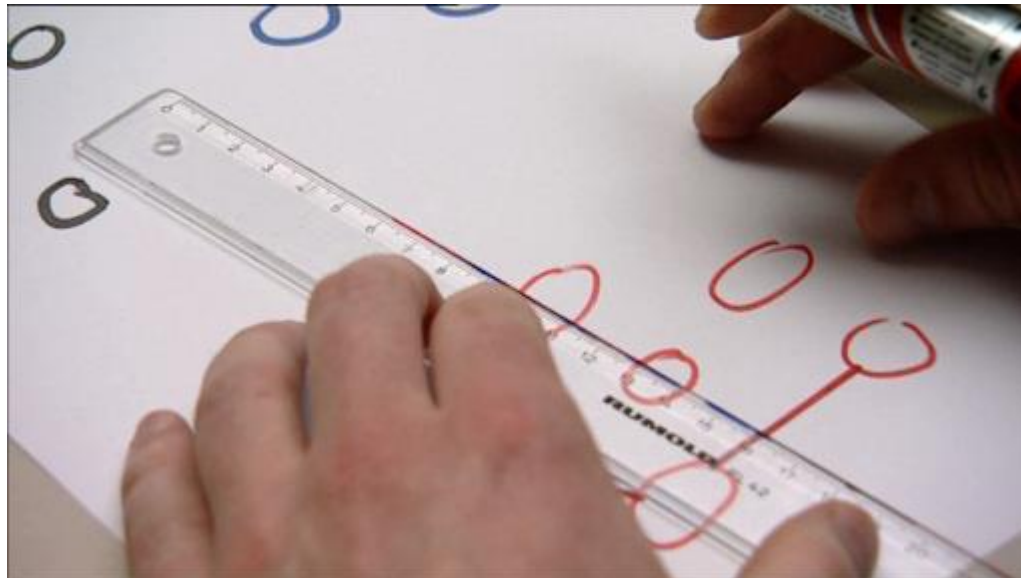


Visuelle Analyse großer Datenmengen

Heidrun Schumann
Institut für Informatik, Universität Rostock

Visuelle Datenanalyse – Warum?

- Verbesserte Kommunikation von Daten, Modellen und Konzepten



Film im SciencExpress „Expedition Zukunft“

Visuelle Datenanalyse – Wie?

- Geeignete **Visualisierung** + Interaktion

Visualisierung von

Daten

+ Zeit

+ Raum

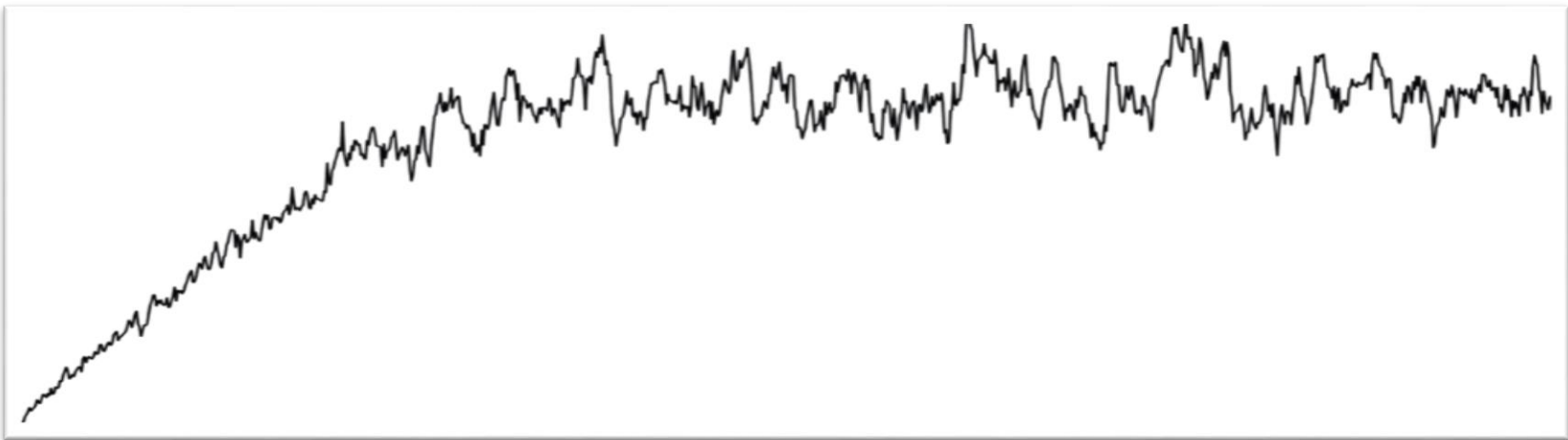
+ Datenherkunft & -qualität

+ Struktur

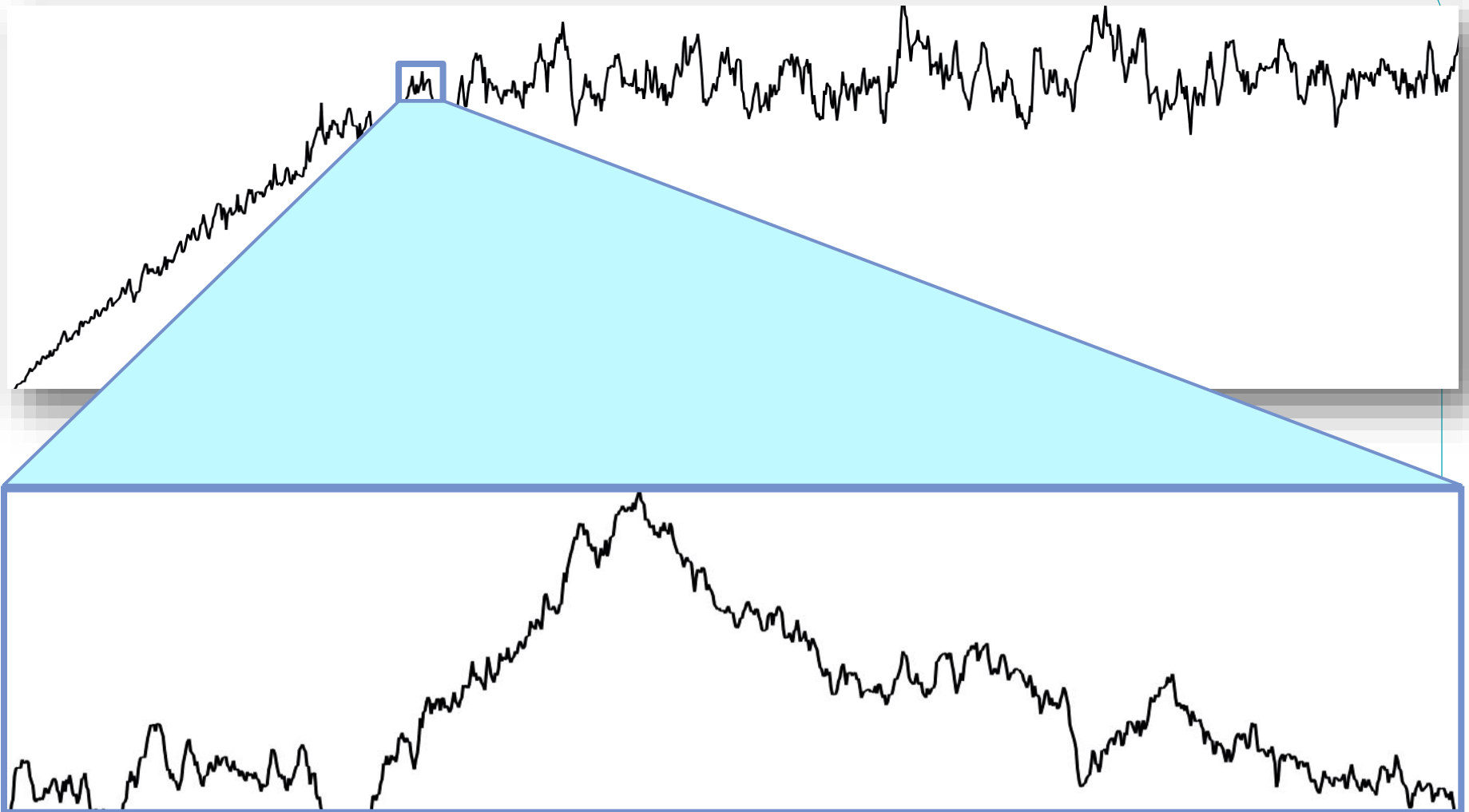
Präsentation & Diskussion

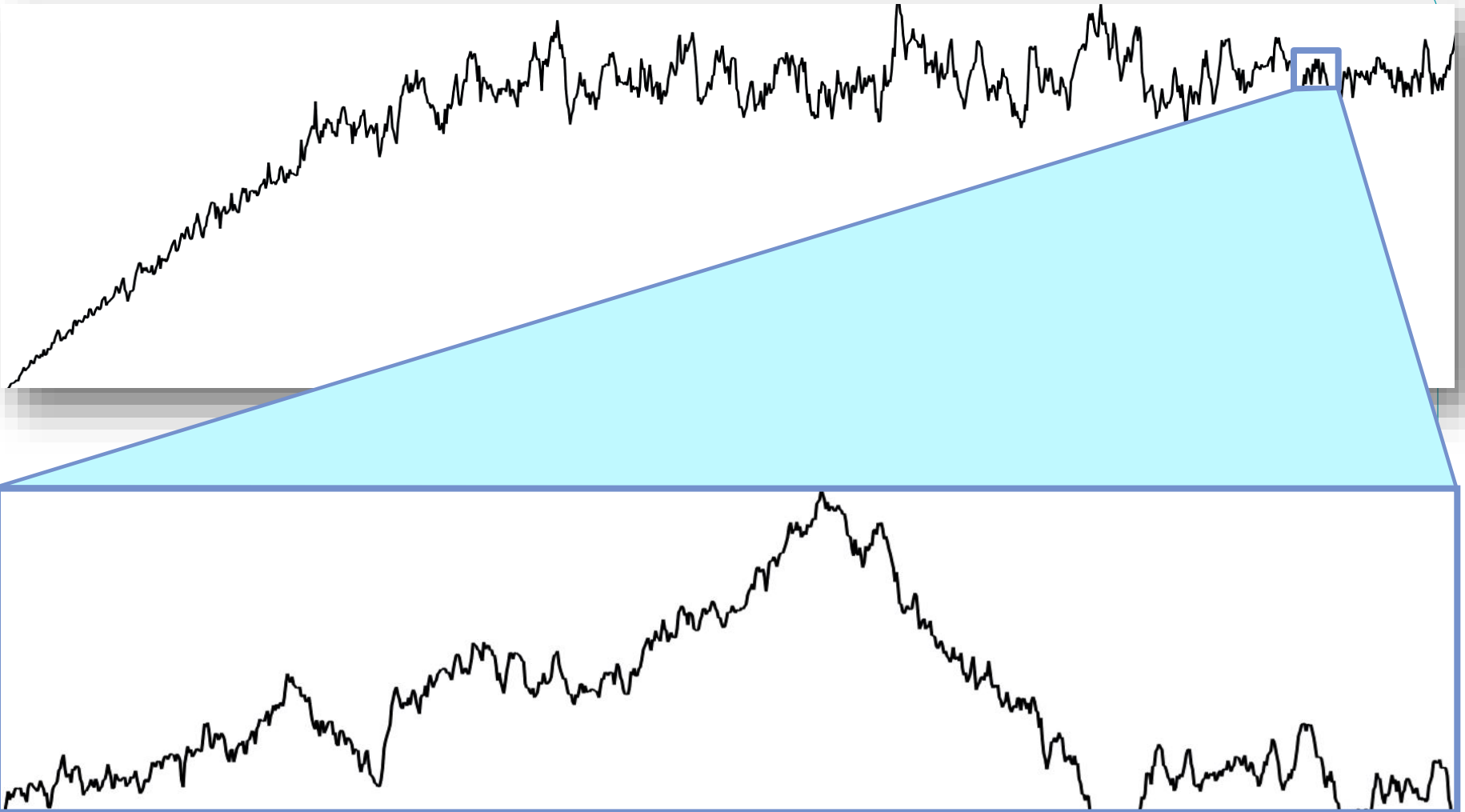
Visualisierung der Daten (1)

- **Beispiel:** Zeitdiagramm



Zeitreihe mit etwa 10^6 Zeitpunkten,
d.h. mehr als 1000 Zeitpunkte pro Pixel!!!





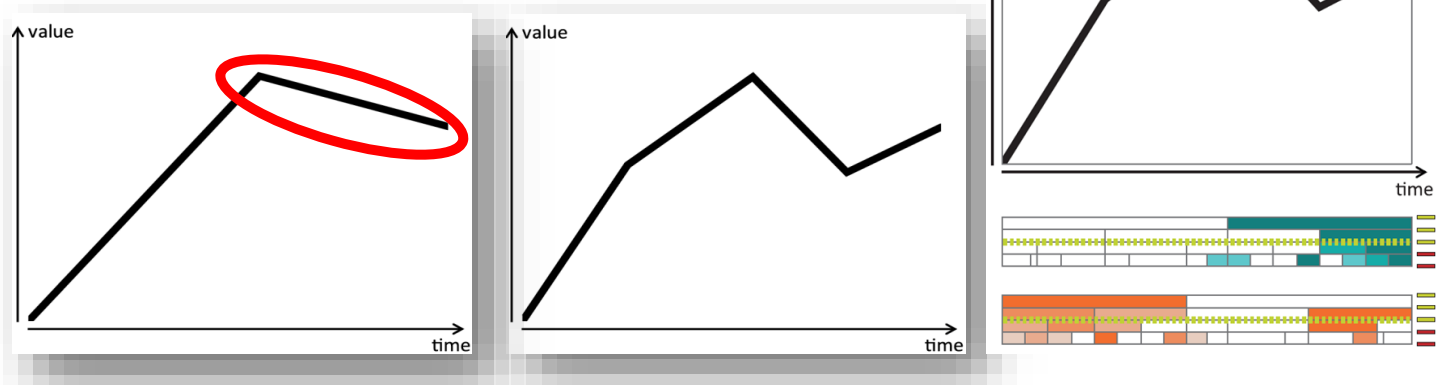
Visualisierung der Daten (2)

- **Unsere These:**

ZoomIn dort notwendig, wo sich low-level Daten von high-level Daten unterscheiden → **interessante Informationen** auf feiner aufgelöster Skala

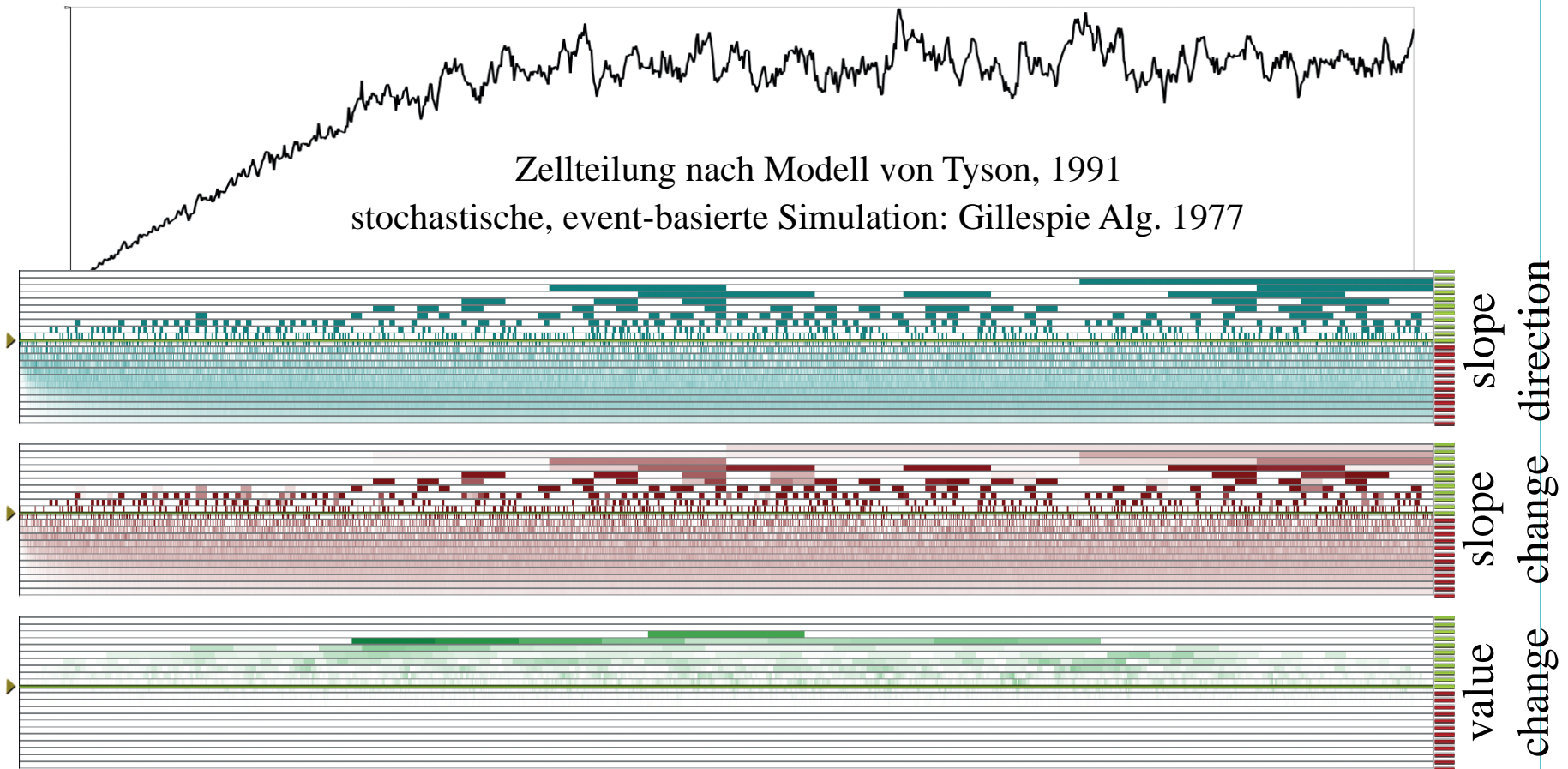
- **Unser Ansatz** (Luboschik et al., BioVis 2012):

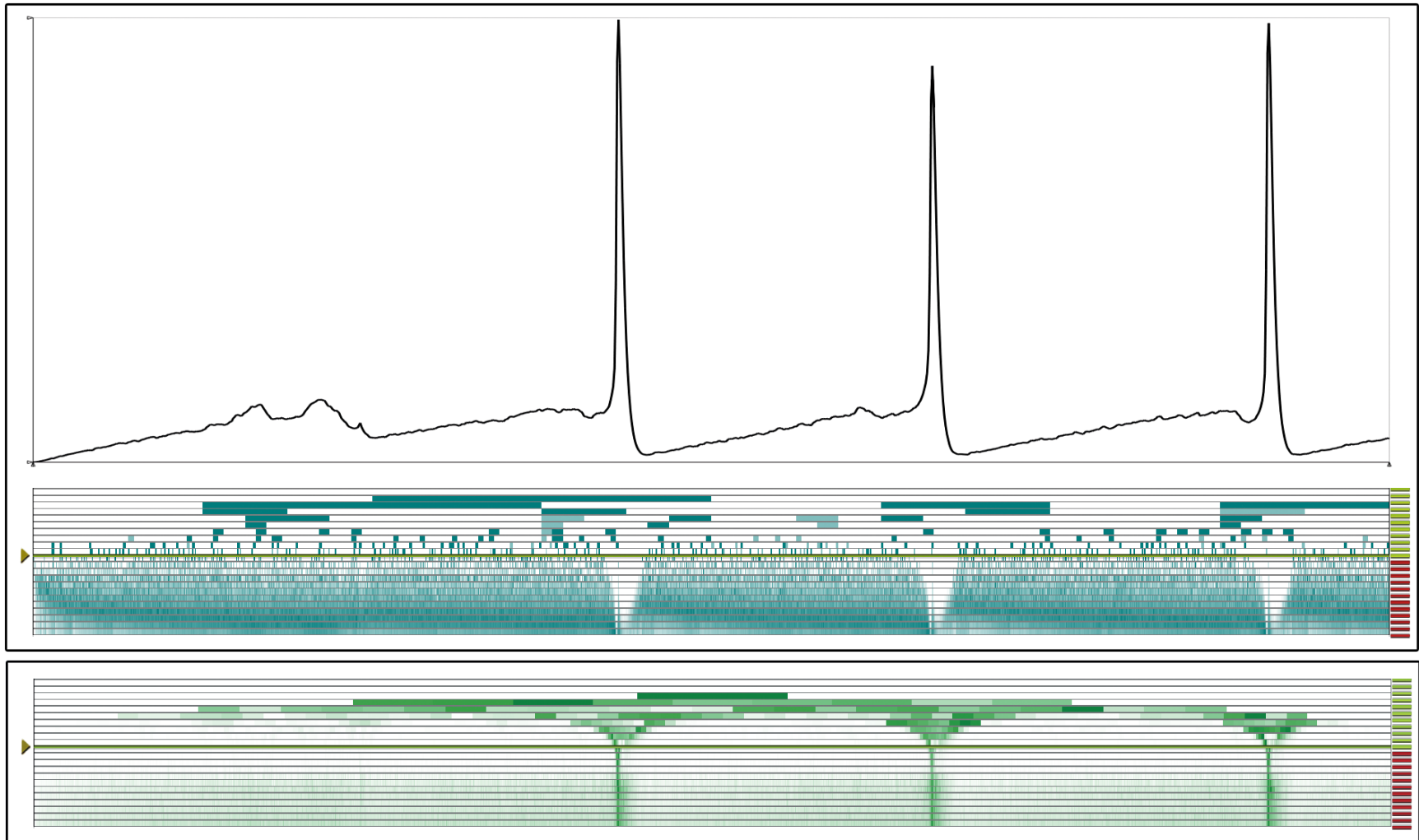
- Berechnen von Heterogenitätswerten zwischen benachbarten Skalen
- Visualisieren der Daten & Heterogenitätswerte
- Navigation auf Basis der Heterogenitätswerte.



Visualisierung der Daten (3)

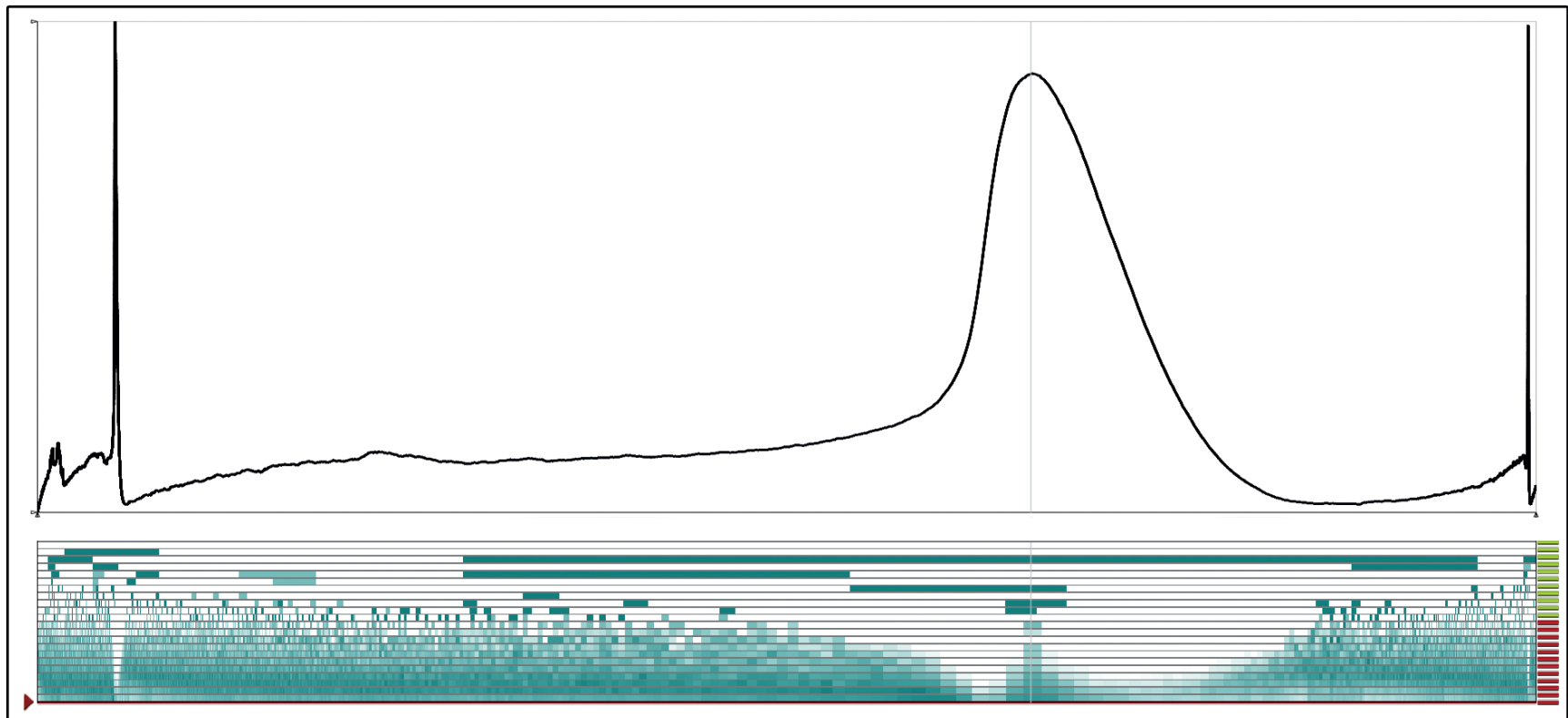
Zellteilung nach Modell von Tyson, 1991
stochastische, event-basierte Simulation: Gillespie Alg. 1977





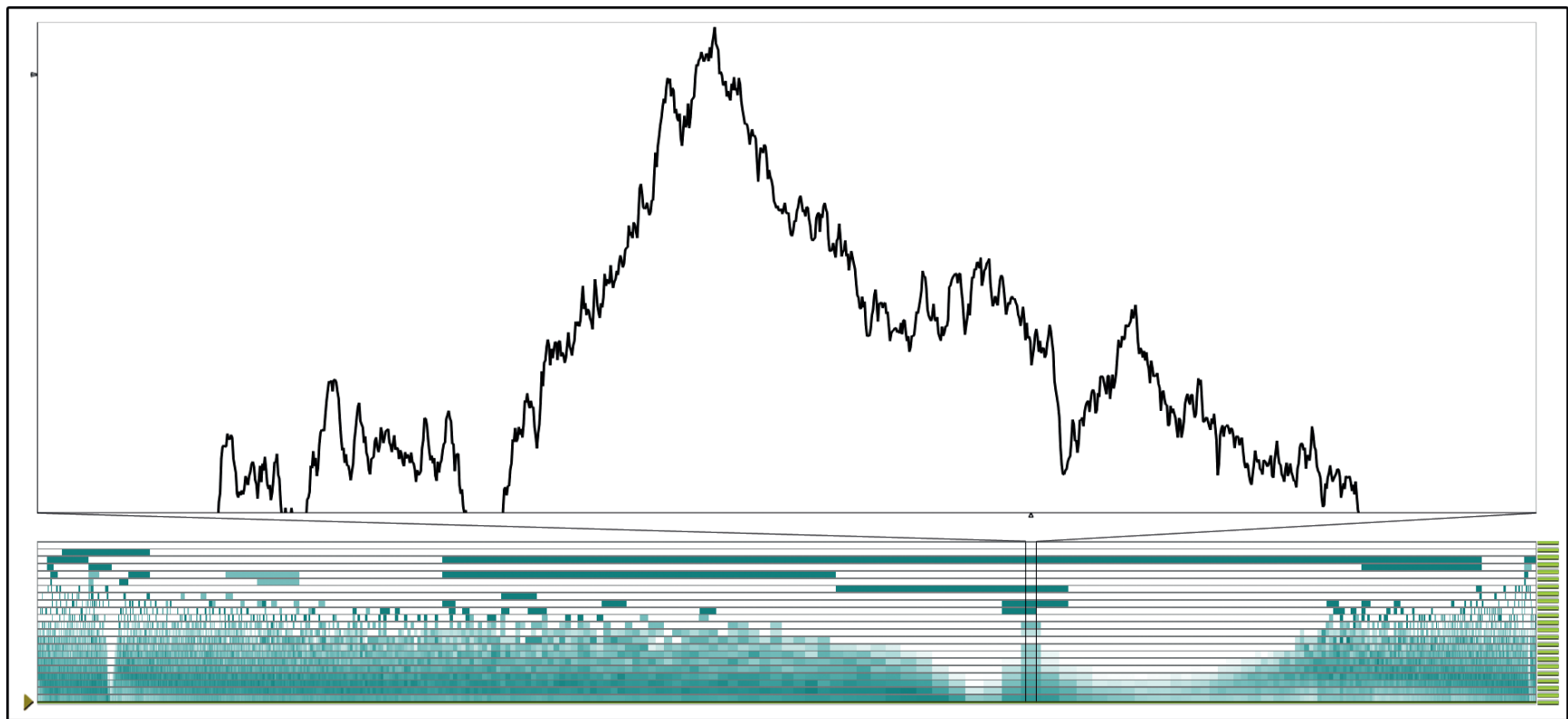
value
change
slope
direction

Visualisierung der Daten (4)



slope
direction

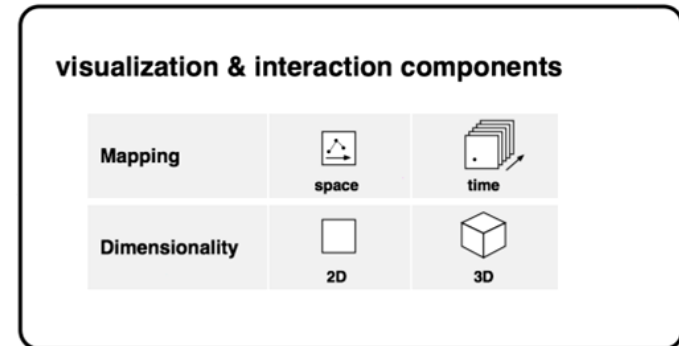
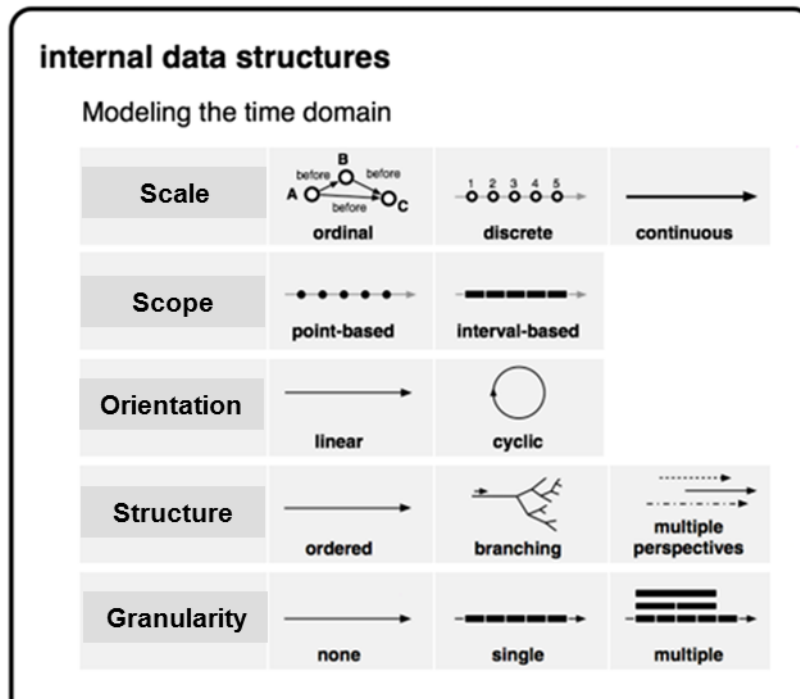
Visualisierung der Daten (5)



slope
direction

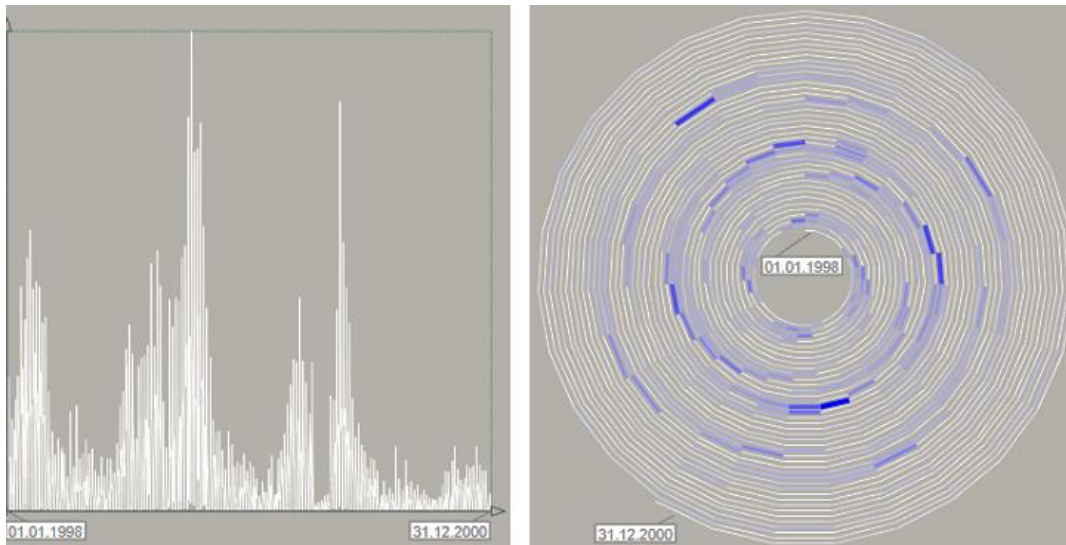
... Daten + Zeit (1)

- ... + Eigenschaften der Zeitachse



... Daten + Zeit (2)

- **Beispiel:** Berücksichtigung der Orientierung der Zeitachse: Linear vs. Zyklisch



Veranschaulichung des zyklischen Zeitcharakters am Beispiel von AOK-Daten M-V

... Daten + Zeit (3)

- TimeViz-Browser: <http://survey.timeviz.net/>

The TimeViz Browser
A Visual Survey of Visualization Techniques for Time-Oriented Data
by Christian Tominski and Wolfgang Aigner

of Techniques: 112

Search:

Data

Frame of Reference

Abstract ON

Spatial ON

Number of Variables

Univariate ON

Multivariate ON

Time

Arrangement

Linear ON

Cyclic ON

Time Primitives

Instant ON

Interval ON

Visualization

Mapping

Static ON

Dynamic ON

Dimensionality

2D ON

3D ON

Our book:



... Daten + Raum (1)

- ... + räumliche Position der Daten

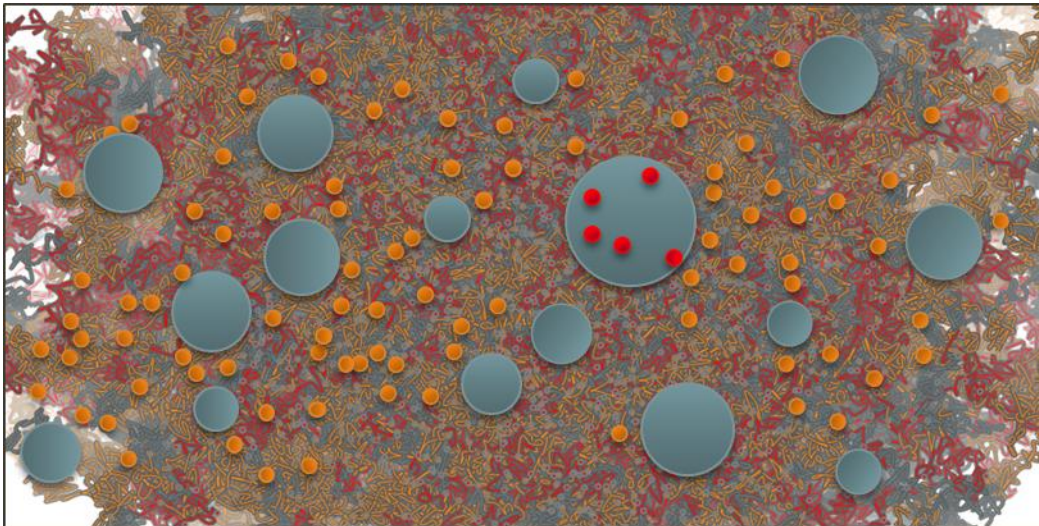


Beispiel:
Stochastische räumliche
Simulation in der Zell-Biologie

Brownsche Bewegung von
Proteinen und Lipid Rafts

... Daten + Raum (2)

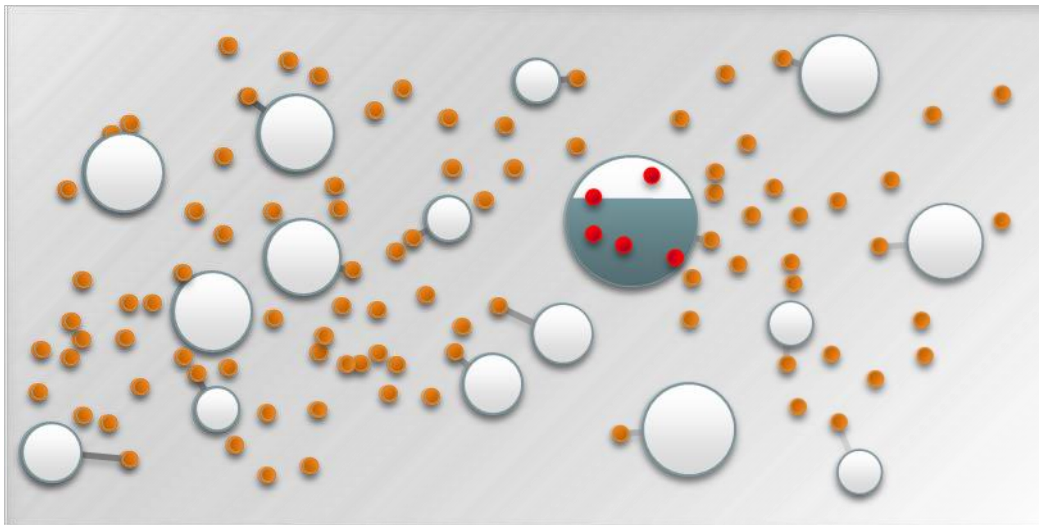
- Bewegungen auf der Zelloberfläche von Lipid Rafts und Proteinen



Tausende Objekte,
Tausende Zeitschritte

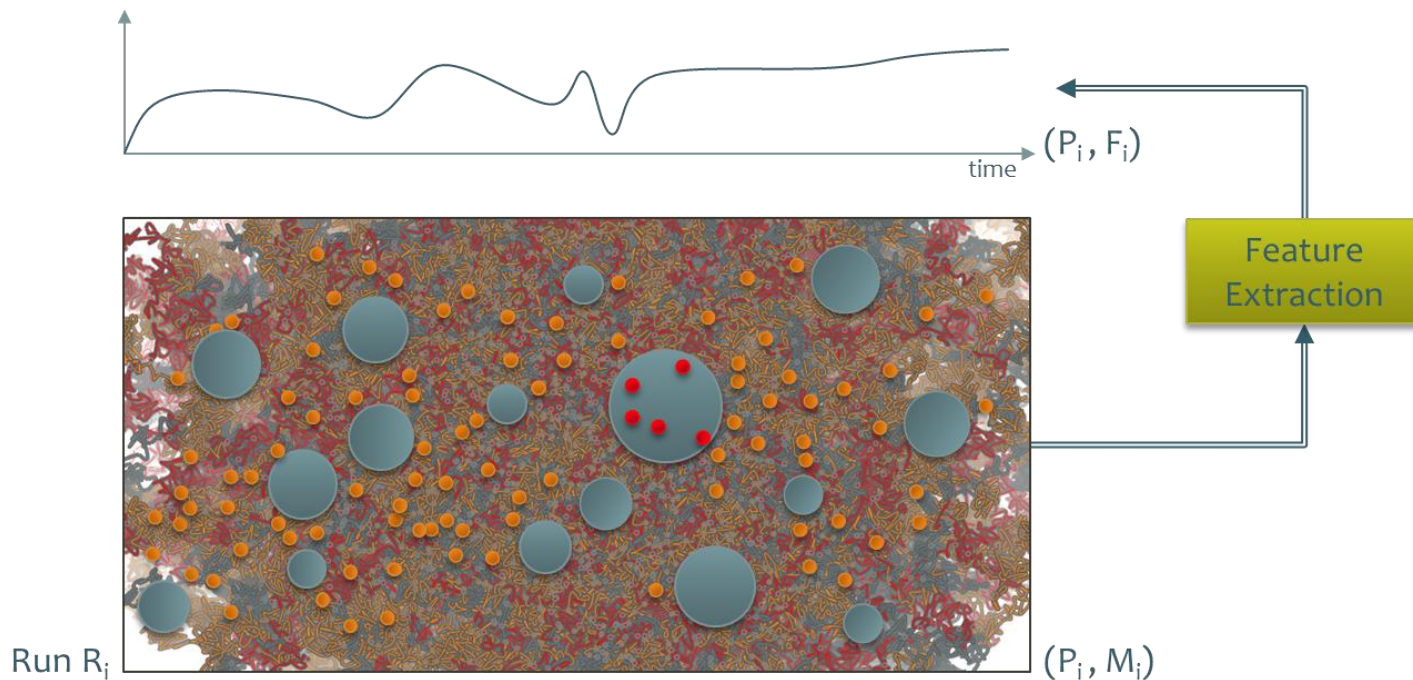
... Daten + Raum (3)

- Unser Ansatz, Luboschik et al., 2015:
Berechnung von Features
 - Basic features: Richtung, Geschwindigkeit, kleinste Distanz
 - Group features:, Anzahl, Füllstand,.....



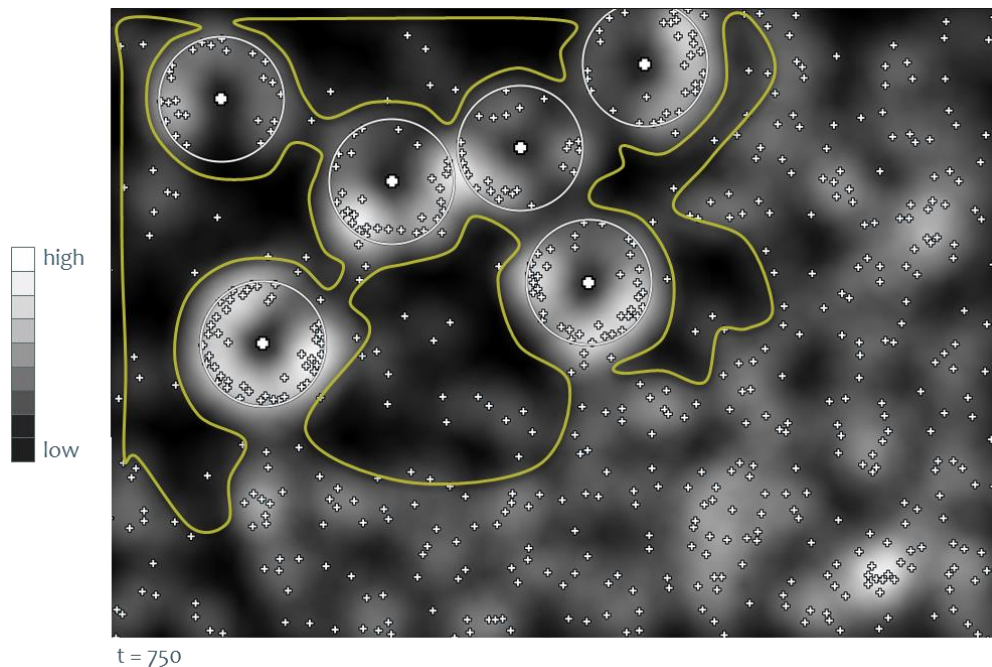
... Daten + Raum (4)

- Berechnung von Features
 - Advanced features: ...temporal clustering, statistische Daten, ...



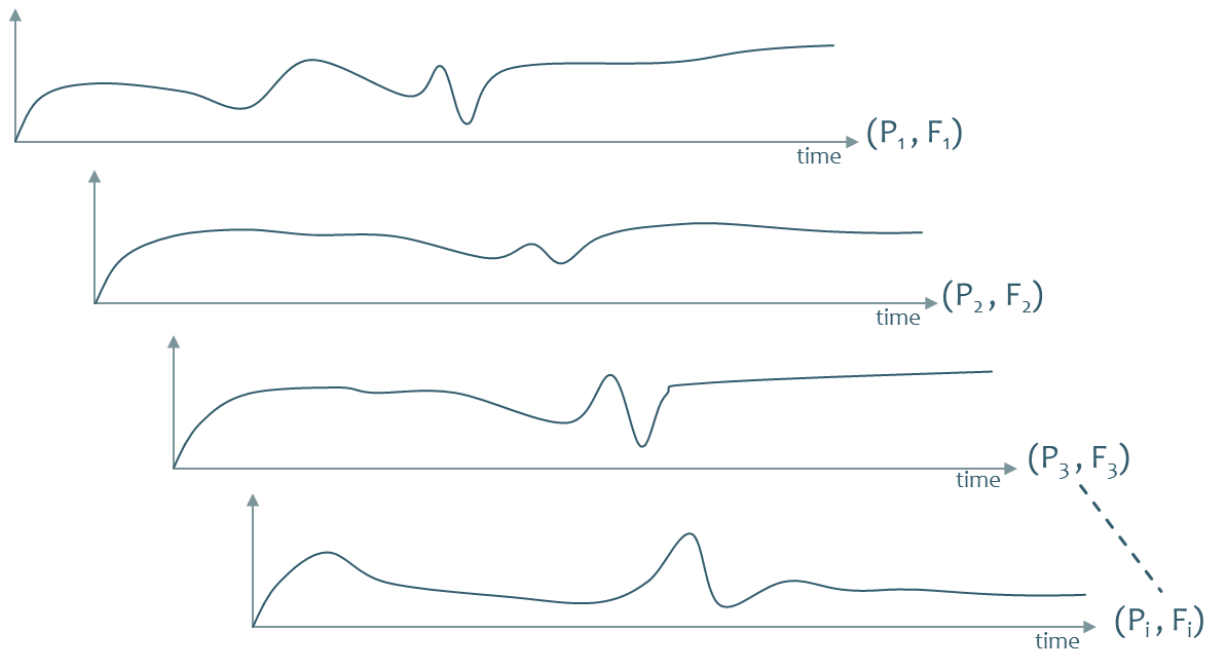
... Daten + Raum (5)

- Berechnung von Features
 - Region features: ... high/low density count, high/low density number, ...



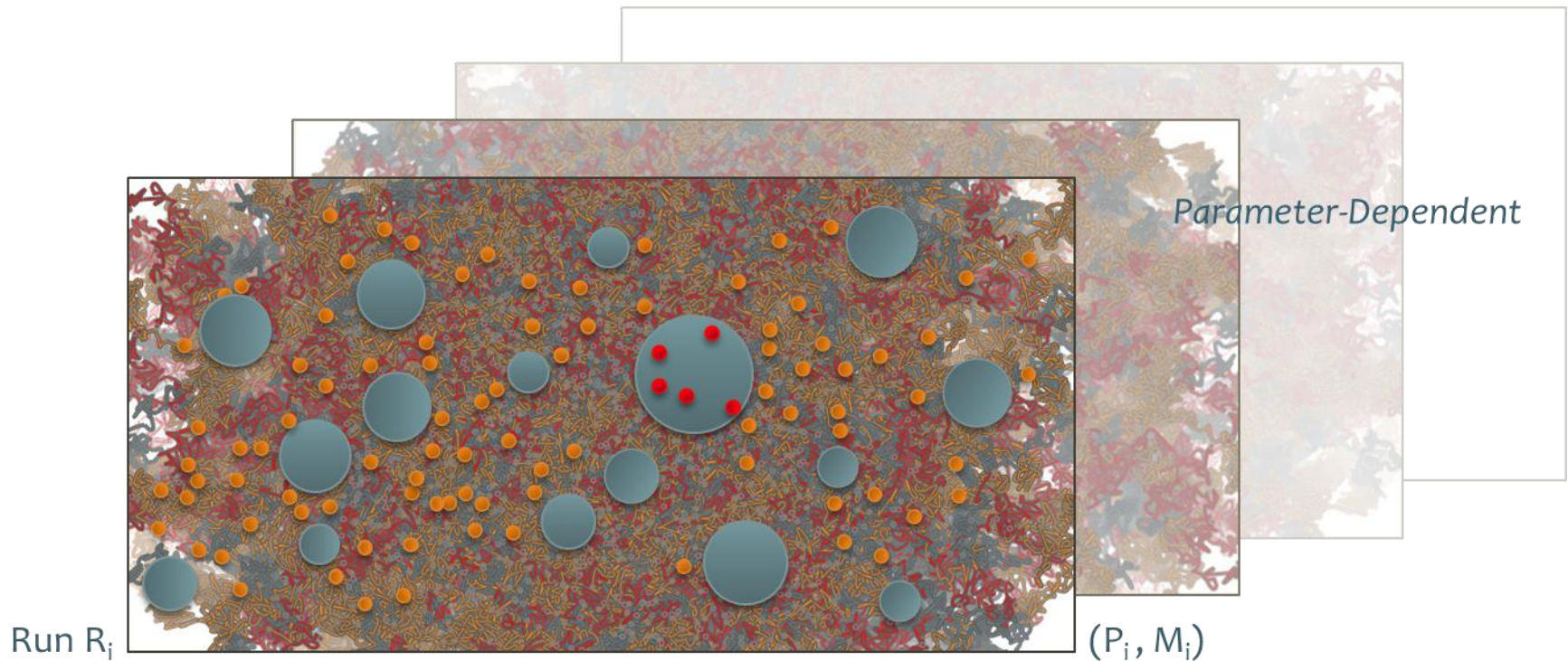
... Daten + Raum (6)

- Berechnung von Features



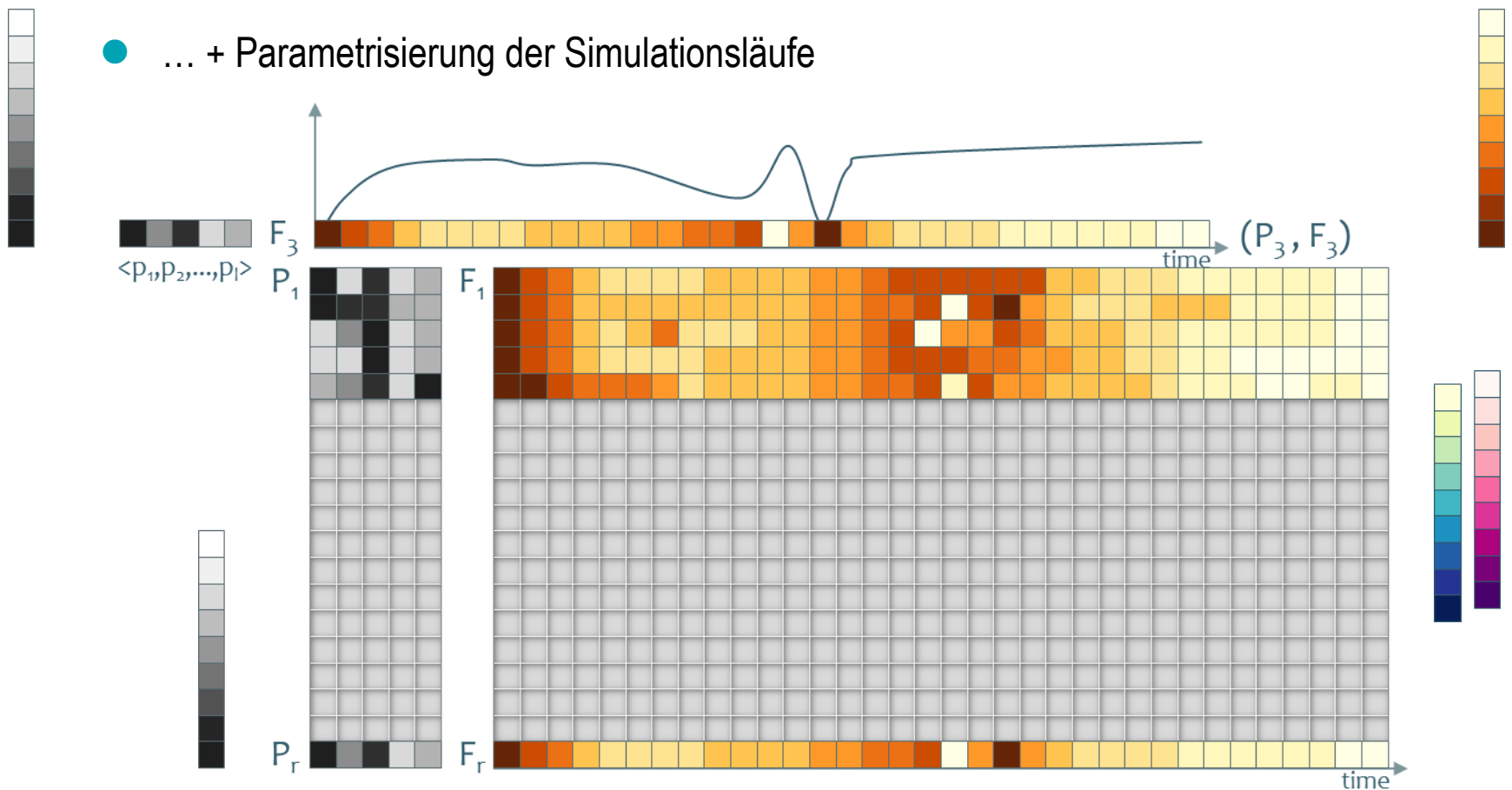
... Daten + Datenherkunft (1)

- ... + Parametrisierung der Simulationsläufe



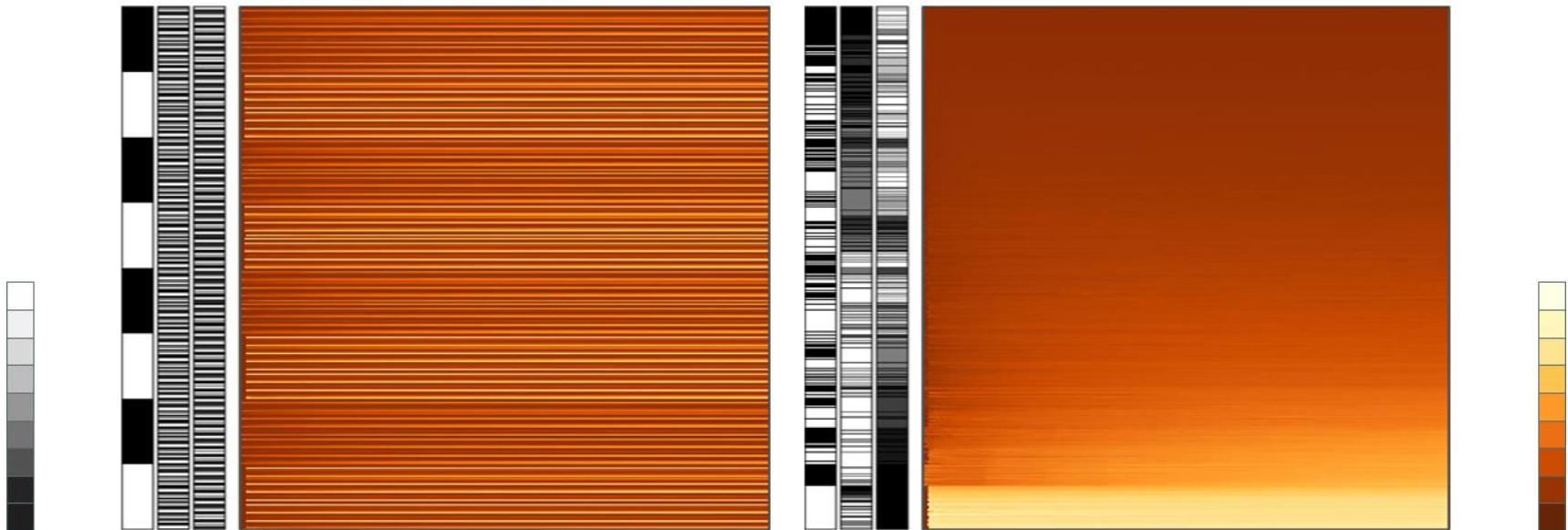
... Daten + Datenherkunft (2)

- ... + Parametrisierung der Simulationsläufe



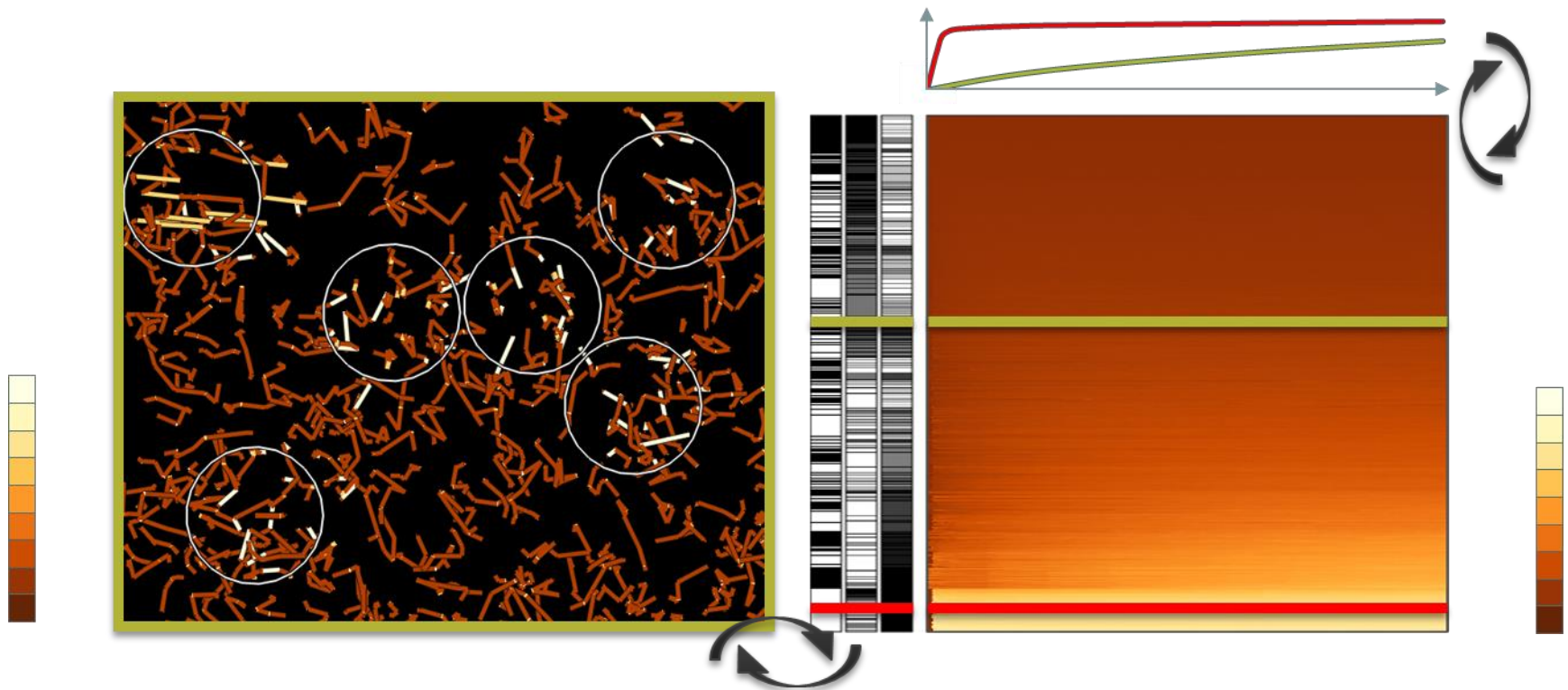
... Daten + Datenherkunft (3)

- Sortierung von Parameter- oder Datenwerten



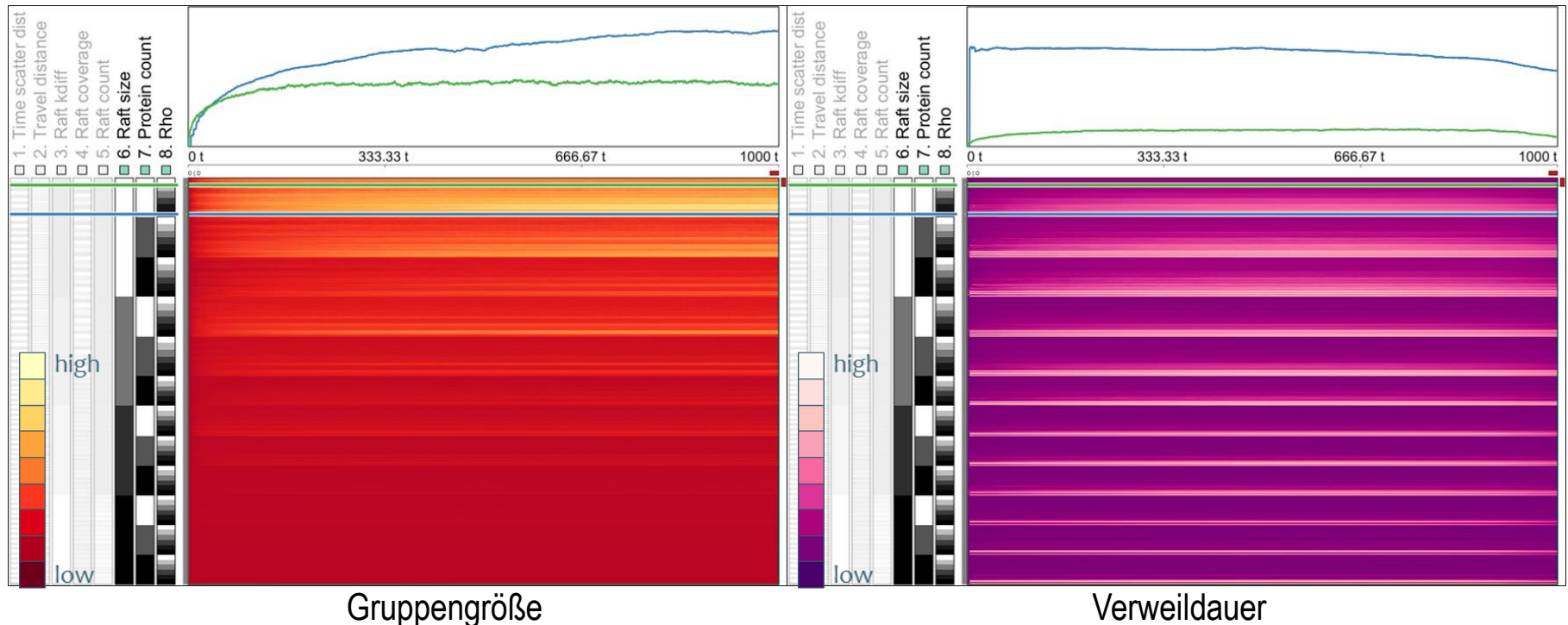
... Daten + Datenherkunft (4)

- Inspektion einzelner Parametrisierungen in Raum und Zeit



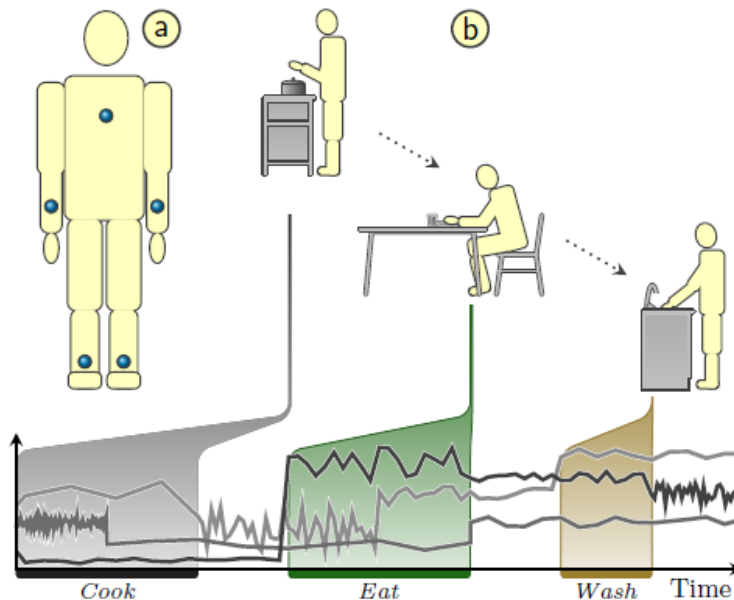
... Daten + Datenherkunft (5)

- Vergleich von Parametersettings
(8 Parameters: Raft size, Fluidity, Count..., ~ 2,000 Parameterizations, ~ 1,000 lipid rafts, ~ 5,000 proteins, ~ 4,000 time steps, ~ **180 GB raw data** → ~ **3 GB feature data**)



... Daten + Datenqualität (1)

- Welche Parametersettings führen zu sicheren Ergebnissen?
Beispiel: Aktivitätserkennung mit CSSM, Krüger et al. 2015



16 Aktivitäten (z.B. nehmen)
99 Aktionen (Messer nehmen)
→ 146 M Zustände

5 Parameter,
432 Parametersettings

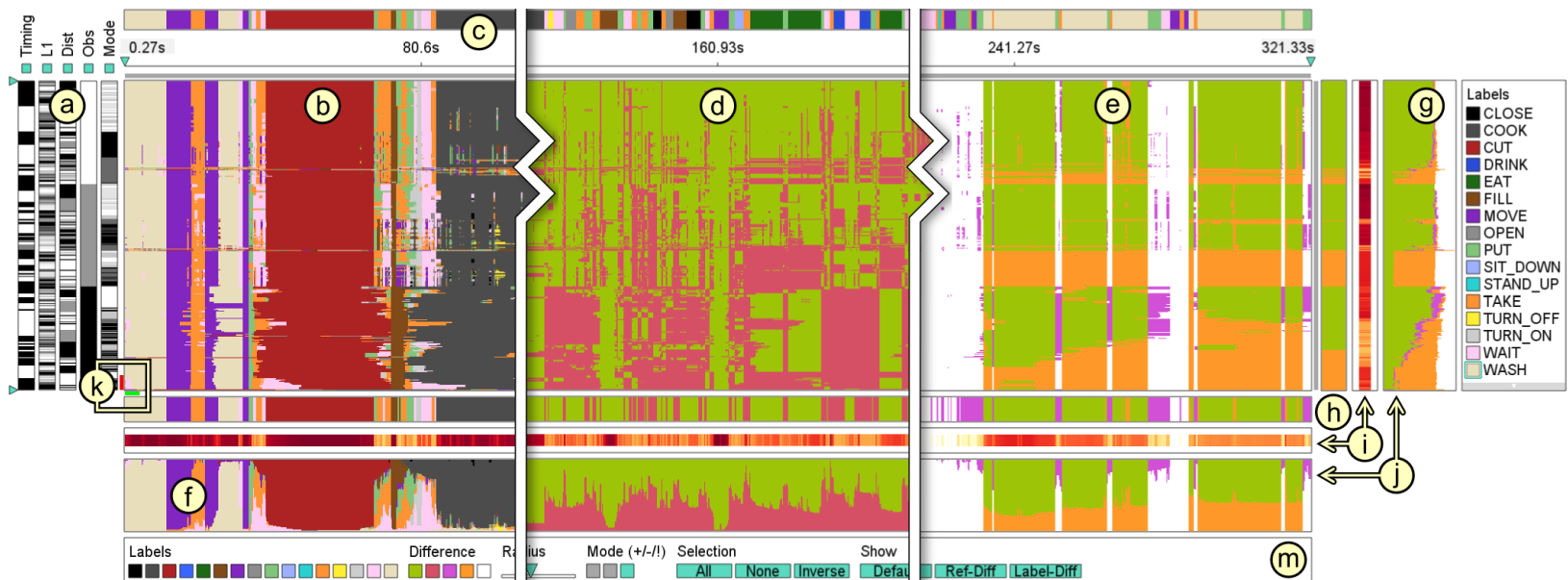
3004 Runs zur Abschätzung der
Wahrscheinlichkeiten für eine
Aktivität auf Basis der gemessenen
Sensordaten

600 – 1200 Zeitpunkte

Begleitende Videodaten

... Daten + Datenqualität (2)

- Einfluss der Parameter & Wertebelegungen → Qualität der Aktivitätserkennung
Röhlig et al., IEEE VAST 2015



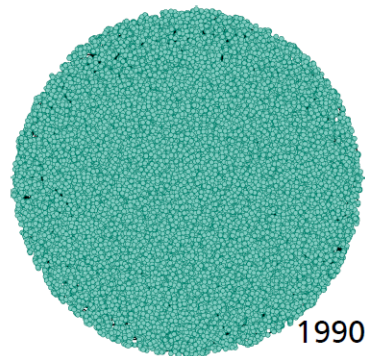
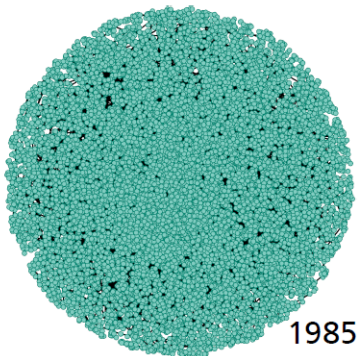
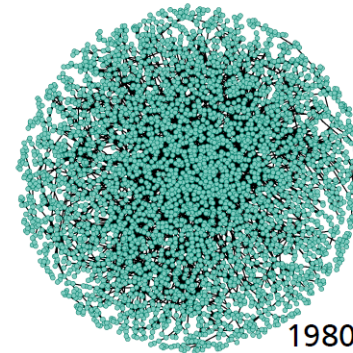
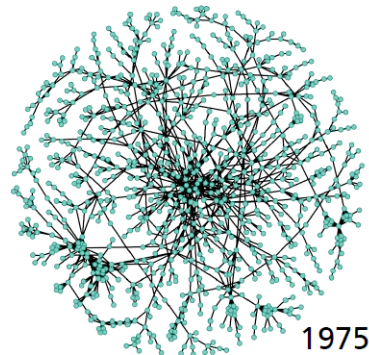
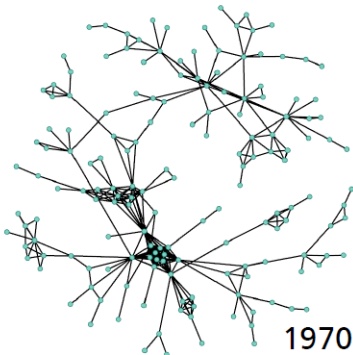
Farbkodierung der
erkannten Aktivitäten:

alle Aktivitäten:
Grün → richtig
rot → falsch

eine Aktivität:
TP → grün; FP → violett
FN → orange; TN → transp.

... Daten + Struktur (1)

- ... + Relationen zwischen den Daten



DBLP:
Co-Autoren Netzwerk
21 time points:
1990-2011
914 492 nodes
3 802 317 edges

$$G_i = \{ V_i, A_{vi}, E_i, A_{ei}, t_i \}$$

V_i : Menge von Knoten
 A_{vi} : Menge von Knotenattributen
 E_i : Menge von Kanten
 A_{ei} : Menge von Kantenattributen
 t_i : Zeitpunkt

... Daten + Struktur (2)

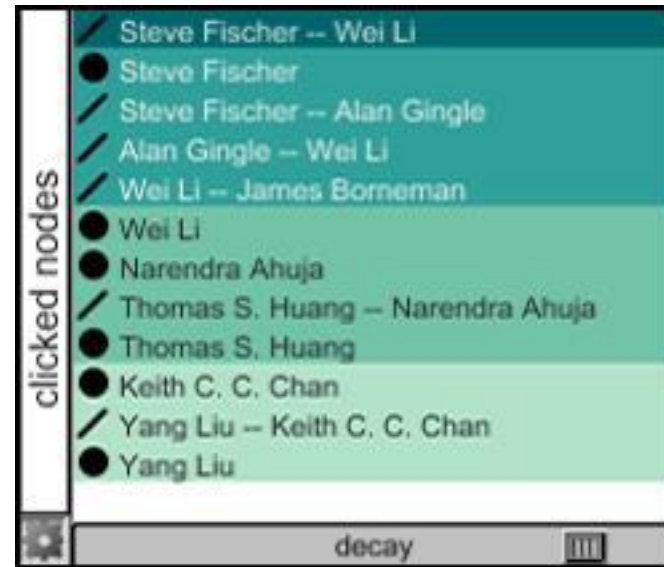
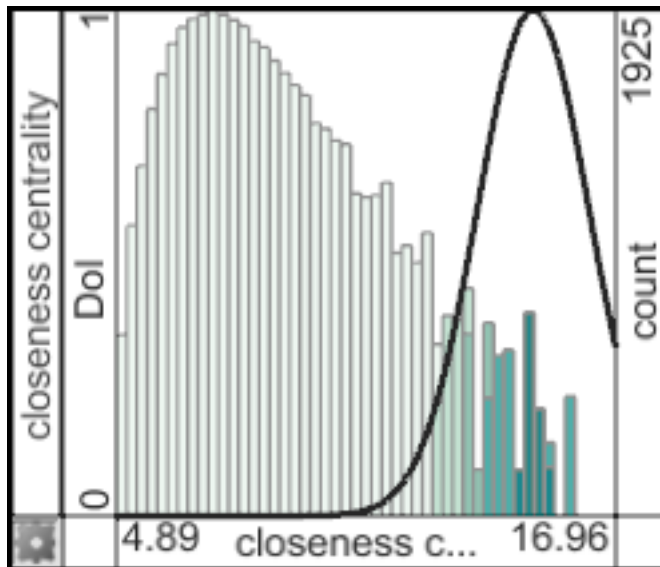
- Unser Ansatz (Hadlak et al. 2014): Flexible Definition interessanter Informationen durch
→ modulare DoI (degree of interest) Spezifikation

4 Schritte:

1. Definition von Komponenten
2. Transformation von Komponenten
3. Kombination von Komponenten
4. Propagation von Komponenten

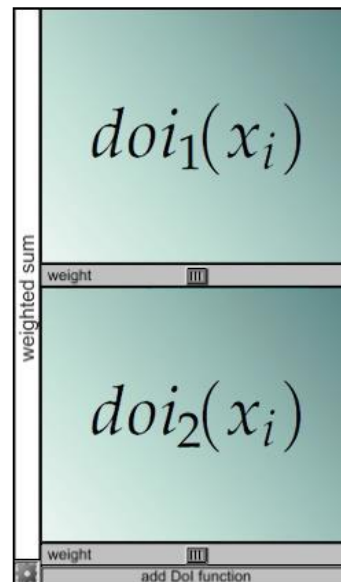
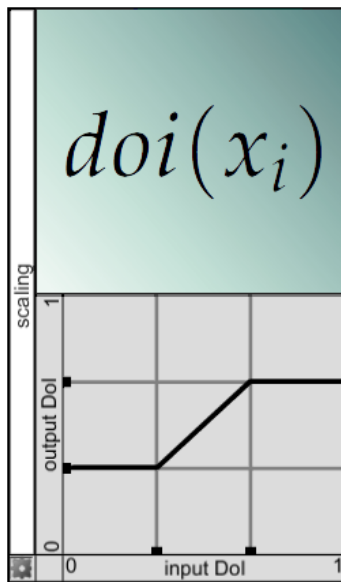
... Daten + Struktur (3)

- Definition von Komponenten
(beschreibt das grundlegende Interesse mit einer Dol-Funktion)



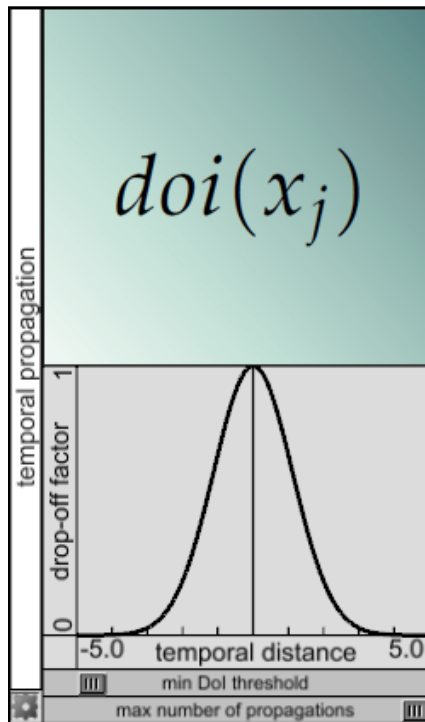
... Daten + Struktur (4)

- Transformation von Komponenten
(modifiziert eine DoI-Funktion, z.B. zur Inversion oder Hervorhebung)
- Kombination von Komponenten
(erlaubt verschiedene Aspekte gleichzeitig zu betrachten)



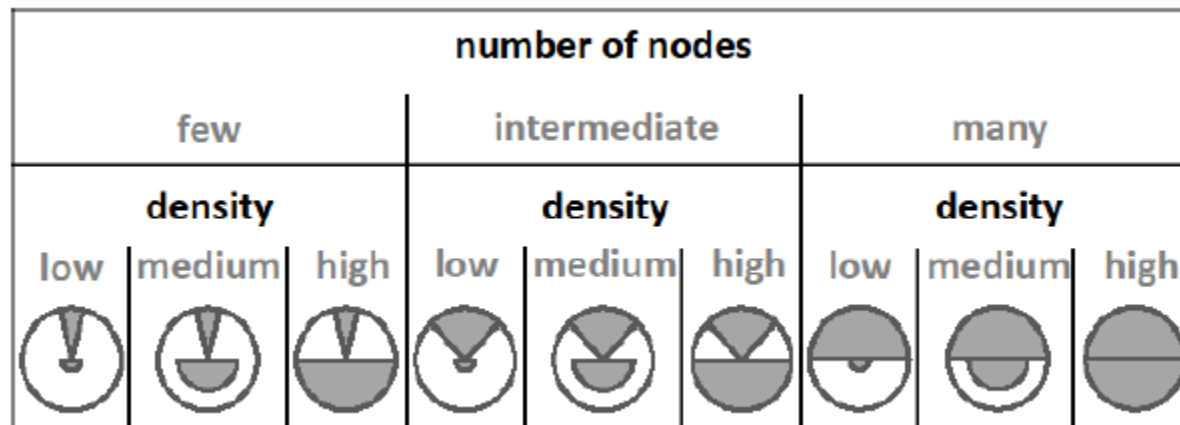
... Daten + Struktur (5)

- Propagation von Komponenten
(propagiert hohe DoI-Werte über Graphenelemente und Zeitpunkte)



... Daten + Struktur (6)

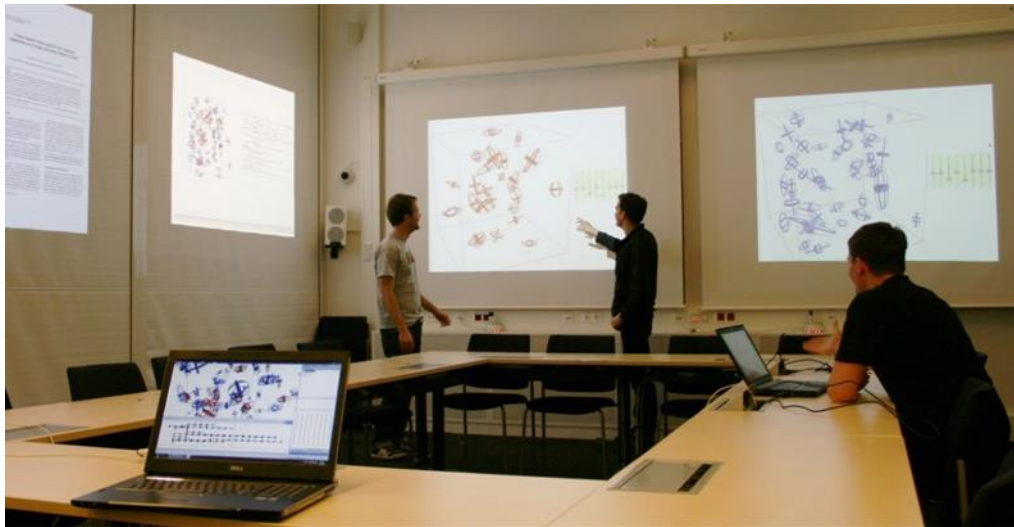
- Visualisierung des Graphen
 - Graphenelemente mit hohen Dol-Werten werden explizit dargestellt (kleine Kreise mit farb-kodierten Dol-Werten)
 - Teilgraphen mit niedrigen Dol-Werten werden in Metaknoten zusammengefasst.



Präsentation & Diskussion

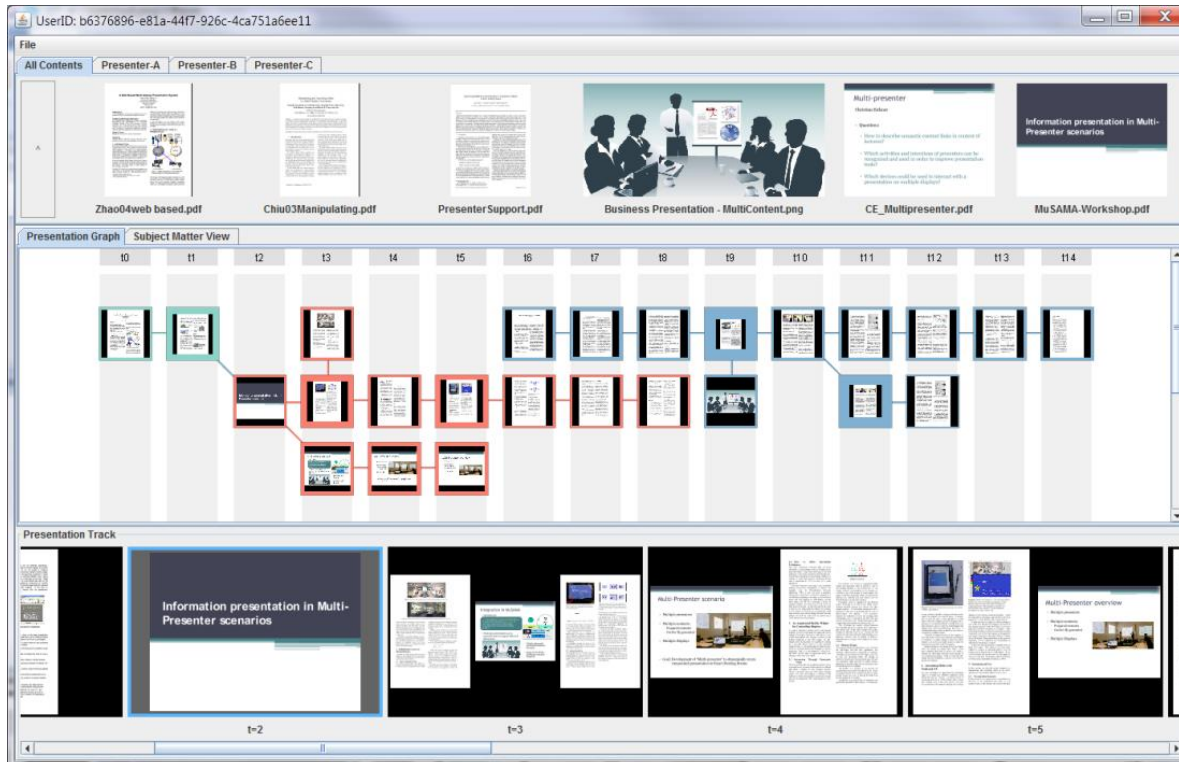
- Multi-user und multi-display-Umgebungen erfordern neue Ansätze für die Visualisierung von und Interaktion mit Informationen.

Beispiel: Multipresenter-Tool, Eichner et al., 2015



Präsentation & Diskussion (2)

- Einfache Komposition von Präsentationen



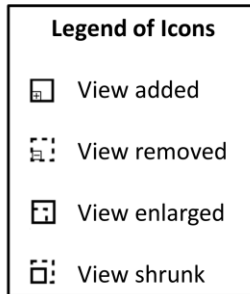
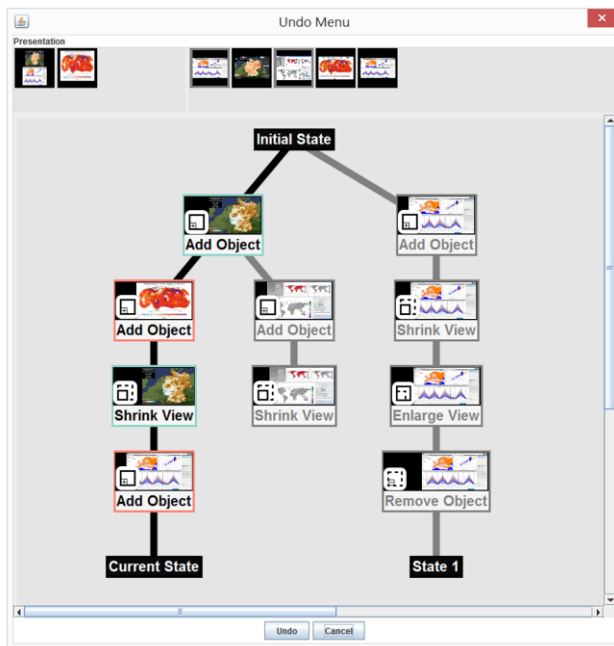
Content Pool

Presentation
Graph

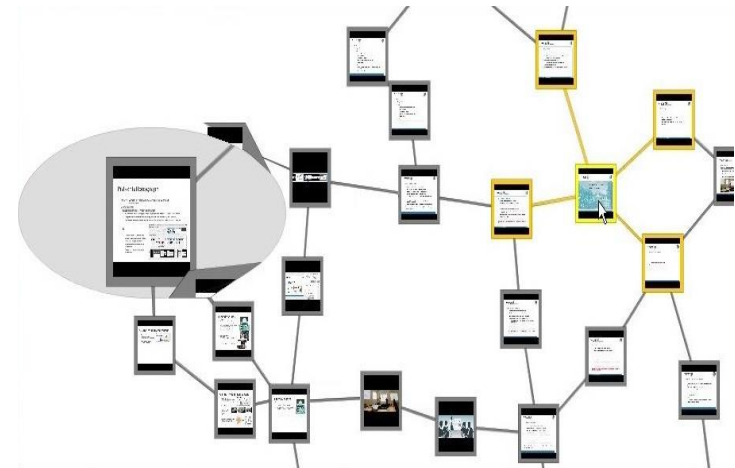
Time Bar

Präsentation & Diskussion (3)

- Protokollierung von Aktionen pro User, Display und Objekt zur Unterstützung von Diskussionen



Interaktion
Graph:
Selektives Undo
& Redo



Semantic Graph:
Suche nach ergänzenden Inhalten



Visualisierung der Daten (3)

- Berechnen der Heterogenitätswerte
 - Bestimme die Punkte, für die Werte berechnet werden sollen.
 - Berechne die Heterogenität an bzw. zwischen den Punkten entsprechend einer gegebenen Metrik
 - Aggregiere die Heterogenitätswerte

