Fachprojekt
Algorithm Engineering

Vorbesprechung

Carsten Gutwenger

LS 11 Algorithm Engineering

SoSe 2011 • 2. Februar 2011

technische universität dortmund

department of computer science

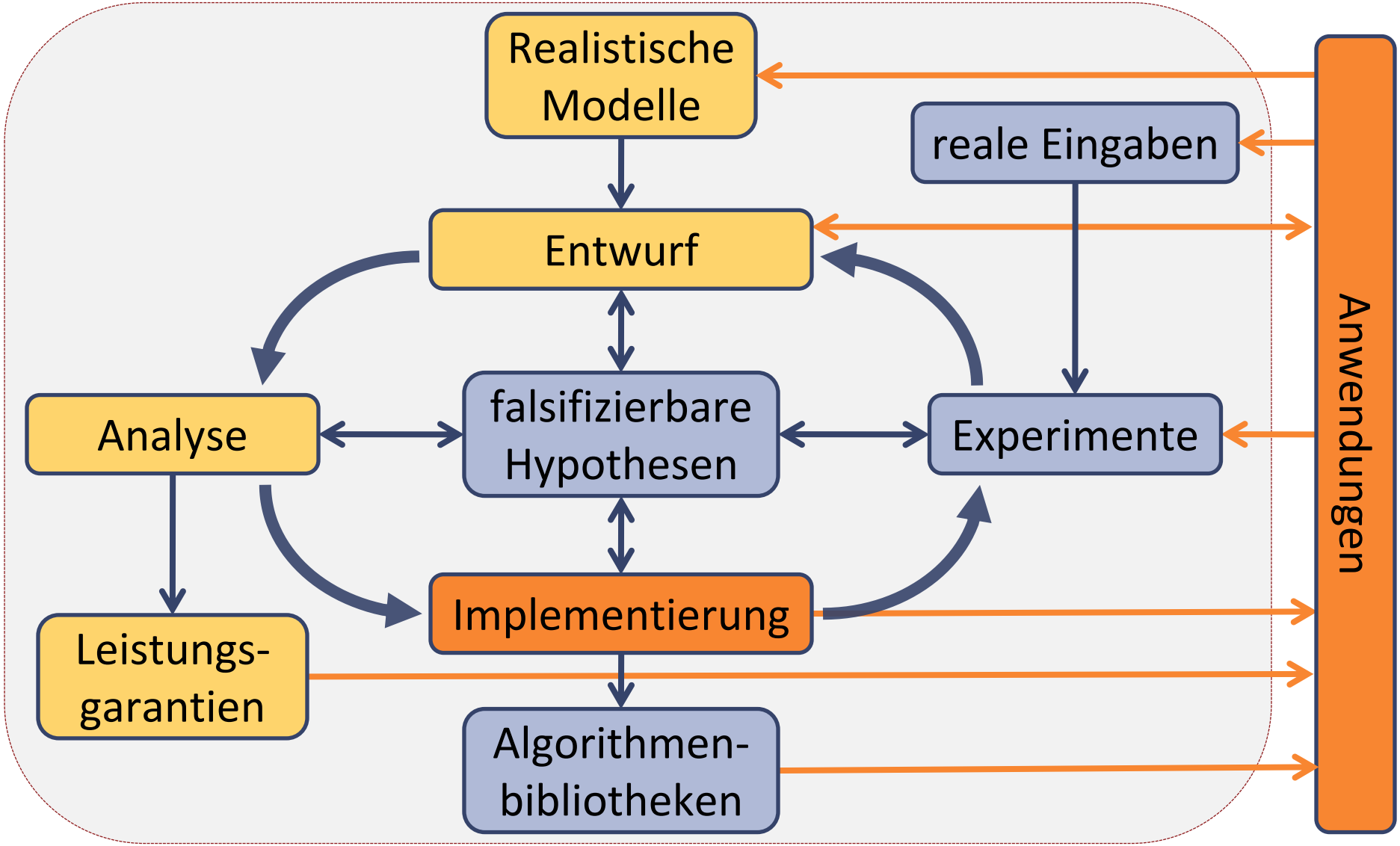
Agenda

- Einführung
- Terminplanung und Ablauf
- Formales
- Vorstellung und Vergabe der Themen

Einführung

- Algorithm Engineering
  - Design von Algorithmen
  - theoretische Analyse
  - Implementierung
  - experimentelle Evaluierung
- Praxisorientiert!

Algorithm-Engineering-Kreislauf



Termine und Ablauf

- Vorbesprechung und Themenvergabe (heute)
  - pro Thema eine Gruppe à 2-4 Teilnehmer
- Ein Betreuer pro Thema/Gruppe
  - genaue Aufgabenstellung
  - Zusammenarbeit in der Gruppe
- Gemeinsame Treffen (ca. 3 im Semester)
  - erstes Treffen: 13. April
  - jeweils Mittwoch, 10:15-12 Uhr, OH14 R202
- Regelmäßige Treffen mit Betreuer (wöchentlich)
- Selbständige Arbeit!

Formales

- Anmeldung über BOSS (für Bachelor-Studenten)
  - Prüfungsnummer 66991
  - bei Fragen: Frau Schiller (Prüfungsamt)

Themen

(1) Implementierung und Evaluierung von Datenstrukturen für Priority-Queues

(2) Implementierung von Varianten für die Union-Find Datenstruktur und experimentelle Evaluation dieser Strategien für ausgewählte Graphalgorithmen

(3) Neue Schichtungsverfahren für Sugiyama-Zeichnungen

(4) Integration von Nebenbedingungen in Multilevel-Verfahren für das Zeichnen von Graphen

(5) Indexstrukturen für Fingerprints

Thema (1)

Priority-Queues

„Implementierung und Evaluierung von Datenstrukturen für Priority-Queues“

Betreuer: **Dr. Carsten Gutwenger**

Priority-Queues

- Bekannt aus z.B. DAP2
- Implementierungen
  - binäre Heaps
  - Fibonacci-Heaps
  - …
- Anwendungen
  - Kürzeste Wege (Dijkstra-Algorithmus)
  - Minimum-Cut (Nagomochi-Ibaraki-Algorithmus)
  - MST (Prim)
  - Sortieren, Huffman-Coding
- DIMACS Implementierungs-Challenge (1995-1997)

Aufgaben

- Implementierung verschiedener Datenstrukturen für PQs
  - Recherche(!), Auswahl
  - Programmiersprache: **C++** oder (zur Not) Java
- Entwurf einer Testumgebung
  - Benchmarks aus Anwendungen
- Evaluierung der Datenstrukturen
  - Sind die Implementierungen kompetitiv?
  - Welche sind besser? Warum? Werden DSen dominiert?
  - Bei welchen Anwendungsszenarien?
  - Kann man Implementierungen / DSen verbessern?

Thema (2)

Union-Find-Algorithmen

„Implementierung von Varianten für die Union-Find Datenstruktur und experimentelle Evaluation dieser Strategien für ausgewählte Graphalgorithmen“

Betreuer: **Bernd Zey**

MST, Union-Find

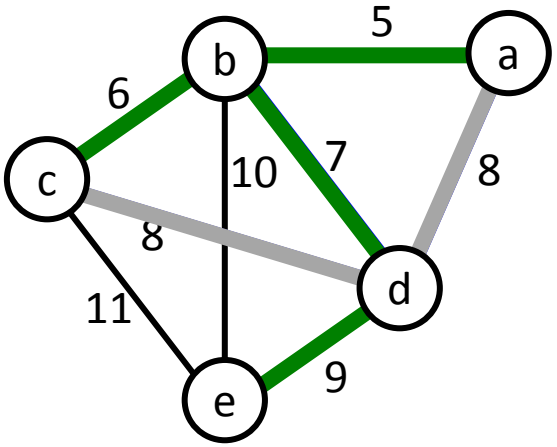
Minimum Spanning Tree (MST)

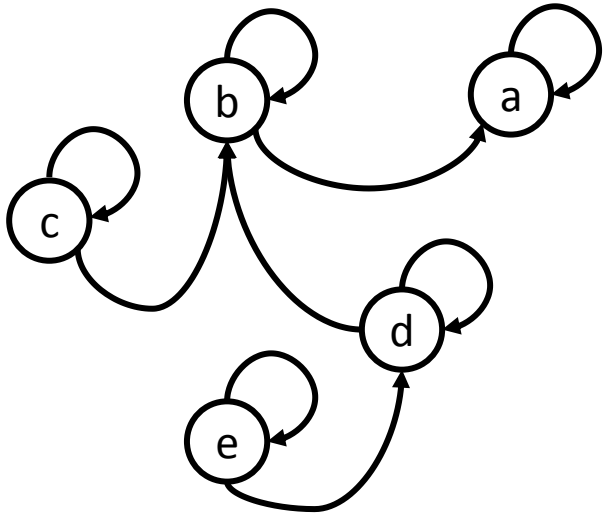
Gegeben: ungerichteter Graph G=(V,E), Kosten c(e)

Gesucht: minimaler Spannbaum G‘=(V,E‘) so dass E‘ alle Knoten mit minimalen Kosten verbindet

(Hoffentlich) bekannte Algorithmen (DAP2):

- Prim (siehe Thema *Priority-Queue*)
- Kruskal





MST, Union-Find

Minimum Spanning Tree (MST)

Gegeben: ungerichteter Graph G=(V,E), Kosten c(e)

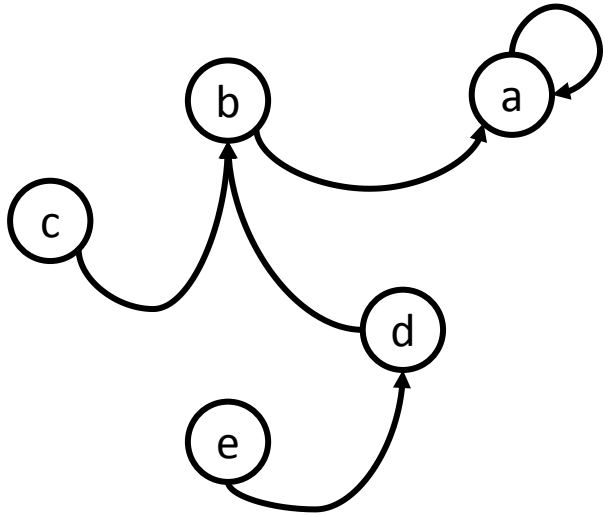
Gesucht: minimaler Spannbaum G‘=(V,E‘) so dass E‘ alle Knoten mit minimalen Kosten verbindet

(Hoffentlich) bekannte Algorithmen (DAP2):

- Prim (siehe Thema *Priority-Queue*)
- Kruskal

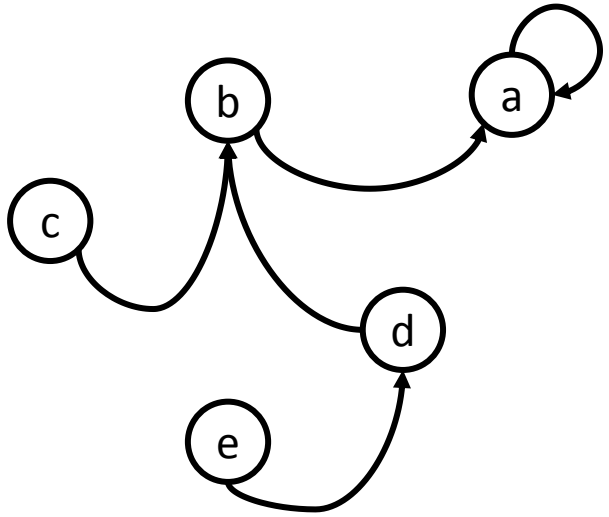
⟶ **Union-Find-Datenstruktur**

- Union-Regeln
- Kompressions-Regeln
- „Implementierungs-Tricks“



Aufgaben

- Implementierung unterschiedlicher Union-Find-Strategien
  - Literatur-Recherche
  - Auswahl der Varianten/Kombinationen
    - *(Eigene Ideen?!)*
  - C++
  - Anbindung an das **OGDF** (**O**pen **G**raph **D**rawing **F**ramework)
- Experimenteller Vergleich
  - Auswahl eines geeigneten Instanz-Sets
  - Auswahl der Rahmenalgorithmen
    - Kruskal, ZSHK verwalten, *(sonstiges?)*
  - Vergleich der implementierten Varianten
- Präsentation der Ergebnisse
- Endbericht



Thema (3)

Schichtungsverfahren

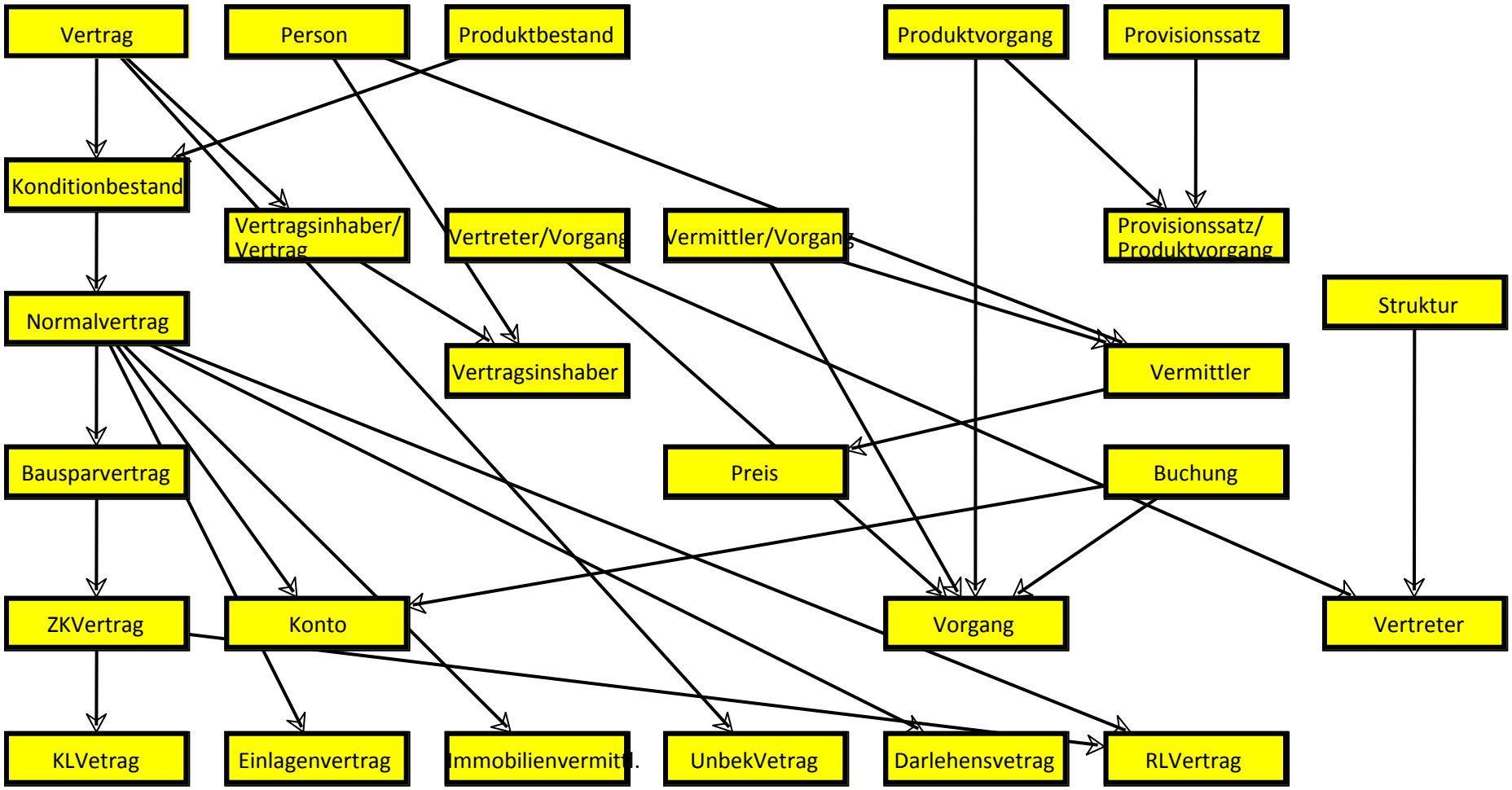
„Neue Schichtungsverfahren für Sugiyama-Zeichnungen“

Betreuer: **Prof. Dr. Petra Mutzel, Hoi-Ming Wong**

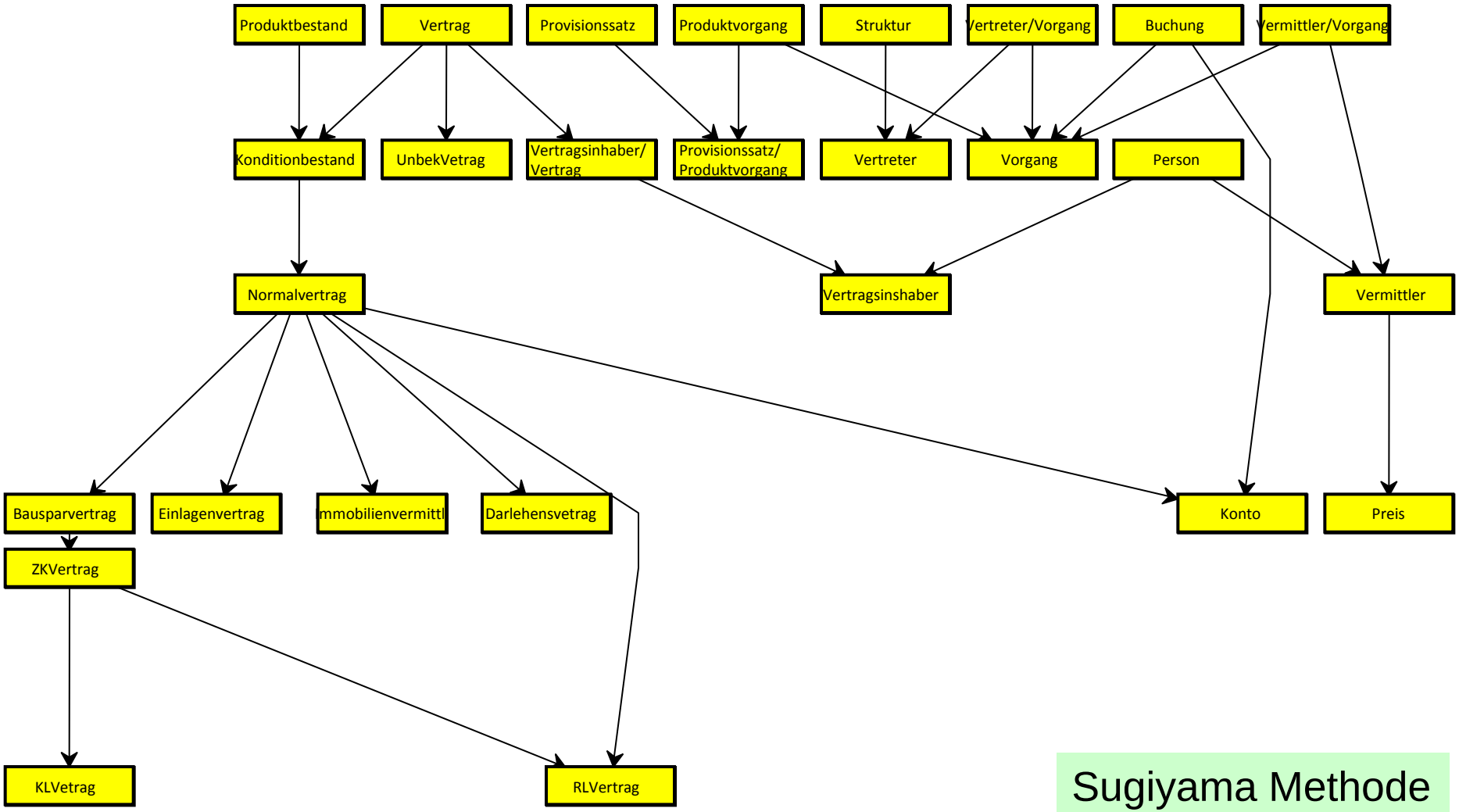
Neue Schichtungsverfahren für Sugiyama-Zeichnungen

- Zeichnen von Graphen
  - Motivation
  - Ästhetikkriterien
- Layoutverfahren von Sugiyama
- Ziel und Aufgabenstellung

Datenbank-Modell: Original



Datenbank-Modell: Autom. Layout



Sugiyama Methode

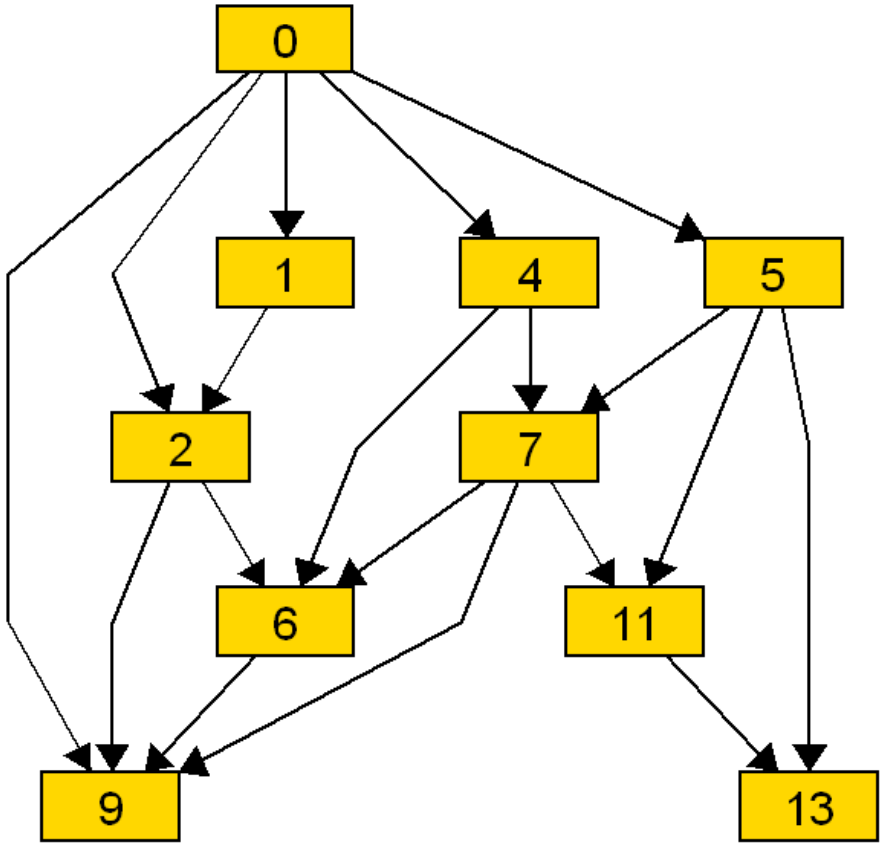
Wichtige Aesthetikkriterien

- Wenige Kreuzungen
- Wenige Kantenknicke
- Kurze Kantensegmente
- Kleine Zeichenfläche

- speziell: Hierarchie-Erhaltung, Symmetrien

Hierarchische Methoden

- Zahlreiche Anwendungen, z.B.
  - gerichtete azyklische Graphen
  - Datenflussdiagramme
  - Workflow Diagramme
  - Organigramme
  - UML Sequenzen Diagramme
  - UML Aktivitäts Diagramme
  - Metabolische Pathways in Bioinformatik
  - ...



■ Mysterien und Leberwurstbrote oder: Die wirrsten Grafiken der Welt (8)

Die Verflechtung der Stromwirtschaft (nach Michael Stelte)

Wie stark verflochten die Stromwirtschaft in Deutschland ist, zeigt eine bildschöne Grafik aus Michael Steltes Berliner Energiedatenbank: Die Hamburgische Electricitäts-Werke AG ist beispielsweise mit 15,7 Prozent an der Sydkraft AB Malmö beteiligt, diese mit einem Prozent an der Veba AG, diese mit 100 Prozent an der PreussenElektra AG, diese mit 17,6 Prozent an der Sydkraft AB Malmö und diese wiederum mit 21,8 Prozent an der Hamburgischen Electricitäts-Werke AG. Ob es wohl Steuerprüfer gibt, die es wagen, sich solche komplizierten Besitzverhältnisse einzuprägen und mit Taschenrechner, Aspirin und Leberwurstbroten bewaffnet in die Mysterien der Stromwirtschafts-Buchhaltung einzudringen? Oder überlässt das Finanzamt diese Arbeit lieber Sisyphus? Um sich stattdessen wehrlosen Kleinsparern zuzuwenden?

Das wäre verständlich, denn die Verflechtung der Stromwirtschaft ist nahezu unentwirrbar, auch wenn die Grafik auf den ersten Blick sehr übersichtlich wirkt. Einmal angenommen, die Veag erwirtschaftet zehn Mark Gewinn. Dann gehören 25 Prozent davon der EBH GmbH (= 2,50 DM), davon 25 Prozent der VEW Energie AG (= 0,625 DM), davon 100 Prozent der VEW AG (= 0,625 DM) und davon 12 Prozent der Bayernwerk AG (= 0,075 DM). Aber 25,3 Prozent von den 0,625 DM gehören der Energie-Verwaltungs-GmbH (= 0,158125 DM), davon 30 Prozent der Contigas AG (= 0,0474375 DM) und davon 91,1 Prozent abermals der Bayernwerk AG (= 0,0432155625 DM). Damit hat die Bayernwerk AG ihren Gewinn also bereits auf satte 0,1182155625 DM aufgestockt. Aber die Bayernwerk ist ja zudem auch noch mit weiteren 22,5 Prozent direkt an der Veag beteiligt, so dass von den zehn Mark noch einmal 2,25 DM abfallen und sich somit eine Gewinnsumme von 2,3682155625 DM ergibt.

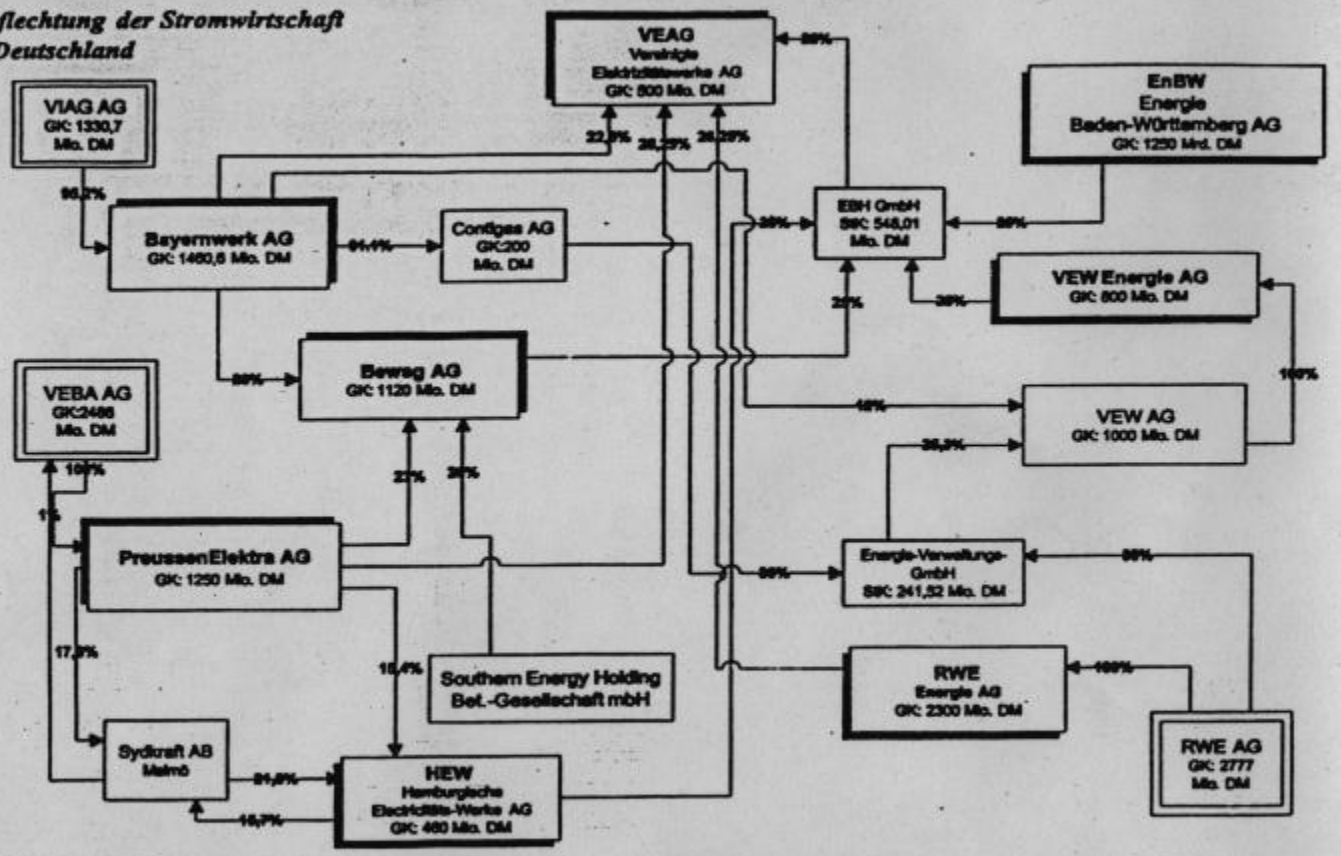
Aufgaben:

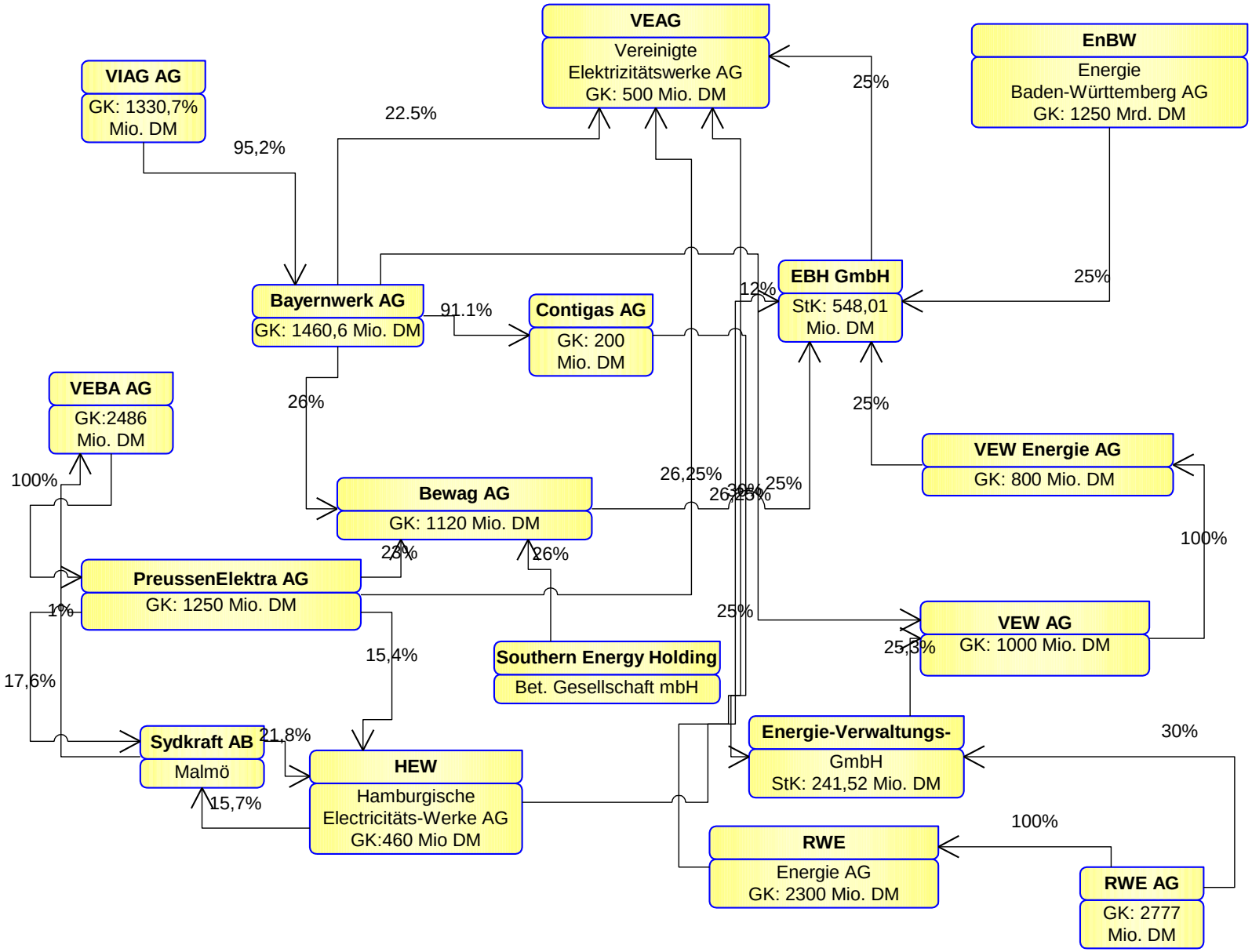
1) Rechne aus, wieviel noch dazukommt, da die Bayernwerk AG ja auch über ihre Beteiligung an der Bewag AG an der EBH GmbH beteiligt ist!

2) Wieviel schuldet die Bayernwerk AG letztlich der Viag AG?

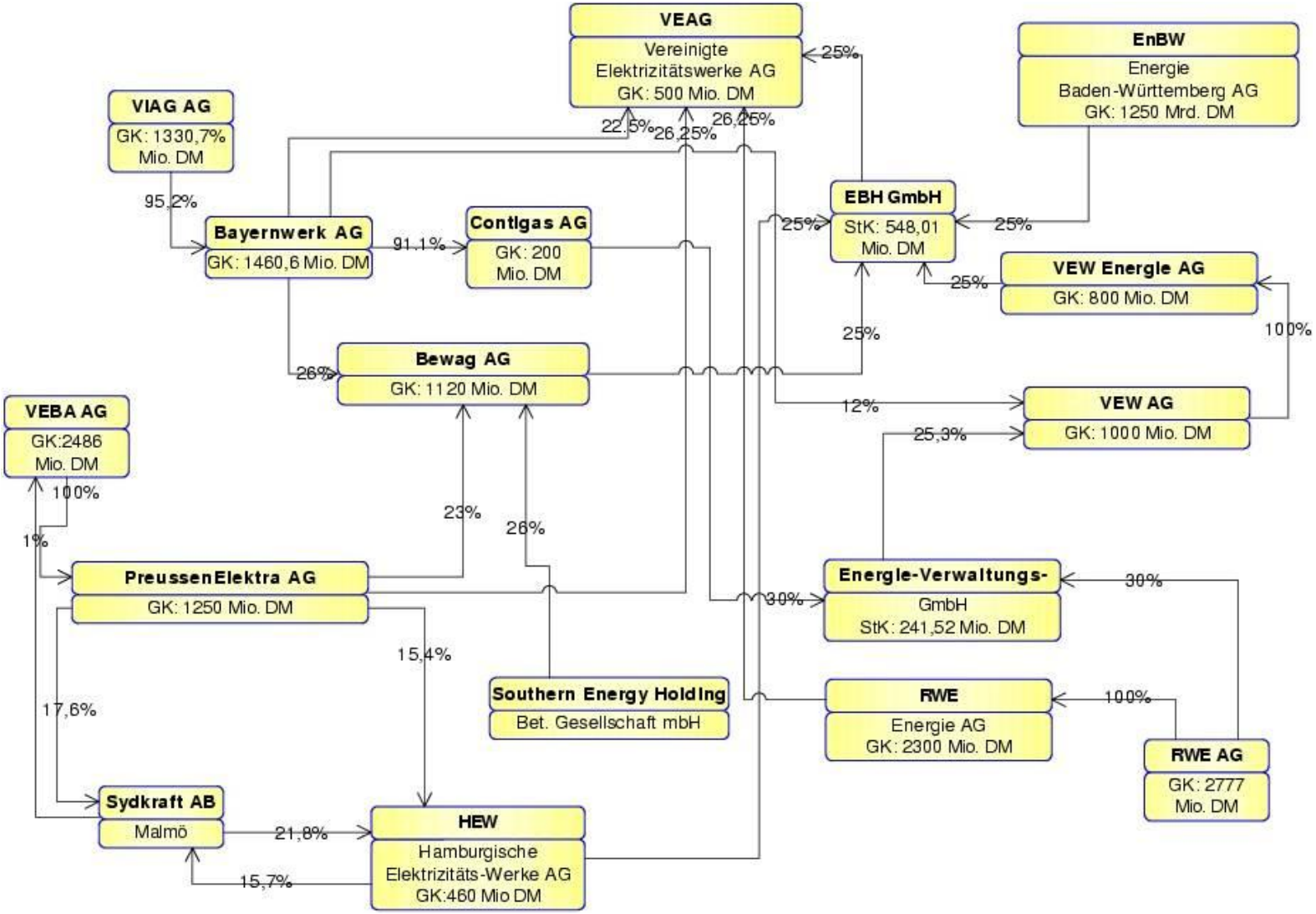
3) Ruf bei der Bayernwerk AG in München an (Durchwahl Hauptverwaltung [089] 12541) und erkundige dich danach, ob deine Rechnung stimmt!

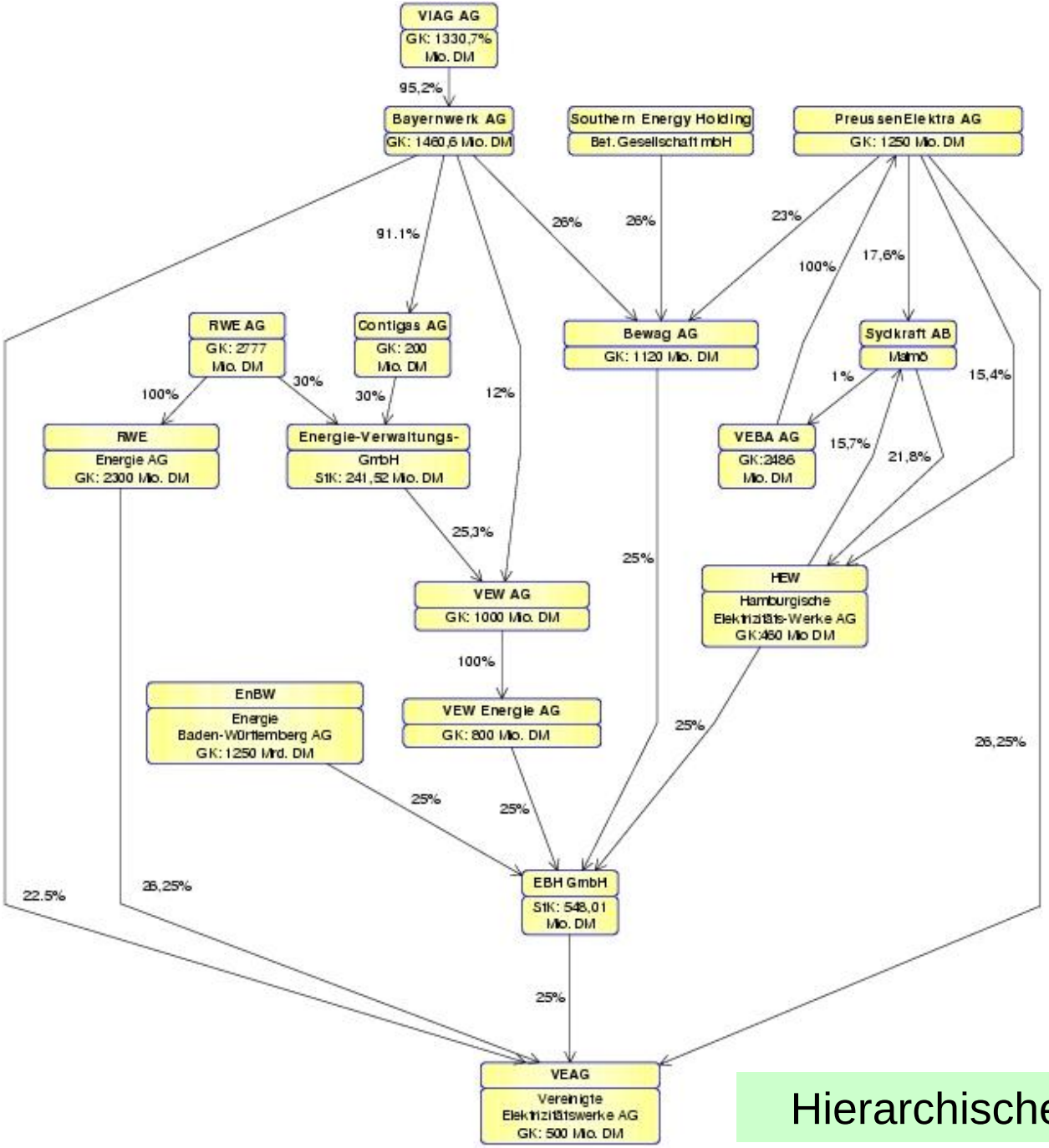
Gerhard Henschel




VEAG
Vereinigte Elektrizitätswerke AG
GK: 500 Mio. DM
EnBW
Energie Baden-Württemberg AG
GK: 1250 Mrd. DM
VIAG AG
GK: 1330,7% Mio. DM
95,2%
22.5%
25%
Bayernwerk AG
GK: 1460,6 Mio. DM
91.1%
Contigas AG
GK: 200 Mio. DM
12%
EBH GmbH
StK: 548,01 Mio. DM
25%
VEBA AG
GK:2486 Mio. DM
26%
25%
100%
26,25%
Bewag AG
GK: 1120 Mio. DM
VEW Energie AG
GK: 800 Mio. DM
100%
PreussenElektra AG
GK: 1250 Mio. DM
1%
26%
25%
VEW AG
GK: 1000 Mio. DM
25,3%
15,4%
17,6%
Southern Energy Holding
Bet. Gesellschaft mbH
Sydkraft AB
Malmö
21,8%
HEW
Hamburgische Electricitäts-Werke AG
GK:460 Mio DM
15,7%
Energie-Verwaltungs-GmbH
StK: 241,52 Mio. DM
30%
RWE
Energie AG
GK: 2300 Mio. DM
100%
RWE AG
GK: 2777 Mio. DM

Sugiyama Verfahren

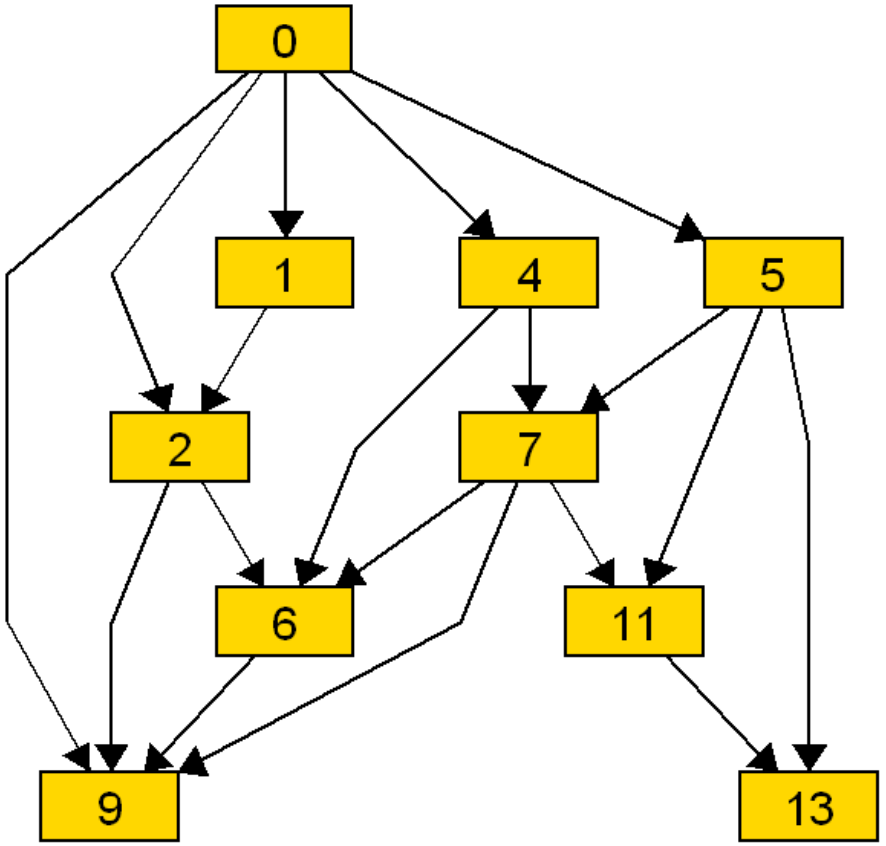

VEAG
Vereinigte Elektrizitätswerke AG
GK: 500 Mio. DM
EnBW
Energie Baden-Württemberg AG
GK: 1250 Mrd. DM
VIAG AG
GK: 1330,7% Mio. DM
95,2%
Bayernwerk AG
GK: 1460,6 Mio. DM
91,1%
Contigas AG
GK: 200 Mio. DM
22,5%
26,25%
26,25%
25%
EBH GmbH
StK: 548,01 Mio. DM
25%
25%
VEW Energie AG
GK: 800 Mio. DM
25%
100%
26%
Bewag AG
GK: 1120 Mio. DM
25%
12%
VEW AG
GK: 1000 Mio. DM
25,3%
VEBA AG
GK:2486 Mio. DM
100%
1%
23%
26%
PreussenElektra AG
GK: 1250 Mio. DM
30%
Energie-Verwaltungs-GmbH
StK: 241,52 Mio. DM
30%
15,4%
17,6%
Southern Energy Holding
Bet. Gesellschaft mbH
RWE
Energie AG
GK: 2300 Mio. DM
100%
RWE AG
GK: 2777 Mio. DM
Sydkraft AB
Malmö
21,8%
HEW
Hamburgische Elektrizitäts-Werke AG
GK:460 Mio DM
15,7%



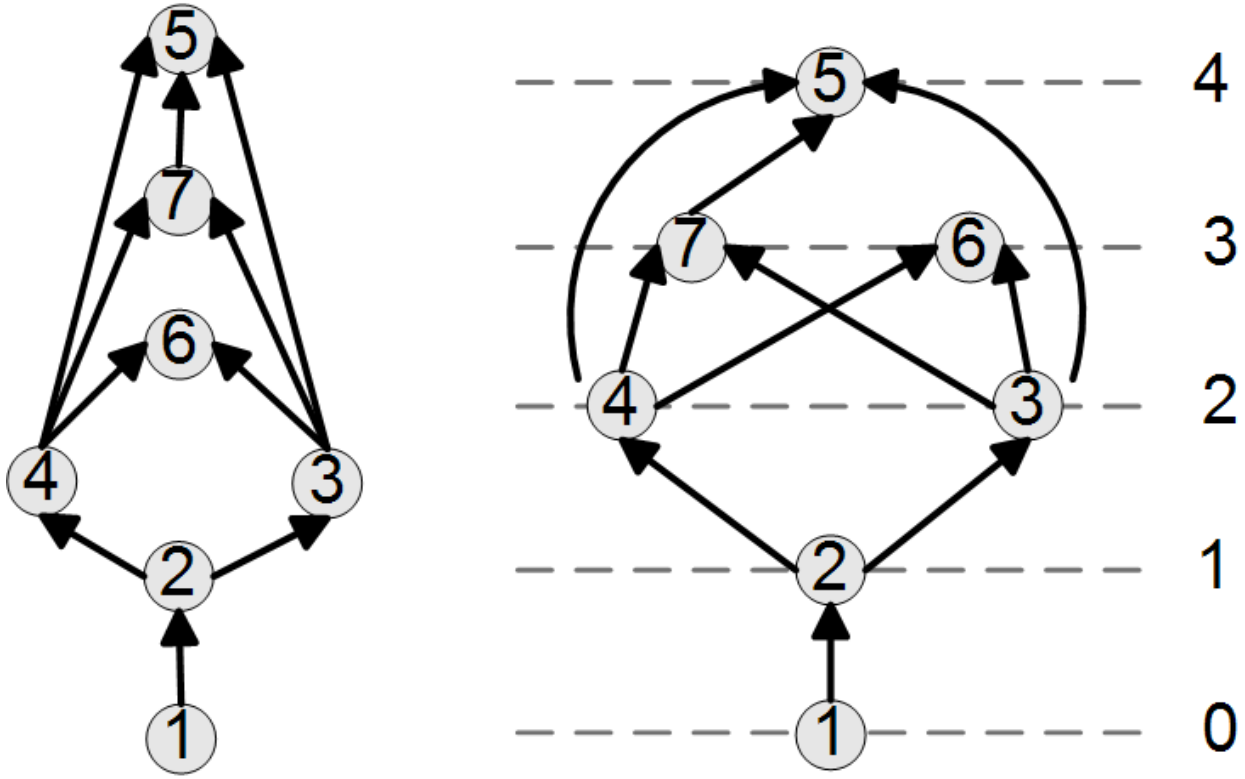
Hierarchische Zeichnung

Sugiyama-Verfahren

1. **Schichtenzuweisung**
2. Kreuzungsminimierung
3. Knotenplatzierung

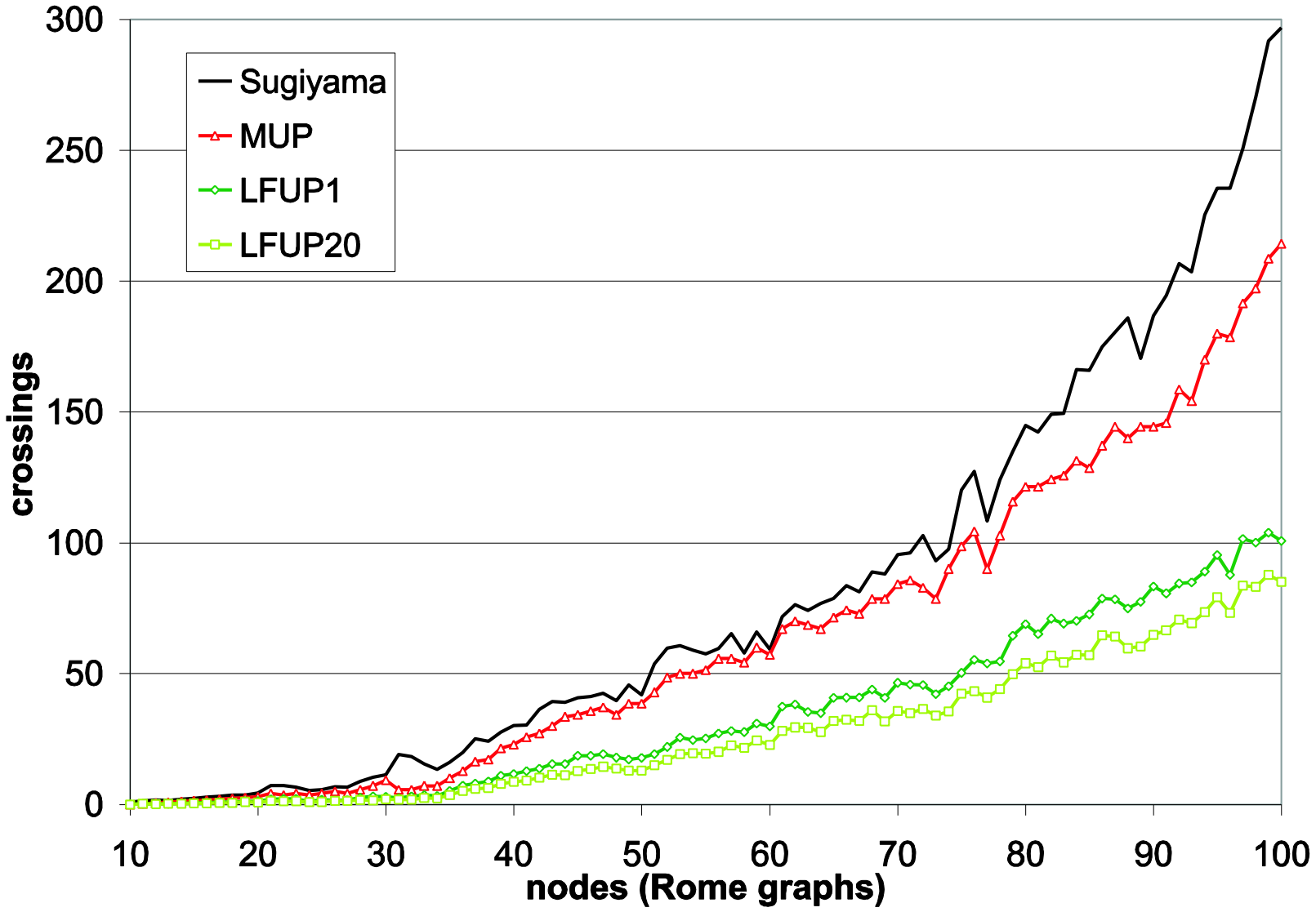


Verschiedene Schichtungen

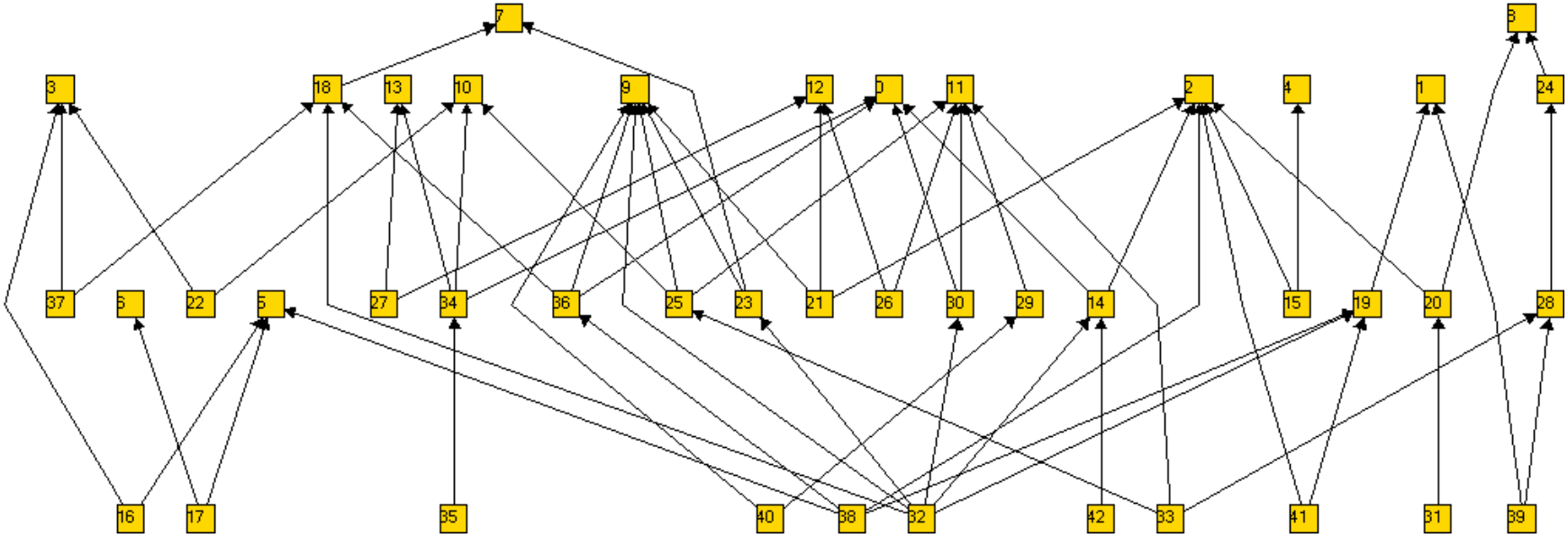


Eine schlechte Schichtung kann zu unnötigen Kreuzungen führen, die bei Kreuzungsminimierung nicht entfernt werden können

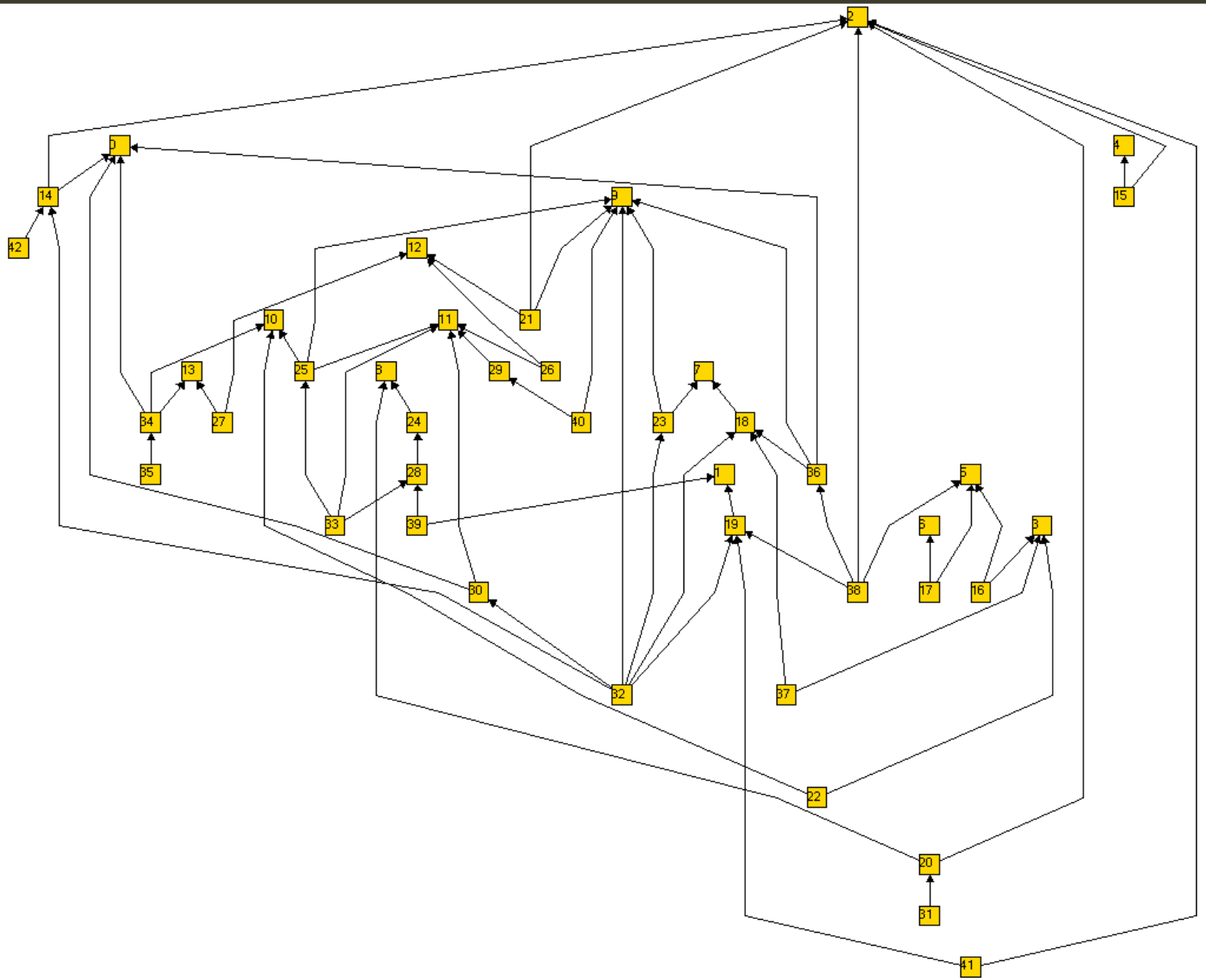
Experimente: Benchmark Graphen



Beispiel: Sugiyama



Upward Planarisierung



Ziele dieser Fachprojektgruppe

- Entwicklung, Implementierung (in OGDF), theoretische und experimentelle Analyse von neuen Schichtungsverfahren im Sugiyama-Verfahren

- Verfahren soll konzeptionell einfach sein
- Weniger Kreuzungen als Standard Sugiyama

Einzelne Aufgaben

- Einarbeitung in das Sugiyama-Verfahren
- Studium der bestehenden Schichtungsverfahren
- Entwicklung (einfacher) neuer Schichtungsverfahren mit Fokus auf wichtige Aesthetikkriterien, insbesondere Kreuzungszahl, basierend auf:
  - theoretischen Überlegungen, z.B. Vermeidung überflüssiger Kreuzungskonfigurationen, Studium von 2-level planaren Graphen
- Theoretische Analyse, wenn möglich
- Implementierung in OGDF (open source software)
- Experimentelle Evaluation anhand Benchmark Graphen

Thema (4)

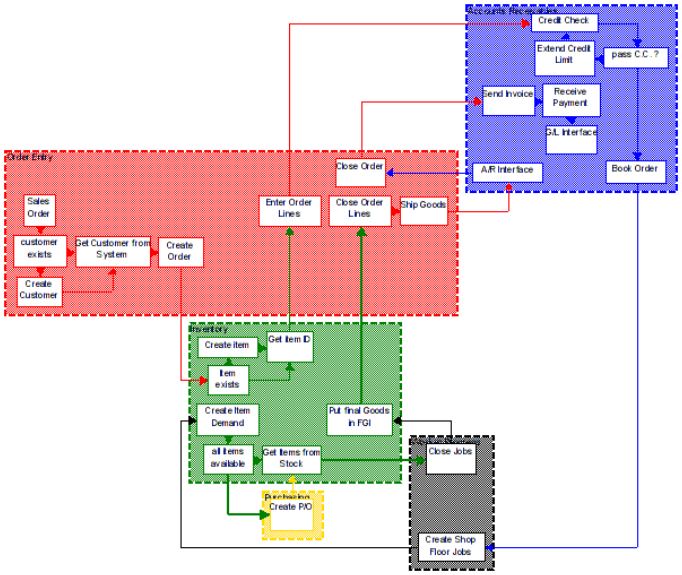
Nebenbedingungen in Multilevel-Verfahren

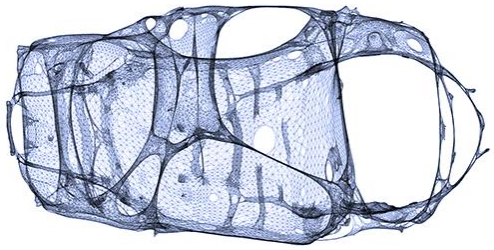
„Integration von Nebenbedingungen in Multilevel-Verfahren für das Zeichnen von Graphen“

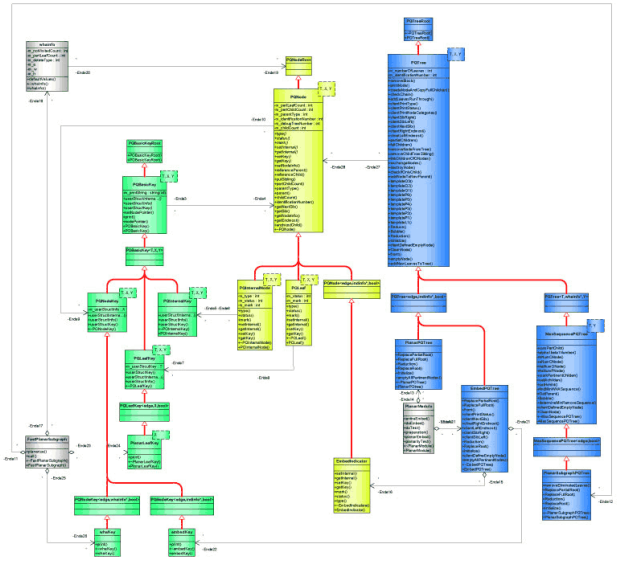
Betreuer: **Karsten Klein**

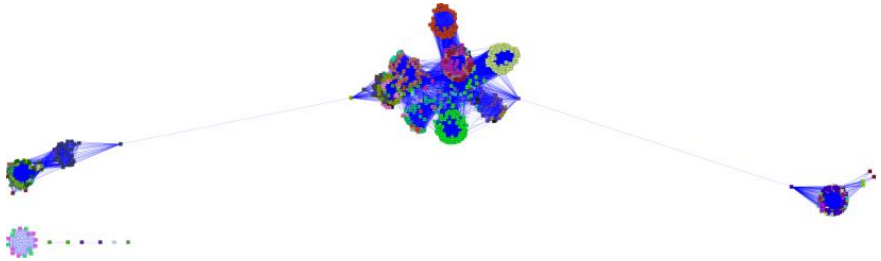
Automatic Graph Drawing

Übersichtliche Darstellung der Graphstruktur:





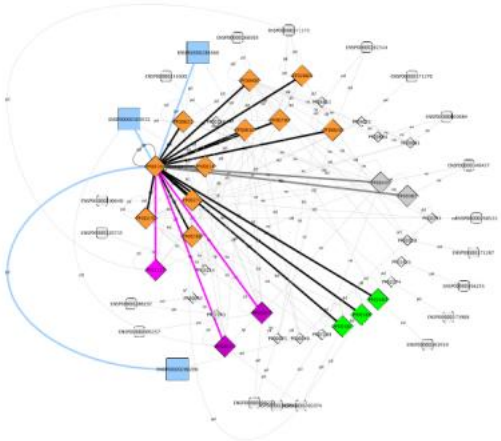


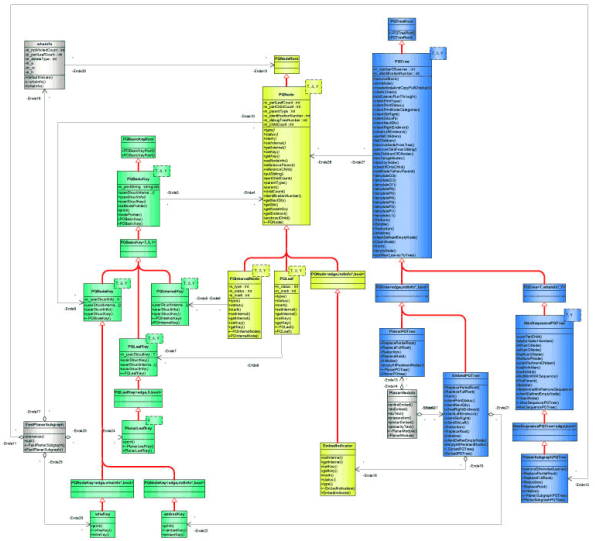


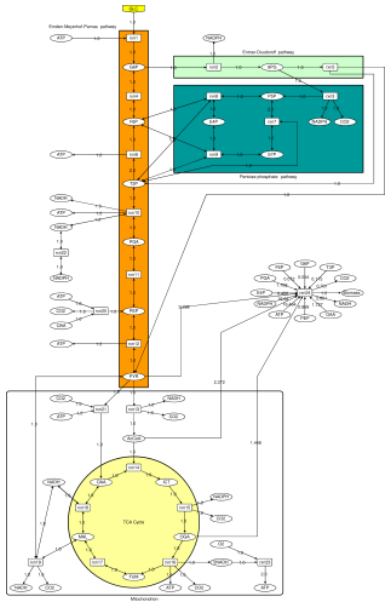
Praktische Anwendung

- Visual analytics in Biologie, Chemie, Software Engineering,... :
- Nicht nur Struktur, sondern auch semantische Information und unterschiedliche Zeichenkonventionen

⇒ Constraints!



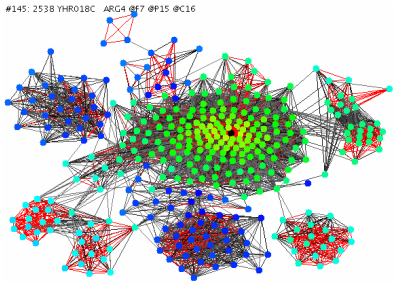


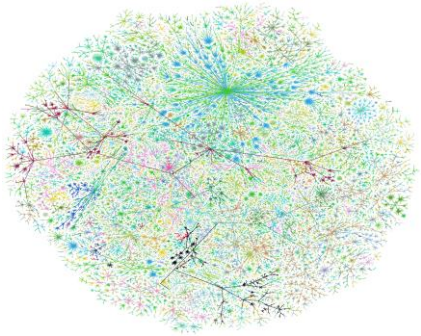


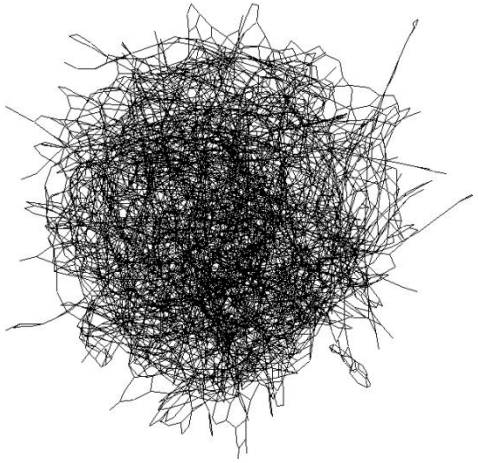
Praktische Anwendung

Große Graphen! > 10.000 Knoten

⇒ Multilevel Verfahren

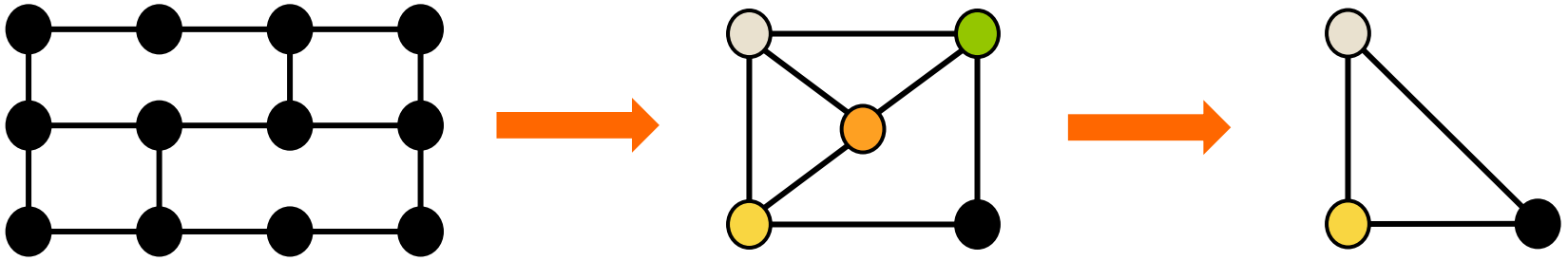




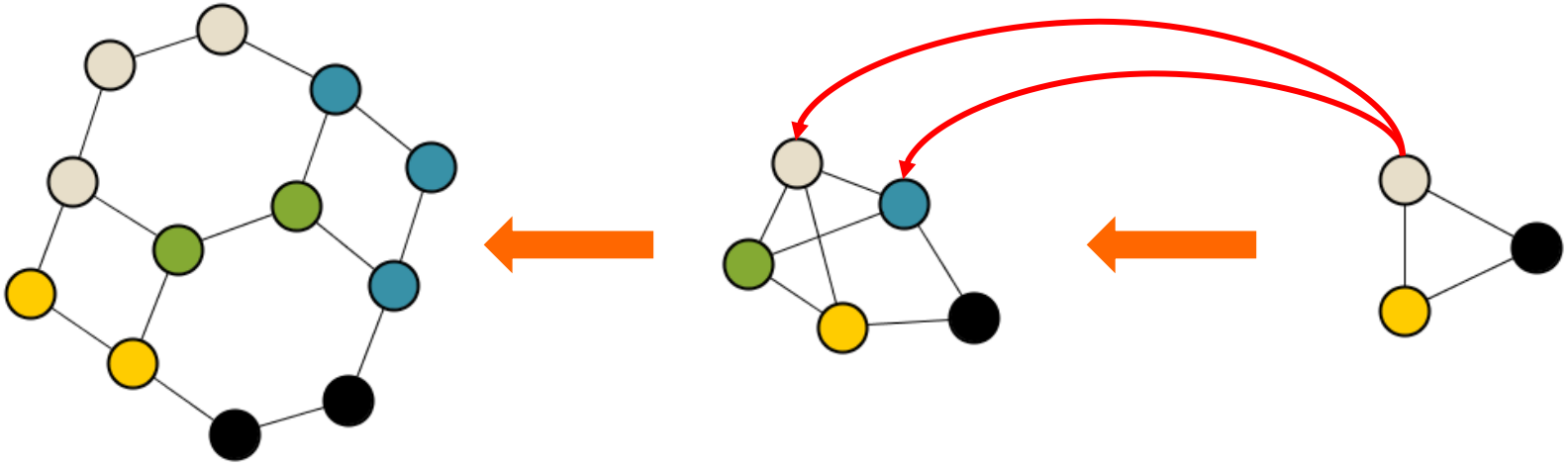


Multilevel-Verfahren

1. *Coarsening:* Sequenz von gröberen Repräsentationen
   $G = G_0 \rightarrow \ldots \rightarrow G_n$

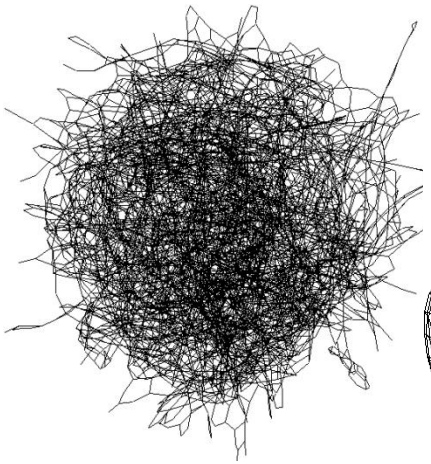


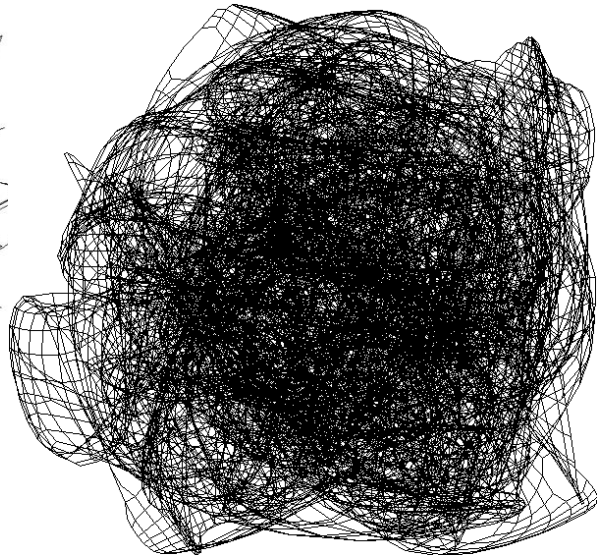
2. *Single Level Layout:* Auf jeder Stufe, $G = G_0 \leftarrow \ldots \leftarrow G_n$

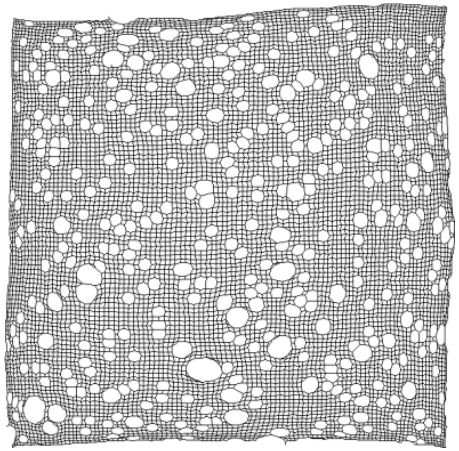


3. *Placement:* Bestehendes Layout erweitern (input)

Multilevel-Verfahren

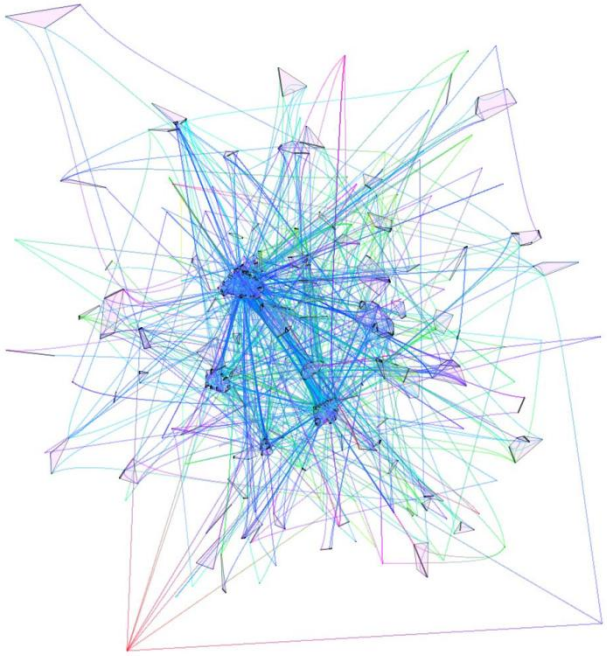




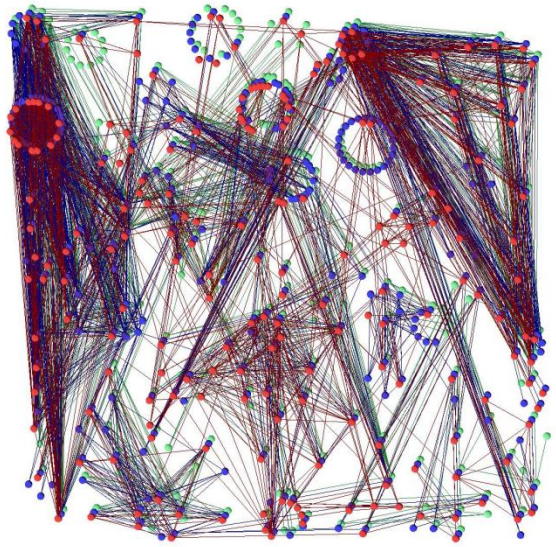


Multilevel-Verfahren

Großartig, aber...



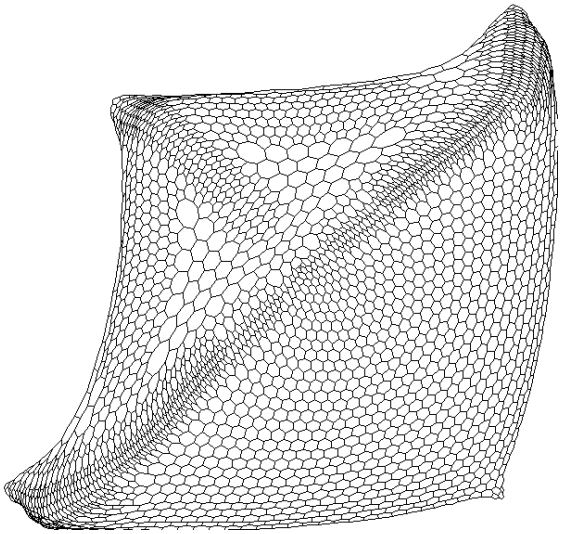
[Bourqui et al. IV2007]



[Brasch et al. 2009]

Aufgabe

- Untersuche Multilevel Verfahren auf Brauchbarkeit für Constraints.
- Implementierungen von Verfahren existieren bereits (OGDF).
- Gute Strategien für Constraints?



Thema (5)

Fingerprints

„Indexstrukturen für Fingerprints“

Betreuer: **Nils Kriege**

Problemstellung

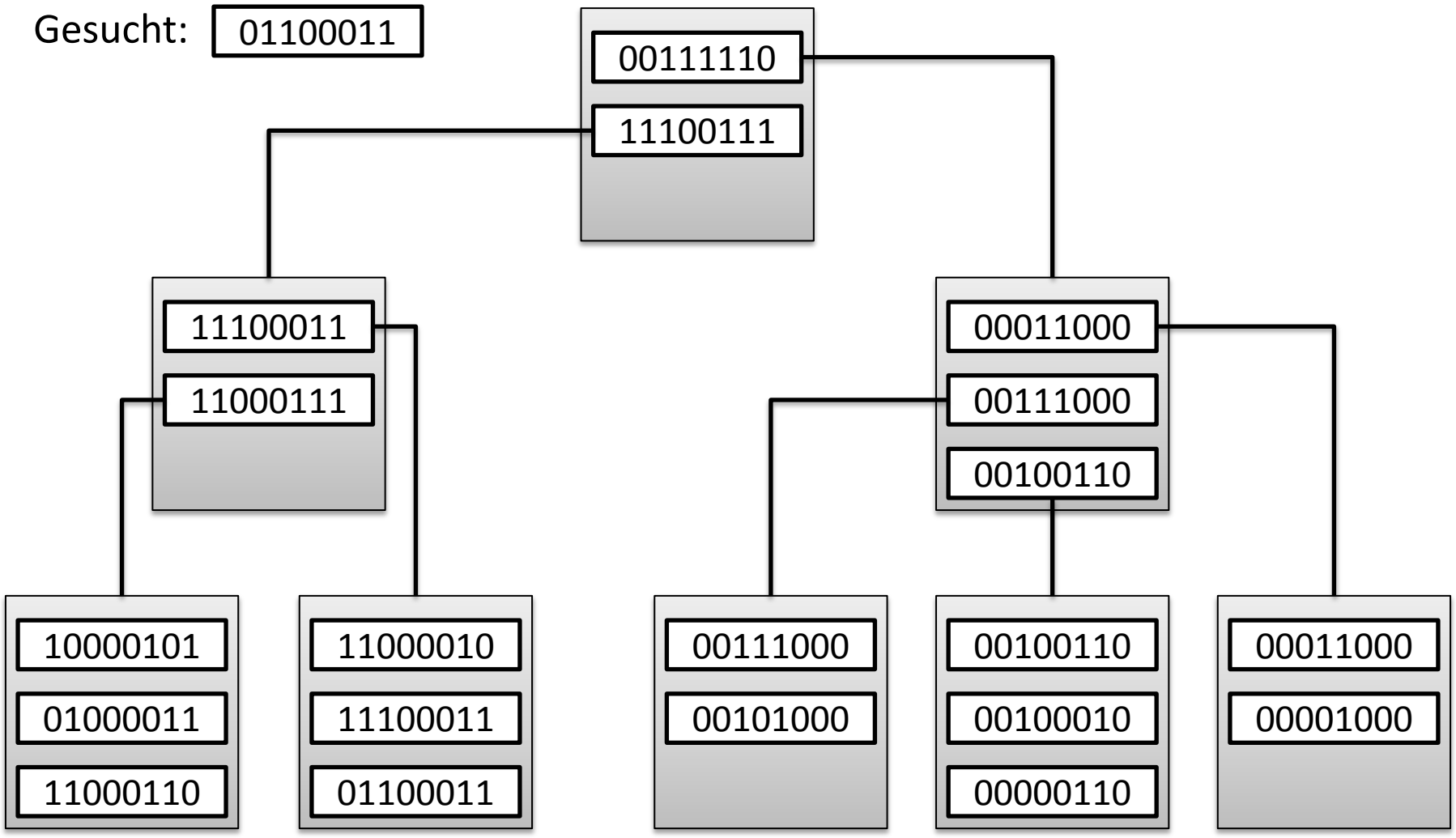
- Gegeben:
  - Grundmenge M
  - Datensatz $D = (D_1, \ldots, D_n)$ mit $D_i \subseteq M$
  - Suchmuster $Q \subseteq M$
- Gesucht:
  - $D(Q) = \{ i : Q \subseteq D_i \}$ (bzw. = oder $\supseteq$)
- Fingerprint:
  - Bitvektor konstanter Länge zur Repräsentation einer Menge
  - Für jedes Element e der Menge: Setze das Bit an Position h(e) auf 1

Beispiel

- D_1 = {DAP1, DAP2, RS} — 01001010
- D_2 = {DAP1, M1} — 10000010
- D_3 = {DAP1, DAP2, RS, M1} — 11001010
- D_4 = {DAP1, BS} — 00010010

- Suchmuster Q = {DAP1, RS} — 01000010
  - D(Q) = {1, 3}

Indexstruktur: S-Tree



Anwendungen / Organisation

- Filtern bei der Suche in großen Datenmengen
  - Textdokumente
  - Moleküldatenbanken
    - Fingerprintgröße typischerweise 512 - 4096 Bit
    - Datenbankgröße: 1000 - 10.000.000 Einträge
- Material:
  - Improved Methods for Signature-Tree Construction; Tousidou, Nanopoulos, Manolopoulos; 2000
  - Eigene Recherche, eigene Ideen
- Programmiersprache: **Java** oder C++