

# Skalierbare Graph-basierte Analyse und Business Intelligence

André Petermann

Universität Leipzig & ScaDS Dreden/Leipzig



UNIVERSITÄT LEIPZIG



Graphs are everywhere

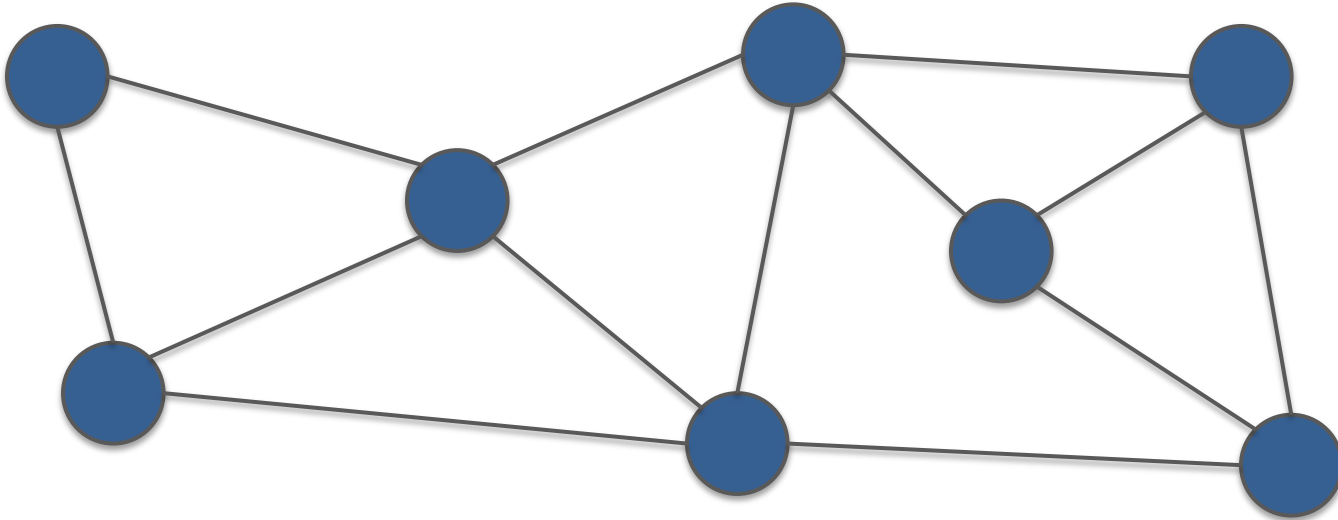
GRADOOP

Business Intelligence

Zusammenfassung

**Graphs are everywhere<sup>1</sup>**

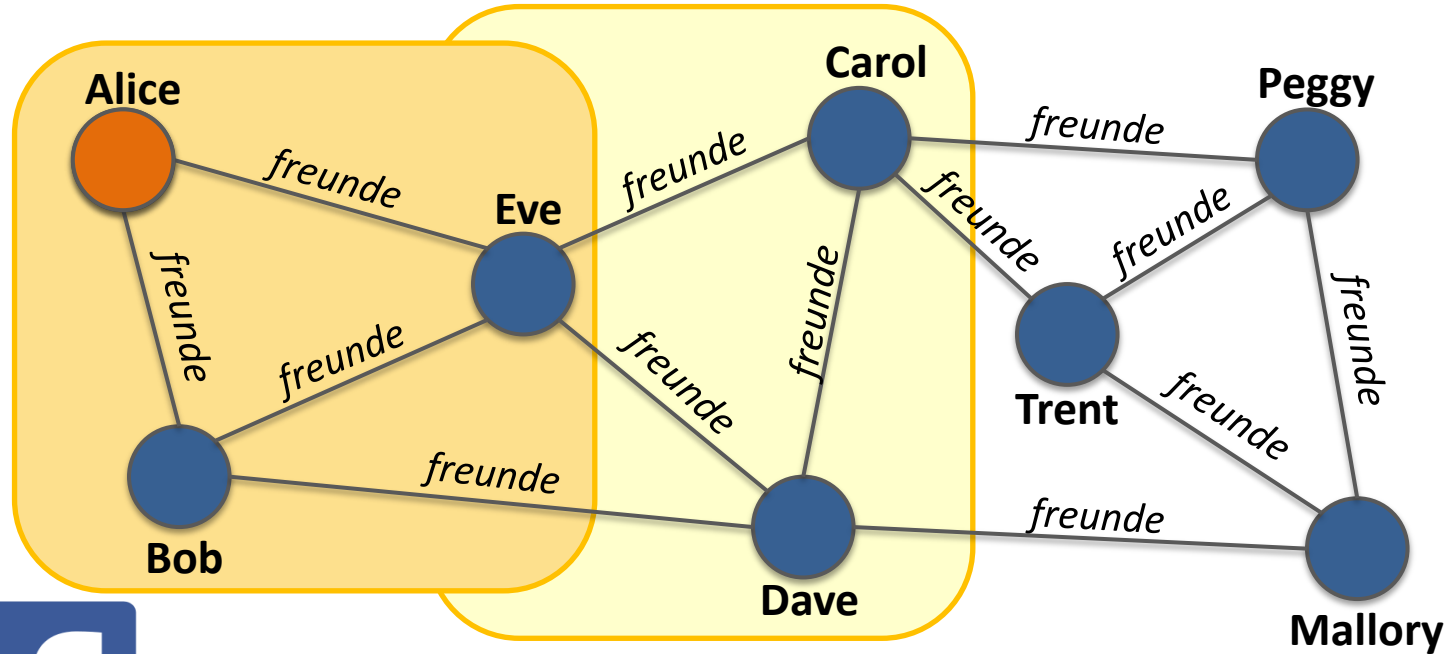
<sup>1</sup>The Neo4j guys



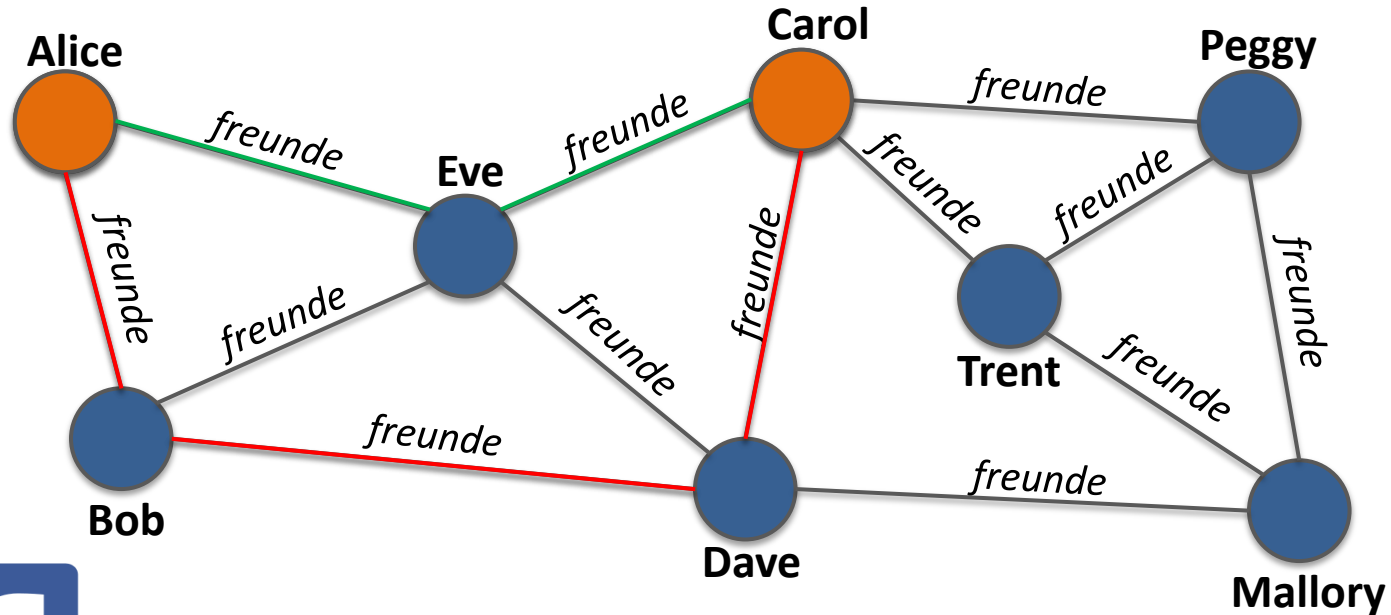
*Graph = (Knoten, Kanten)*



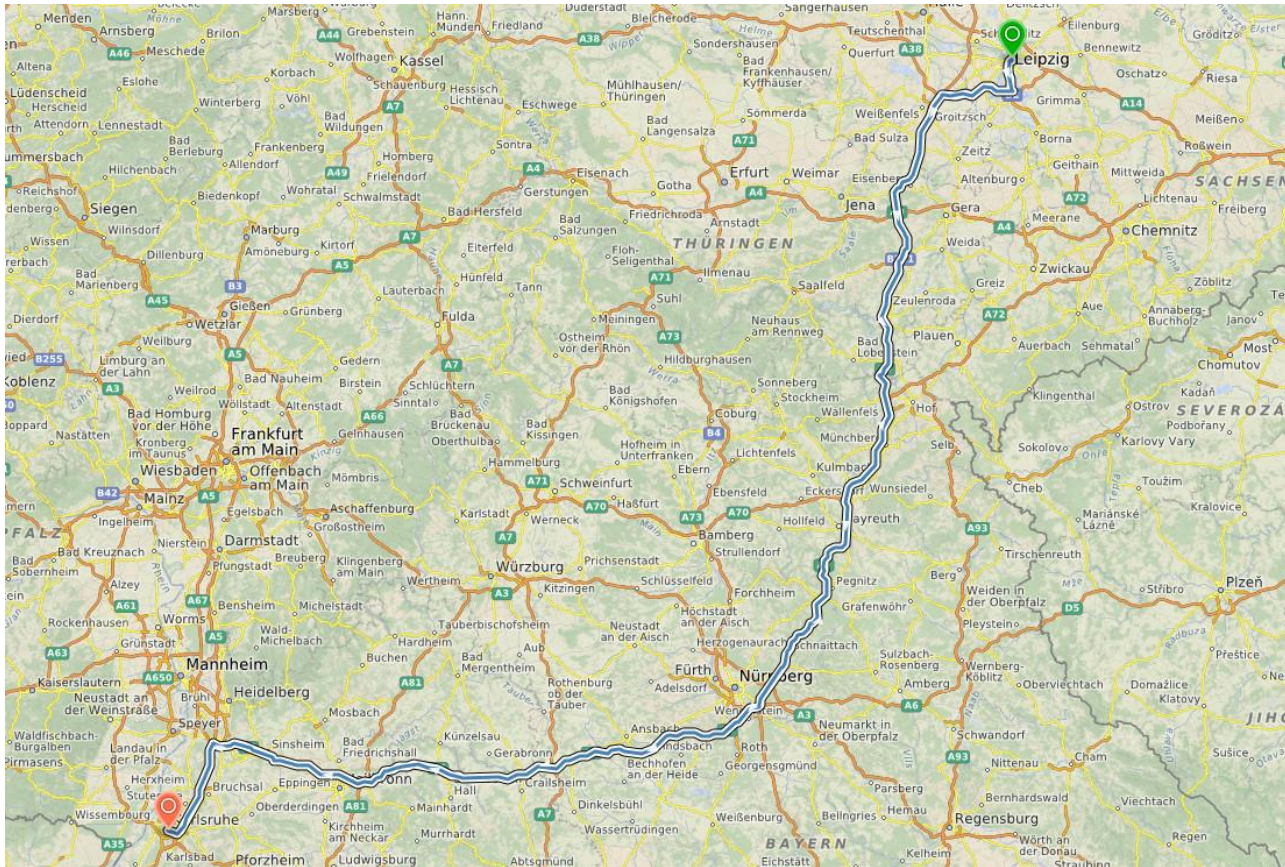
Quelle: <https://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919/>



*Graph = (Nutzer, Freundschaften)*

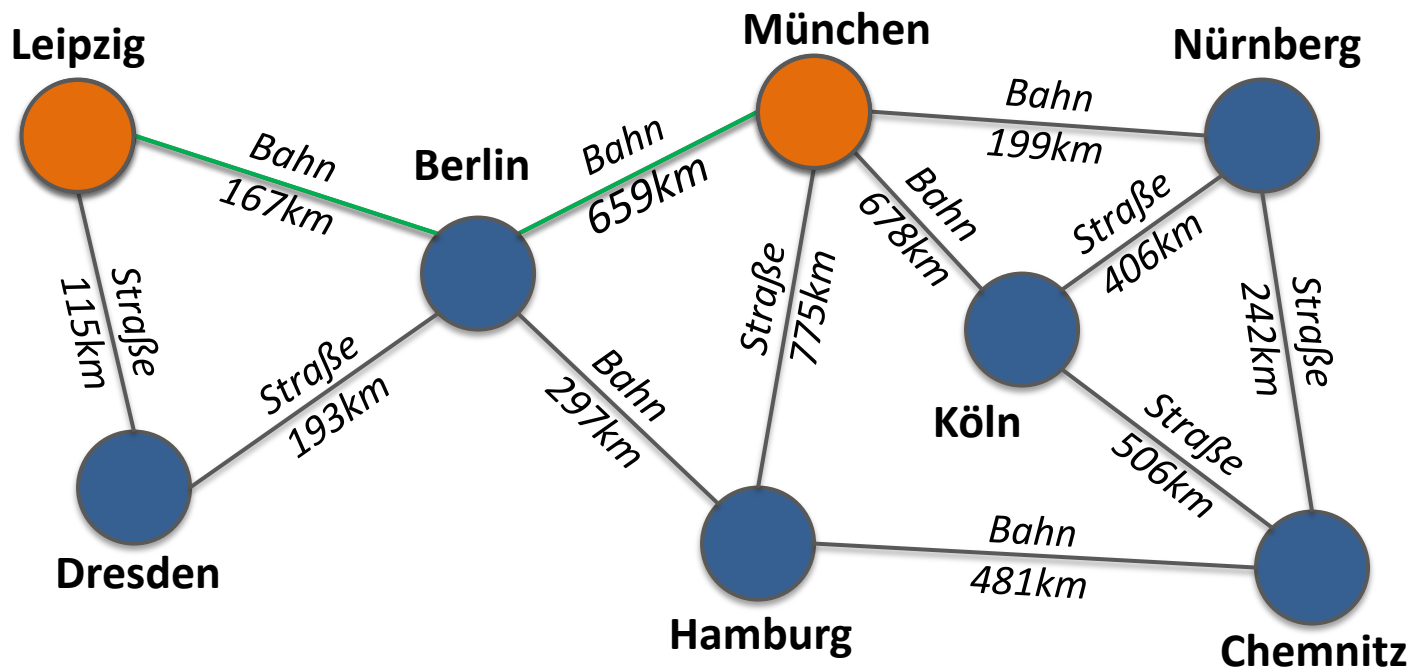


$Graph = (Nutzer, Freundschaften)$



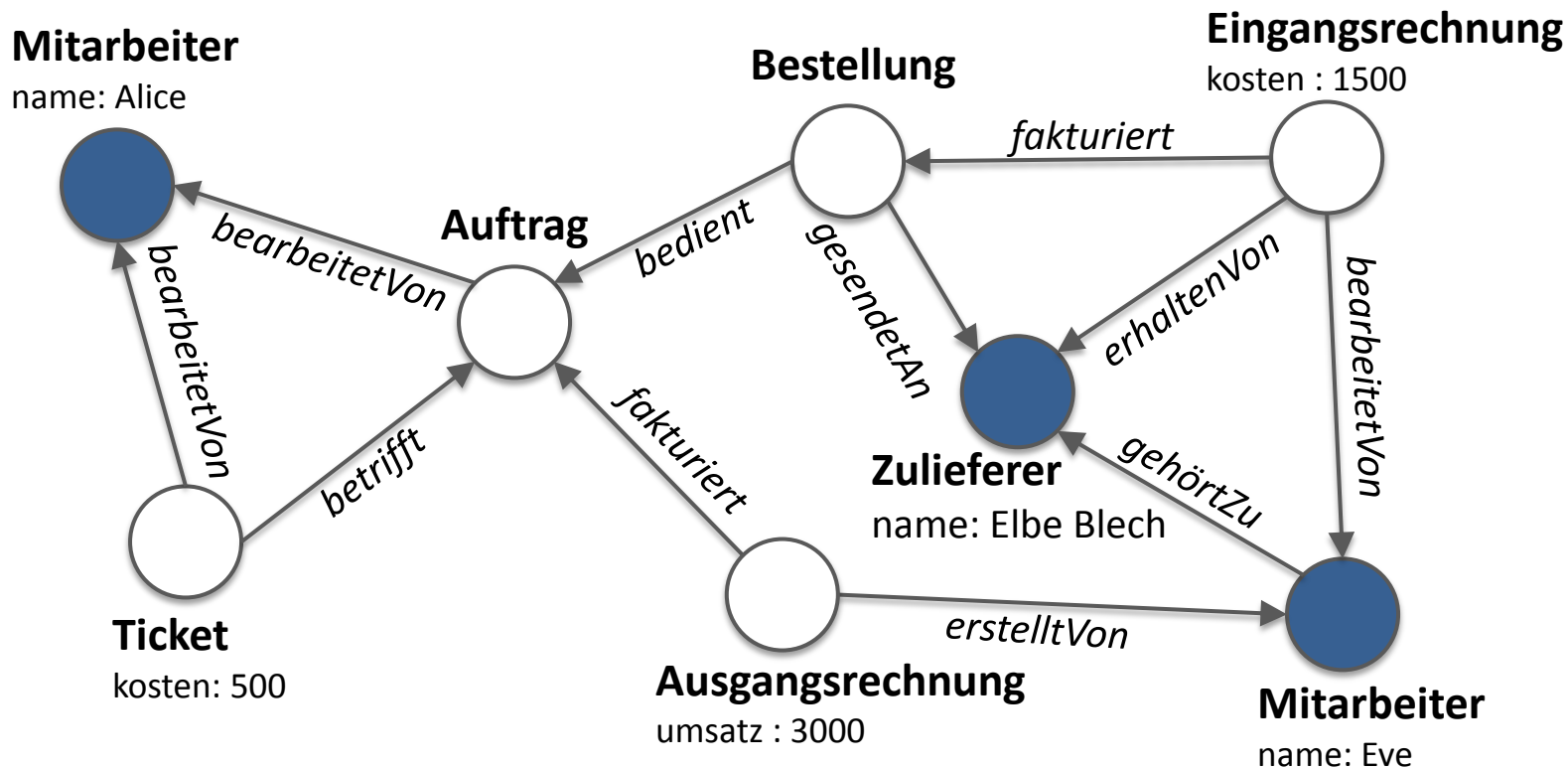
Quelle: <https://www.openstreetmap.de/>





*Graph = (Städte, Verbindungen)*

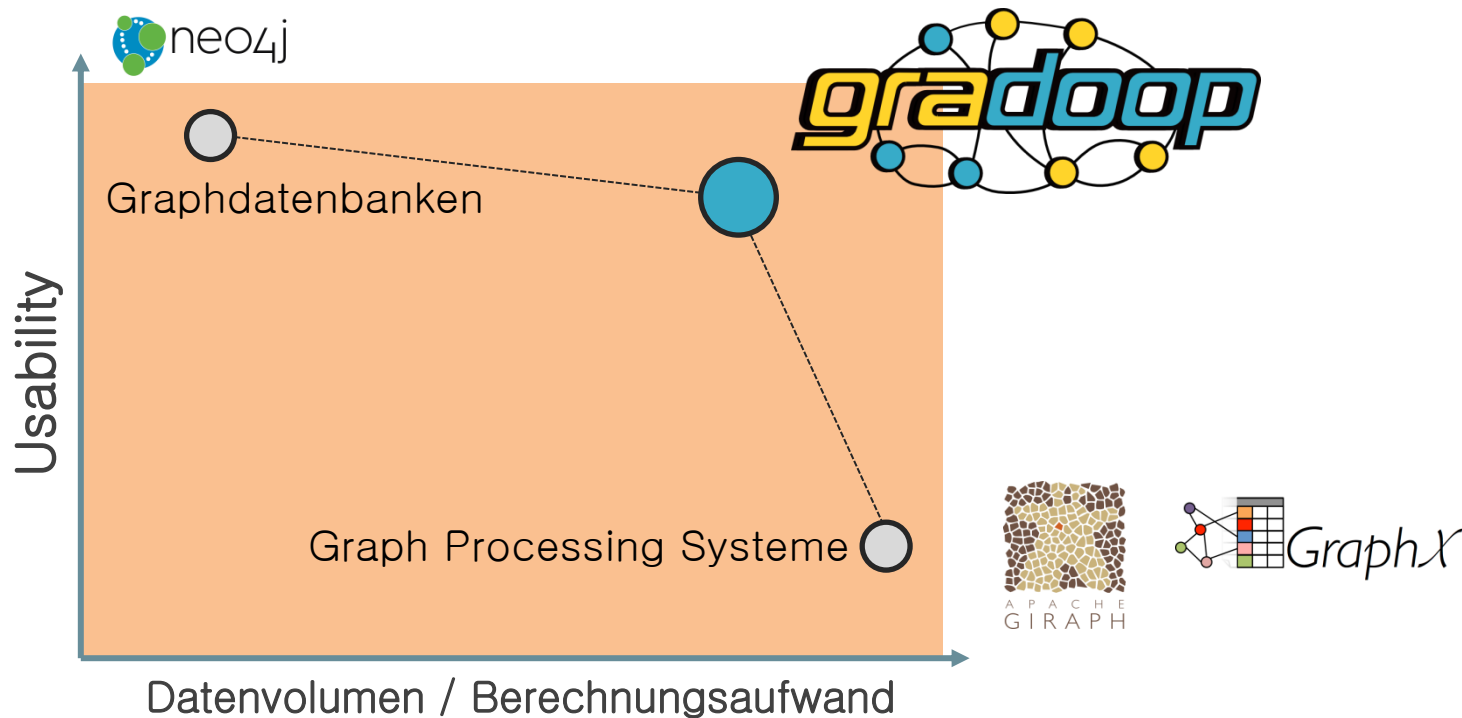




*Graph = (Geschäftsobjekte, Kontext)*

# GRADOOP

## Graph Analytics with Hadoop





Ein Open Source Framework für die  
**deklarative** Analyse **verteilter** Graphdaten

- Ausdrucksstarkes Datenmodell (Extended Property Graph Model)
- Analytische Algorithmen und Operatoren
- Analytical Program API (Deklarativ)
- Implementiert auf Big Data Technology





Analytical Program API

Extended Property Graph Model (EPGM)

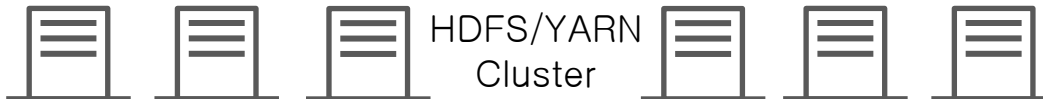
Apache Flink: Operator Implementation



Apache Flink: Distributed Operator Execution

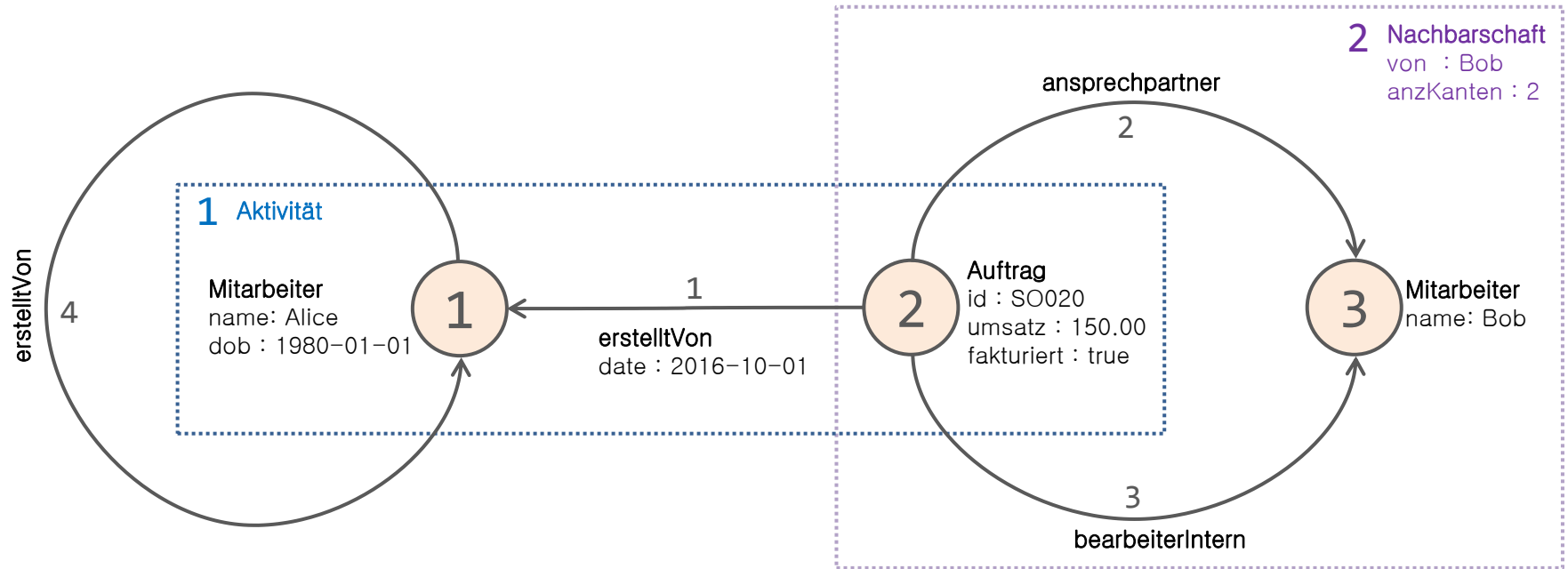


Apache HBase: Distributed Graph Store





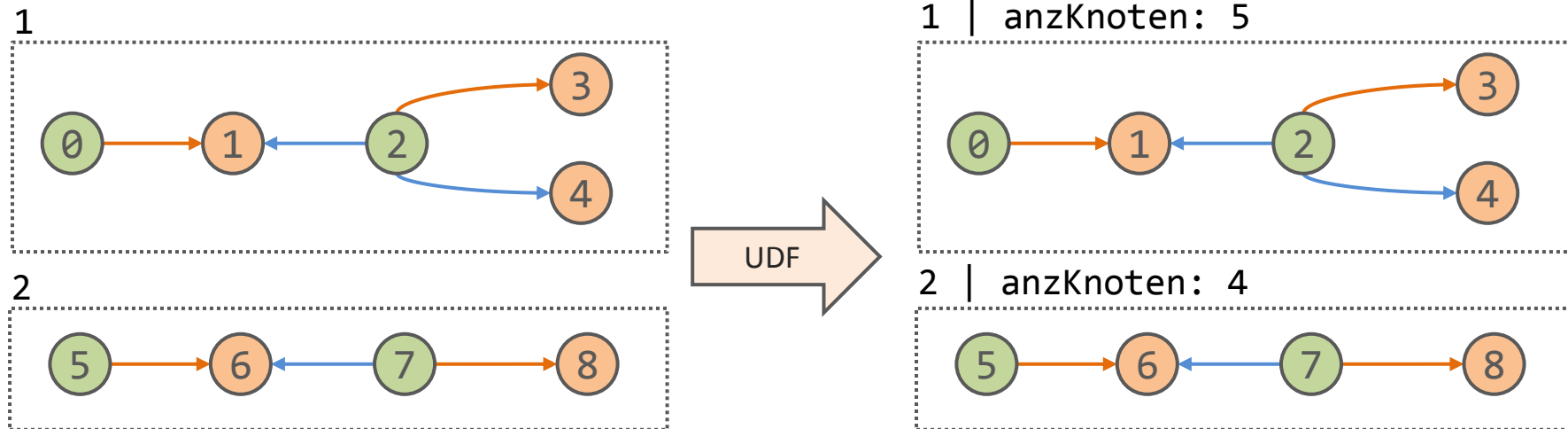
# Extended Property Graph Model (EPGM)



## Analytical Program API

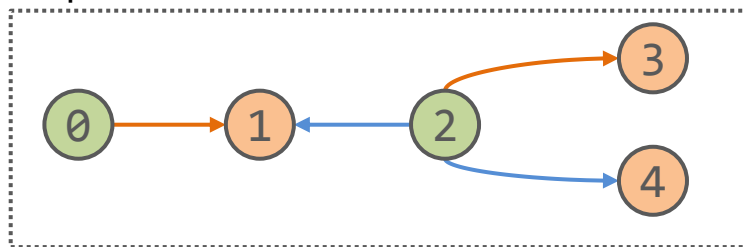
- Deklarative Java-API für das EPGM
- Operatoren für Logische Graphen und Graphmengen
- Anwendungslogik in benutzerdefinierten Funktionen
- Plugin-Konzept für externe Graphalgorithmen

## Aggregation

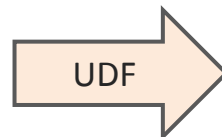


## Selection

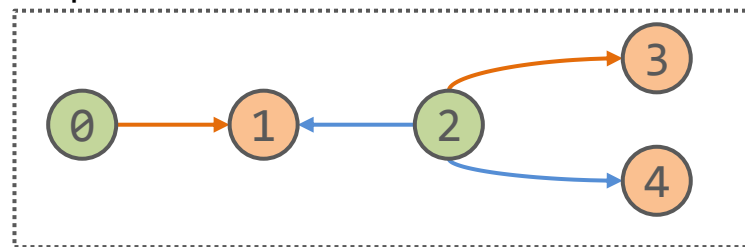
1 | anzKnoten: 5



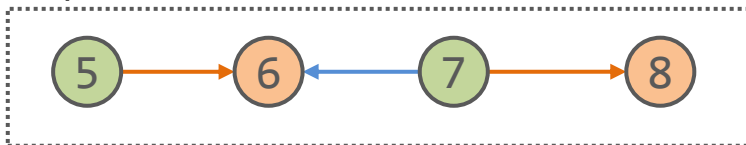
anzKnoten &gt; 4



1 | anzKnoten: 5



2 | anzKnoten: 4

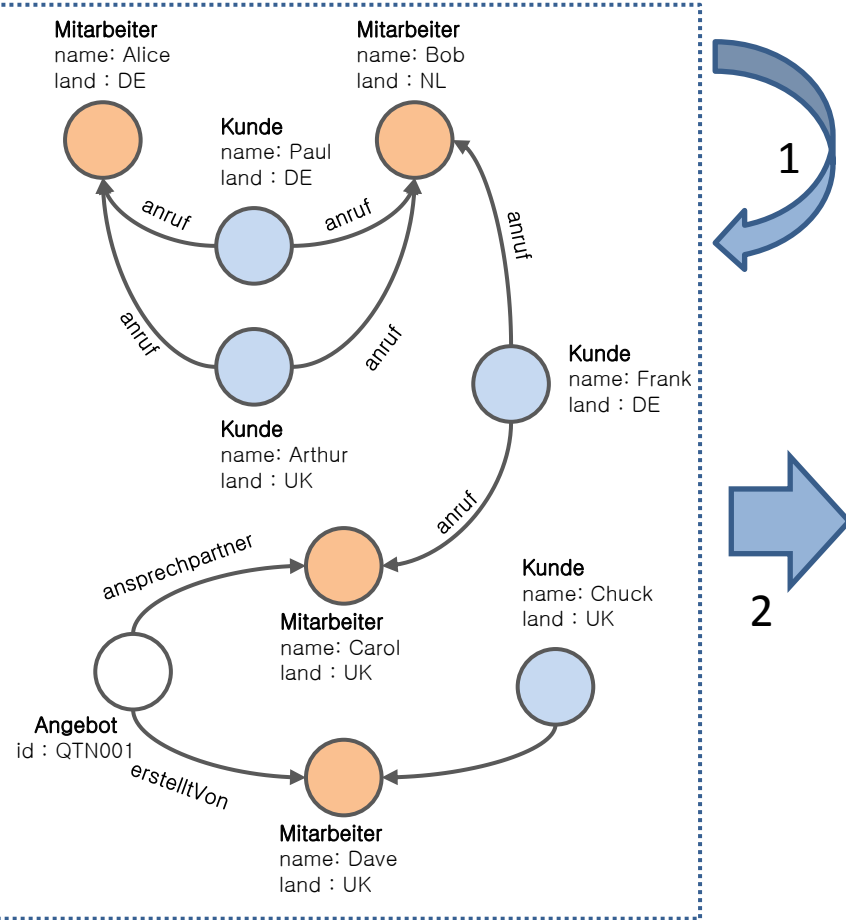


# Business Intelligence

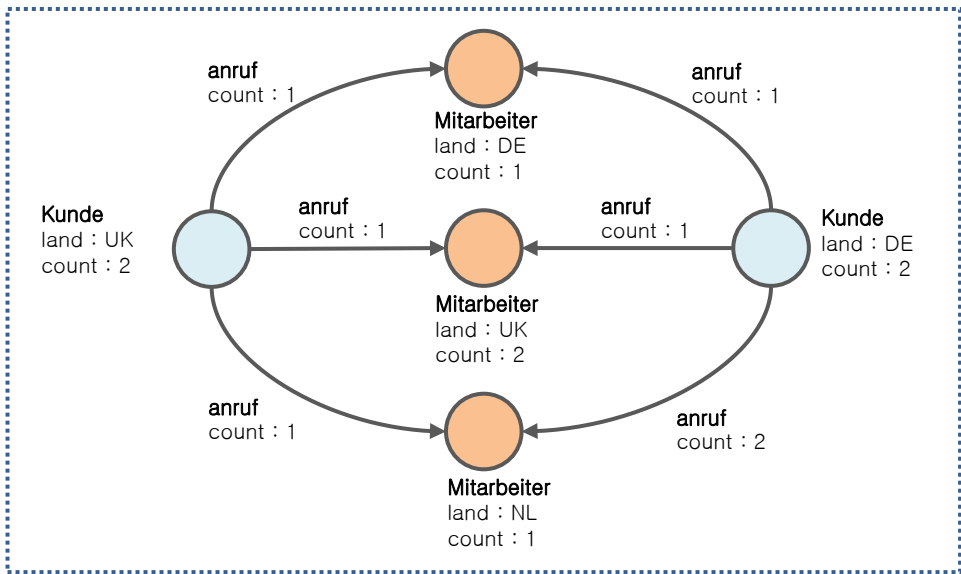


Beispiel 1 : Geographische Kundenkontakte mit Graph Grouping

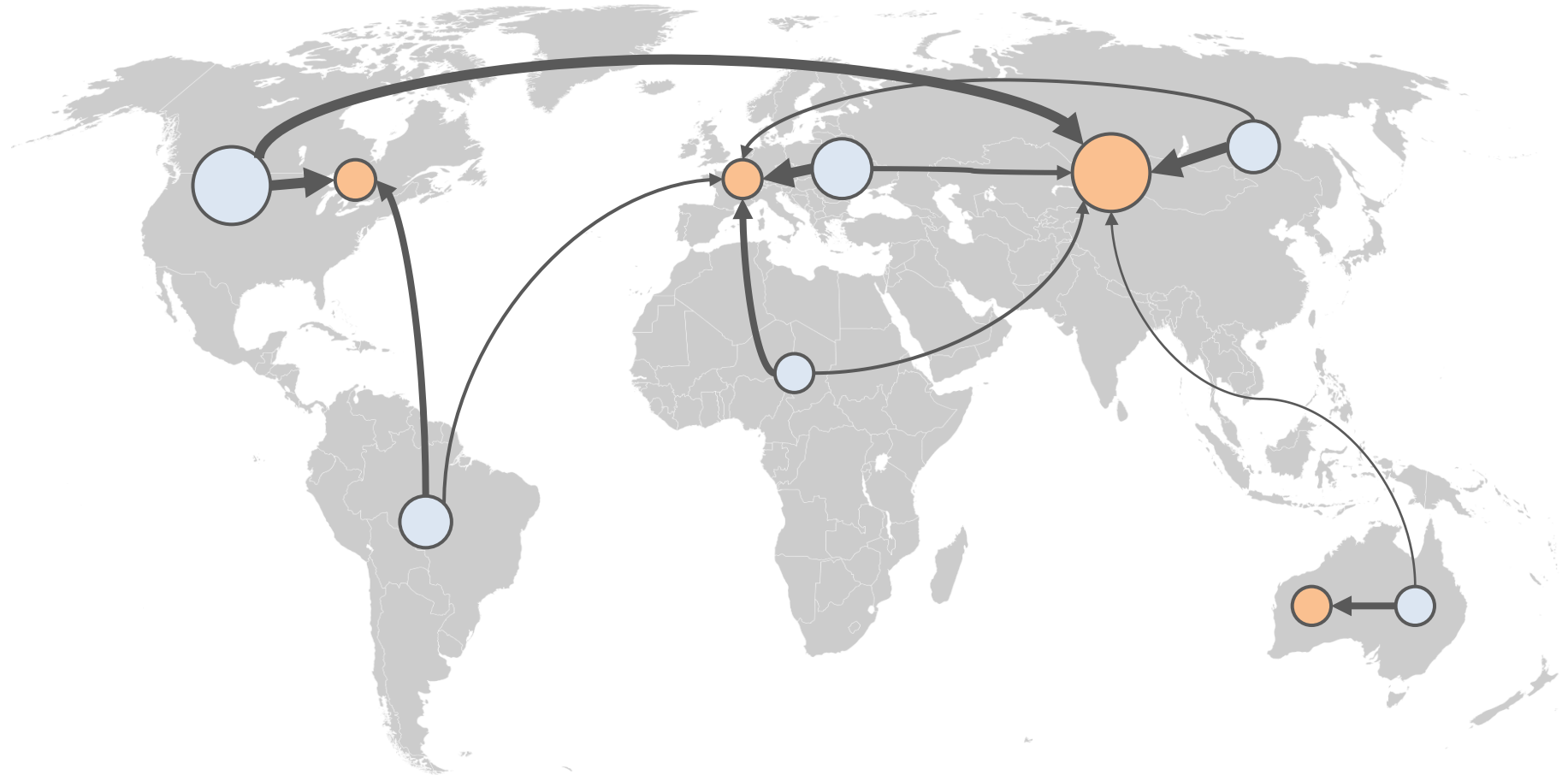
Wie oft telefonieren Mitarbeiter und Kunden aus verschiedenen Ländern miteinander?



# 1. Entferne uninteressante Knoten (und Kanten)




# 2. Gruppierere Knoten nach Typ und Land (und Kanten nach Typ)






Wie oft telefonieren Mitarbeiter  und Kunden  aus verschiedenen Ländern miteinander?

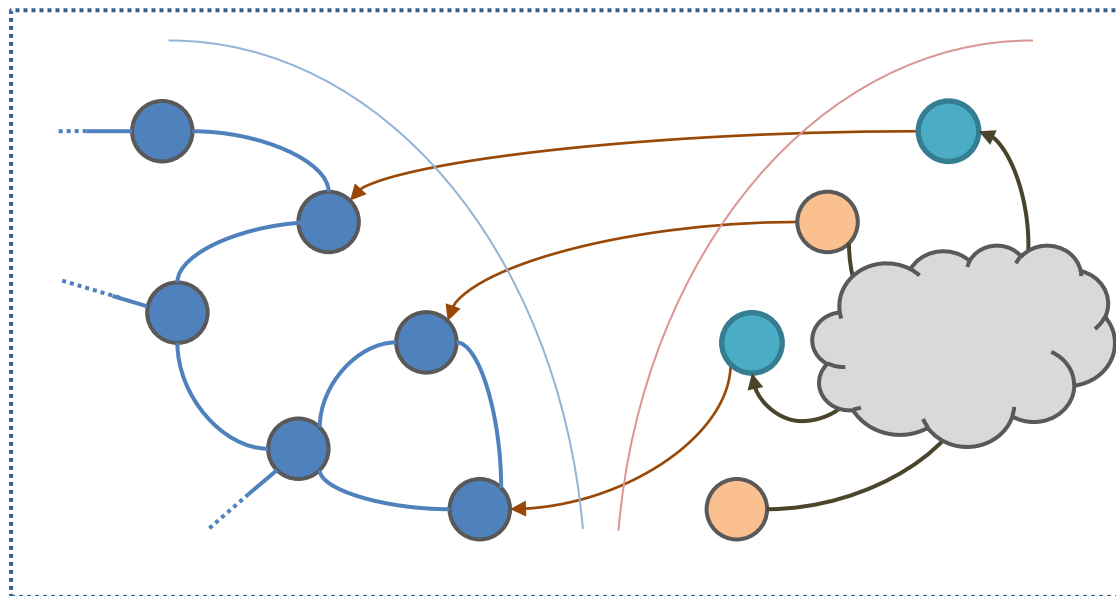






Beispiel 2:  
Nach verdächtigen Mustern suchen mit Pattern Matching

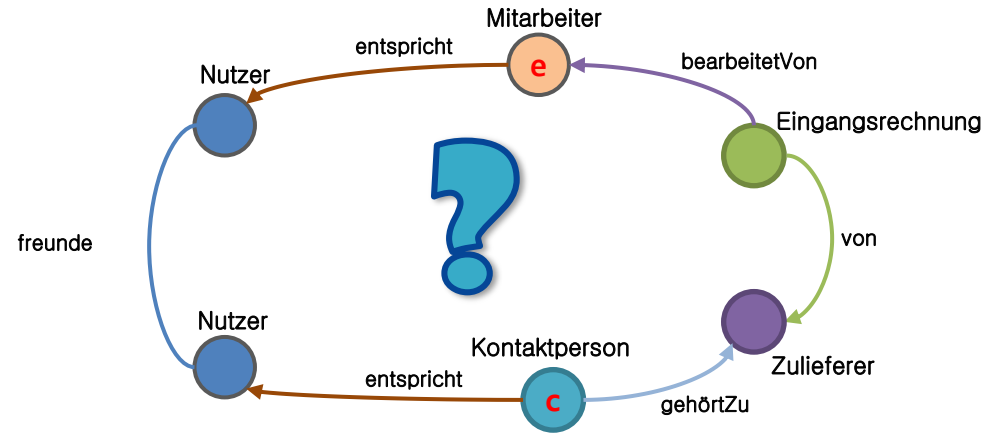
### Soziales Netzwerk

-  Nutzer
-  freunde
-  entspricht



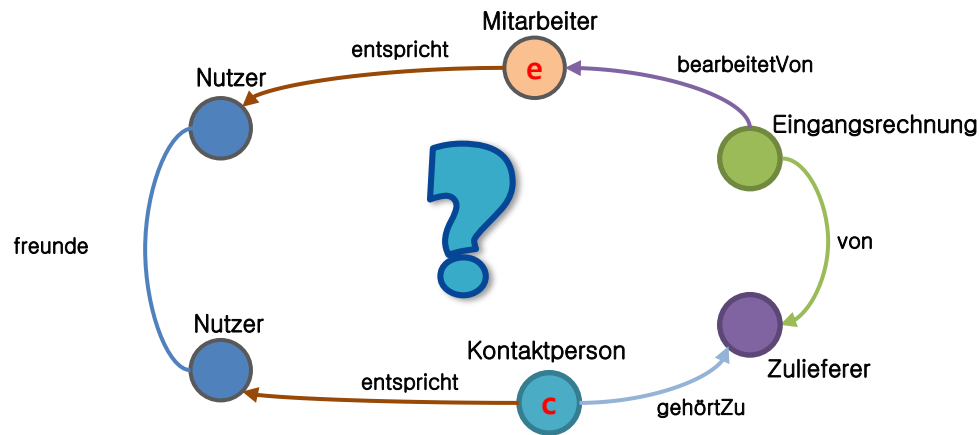
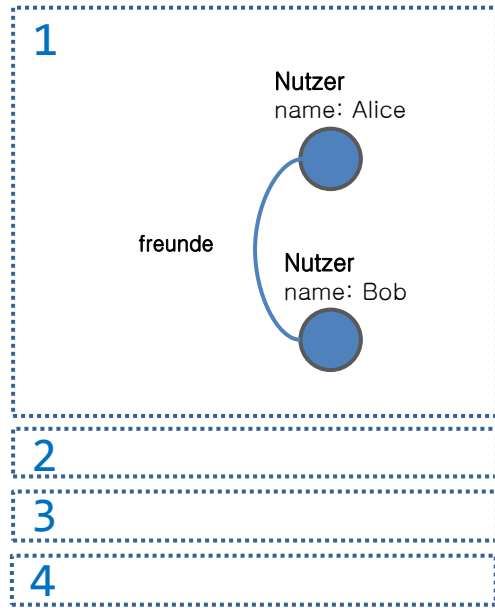
### Geschäftsdaten

-  Mitarbeiter
-  Kontaktperson



```
GraphCollection matches = graph.match(
```

```
(e:Mitarbeiter)<-[:bearbeitetVon]-(:Eingangsrechnung)-[:von]->(:Zulieferer)<-[:gehörtZu]-(:Kontaktperson)
(c)-[:entspricht]->(:Nutzer)-[:freunde]->(:Nutzer)<-[:entspricht]-(:e))
```



SOURCE	EDGE	TARGET
Alice	freunde	Bob
Alice	freunde	Bob
Eve	freunde	Frank
Alice	freunde	Bob

SRC	TRGT	CNT
Alice	Bob	3
Eve	Frank	1

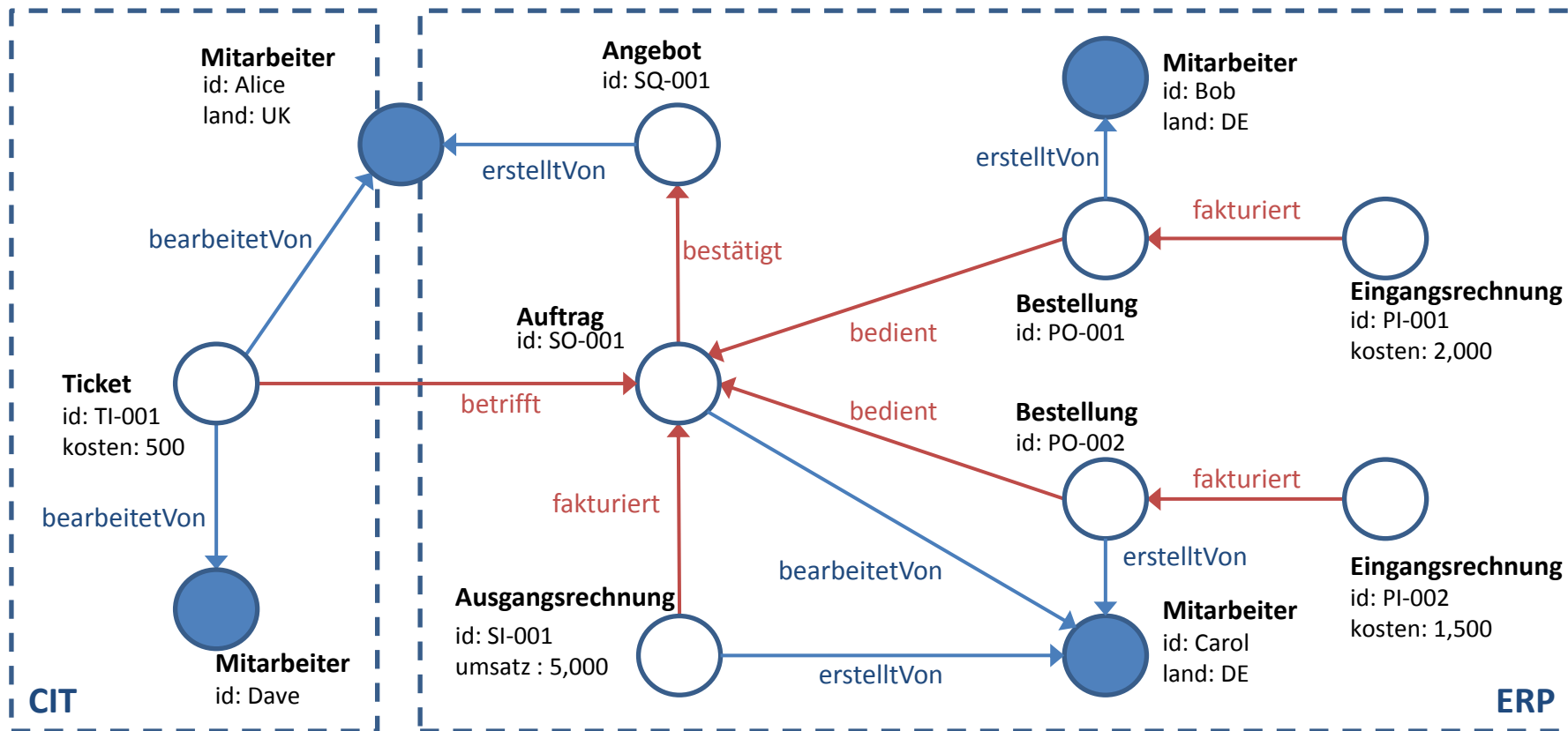
Beispiel 3:  
Graph Pattern Mining  
in Geschäftsprozessen

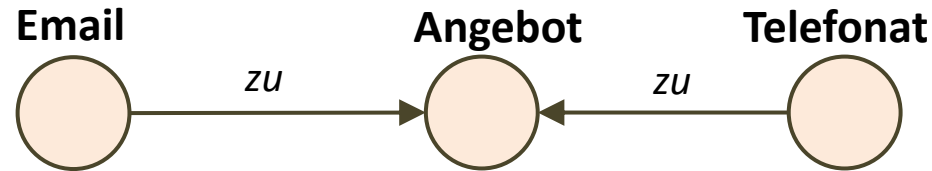
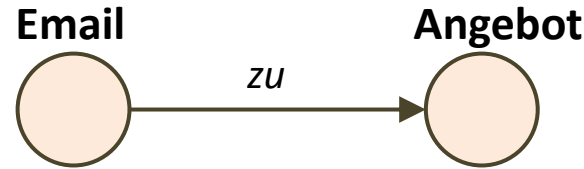


## Graph Pattern Mining

- Eingabe: Graphmenge, Kennzahl und Kategorisierung
- Muster werden automatisch gefunden
- Keine Vermutung notwendig (z.B. überprüfe E-Mails)
- Muster können weiter verwendet werden (Filtern, Anzeigen,...)

# Prozessausführung als Graph





Welche Muster sind charakteristisch für Erfolg und Misserfolg eines Geschäftsprozesses?



# Zusammenfassung

- Open Source Framework
- Deklarative Graphanalyse
- Ausdrucksstarkes und flexibles Datenmodell
- Erweiterung von BI durch neuartige Netzwerkanalysen
- Big Data ready



Junghanns, M.; Petermann, A.

### **Verteilte Graphanalyse mit Gradoop**

JavaSPEKTRUM 05/2016

Junghanns, M.; Petermann, A.; Teichmann, N.; Rahm, E.;

### **The Big Picture: Understanding large-scale graphs using Graph Grouping with GRADOOP**

Demonstration @ Datenbanksysteme für Business, Technologie und Web (2017)

Petermann, A.; Junghanns, M.; Kemper, S.; Gomez, K.; Teichmann, N.; Rahm, E.

### **Graph Mining for Complex Data Analytics**

Demonstration @ IEEE International Conference on Data Mining (2016)

Petermann, A.; Junghanns, M.

### **Scalable Business Intelligence with Graph Collections**

it – Information Technology, Special Issue: Big Data Analytics (2016)

Junghanns, M.; Petermann, A.; Teichmann, N.; Gomez, K.; Rahm, E.

### **Analyzing Extended Property Graphs with Apache Flink**

International Workshop on Network Data Analytics @ SIGMOD (2016)

Petermann, A.; Junghanns, M.; Müller, R.; Rahm, E.

### **Graph-based Data Integration and Business Intelligence with BIIG**

Demonstration @ International Conference on Very Large Databases (2014)

[www.gradoop.com](http://www.gradoop.com)



UNIVERSITÄT LEIPZIG

ScaDS   
DRESDEN LEIPZIG