

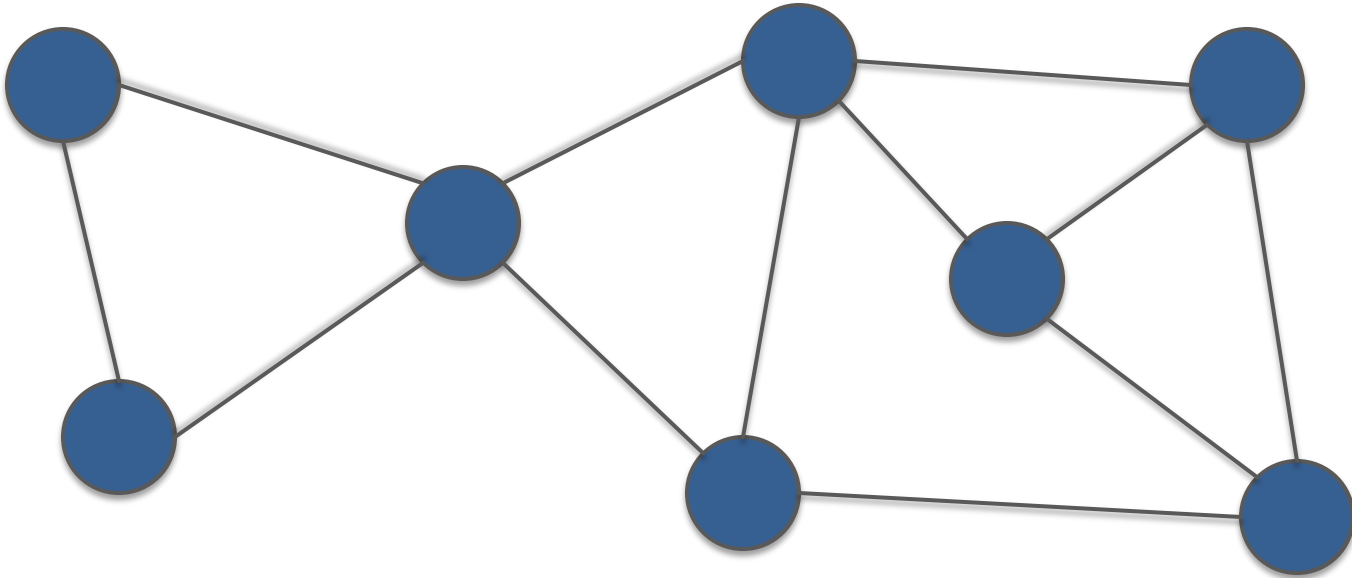
Seminar WS 2014/15

# Graph Data Management

Abteilung Datenbanken

Universität Leipzig

*„Graphs are everywhere“*



*Graph = (Vertices, Edges)*

# „Graphs are everywhere“

*(Online) Soziale Netzwerke*

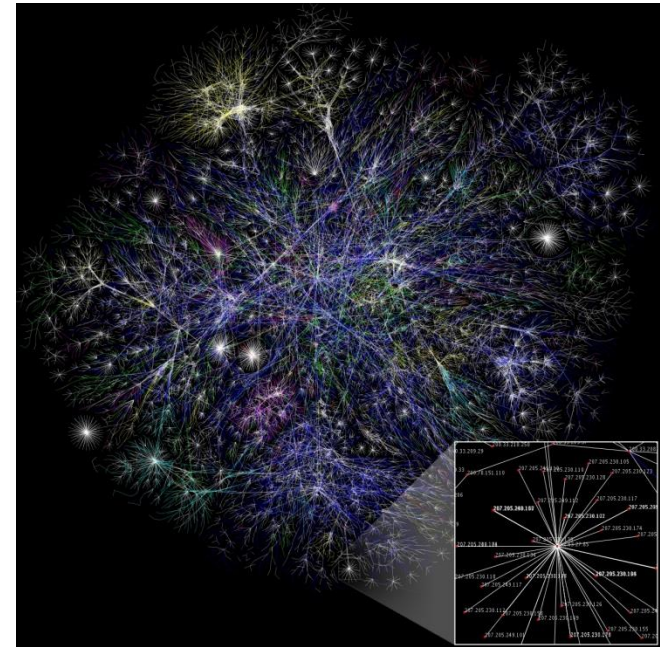
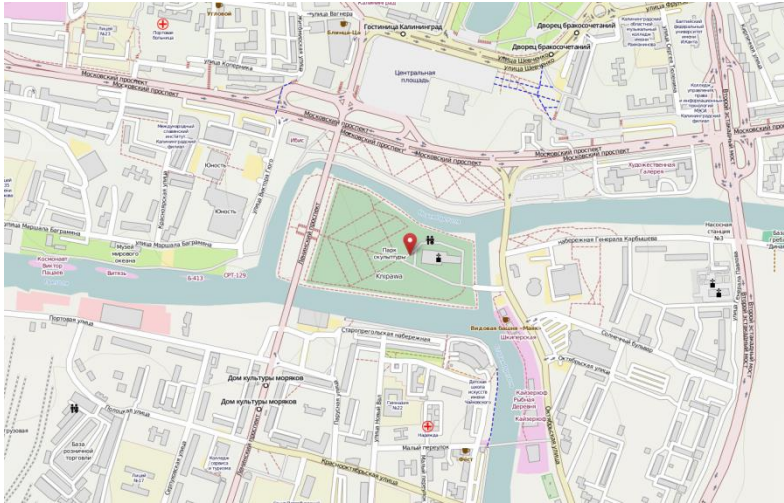


- Facebook
  - ca. 1.3 Milliarden Nutzer
  - ca. 340 Freunde pro Nutzer
- Twitter
  - ca. 300 Millionen Nutzer
  - ca. 500 Millionen Tweets pro Tag

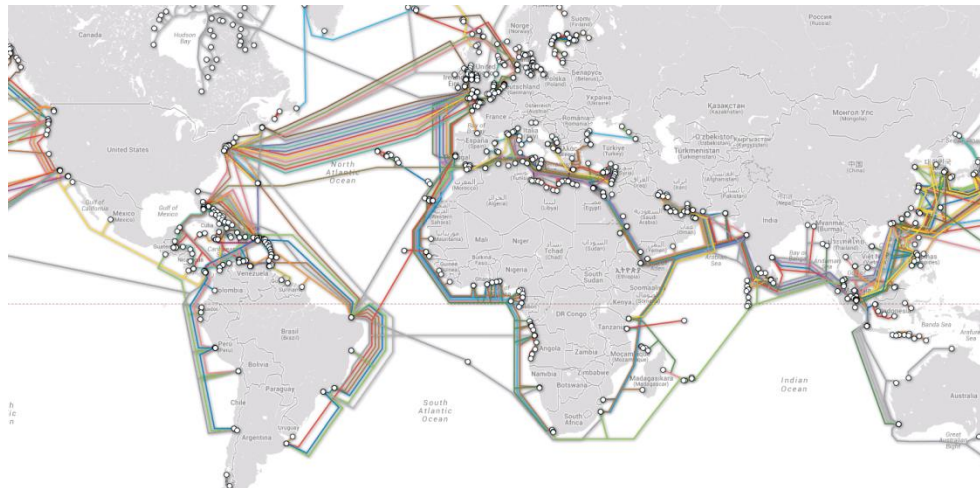


# „Graphs are everywhere“

## Technologische Netzwerke

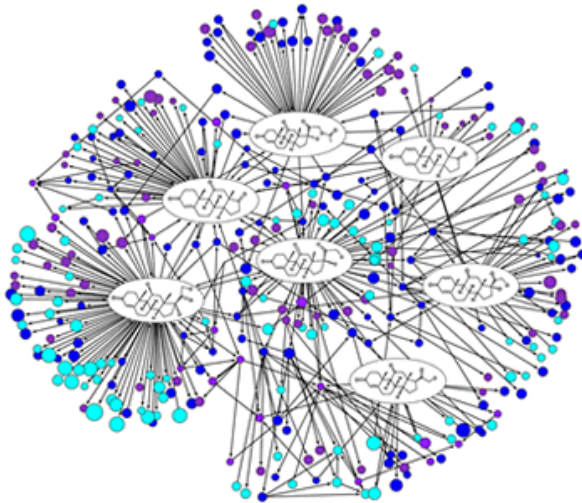


- Internet
  - ca. 2.9 Milliarden Nutzer

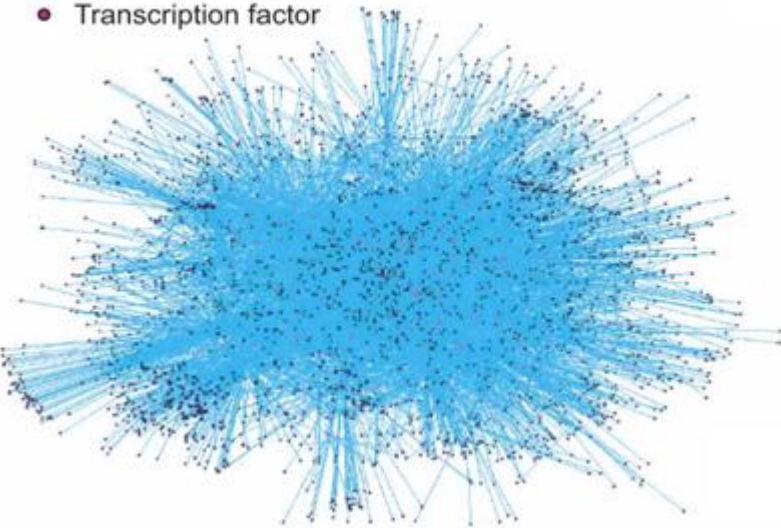


# „Graphs are everywhere“

*Chemische / biologische /medizinische Netzwerke*



- Regulated target
- Transcription factor

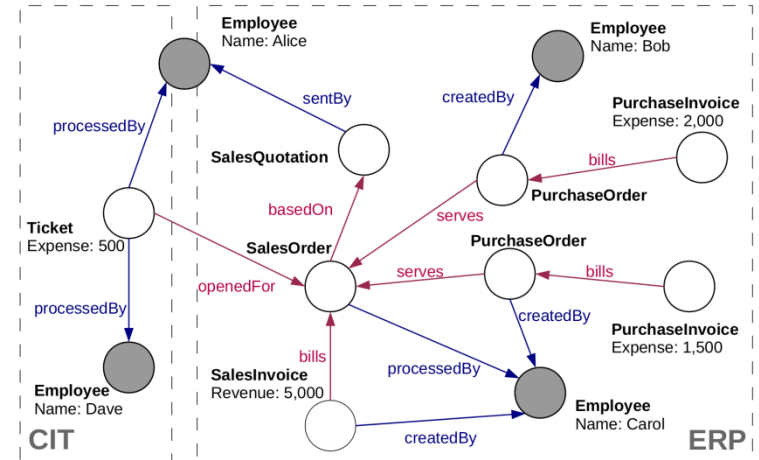


- Gene (Mensch)
  - 20,000-25,000
  - ca. 4 Millionen strukturelle Varianten (Individuen)
- Molekulare Interaktionen (bekannt)
  - IntAct: ca. 450.000 Interaktionen, ca. 80.000 Interaktoren
- Chemie
  - ChEBI: ca. 40000 Entitäten , ca. 70000 Beziehungen in
- Patienten:
  - > 18 Millionen in deutschen Krankenhäusern
- Krankheiten
  - > 30.000

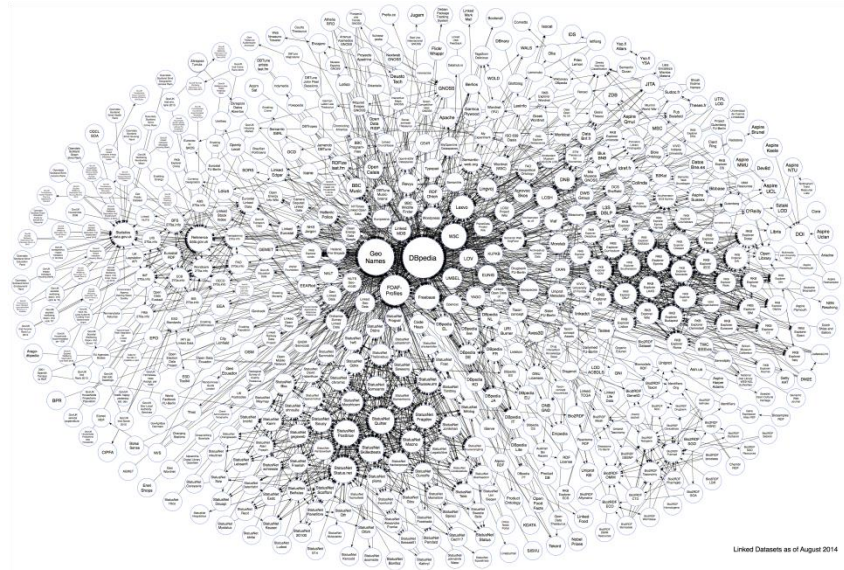


# „Graphs are everywhere“

## Informationsnetzwerke

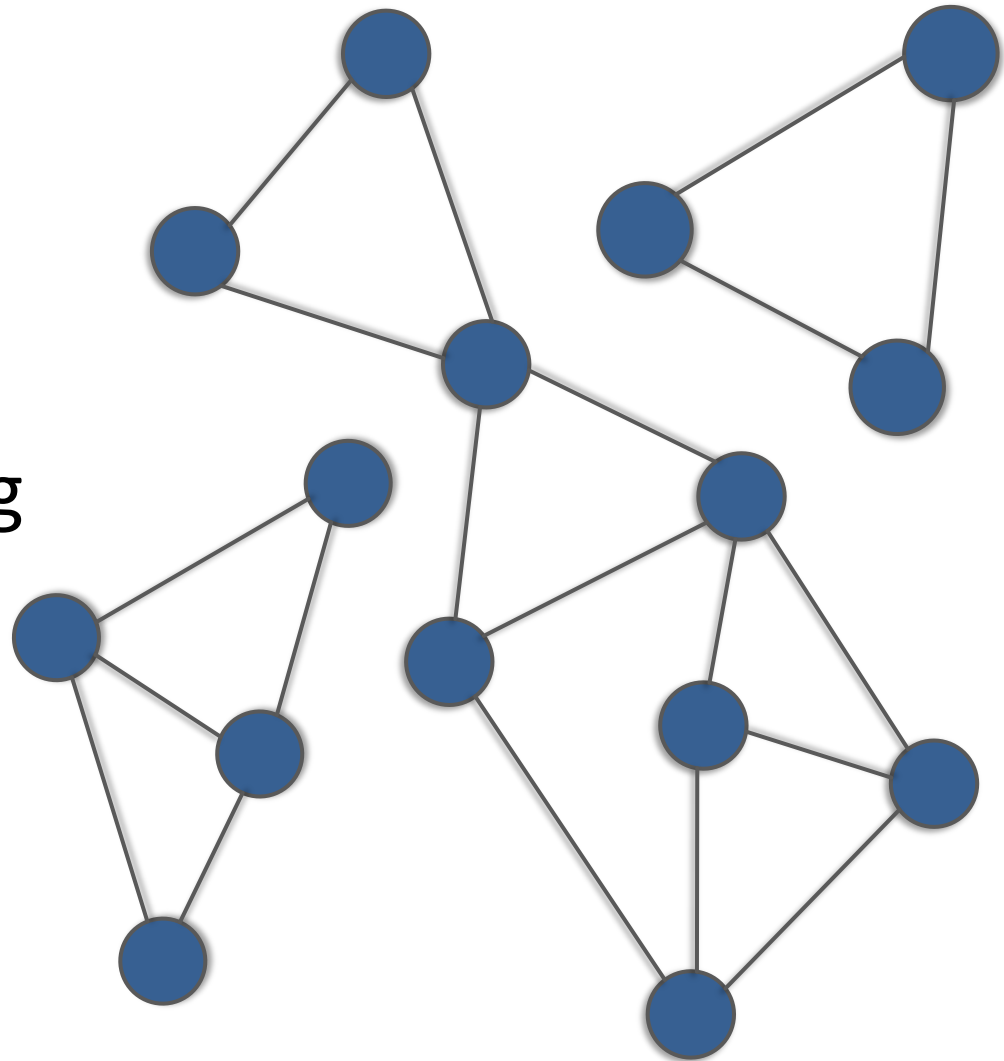


- World Wide Web
  - ca. 1 Milliarde Websites
- LOD-Cloud
  - ca. 31 Milliarden Triples

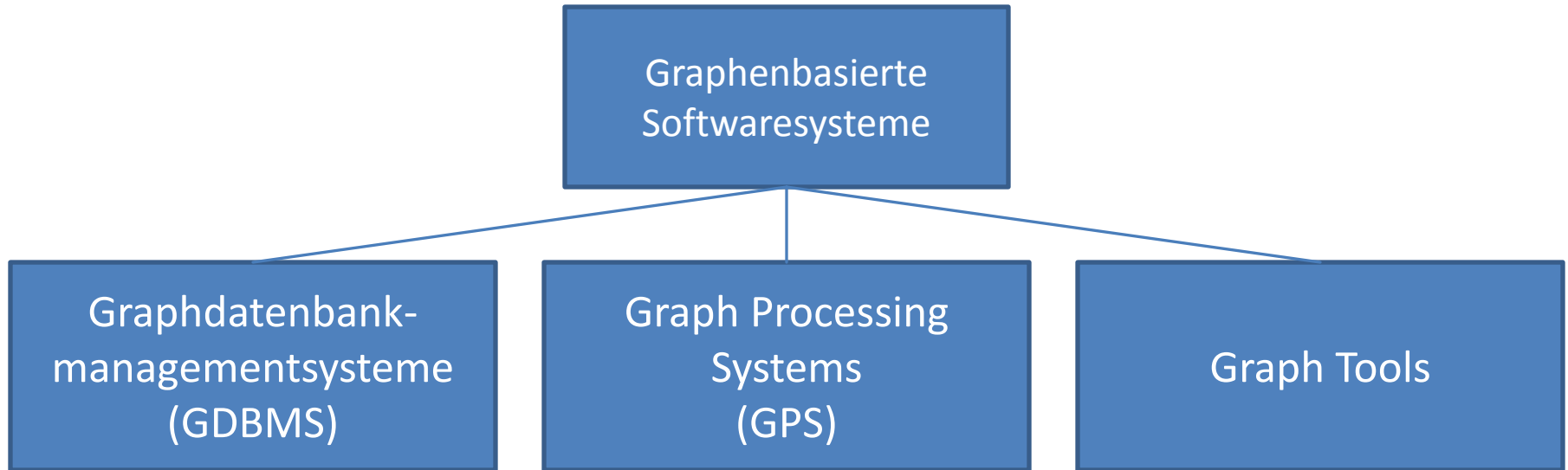


# Graphentheoretische Anwendungen

- Pfadsuchen
- Clustering
- Flüsse
- Partitionierung
- Mustersuche
- Empfehlungen
- ...



# Kategorisierung



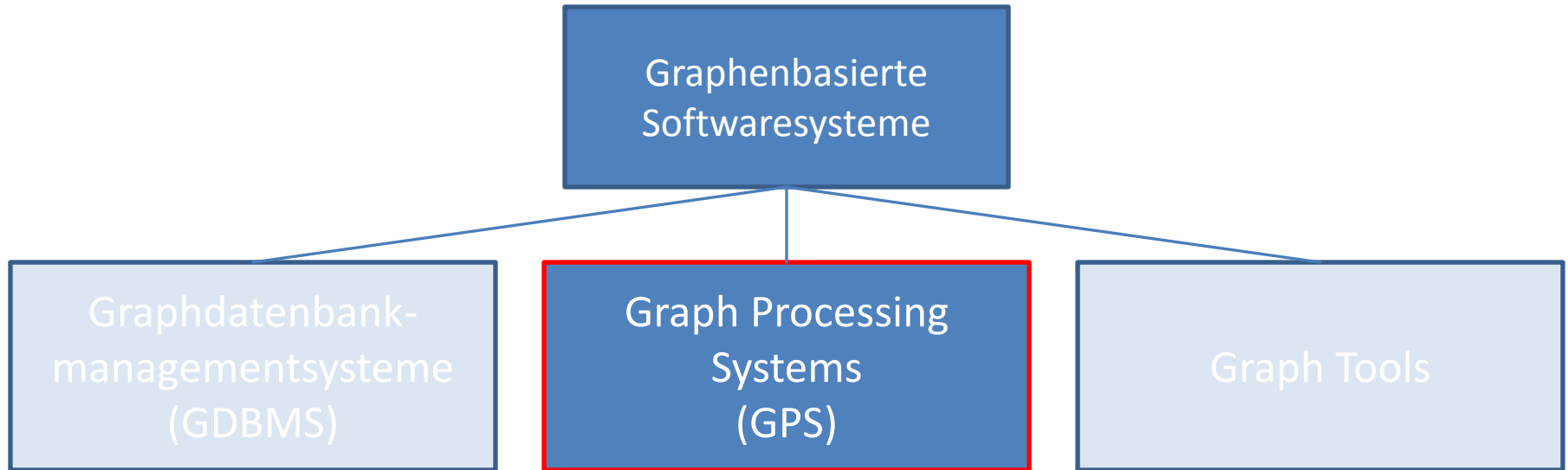
- Interaktiver Zugriff (OLTP)
- Lokale Operationen
- DB-Funktionalitäten
- Vertikale Skalierbarkeit
- Neo4j, OrientDB, ...

- Verteilte Analyse (OLAP)
- Globale Operationen
- Batch Processing
- Horizontale Skalierbarkeit
- Pregel, Giraph, GraphX, ...

- Analyse/Visualisierung
- Lokale/globale Operationen
- SW-Bibliotheken
- Vertikale Skalierbarkeit
- Gephi, JUNG, GraphViz, ...



# Kategorisierung

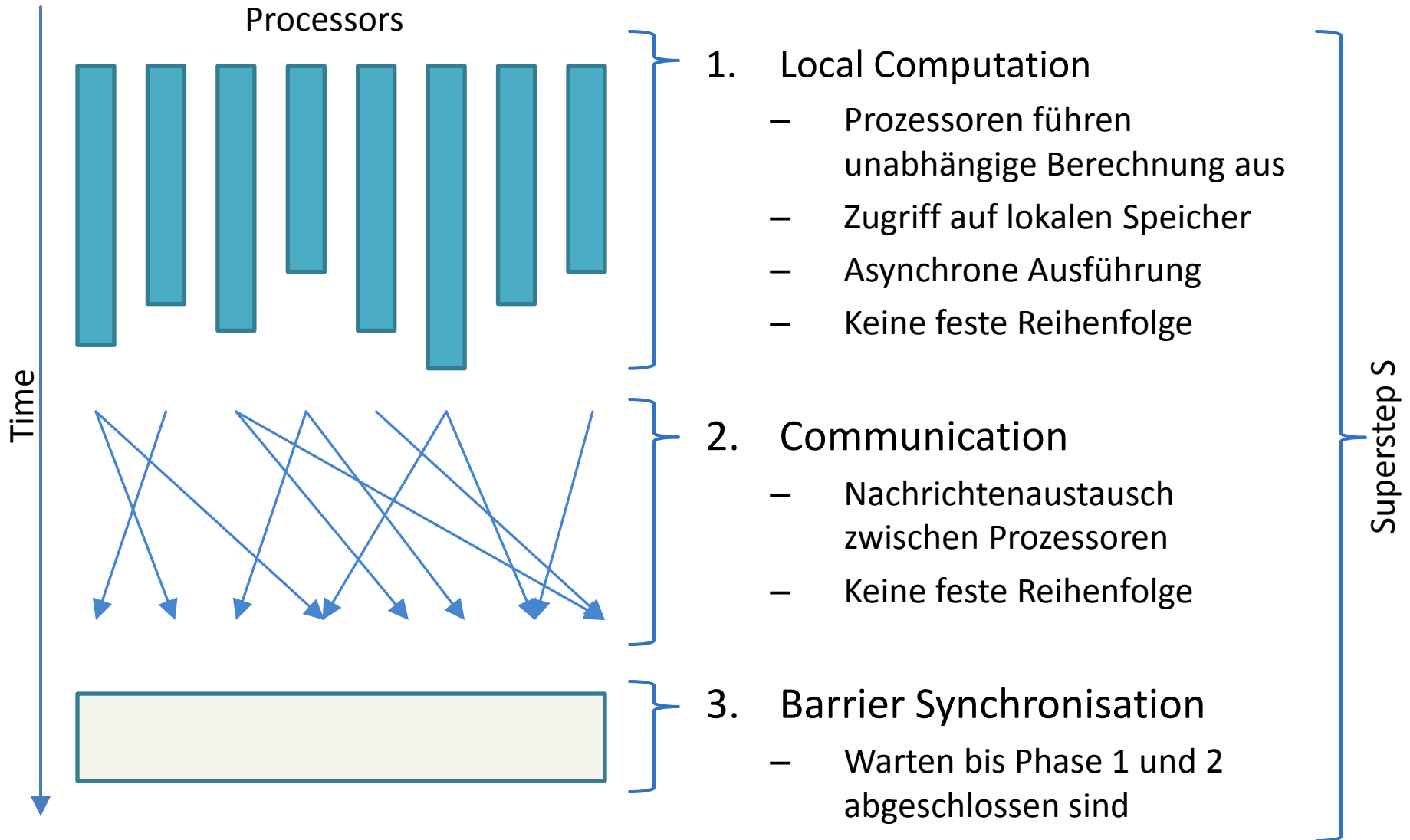


- Interaktiver Zugriff (OLTP)
- Lokale Operationen
- DB-Funktionalitäten
- Vertikale Skalierbarkeit
- Neo4j, OrientDB, ...

- Verteilte Analyse (OLAP)
- Globale Operationen
- Batch Processing
- Horizontale Skalierbarkeit
- Pregel, Giraph, GraphX, ...

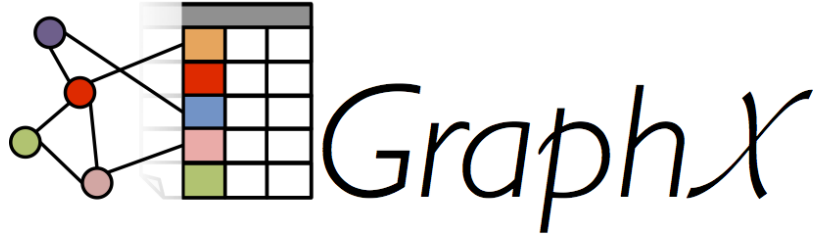
- Analyse/Visualisierung
- Lokale/globale Operationen
- SW-Bibliotheken
- Vertikale Skalierbarkeit
- Gephi, JUNG, GraphViz, ...

# Bulk Synchronous Parallel

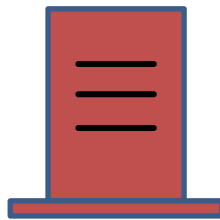
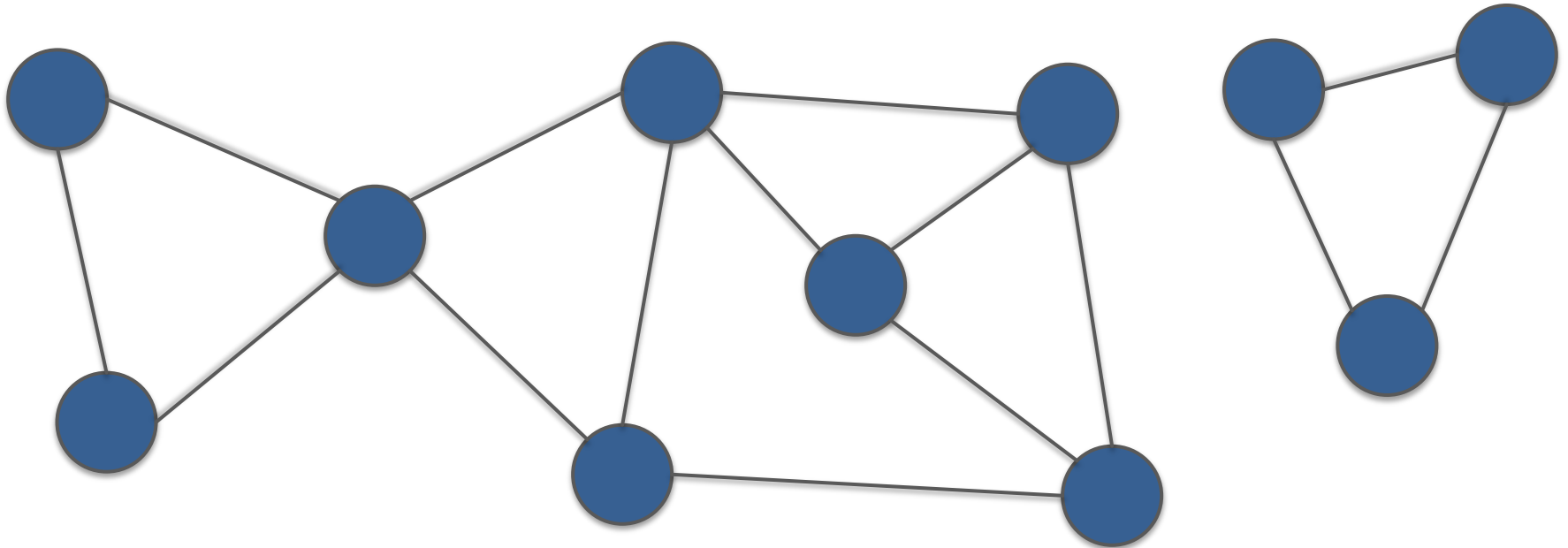




# Graph Processing Systems

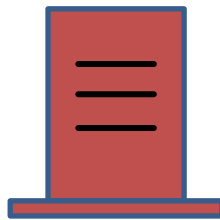
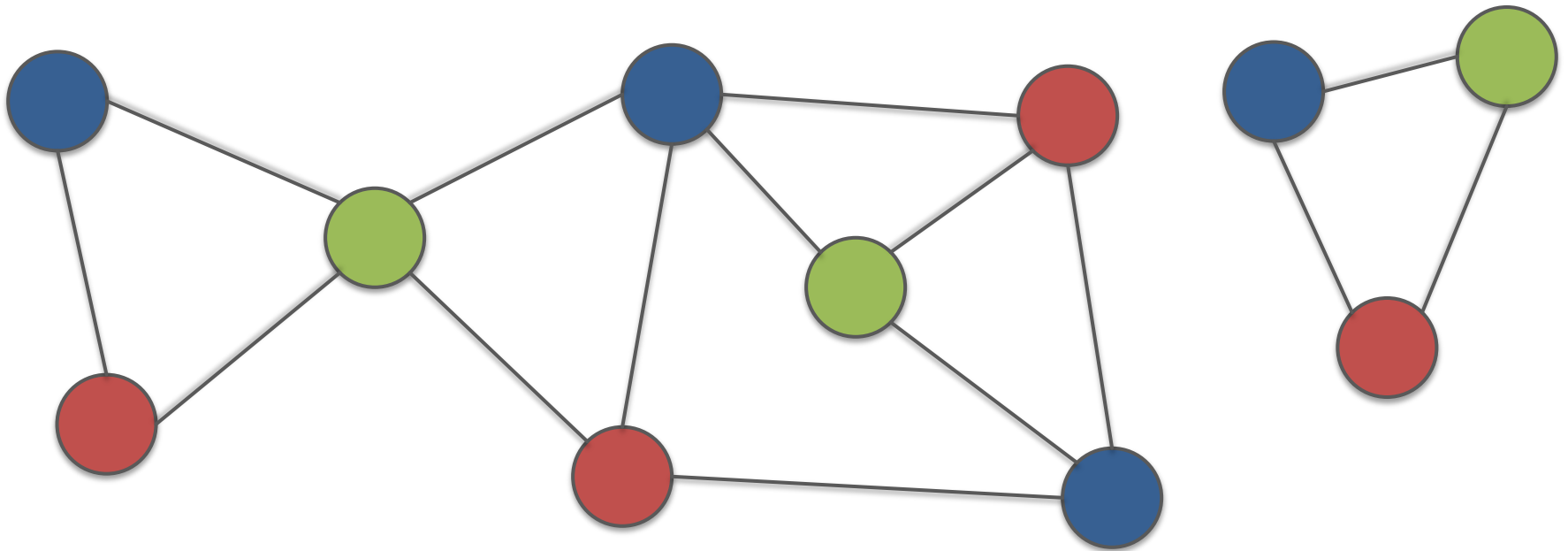


# Algorithmen - Partitionierung

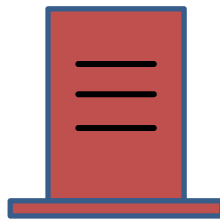
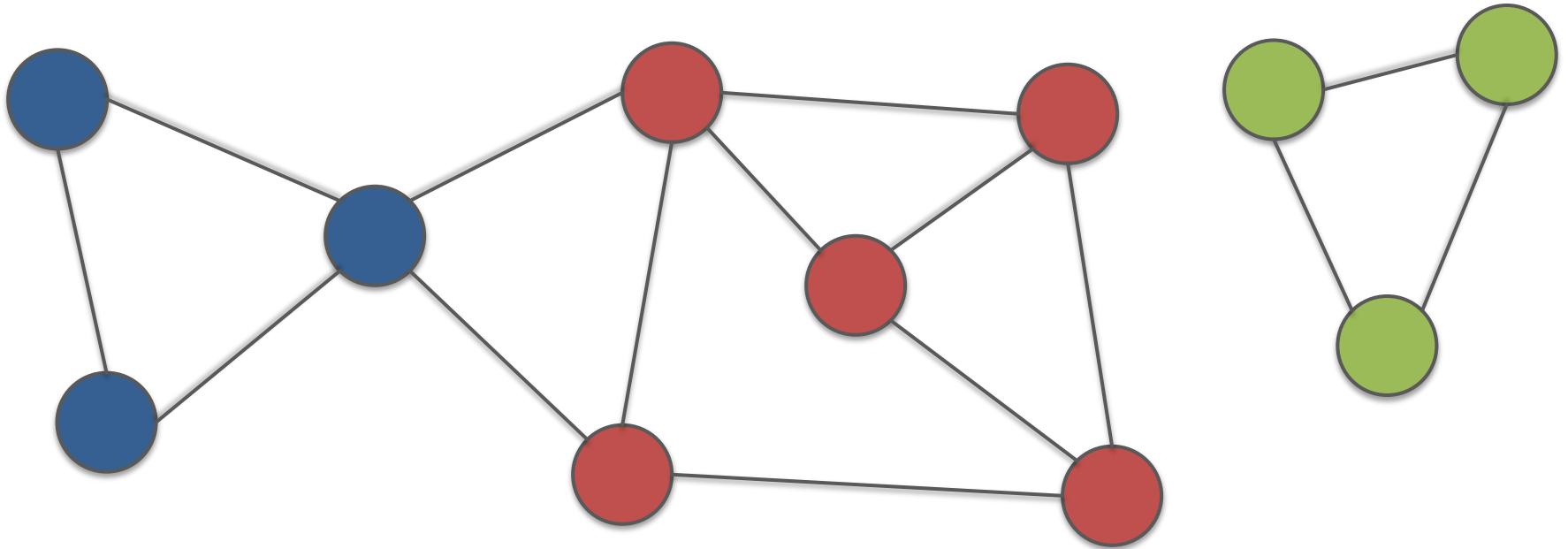




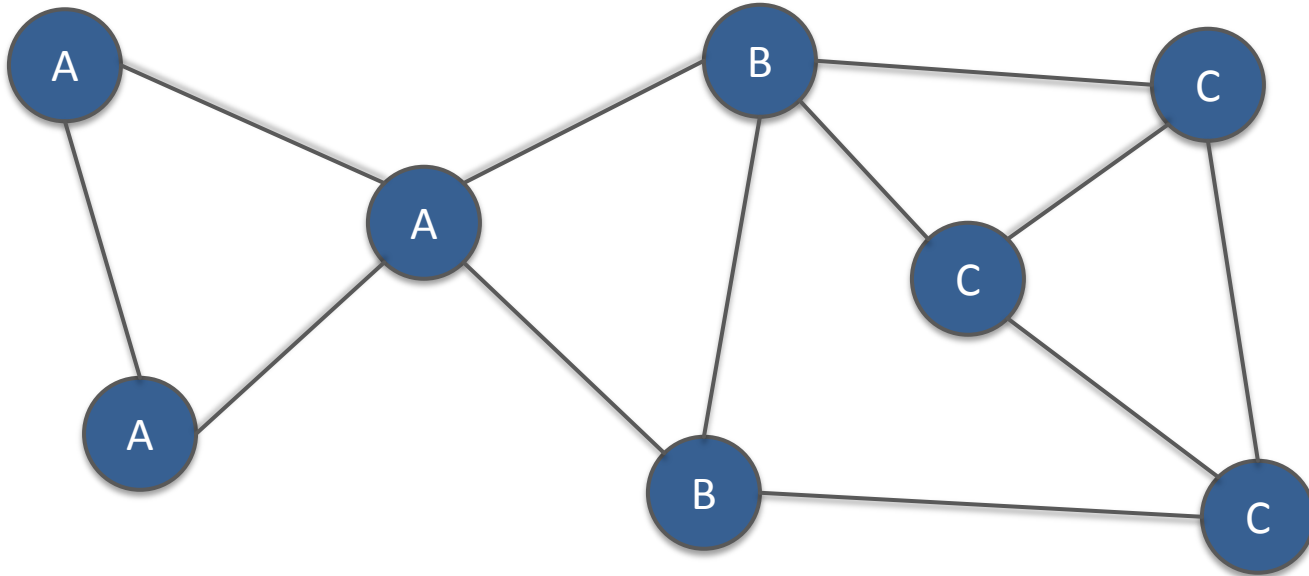
# Algorithmen - Partitionierung



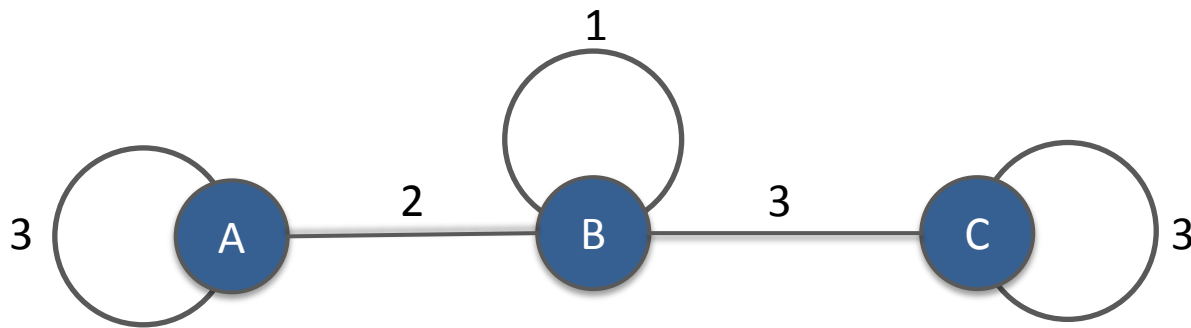
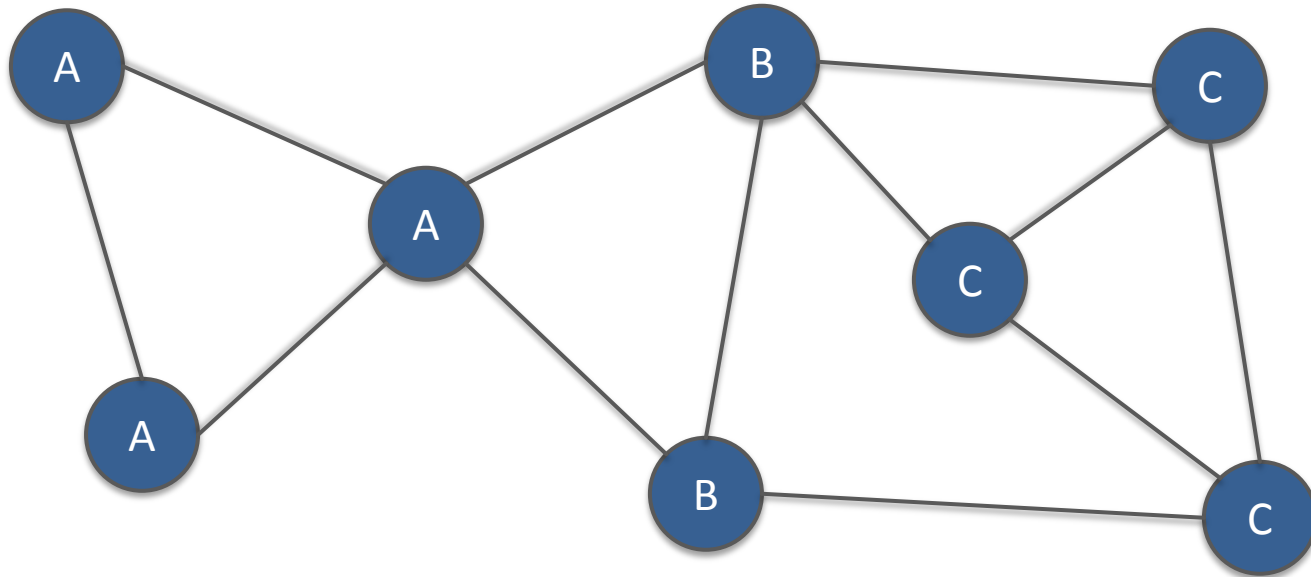
# Algorithmen - Partitionierung



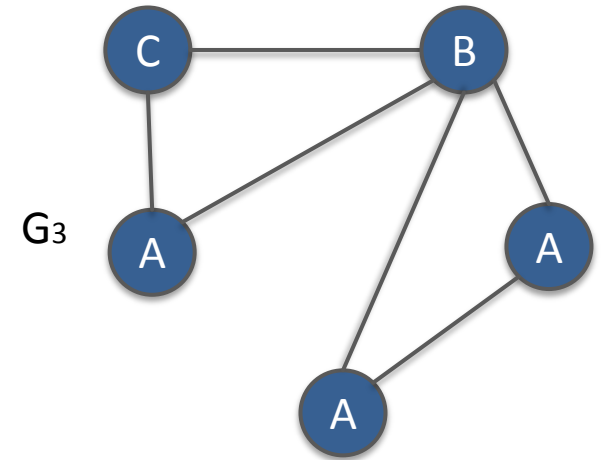
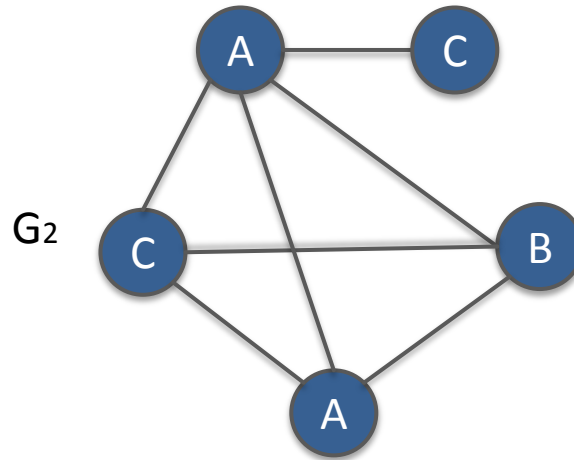
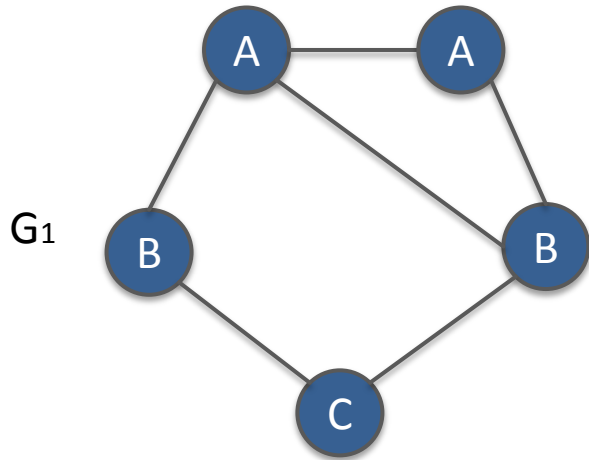
# Algorithmen – Graph OLAP



# Algorithmen – Graph OLAP

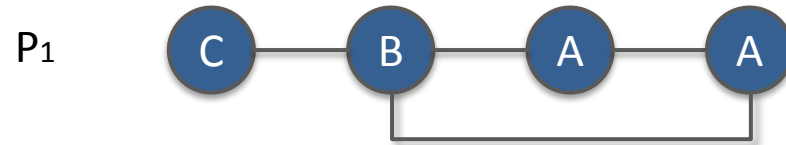
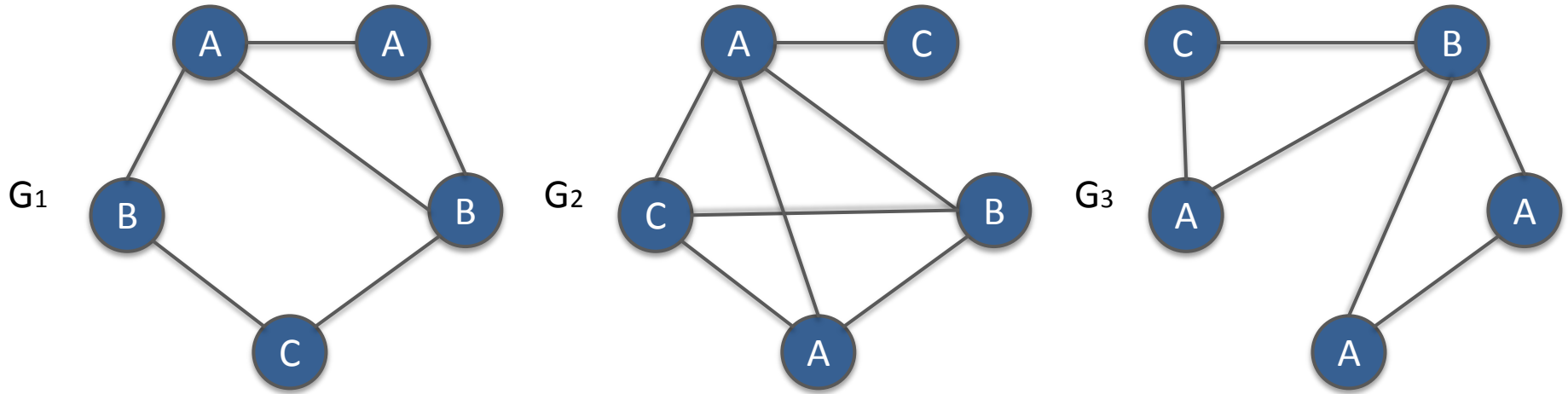


# Algorithmen – Graph Pattern Mining





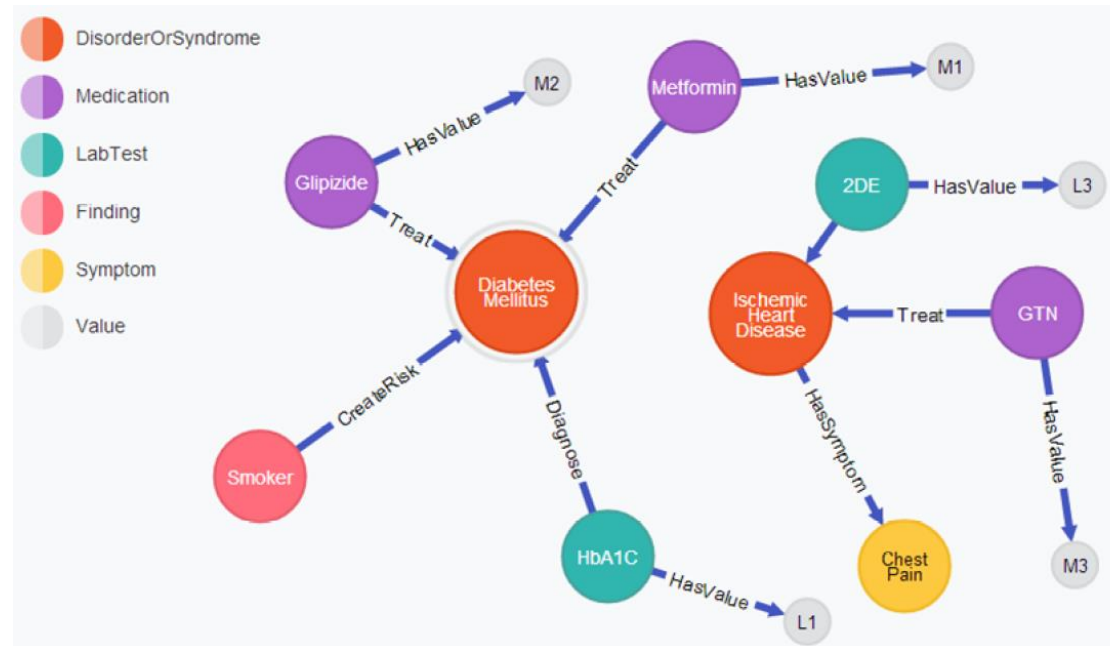
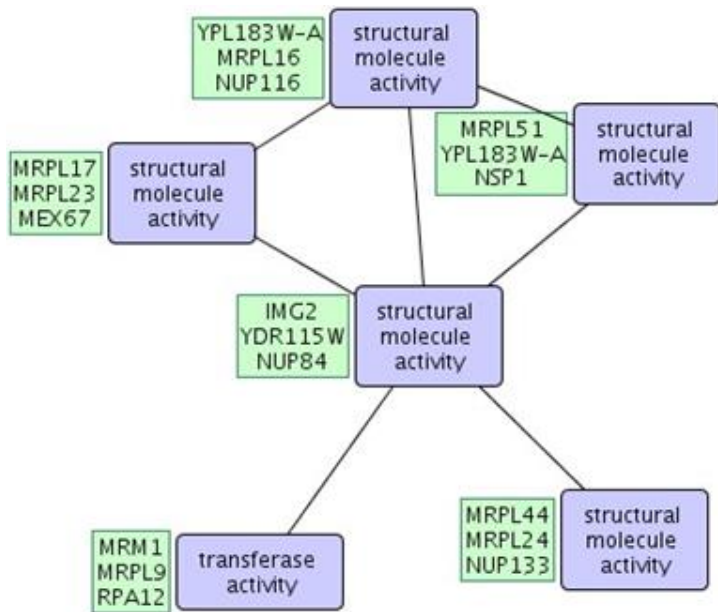
# Algorithmen – Graph Pattern Mining



support  $\geq 3$



# Graph Applications



# Seminar

Allgemeine Informationen

# Seminarziele

- Beschäftigung mit einem praxis- und wissenschaftlich relevanten Thema
  - Kann Grundlage für Abschlussarbeit oder SHK-Tätigkeit sein
- Erarbeitung und Durchführung eines **Vortrags** unter Verwendung wissenschaftlicher (englischer) Literatur
- Diskussion des vorgestellten Themas
- Schriftliche **Ausarbeitung** zum Thema
- Hilfe und Feedback durch zugeteilten Betreuer

# Anrechnungsmöglichkeiten

- Masterstudium
  - Teil der Module Moderne Datenbanktechnologien
  - Seminar modul (oder Masterseminar)
- Bachelorstudium
  - Seminar modul (oder Bachelorseminar)
- Diplom
  - Problemseminar



# Modulprüfung

- Selbständiger **Vortrag** (ca. 30 Minuten) mit **Diskussion** (ca. 15 Minuten)
  - Abnahme der Folie durch Betreuer
  - Abgabe spätestens 1 Woche vor Vortragstermin
- Schriftliche **Ausarbeitung** (ca. 15 – 20 Seiten)
  - Abnahme der Ausarbeitung durch Betreuer
  - Ausarbeitung spätestens zum Vortragstermin
- Aktive Teilnahme an **allen** Vortragsterminen
- Modul-Workload
  - 30h Präsenzzeit
  - 120h Selbststudium

# Termine

- Themenzuordnung
  - Koordinierungstreffen mit Betreuer bis **spätestens** 07.11.2014
  - Ansonsten verfällt Seminaranmeldung
  - Freiwilliger Rücktritt auch bis max. 07.11.2014
- Vortragstermine
  - Montags, 9:15 – 12:00 Uhr, Raum A314
  - Vorträge ab 5.1.2015 (4 – 5 Termine)

# Themenvergabe

#	Thema	#	Thema
1	Google Pregel (+BSP)	11	Adaptive Partitioning of Large-Scale Dynamic Graphs
2	Apache Giraph	12	How to Partition a Billion-Node Graphs
3	GraphLab PowerGraph	13	Discovery-Driven Graph Summarization
4	GraphX	14	Pagrol: Parallel Graph OLAP over Large-scale Attributed Graphs
5	Apache Flink	15	SynopSys: Foundations for Multidimensional Graph Analytics
6	Mizan	16	Large-Scale Frequent Subgraph Mining in MapReduce
7	Optimizing Graph Algorithms on Pregel-like Systems	17	GRAMI: Frequent Subgraph and Pattern Mining in a Single Large Graph
8	From „Think Like a Vertex“ to „Think like a Graph“	18	Graph databases for healthcare analytic systems
9	Streaming graph partitioning for large distributed graphs	19	Mining biological networks for unknown pathways
10	Balanced label propagation for partitioning massive graphs	20	Unicorn: A System for Searching the Social Graph