 2 ms, RDB: ∞ ms

Die Qual der Wahl

Viele verschiedene GDB vorhanden

Wahl einer geeigneten Datenbank?

→ Benchmark



Existierende Benchmarks 1

LDBC

2 Benchmarks

3 Workloads: Interactive, Business Intelligence,
Graph Analytics

Bietet Graphgenerator



Graphalytics

Vor allem für verteilte Systeme

Nutzt LDBC-Graphgenerator

Existierende Benchmarks 2

Graph500

Für HPC

BFS auf verschieden großen Datensätzen

Seit 2010 halbjährlich



XDGBench

Für HPC

7 Workloads

DNA

DNA
DYNAMIC NETWORK ANALYZER



Graphanalyse-Framework

Bietet:

- Graphgeneratoren: zufällig, Barabasi-Albert, ...
- Graphoperationen: BFS, DFS, Clustering-Koeffizient, ...



Einheitliche Schnittstelle

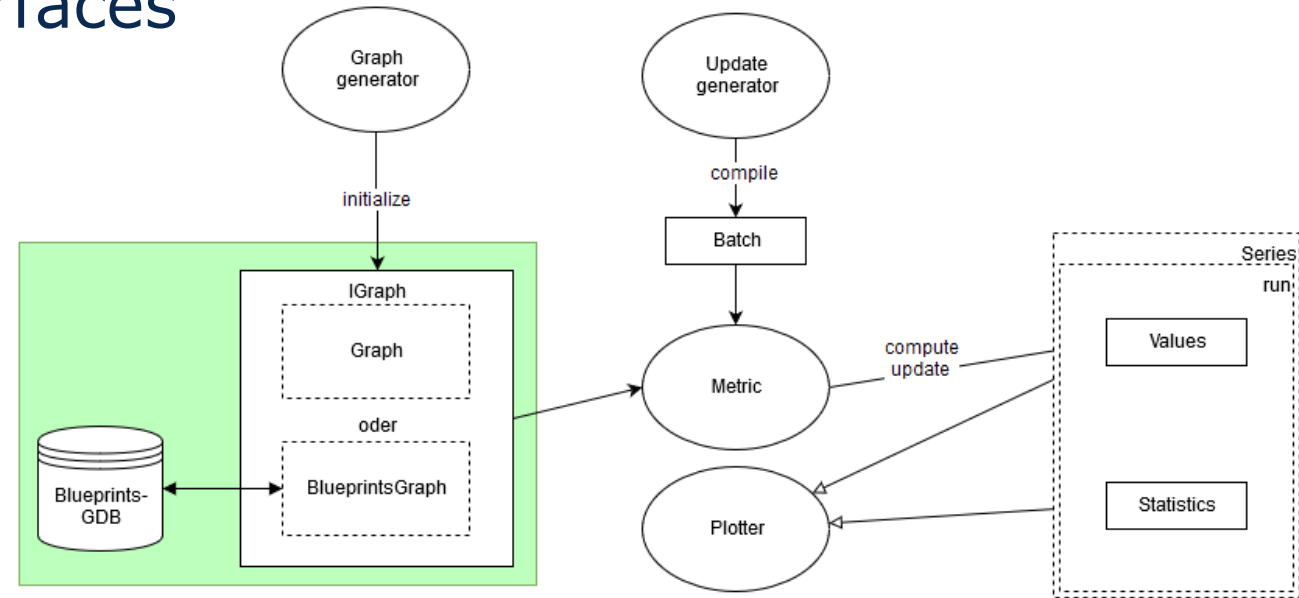
Viele unterstützte Datenbanken:
Neo4J, OrientDB, Bitsy, Tinkergraph

- TitanDB, Sparksee, ...

Implementierung

Einführung eines neuen Interfaces

Eigene Klasse für Blueprints



Wahl der GDB über Konfigurationsdatei möglich

Neue Klasse zur Erstellung von Graphen

Messwerverfassung - Setup

3 DNA-Datenstrukturen: Array, HashMultimap, Empty

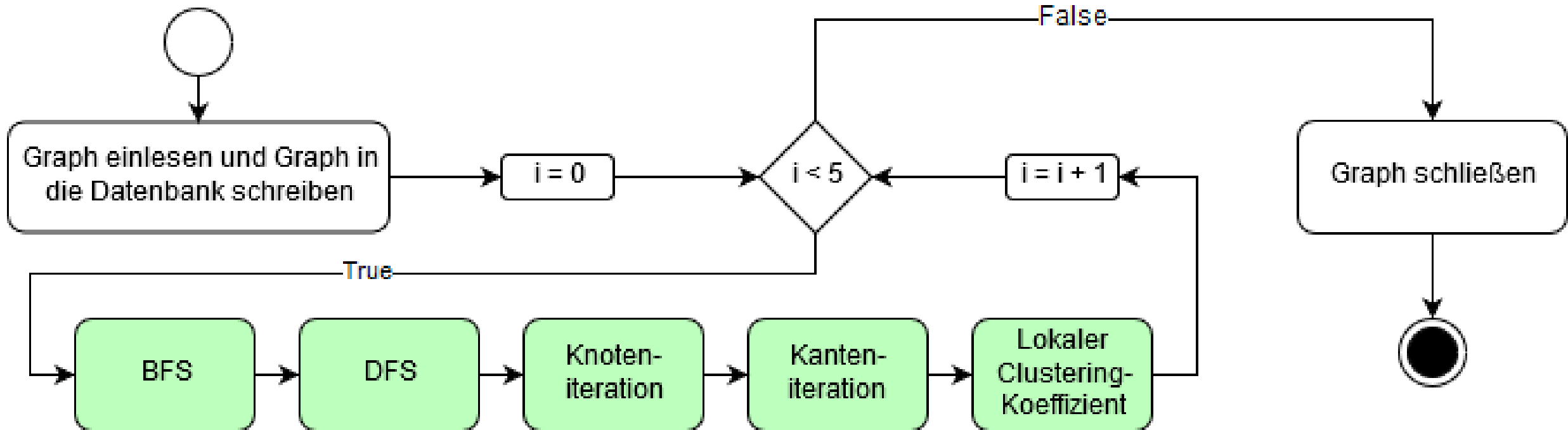
4 Datenbanken: Bitsy, Neo4J, OrientDB, Tinkergraph

3 In-Memory: Bitsy, OrientDB, Tinkergraph

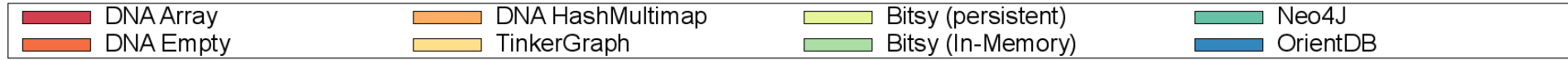
2 persistent: Bitsy, Neo4J

50 verschiedene Graph

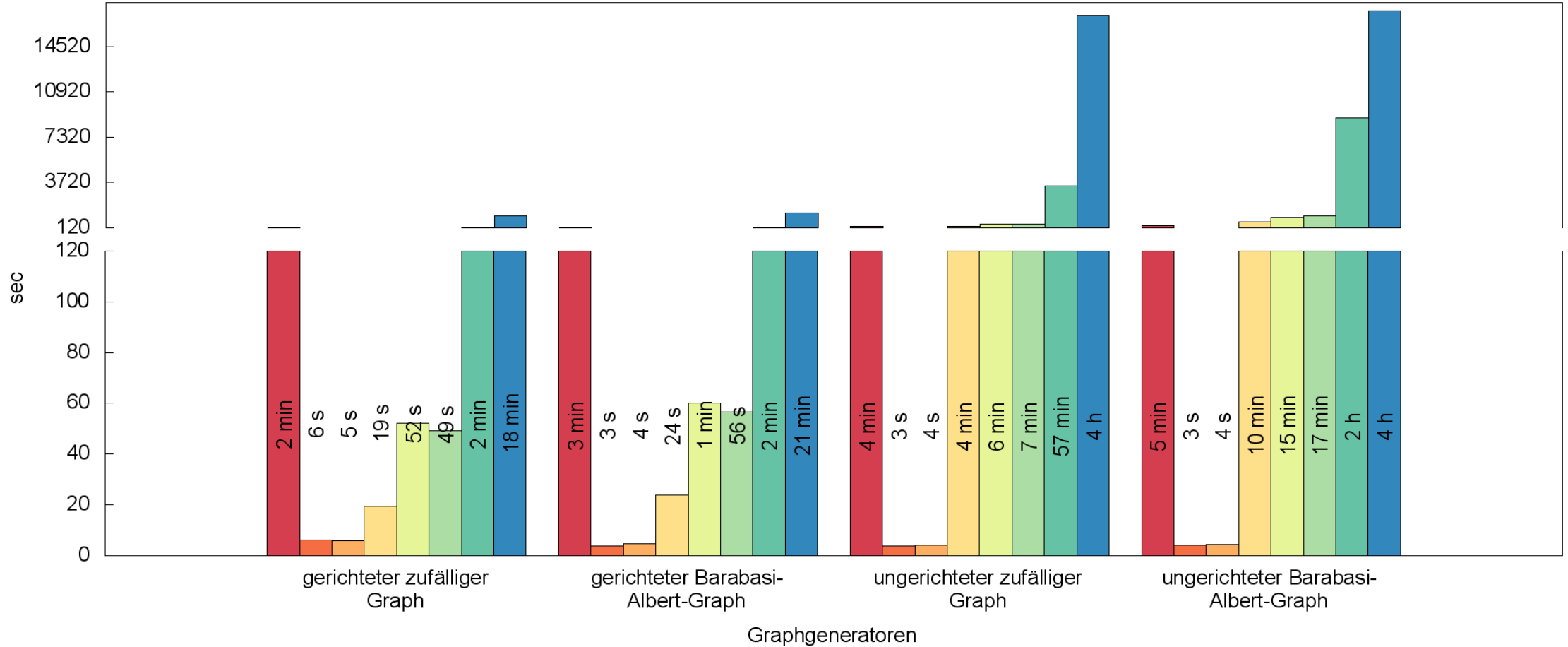
Messwerterfassung - Ablauf



Graphgenerierung

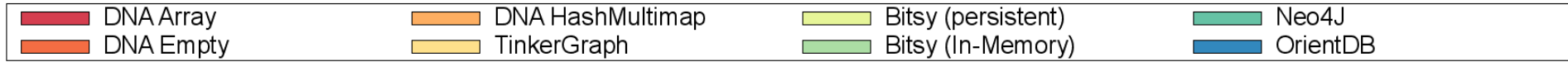


Durchschnitt über 5 Durchläufe

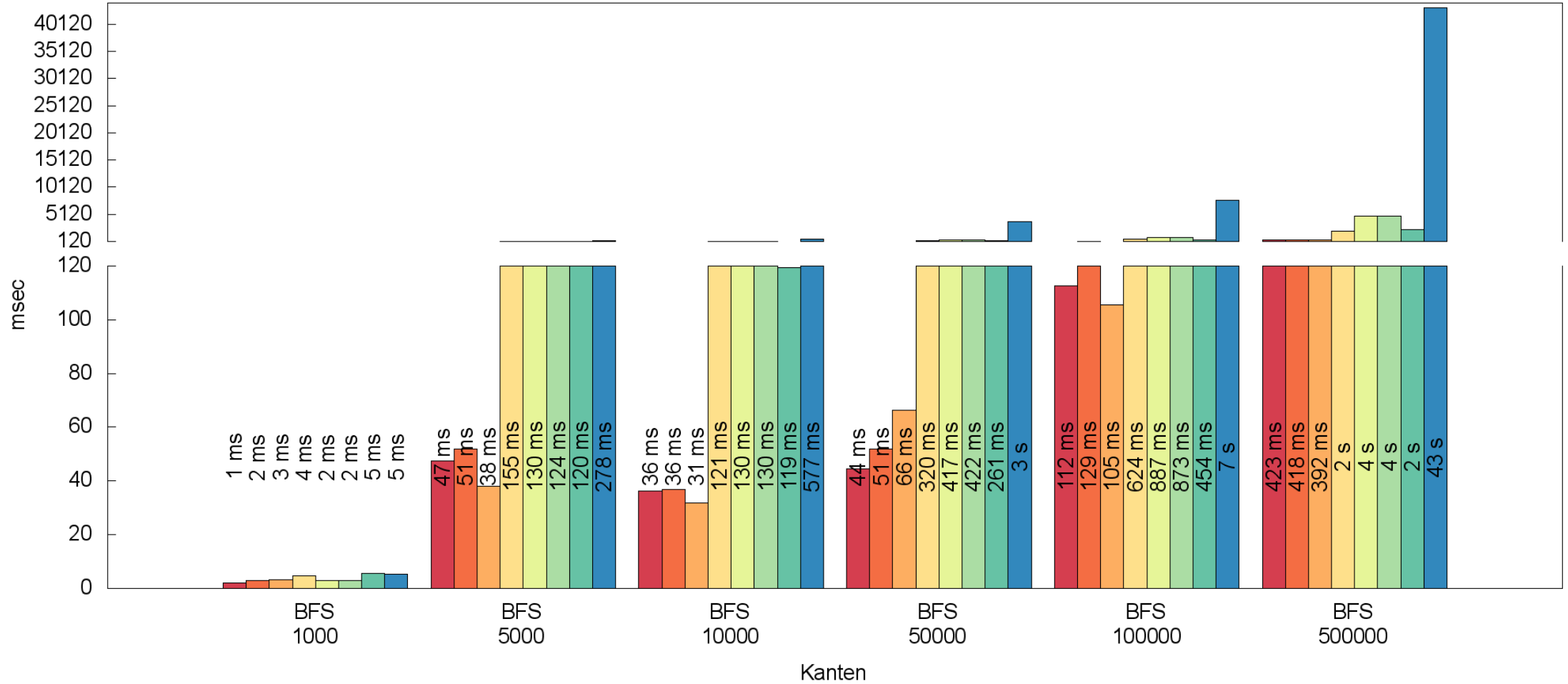


Für 1000 Knoten und 100k Kanten

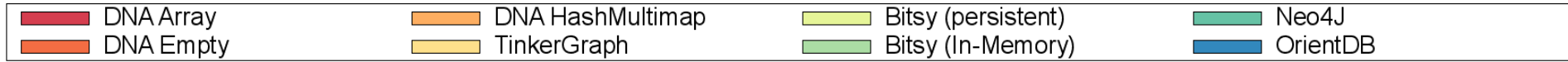
BFS, DFS



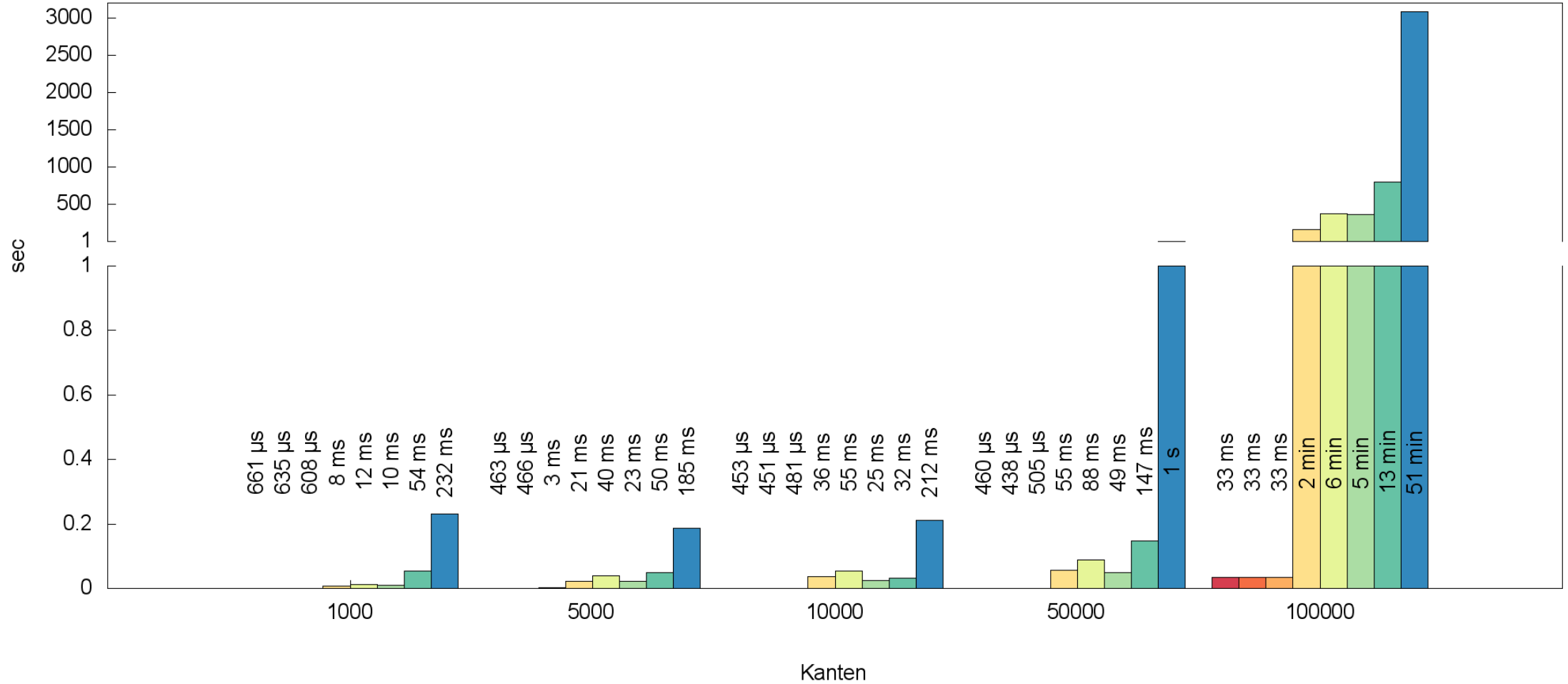
Durchschnitt über 5 Durchläufe



Lokaler Clustering-Koeffizient



Durchschnitt über 5 Durchläufe



Ergebnisse

Graph	Workload	Rangfolge							
		1.	2.	3.	4.	5.	6.	7.	8.
1k Knoten, gerichtet	BFS oder DFS	DNA HashMultimap	DNA Array	DNA Empty	Neo4J	Tinkergraph	Bitsy (persistent)	Bitsy (In-Memory)	OrientDB
10k Knoten, gerichtet	BFS oder DFS	DNA Array	DNA Empty	DNA HashMultimap	Neo4J	Tinkergraph	Bitsy (persistent)	Bitsy (In-Memory)	OrientDB
1k Knoten, ungerichtet	BFS oder DFS	DNA E	DNA Array	DNA HashMultimap	Tinkergraph	Bitsy (persistent)	Bitsy (In-Memory)	Neo4J	OrientDB
10k Knoten, ungerichtet	BFS oder DFS	DNA E	DNA Array	DNA HashMultimap	Tinkergraph	Bitsy (persistent)	Bitsy (In-Memory)	Neo4J	OrientDB
1k Knoten, gerichtet	Lokaler Clustering-Koeffizient	DNA Array	DNA HashMultimap	DNA Empty	Tinkergraph	Bitsy (persistent)	Bitsy (In-Memory)	Neo4J	OrientDB
10k Knoten, gerichtet	Lokaler Clustering-Koeffizient	DNA Array	DNA HashMultimap	DNA Empty	Tinkergraph	Neo4J	Bitsy (In-Memory)	Bitsy (persistent)	OrientDB

Fazit & Ausblick

Vergleich von Graphdatenbanken

Erweiterung möglich

Weitere Datenbanken hinzufügen

Weitere Messungen

Import-Funktion