Seminar Graph Clustering

Seminar 1

15.04.2009

Eckdaten

- Abteilung: Betriebliche Informationssysteme
- Modul: Semantic Web
- Leiter: Axel Ngonga, Frank Schumacher

Ziele

- Überblick über Graph Clustering gewinnen
- Grundgedanken der verschiedenen Algorithmenklassen verstehen
- Metriken zur Evaluation von Clustering

Gliederung

- Organisatorisches
- Kurze Einführung
- Themen
- Lernen

Organisatorisches

- Umfang: 2 SWS
- Ort und Zeit: Mi. 15:15 – 16:45 (Jo 1–22)
- Ablauf
  - Vortrag der Teilnehmer zu dem gewählten Thema (30–45 min)
  - Diskussion (10–15 min)
  - Gruppenarbeit (20–30 min)

Organisatorisches

- Voraussetzungen
  - Grundkenntnisse in Graphtheorie

- Voraussetzungen für Scheinerhalt
  - Teilnahme an Präsenzveranstaltungen
  - Vortrag
  - Hausarbeit (Problemseminarschein)
  - Implementierung (Problemseminarschein)

Organisatorisches

- **Master neu**: benoteter Schein, fließt in Modulnote ein
- **Diplom**: benoteter oder unbenoteter Schein, Problemseminarschein
- **Master alt**, **Bachelor neu** und **alt**: benoteter oder unbenoteter Schein

Organisatorisches

- Verbindliche Anmeldung zum Seminar per Mail bis 21.04. um 10Uhr

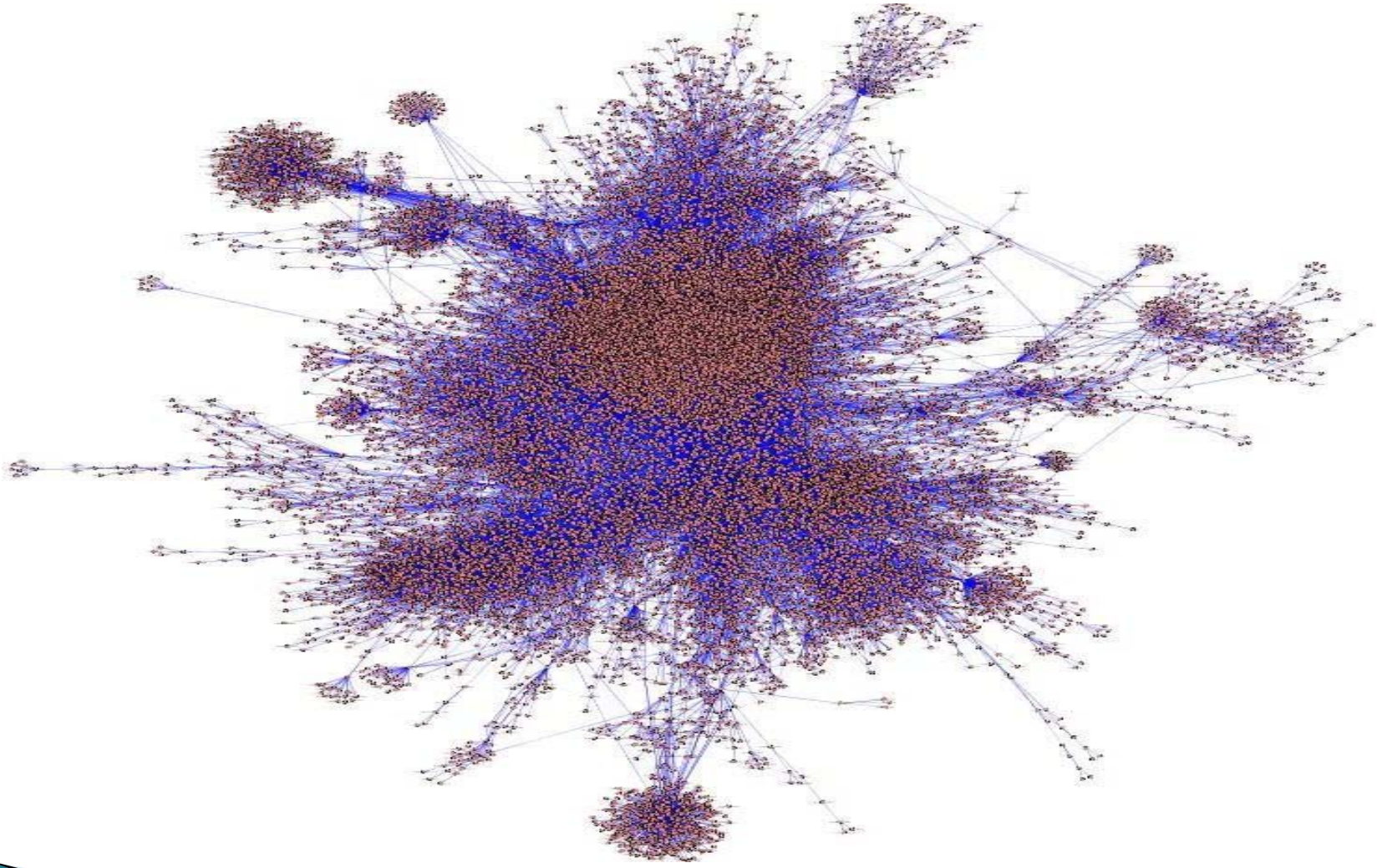
  ngonga@informatik.uni-leipzig.de
  - Name
  - Matrikelnummer
  - Semester
  - Studiengang
  - Gewünsches Thema

Fragen?

Kurze Einführung

- Grundideen
- Globale Algorithmen
- Lokale Algorithmen
- Evaluation von Clustering
- Themen

Grundideen



Grundideen

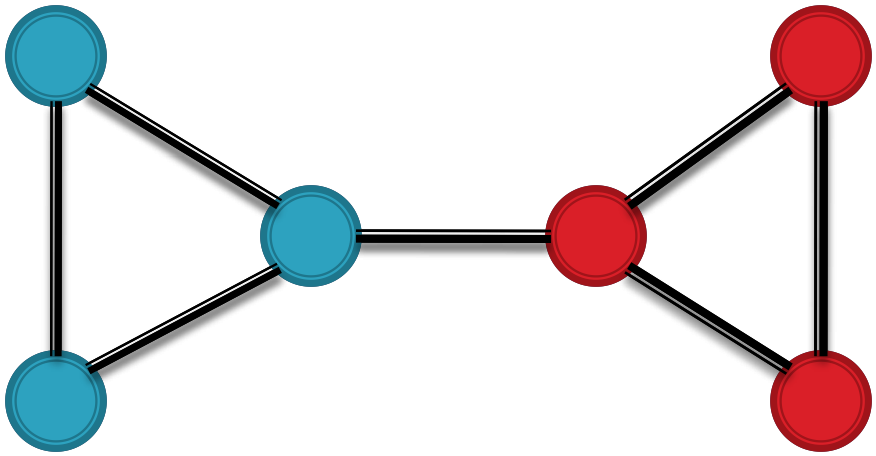
- Auffinden von Gruppen ähnlicher Knoten
- Cluster definiert durch
  - Hohe innere Kantendichte
  - Niedrige Kantendichte nach außen
- 2 Arten von Algorithmen
  - Globale Algorithmen
  - Lokale Algorithmen

Globale Algorithmen

- Gehen von gesamter Graphtopologie aus
- Globales Wissen über Graphtopologie vs. Speicherverbrauch
- Beispiele
  - Minimum-Cut Algorithmen
  - Spectral Clustering (SC)
  - Markov Clustering (MCL)

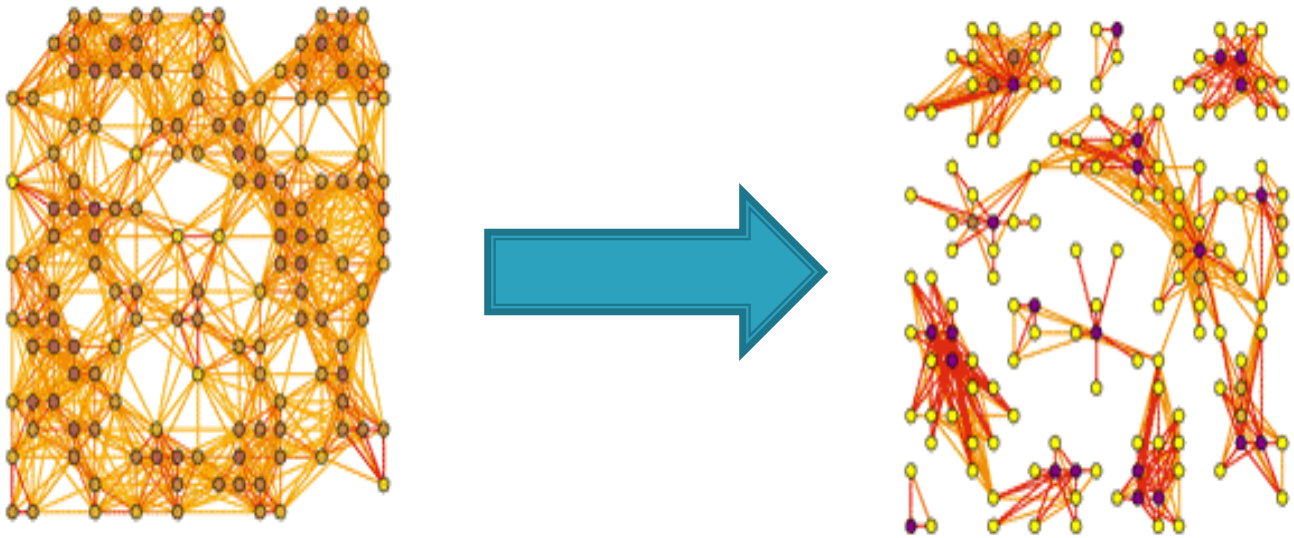
Minimum-Cut Algorithmen

- **Idee**
  Finden einer 2-Partition der Knotenmenge, die die Summe der Gewichte der Kanten zwischen den Partitionen minimiert



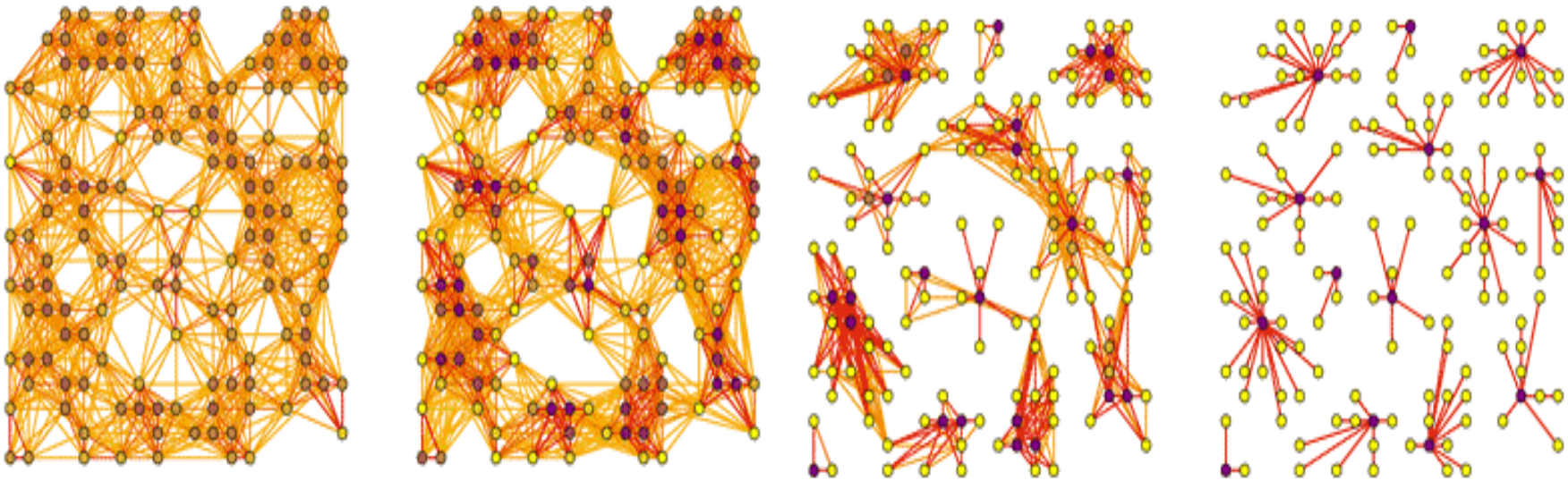
Spectral clustering

- **Idee**
  Verwenden Eigenvektor-Transformation zum Mapping des Graphs auf weniger Features



MCL

- **Idee**
  Zufällige Läufe im Graph tendieren dazu, im gleichen Cluster zu bleiben

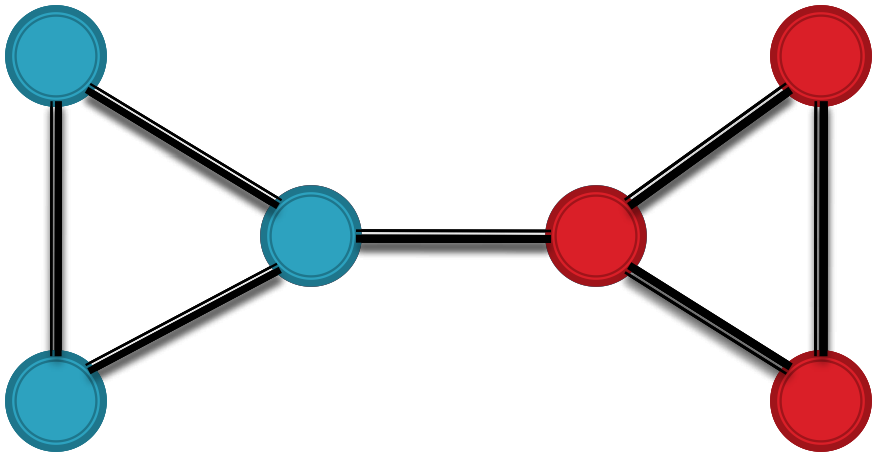


Lokale Algorithmen

- Verwenden vorgegebenem Clustering um Gütekriterium zu maximieren
- Speicherverbrauch vs. lokales Wissen über Graphtopologie
- Beispiele
  - Randomized Neighborhood Search Clustering (RNCS)
  - Chinese Whispers (CW)
  - BorderFlow (BF)

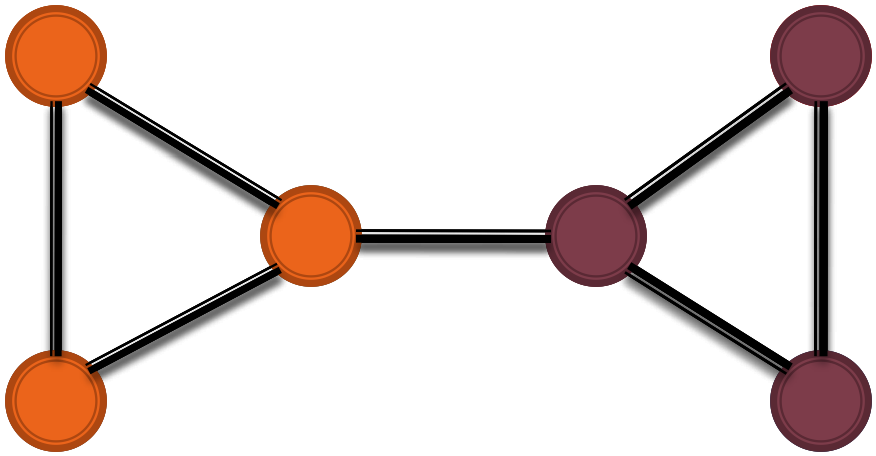
RNSC

- **Idee**
  Verteilt Knoten zufällig um, bis maximale Güte erreicht ist



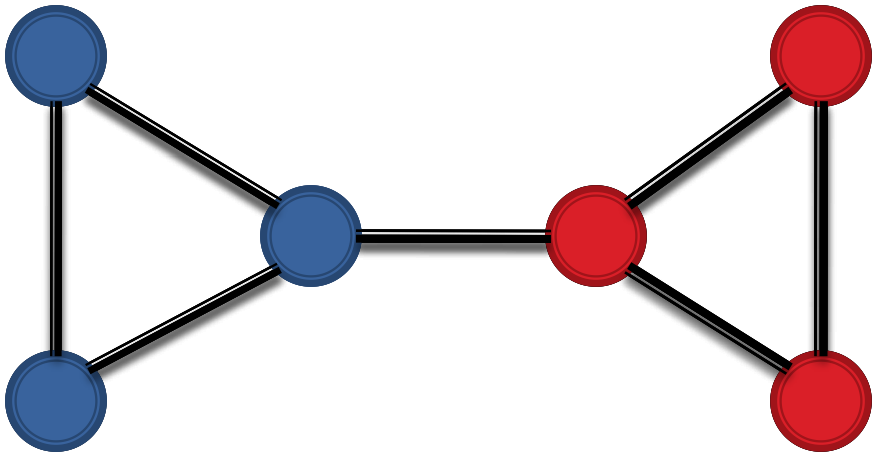
CW

- **Idee**
  Modifikiation der initialen Färbung des Graphs durch zufälliges Ausbreiten



BF

- **Idee**
  Seed-basierte Maximierung der relativen Kantendichte von Clustern



Evaluation

- Quantifizierung der Güte eines Clusterings
- Lokale Metriken
  - Compacteness index
  - Silhouette index
- Globale Metriken
  - Dunn index

Themen

1. Globales Clustering
   1. Minimal Cut
   2. Spectral Clustering
   3. MCL
   4. MCODE
   5. SPC
2. Lokales Clustering
   1. RNSC
   2. CW
   3. BorderFlow
3. Evaluation

Fragen?

Lernen