

## 12. Visualisierung von Tensor- und hochdimensionalen Daten, Informationsvisualisierung

### 12.1. Tensor- und hochdimensionale Daten

#### Tensor- und hochdimensionale Daten:

Erweiterung des Konzepts von Skalaren und Vektoren

- Skalare haben Ordnung 0 (kein Index)
- Vektoren haben Ordnung 1 (ein Index)
- Matrizen haben Ordnung 2 (zwei Indizes)
- Beispiele:
  - Diffusionstensor (Medizin, siehe später)
  - Materialeigenschaften (Werkstoffwissenschaften):
    - Leitfähigkeitstensor
    - Spannungstensor
    - (Dielektrische) Permittivität (-zahl)
    - (Magnetische) Permeabilität

#### Tensor- und hochdimensionale Daten:

- Tensor der Ordnung  $k$  ist gegeben durch

$$t_{i_1, i_2, \dots, i_k}(x_1, \dots, x_n)$$

$k$  Indizes

$n$ -dim Daten

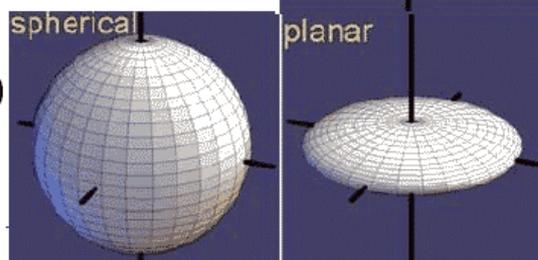
- Schreibweisen für Elemente:  $a_{ijk}$ ,  $a^{ijk}$ ,  $a_i^{jk}$ , ...
- Tensoren 2. Ordnung werden iA. durch **Matrix** repräsentiert

#### Tensor- und hochdimensionale Daten:

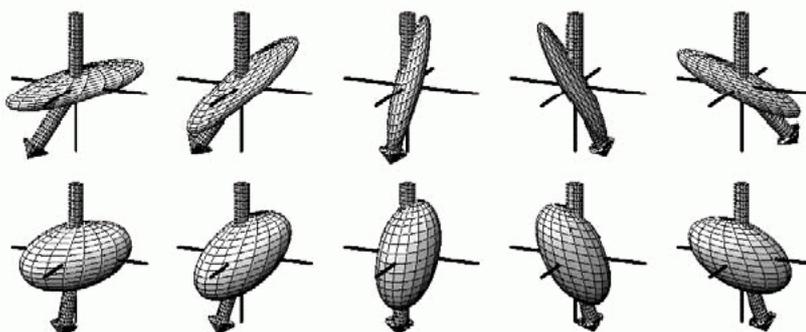
- Nicht symmetrische Tensoren:
  - Eigenvektoren nicht unbedingt orthogonal
  - Eigenwerte können komplex sein
- Deshalb werden Tensoren oft in symmetrischen und nichtsymmetrischen Teil zerlegt
- Symmetrischer Teil: Orientierung
- Nichtsymmetrischer Teil: Scherung, etc.
- Physikalische Größen im nichtsymmetrischen Teil

## Typischer Tensor 2. Ordnung: Diffusionstensor

- Diffusion: Basiert auf **Bewegung** von flüssigen Teilchen auf **mikroskopischer Ebene**
- Stochastisches Phänomen
- Basiert auf **Braunsche Bewegung** der Teilchen
- Messung durch moderne (iA. Hochfeld) **MRT-Scanner**
- Diffusionstensor beschreibt **Diffusionsrate** in **verschiedene Richtungen** via symmetrischen Tensor (Wahrscheinlichkeitsdichtenverteilung)
- Anschaulich: repräsentiert Orientierung
- In 3D: Repräsentation durch **3×3 symmetrische Matrix**
- Symmetrische Diffusionsmatrix kann diagonalisiert werden:
  - Reelle Eigenwerte  $\lambda_1 \geq \lambda_2 \geq \lambda_3$
  - Matrix positiv definit:  $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$
  - Eigenvektoren stehen **senkrecht** zueinander
- Isotropie / Anisotropie:
  - Sphärisch:  $\lambda_1 = \lambda_2 = \lambda_3$
  - Linear:  $\lambda_2 \approx \lambda_3 \approx 0$
  - Planar:  $\lambda_1 \approx \lambda_2$  und  $\lambda_3 \approx 0$



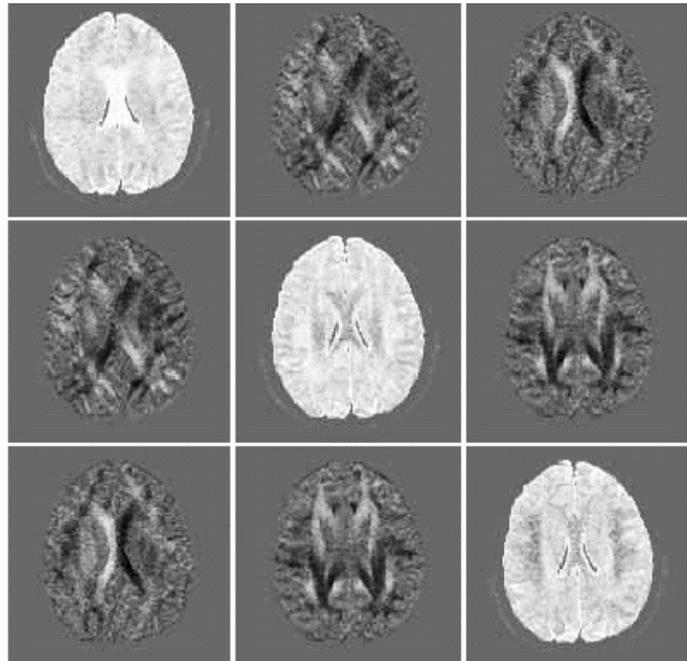
- Beliebige Vektoren sind allgemein nach Matrixmultiplikation abgelenkt
- **Ablenkung** in die Richtung der **Haupteigenvektoren** (größter Eigenwert)



## Visualisierungsansätze:

### Bildmatrix:

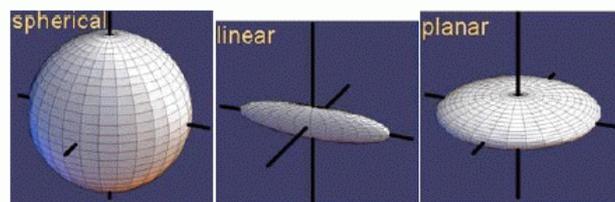
- Schicht durch Volumen
- Jedes Bild zeigt Komponente der Tensormatrix



[Weinstein et al., IEEE Vis 1999]

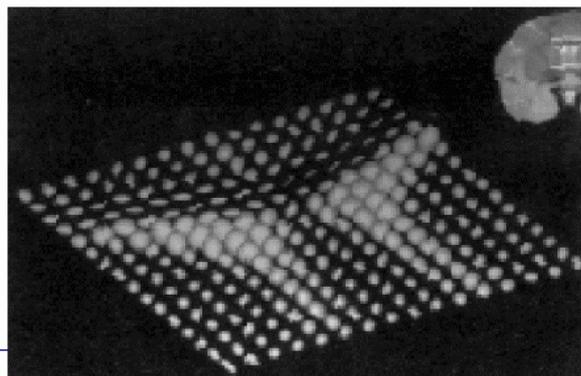
### Symmetrischer Tensor kann diagonalisiert werden

- Repräsentation durch Ellipsoid
- Glyphen-basierter Ansatz



### Uniformes Gitter von Ellipsoiden

- Symmetrischer Tensor 2. Ordnung wird auf Ellipsoid abgebildet
- Schichtdarstellung des Volumen

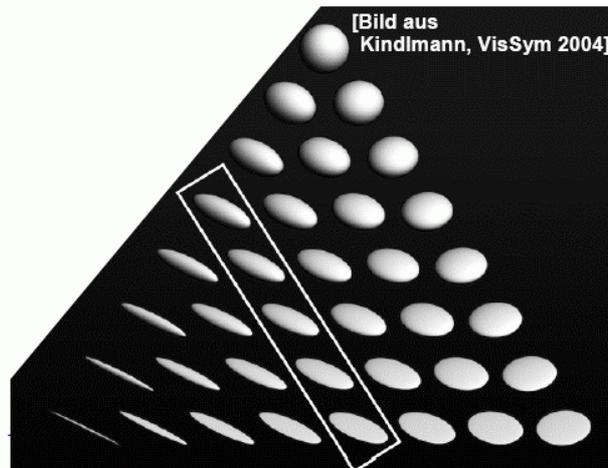


[Pierpaoli et al., MR in Med 1996]

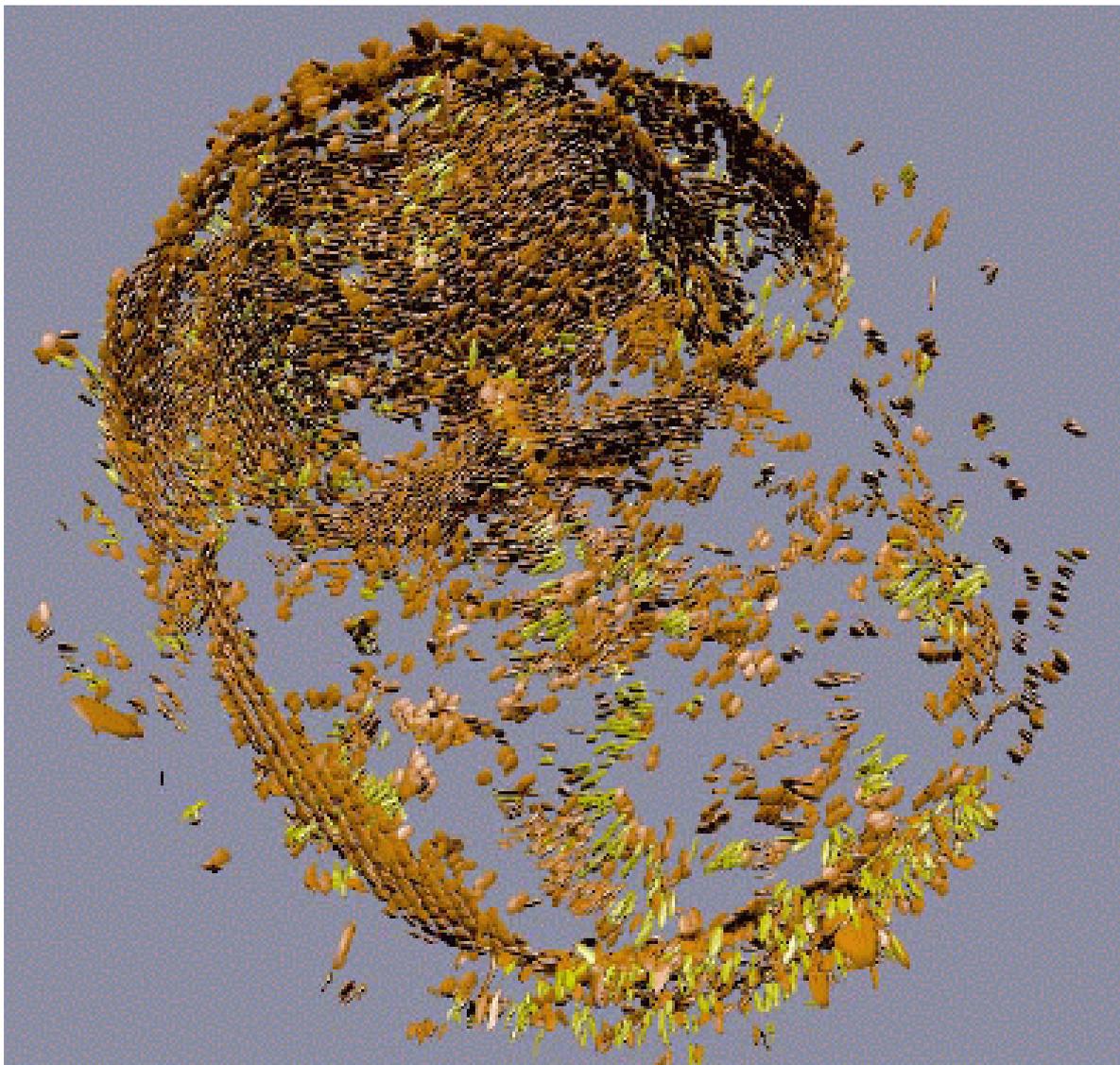
## Ellipsoide in 3D

Probleme:

- Verdeckung
- Fehlende Kontinuität
- Mehrdeutigkeit



*Beispiel:*

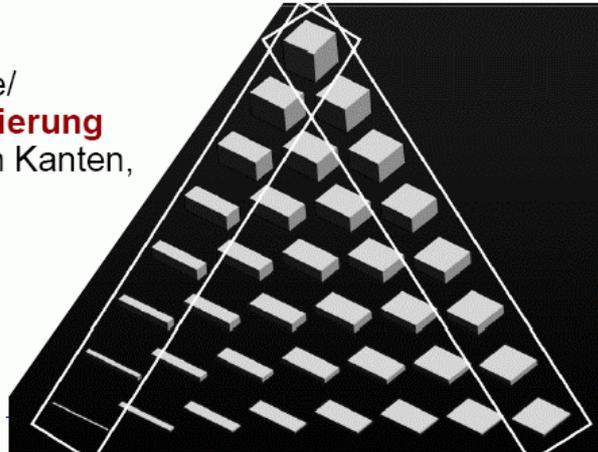


## Cuboid/Box-Glyphen

Probleme:

- Verdeckung
- Fehlende Symmetrie/  
**suggestieren Orientierung**  
durch Ecken auf den Kanten,  
die es **so nicht gibt**

[Bild aus Kindlmann, VisSym 2004]



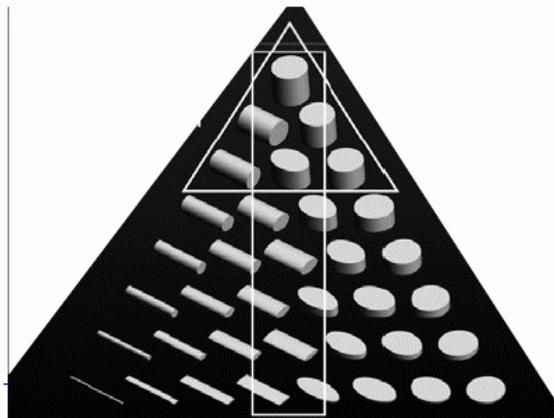
[Worth, et al., TR 1998]

## Zylinder-Glyphen

Probleme:

- Verdeckung
- Unstetige Naht/Saum in der Mitte
- Pseudoorientierung in der Spitze

[Bild aus Kindlmann, VisSym 2004]

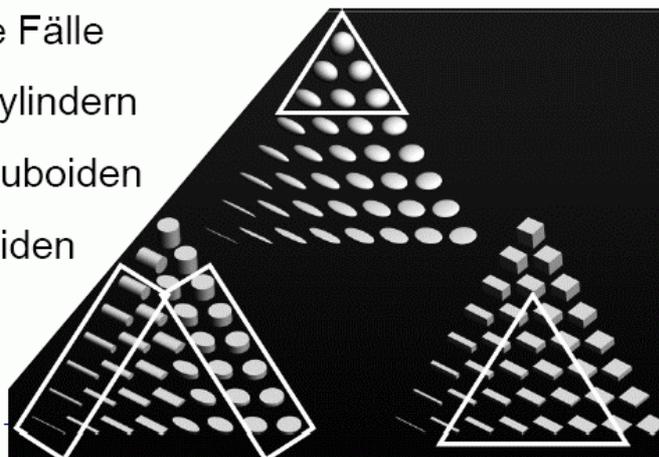


## Superquadratische Glyphen

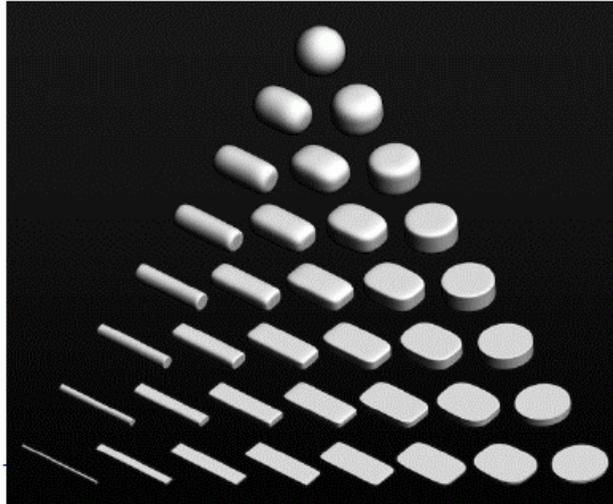
Idee:

- Kombiniere beste Fälle
- Symmetrie von Zylindern
- Naht/Saum mit Cuboiden
- Spitze mit Ellipsoiden

[Kindlmann, VisSym 2004]

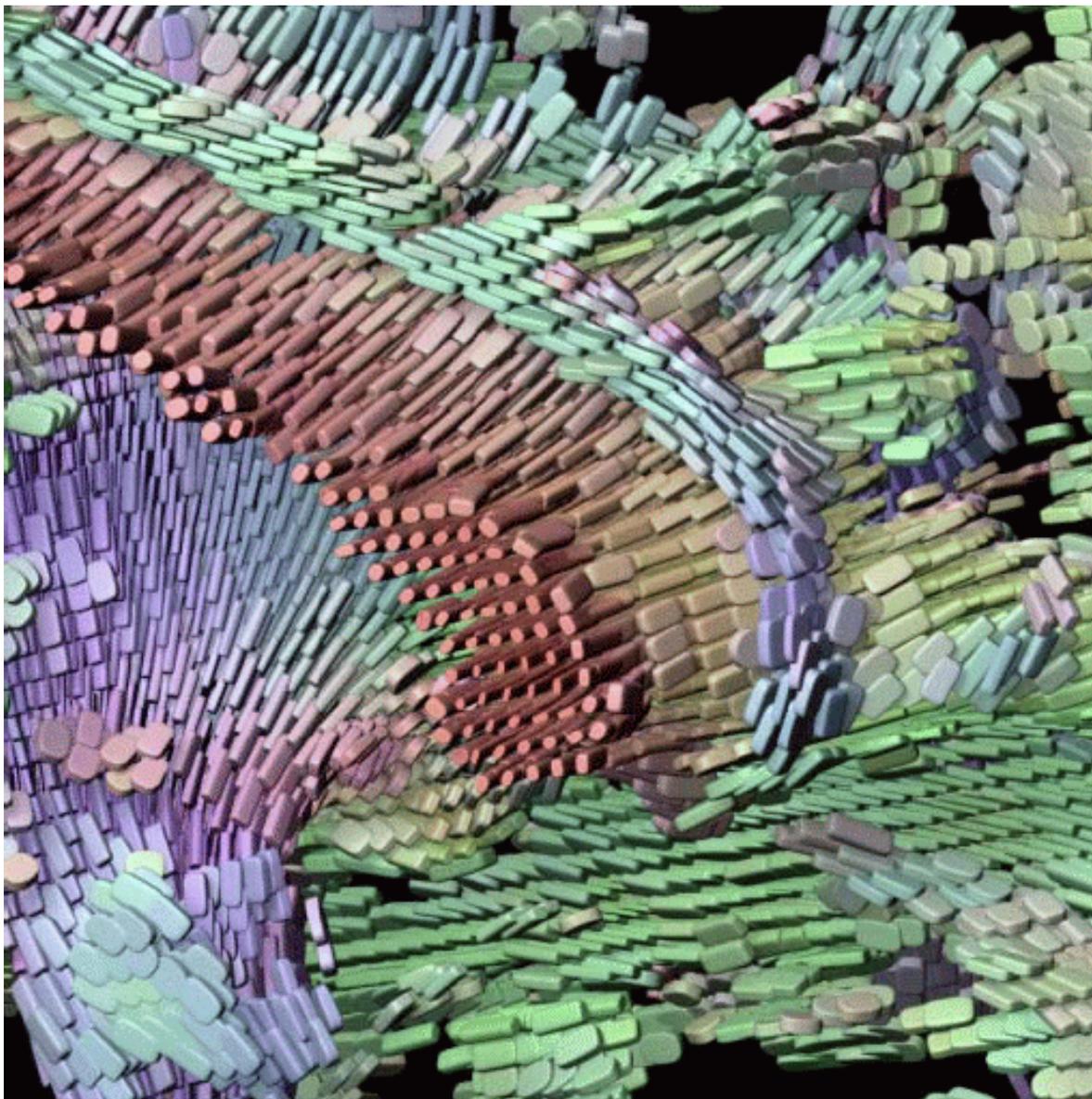


- Kombiniere beste Fälle
- Kantenschärfe wird mit Parameter  $\gamma$  kontrolliert



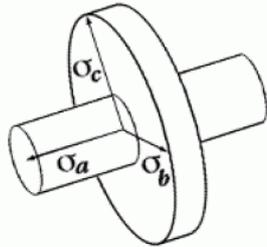
[Kindlmann, VisSym 2004]

*Beispiel:*

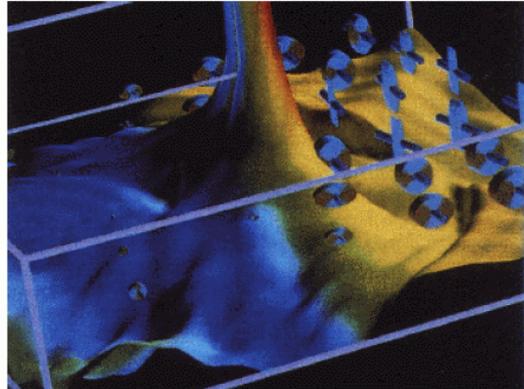


## Haber-Glyphen

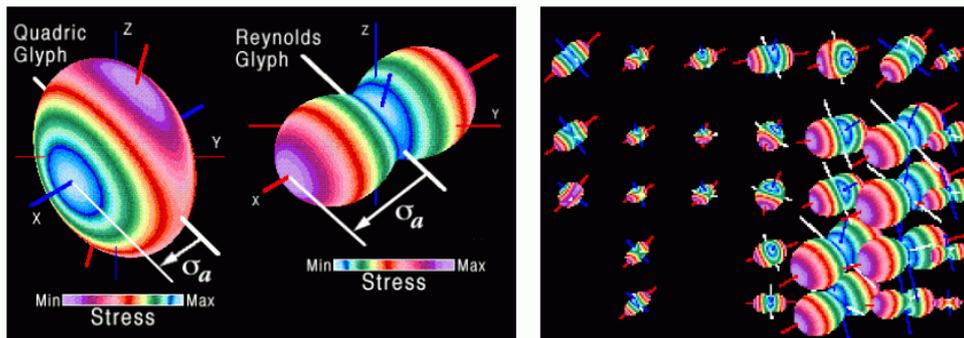
- Stab und elliptische Scheibe
- Besser geeignet für die Darstellung der Größe und der Hauptachse von Tensoren



[Haber, Computing Sys. Eng., 1990]

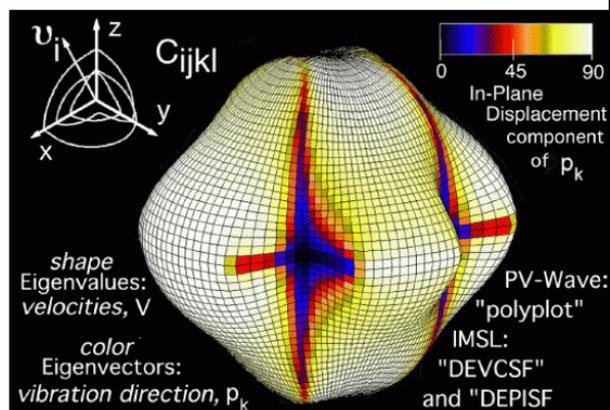


## Reynolds-Glyphen



[Moore et al., ASME Gas Turbine Conf., 1994]

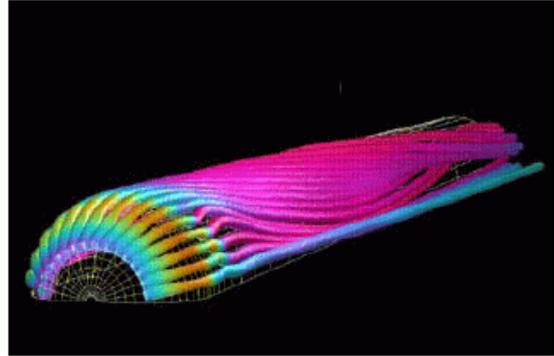
## Glyphen für Tensoren 4. Ordnung (Wellenpropagierung in Kristallen)



[Ledbetter/Kriz, Physica Status Solidi, 1982]

## Hyperstreamlines

- Streamlines werden durch Eigenvektoren definiert
- Richtung der Streamline durch Haupteigenvektor
- Visualisierung des Vektorfeldes, dass durch den Haupteigenvektor definiert ist
- Andere Eigenvektoren definieren Querschnitt



[Delmarcelle/Hesselink, IEEE Vis 1992/93]

(Bartz 2005)

- auch andere Techniken der 3D-Strömungsvisualisierung auf Haupteigenvektor-Feld des Tensors anwendbar

## 12.2. Hochdimensionale Daten und Informationsvisualisierung

Bisher: Voraussetzung, dass die Daten auf einem Gitter in 2D oder 3D gegeben sind

in der Praxis aber häufig auch die Situation, dass

- der Beobachtungsraum mehr als 3 Dimensionen hat
- oder dass er überhaupt nicht algebraisch mit einer Dimension versehen werden kann

Beispiele:

- Bioinformatik
  - Sequenzdaten (Genomräume), hochdimensional
  - Gen-Aktivierungsmuster (Microarrays)
- Datenbanken, wo zahlreiche Merkmale gespeichert sind (z.B. Personendaten, Kauf-Transaktionen)
- Textinformationen (nicht strukturiert)
- Hypertext-Daten (z.B. das WWW)

Man spricht auch von "Informationsvisualisierung"

Ziele:

- **Datenexploration** mit geringen Vorkenntnissen ermöglichen
- **Identifizieren** von Mustern, Strukturen, Trends, Beziehungen, Abnormalitäten
- **Entdecken** von neuen Informationen

→ visuelle Variante des "Data Mining" (Extraktion von Wissen aus großen Datenbanken; vgl. entsprechende Vorlesung "Grundlagen des Data Mining")

## Data Mining:

### Hauptziele:

- Zusammenfassen
- Verknüpfen
- Klassifizieren
- Vorhersagen
- Gruppieren
- Analysen des zeitlichen Verhaltens

verwendet →

### Wichtige Techniken:

- Lineare Regression
- Nicht-lineare Regression
- Naïve Bayes
- K-Mittel, K-Median
- Neuronale Netze/Heuristische Methoden
- Verknüpfungsregeln
- Entscheidungsbäume
- Hauptkomponentenanalyse (PCA)
- Dimensionsreduktion

### Statistische Methoden:

- Normalisierung
- Fehler- und Abweichungsanalyse
- Stat. Sicherheit

← beruht auf

## Visualisierung:

- **Interaktive** Exploration von hochdimensionalen Daten
- **Flexible** Exploration
- **Visuelle** Exploration
- Bestätigende Analyse (einer Hypothese)
- Explorierende Analyse (finden einer Hypothese)

### Vis' und Analyse:

- Dimensional Stacking
- PCA und Principal Curve Analyse
- Selbstorganisierende Abbildungen/Karten

### Interaktionen:

- Auswahl
- Testen/probing
- Übersicht (Grand-Tours)
- Nicht-lineares Zoomen

### Visualisierungstechniken:

- **Histogramme**
- **2D und 3D Scatterplots**
- **Scatterplot-Matrix**
- Statistische Grafiken
- Graphen
- **Parallele Koordinaten**
- Kreissegmente
- Polargrafiken
- Übersichtsabbildungen
- Heatmaps
- **Höhenfelder**
- **Piktogrammschaubilder**
- ...

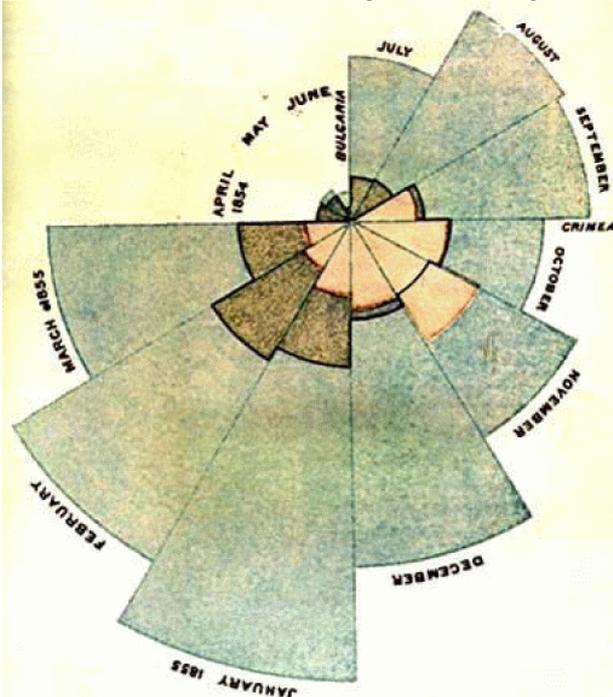
Visualisierungstechniken:

z.T. schon bekannt

- Histogramme (auch 3D, aber mit Nachteilen)
  - Auswahl der Merkmale zu treffen

- **Polardiagramme**

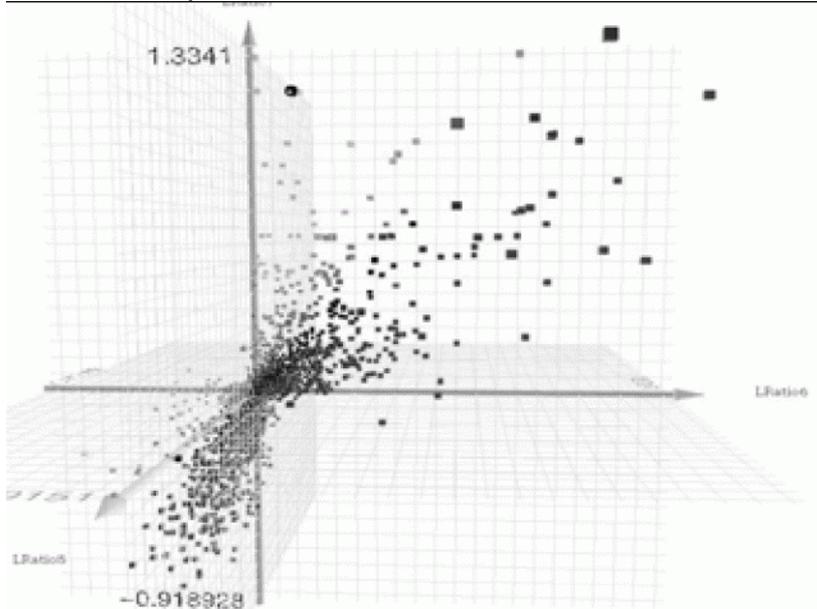
historisches Beispiel: Napoleons Russlandfeldzug



(F. Nightingale, in Bartz 2005)

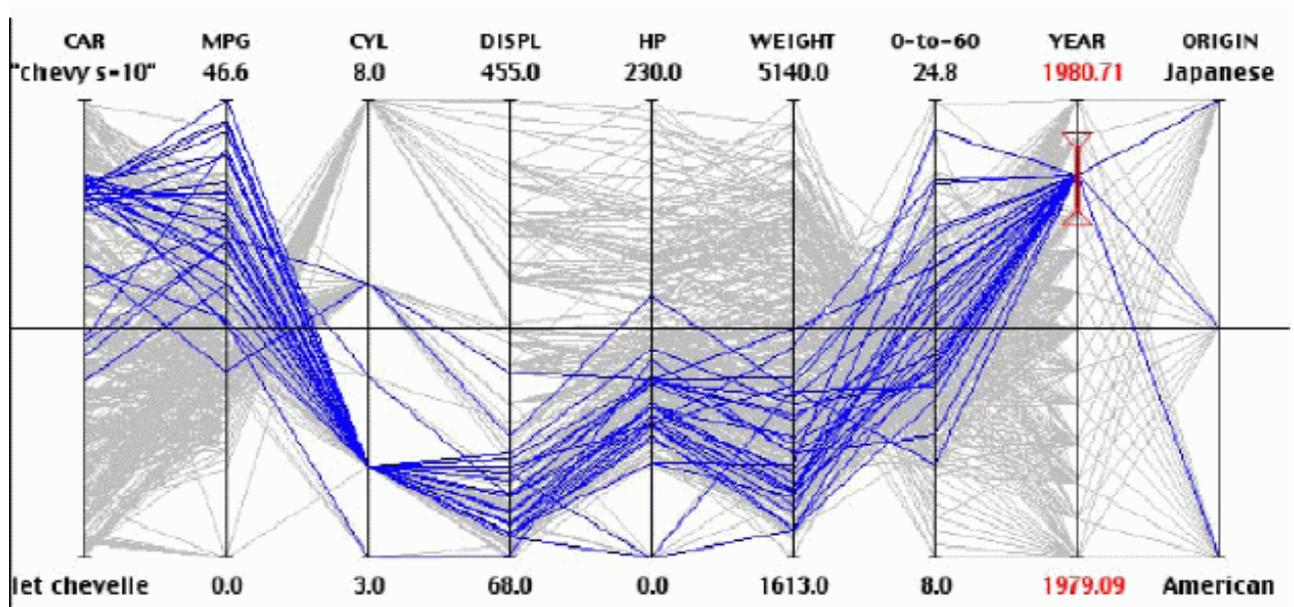
- Scatterplots (für visuelle Korrelationsanalyse)

3D-Scatterplot





Beispiel:



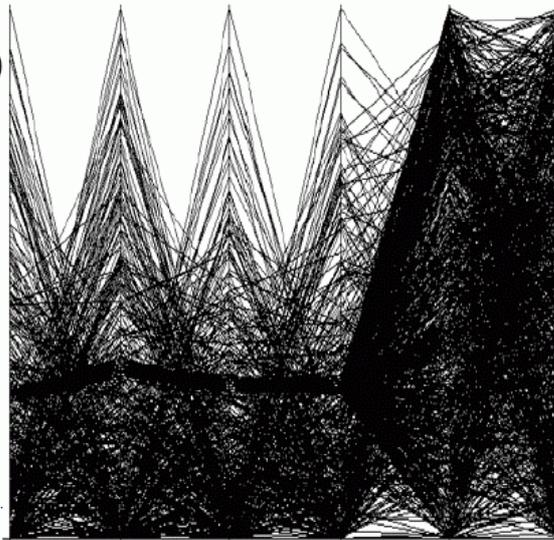
Beispiel, wo Methode fehlschlägt:

- 15K Datenelemente
- Mit Rauschen

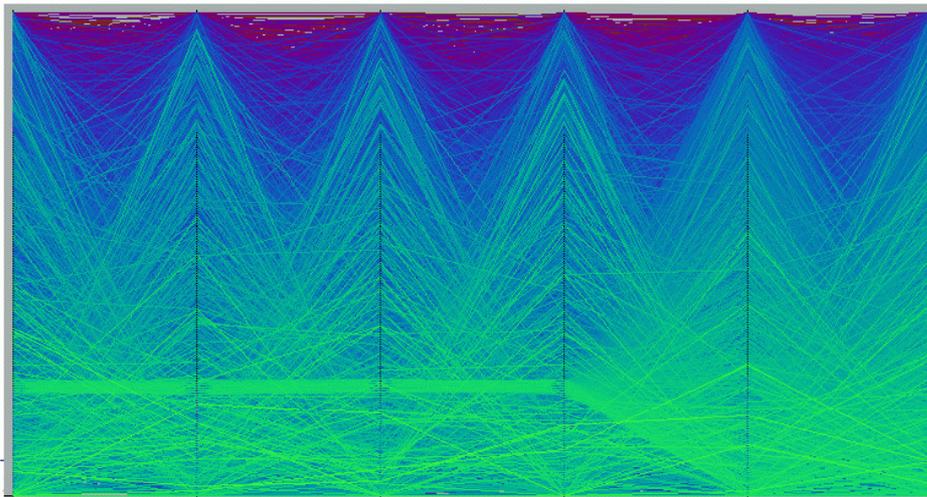


## Abhilfe?

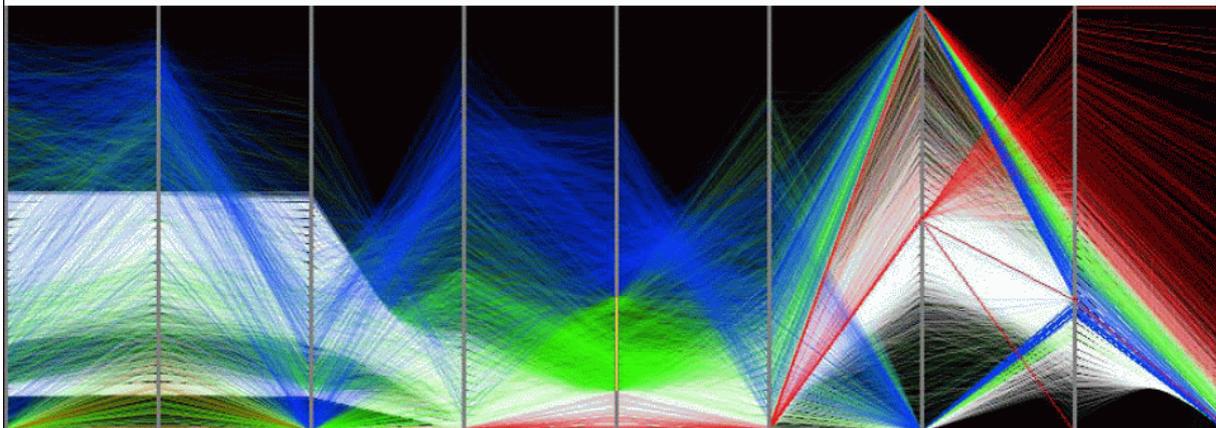
- 5% der Daten (750)



- 15K Datenelemente und Anwendung einer Farbabbildung

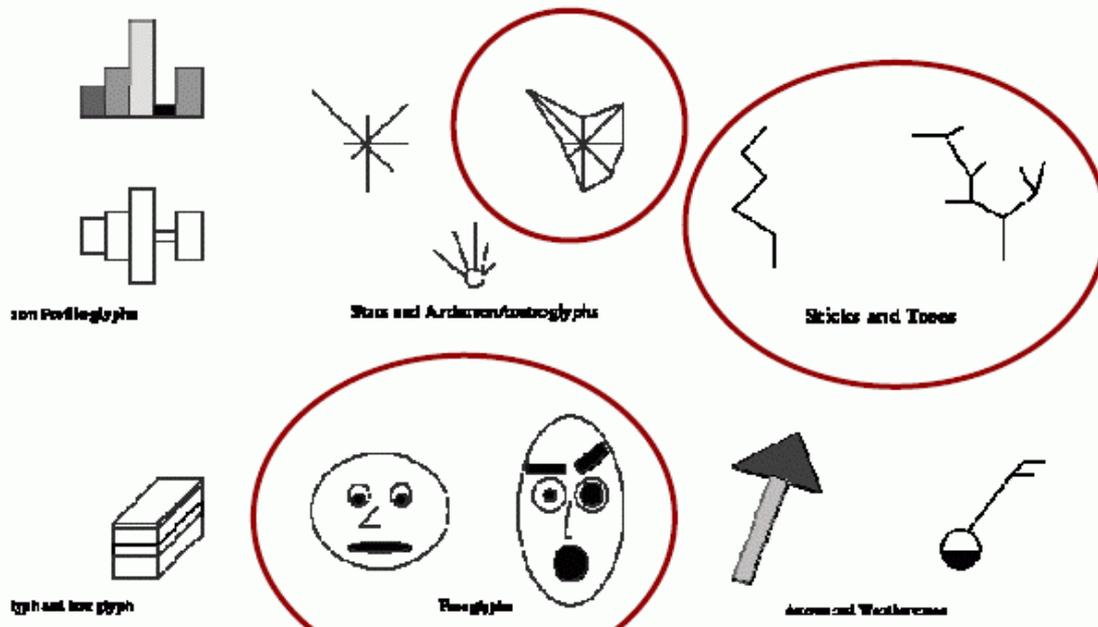


- 66K Datenelemente und Anwendung einer **Opazitätsmodulierung** und einer **Farbabbildung**

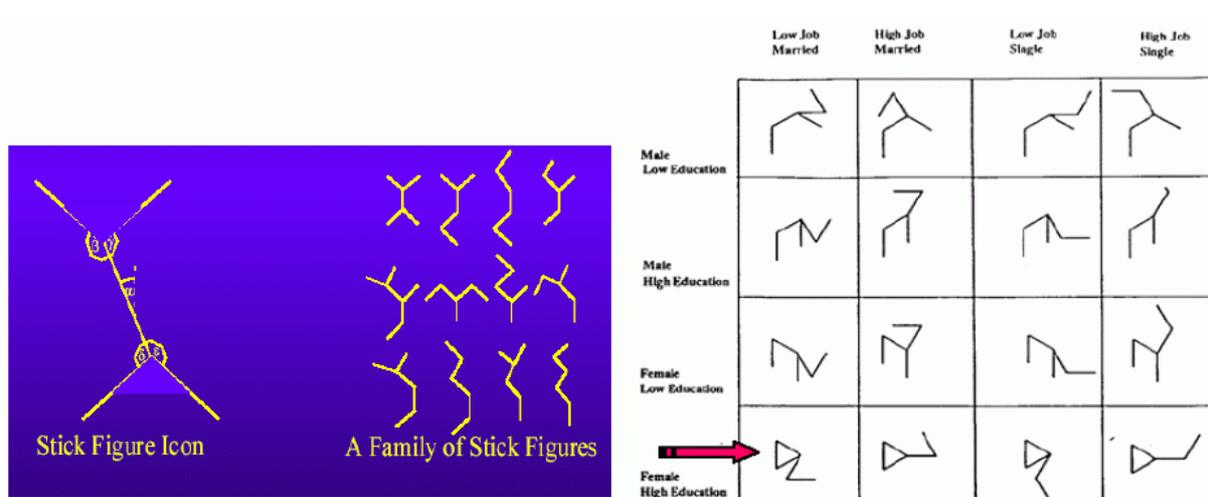


## Piktogrammdiagramme (Icons):

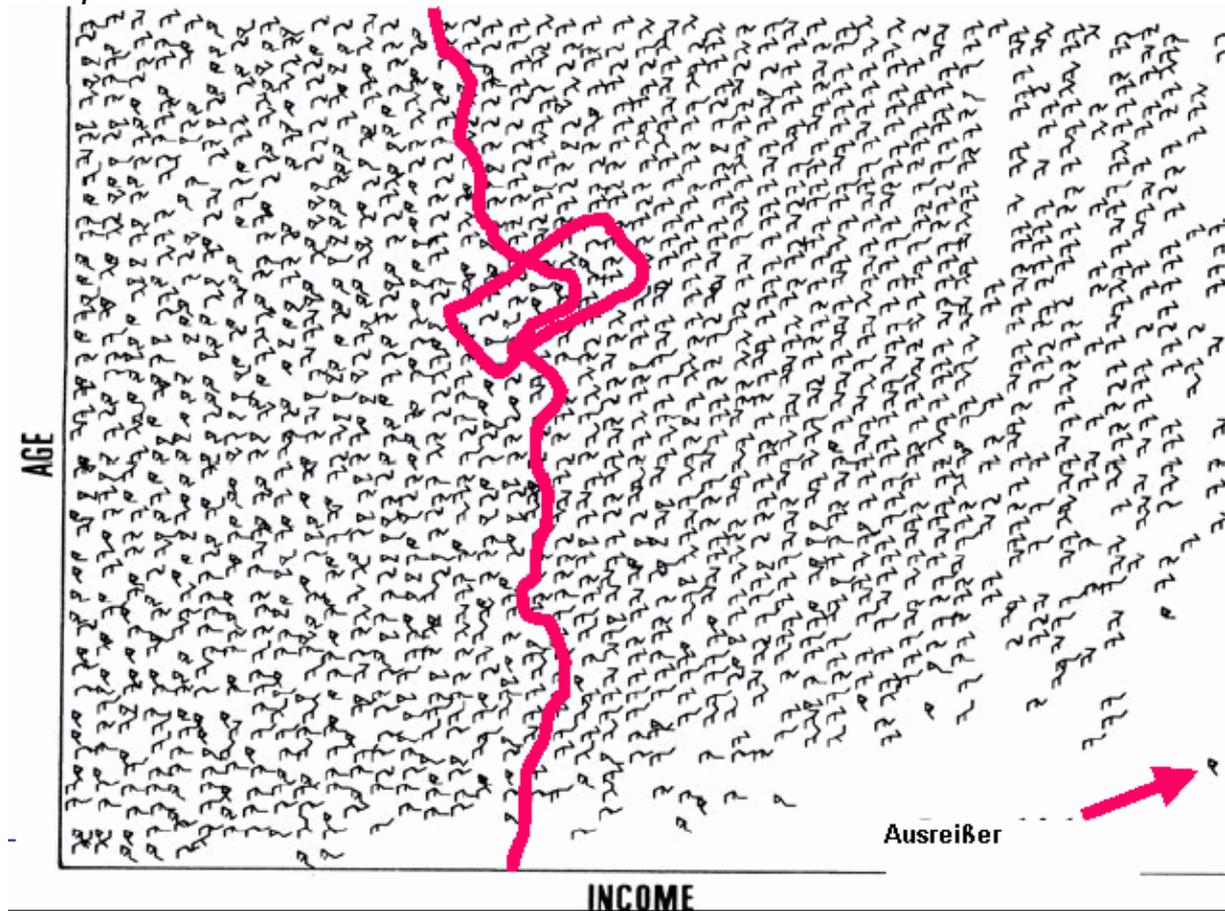
- Piktogrammeigenschaften werden Dimensionen zugeordnet
- Bei vielen Datenwerte schnell unlesbar
- Problem auch bei vielen Dimensionen



- Stockfiguren (Stick Figures) – Attribute werden mit Kanten und deren Abwinkelung dargestellt
- Strukturen können auch in großen Mengen sichtbar sein (können!)



Beispiel:

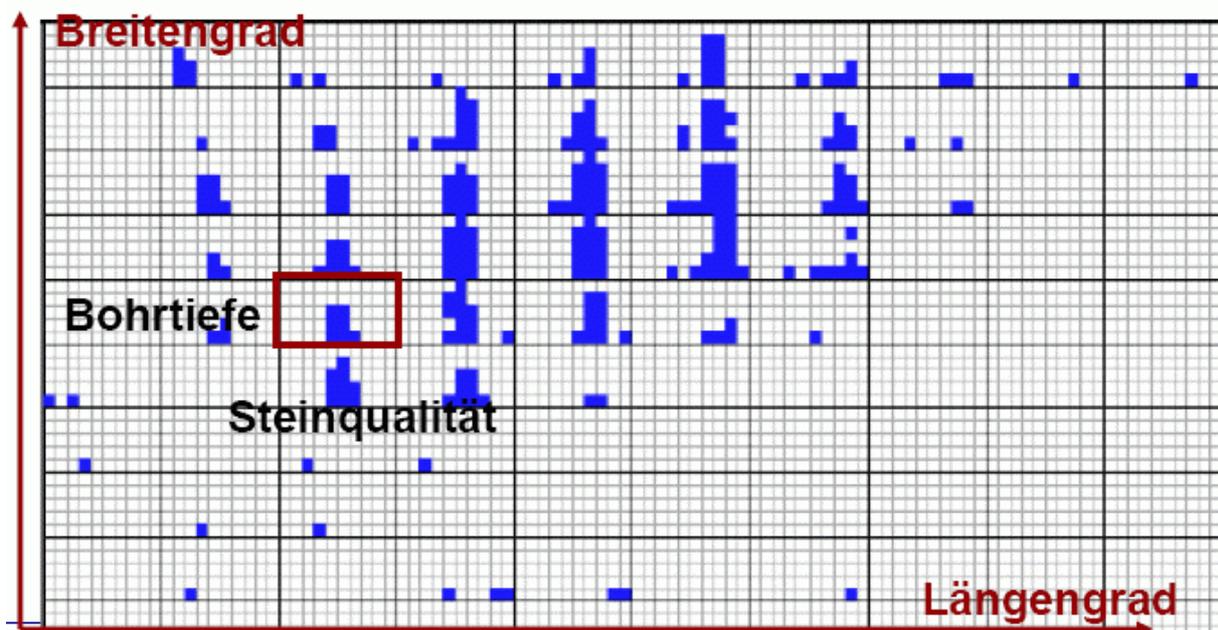


- Hochdimensionale Daten sind sehr komplex
- Für eine „verstehbare“ Darstellung müssen Daten aufbereitet werden durch
  - Gruppierung (Clustering)
  - Klassifikation
  - Dimensionsreduktion
  - Hierarchisierung
- Hierzu werden oft statistische Methoden verwendet:
  - Principal Component Analysis (PCA)
  - Distanzmetriken
  - Regression
  - ...

## Dimensional Stacking:

- Unterteilung in Untergruppen für Dimensionen
- Auf den Hauptachsen die „Hauptdimensionen“
- In Teilabschnitten die „Nebendimensionen“
- Gut bei einer geringen Kardinalität (Anzahl von Datenpunkten)

*Beispiel:*



## Geschachtelte Darstellungen - Treemaps:

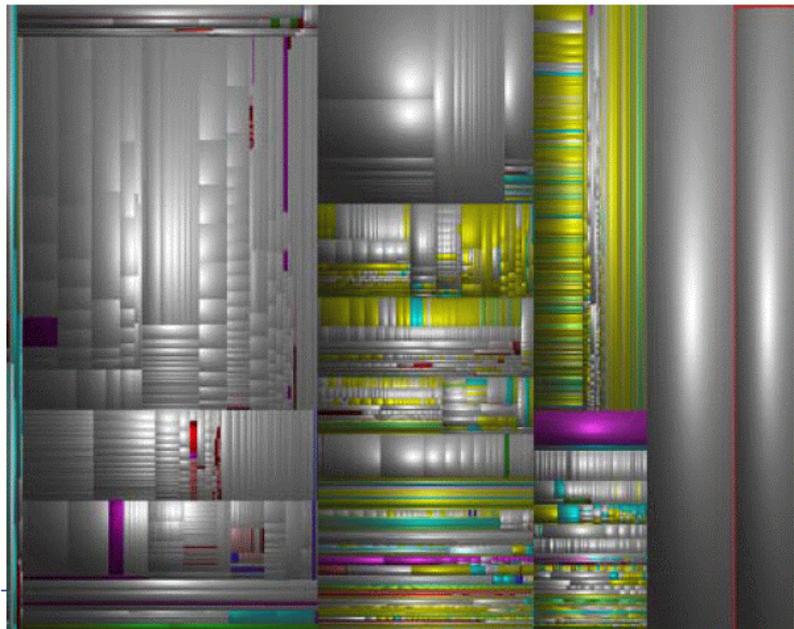
- Rekursive Unterteilung entlang einer Baumstruktur
- Wechsel der Orientierung von Waagrecht und Senkrecht
- Flächengröße entspricht Attributgröße

*Beispiel:*

- Dateibaum

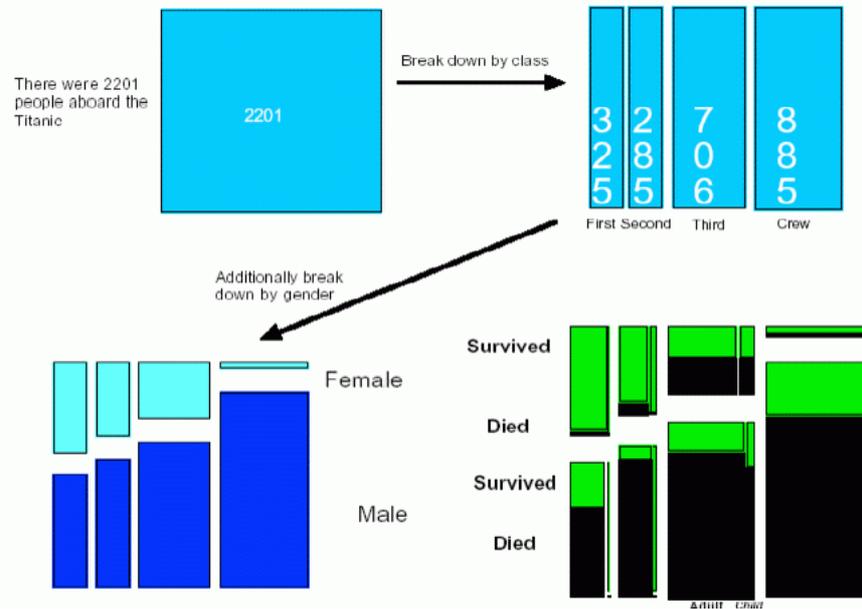


- Dateibaum
- Kontrast-  
verstärkung  
durch  
Beleuchtung



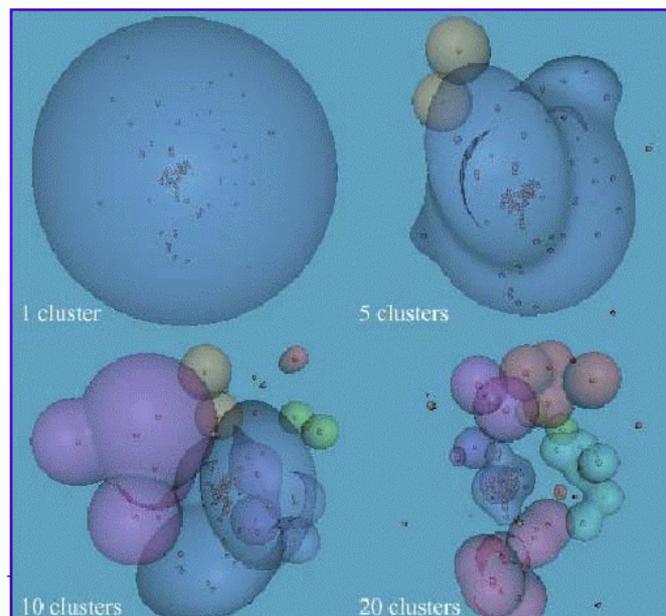
# Geschachtelte Darstellungen - Mosaics:

- Titanic



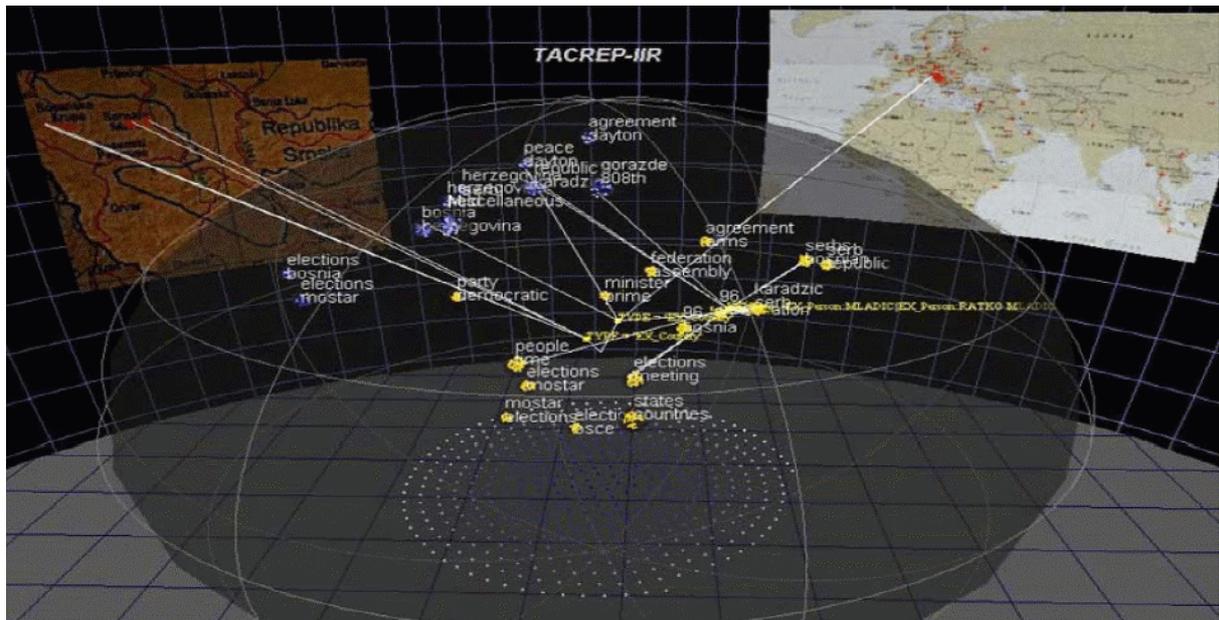
# Darstellungen von Gruppen – H-Blobs:

- Hierarchische Blobs mit 1, 5, 10 und 20 Gruppen (Cluster)



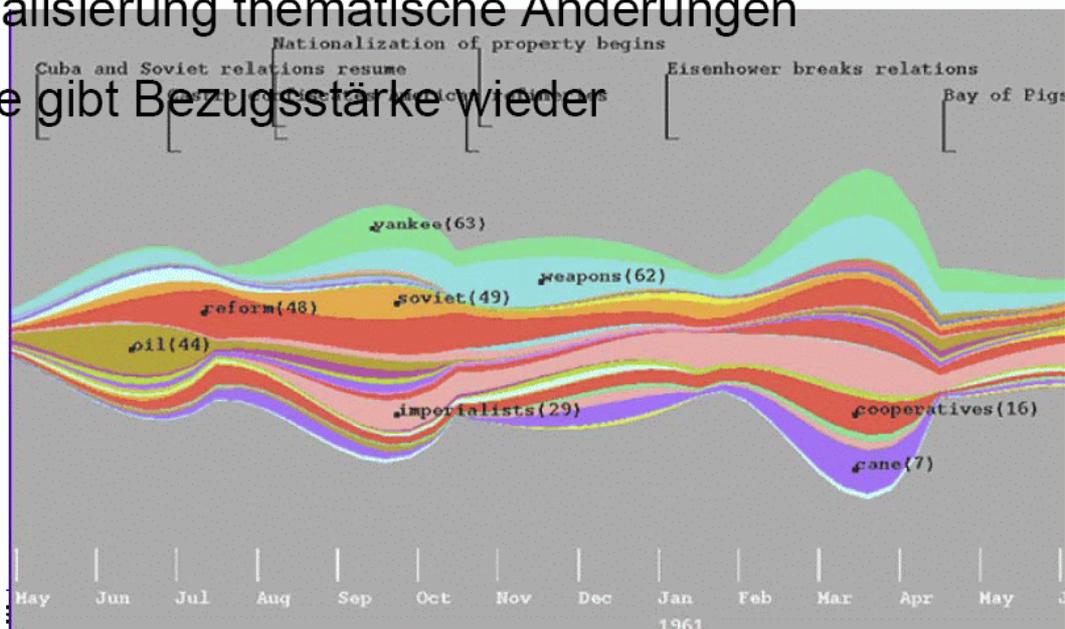
## Text-Mining:

- Fasse Dokumentenstruktur zusammen
- Analyse Inhalt (wie?)
- Klassifikation und Gruppierung



## Themenfluss – Theme-River:

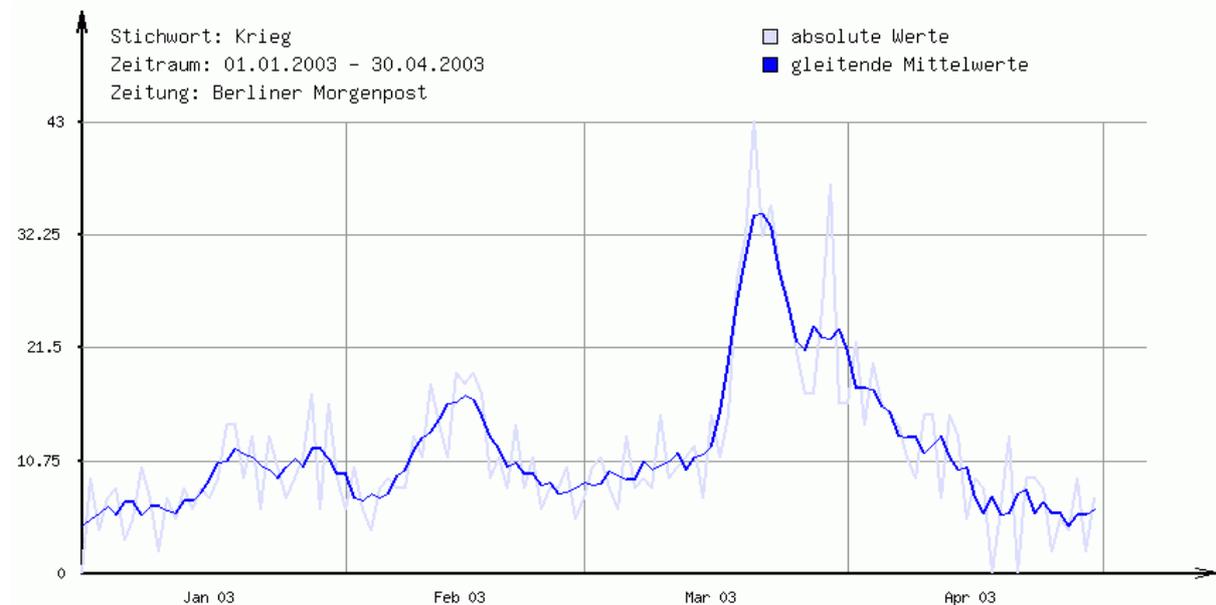
- Visualisierung thematische Änderungen
- Höhe gibt Bezugsstärke wieder



(Bartz 2005)

weiteres Beispiel (Software-Werkzeug aus einer Bachelorarbeit am LS Grafische Systeme – H. Würke):

Chart-Anzeige der Dokumentenhäufigkeit für Zeitungsartikel, die ein Stichwort enthalten, aus WWW-Archiven

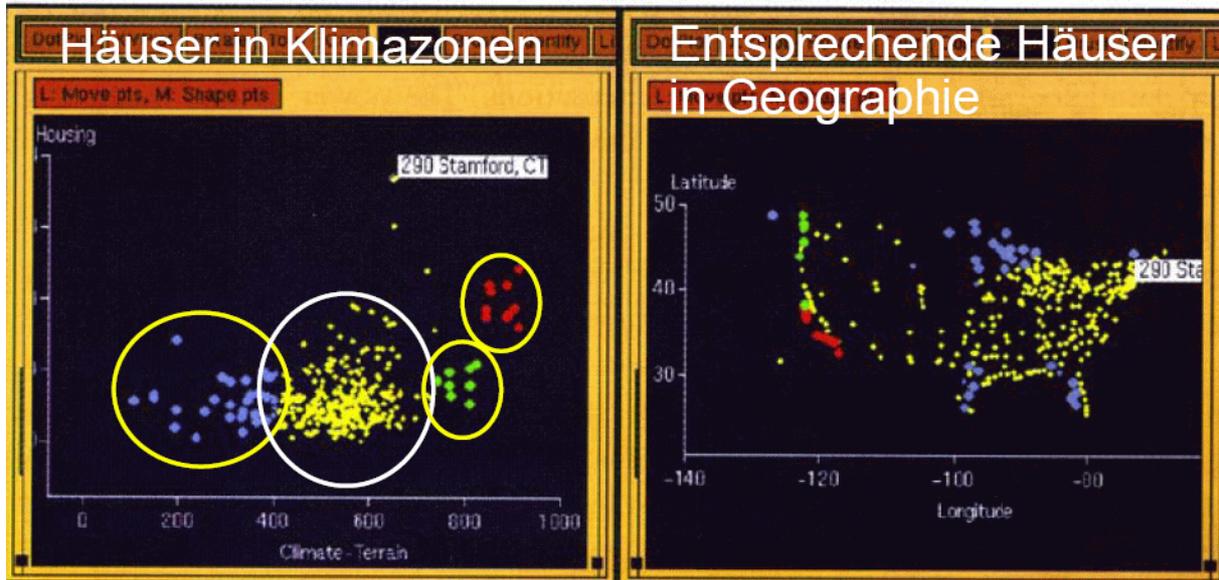


## Interaktionstechniken

- **Brushing**: Markieren von Datenwerten in einer Darstellung und Übertragung in andere Darstellungen
- **Zooming**: Vergrößern von bestimmten Bereichen, die von Interesse sind
- **Filtern**: Mit logischen Operationen Menge definieren/einschränken
- **Projektion**: Interaktive und manuelle Dimensionsreduktion

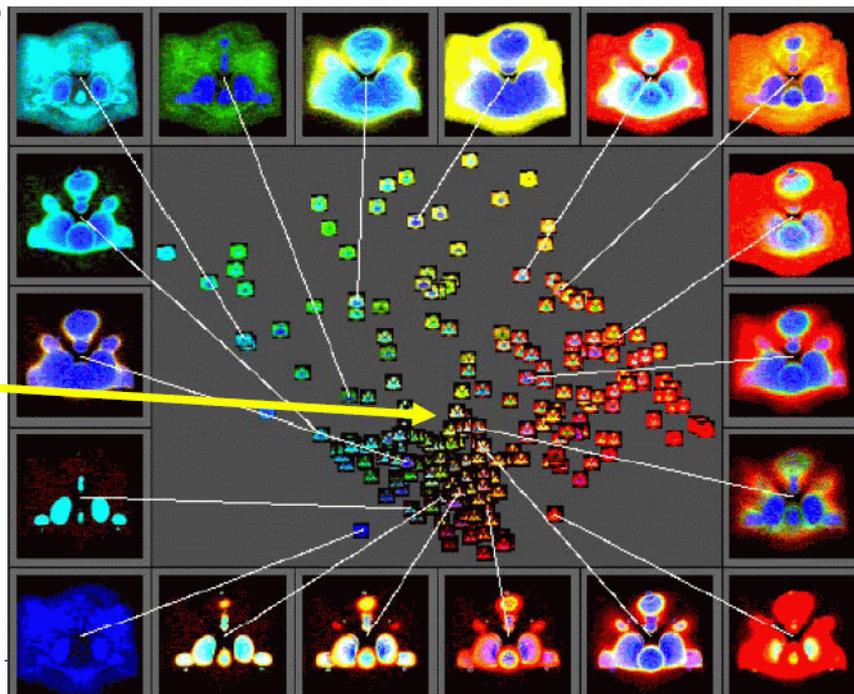
Beispiele:

- **Brushing** und **Linking**



- **Interaktives Zoomen**

Federmodell  
des Ergebnis-  
raums

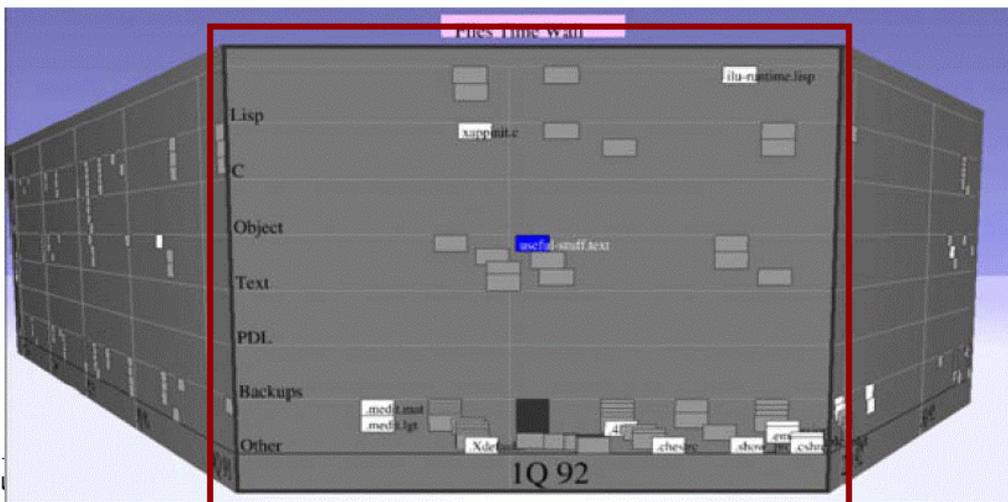


## Verzerrungstechniken (Distorsion):

- Fokus + Kontext
- Ungleichmäßiges Zoomen/Transformieren
- Hyperbolische Darstellungen

## Fokus + Kontext:

- Perspektivische Wand

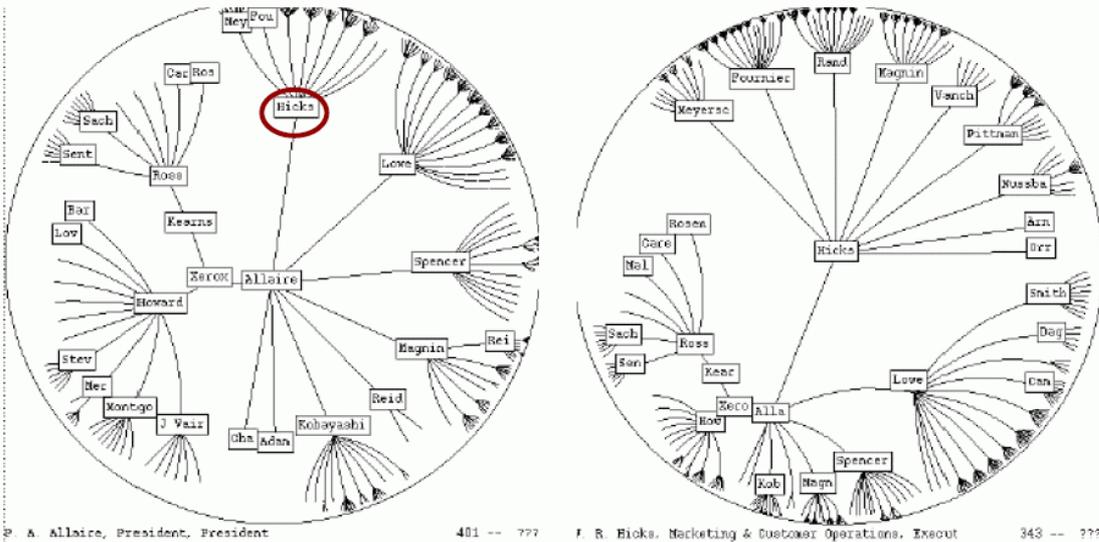


## Fokus + Kontext:

- Blurring



# Hyperbolische Bäume:

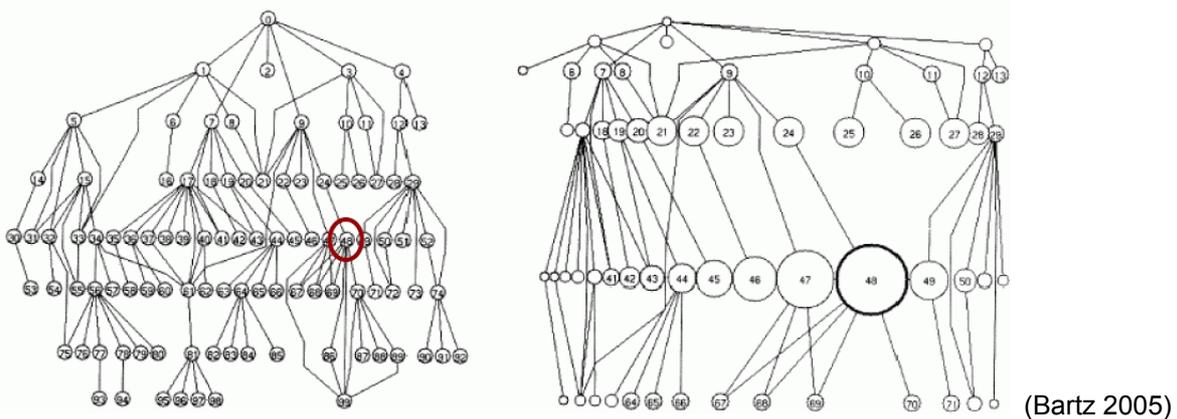


# Ungleichmäßiges Zoomen:

- Tabellenlupe - TableLens



- FishEye



## 12.3. Graphen-Visualisierung

Spezialfall der Informationsvisualisierung

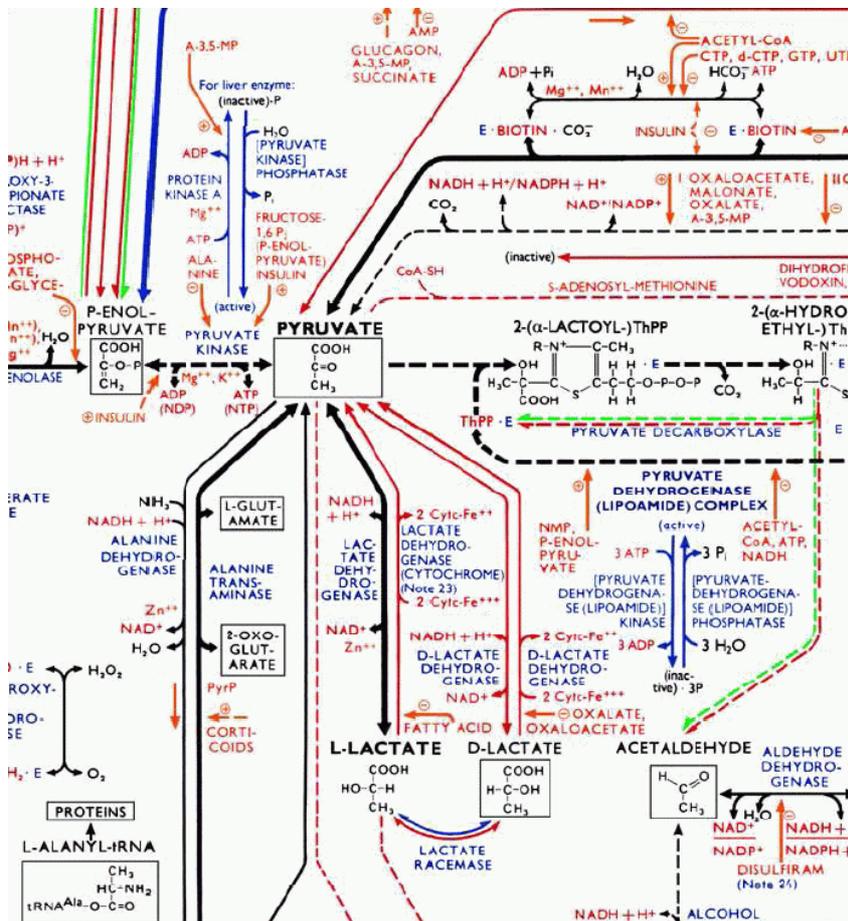
Darstellung von großen Graphen

- Graphen: hier Inzidenzstrukturen aus Knoten und Kanten
- Knoten: oft mit Bezeichnungen (Labels) und evtl. Eigenschaften (Attribute)
- Kanten: gerichtet oder ungerichtet, mit Eigenschaften

Ziele:

- im Graphen vorkommende Strukturen erkennbar machen
- interessante Charakteristika der Daten aufzeigen
- Ablenkung vermeiden, Irrelevantes ausblenden
- Längen, Abstände, evtl. Orientierungen (von Kanten): können Informationen codieren

klassisches Beispiel: Metabolic Pathways (Boehringer)



gesucht: *Graph-Layout* (Abbildung des Graphen in 2D)

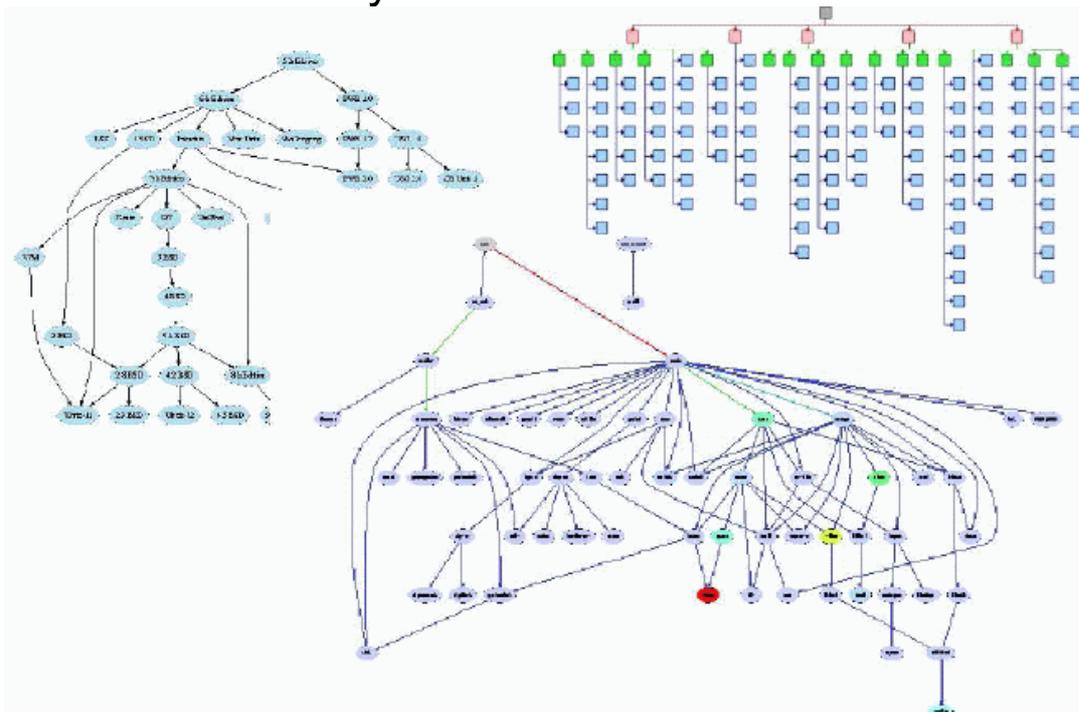
Anforderungen an Layout-Algorithmen:

- benachbarte Knoten sollten nah beieinander liegen
- unverbundene Knoten sollten auseinander liegen
- Vermeiden von Überschneidungen
- gleichmäßige Flächenfüllung
- Wiedergabe vorhandener Symmetrien im Layout
- Anzahl der Richtungen (von Kanten) minimieren
- Blick sollte auf wesentliche Knoten gelenkt werden

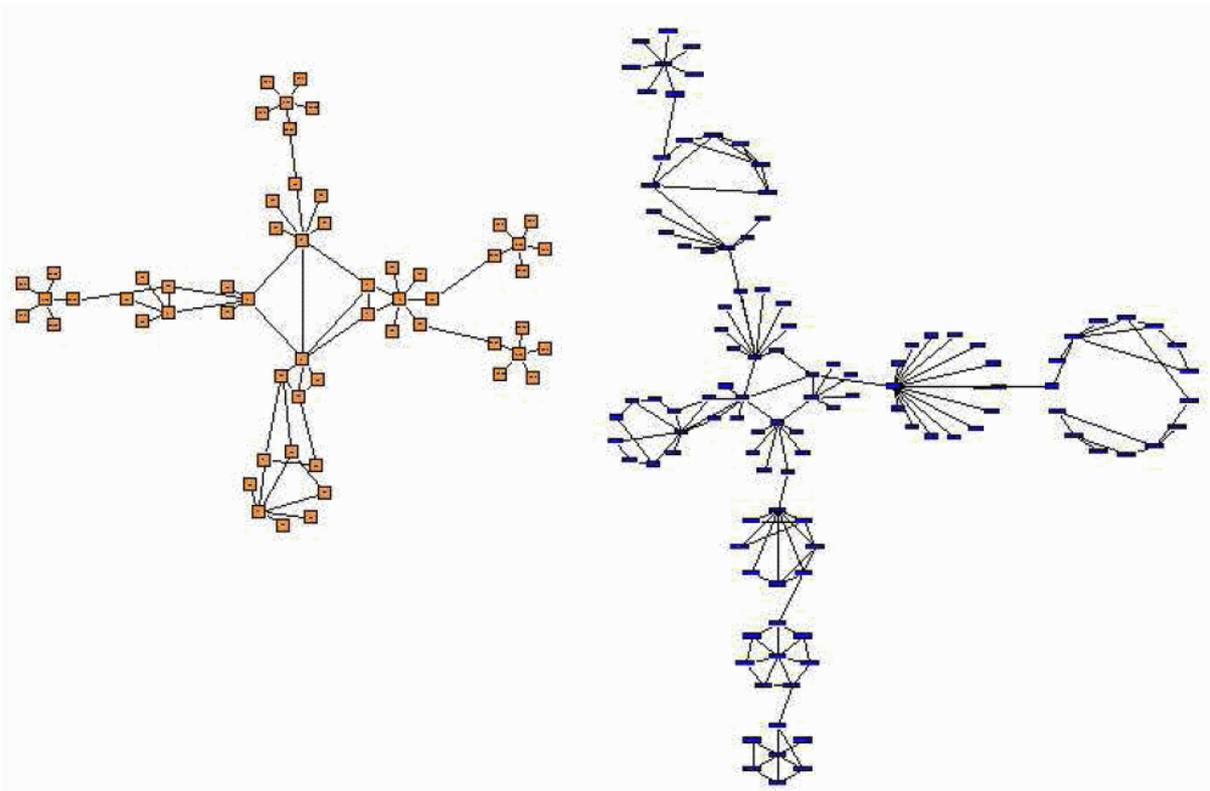
Ansätze:

- Zufalls-Layout
- gitterbasierte Layouts (vgl. Schaltkreisentwurf)
- hierarchische Layouts (verwenden aufspannenden Baum)
- physikalisch basierte Layouts (iteratives Herbeiführen eines Kräftegleichgewichts oder Energieminimums)

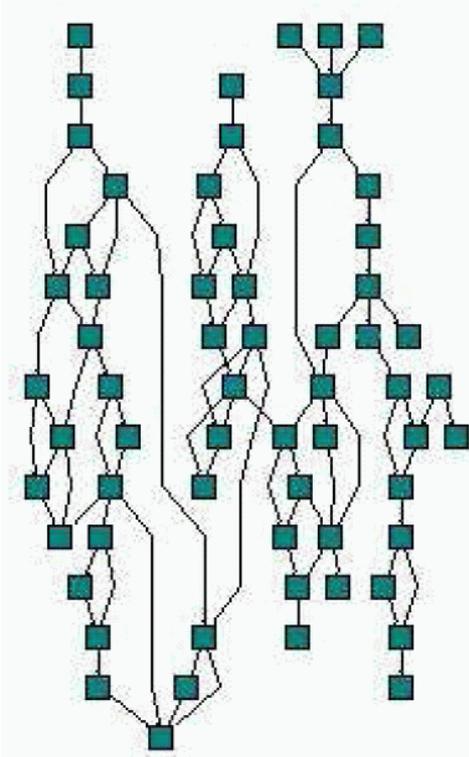
Hierarchische Layouts:



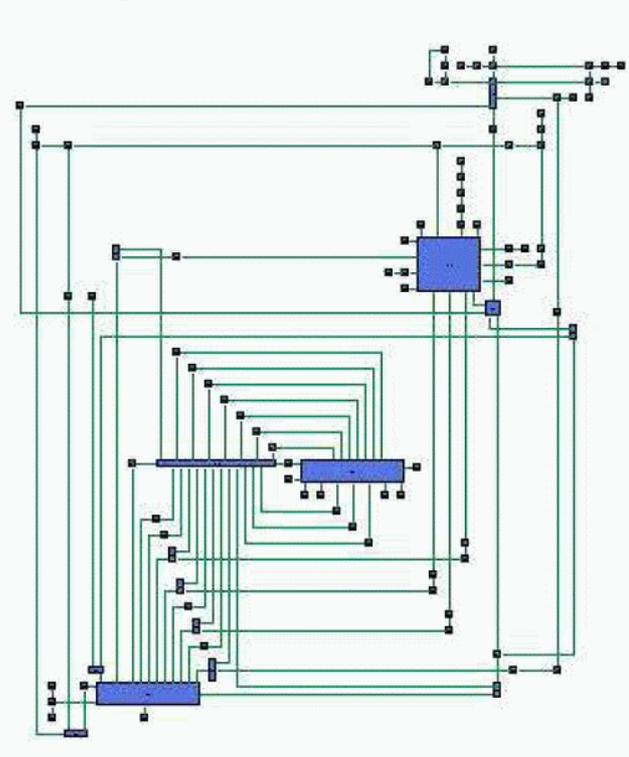
Kreisbasierte Layouts:



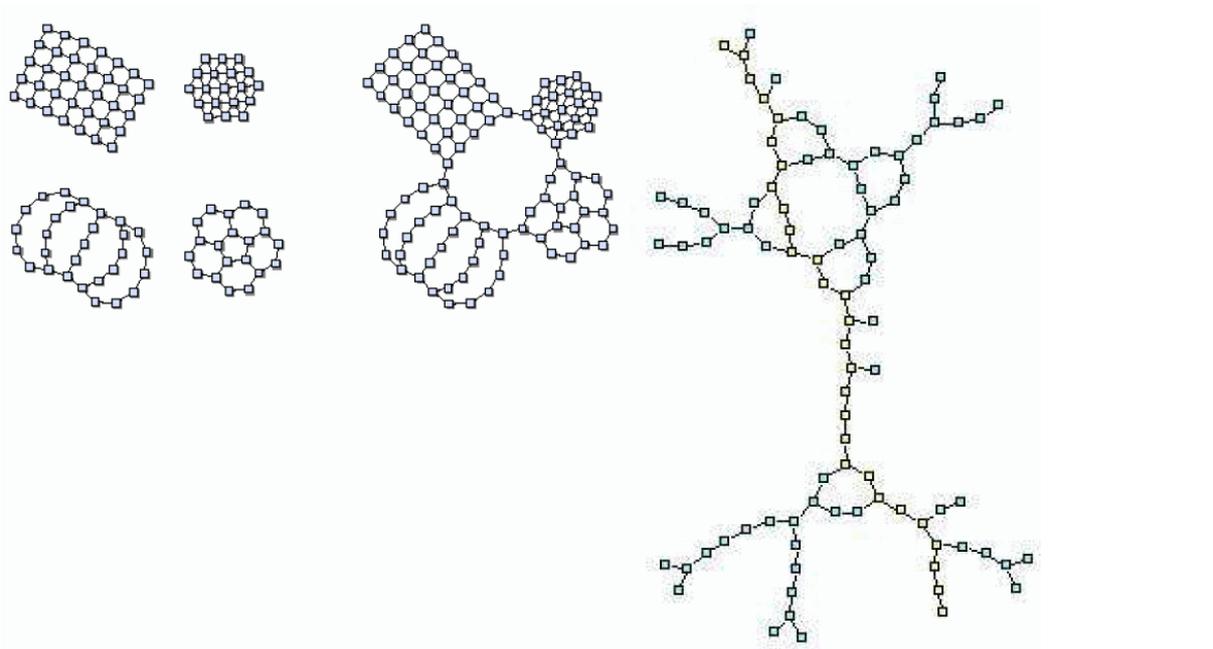
hierarchisch,



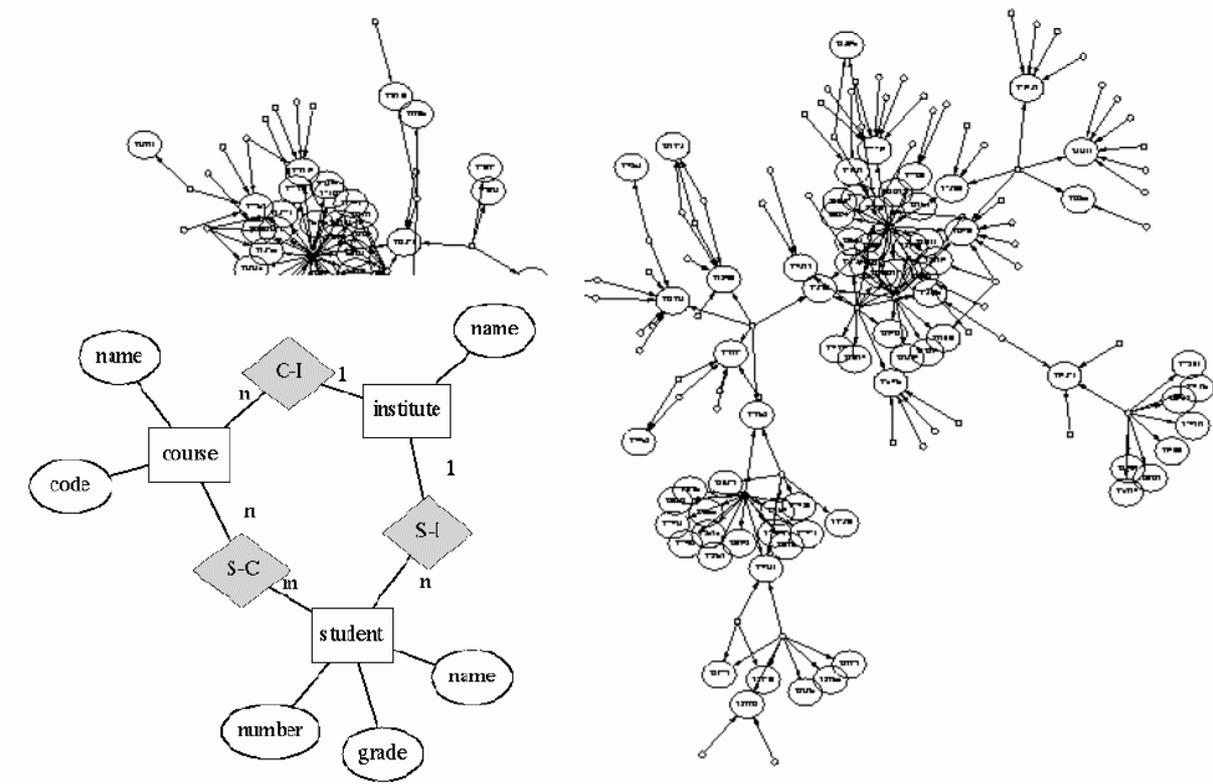
orthogonal:



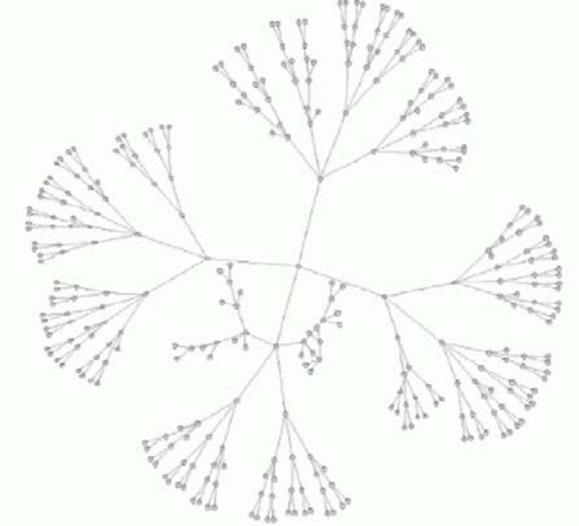
Wiedergabe von (partiellen) Symmetrien im Layout:



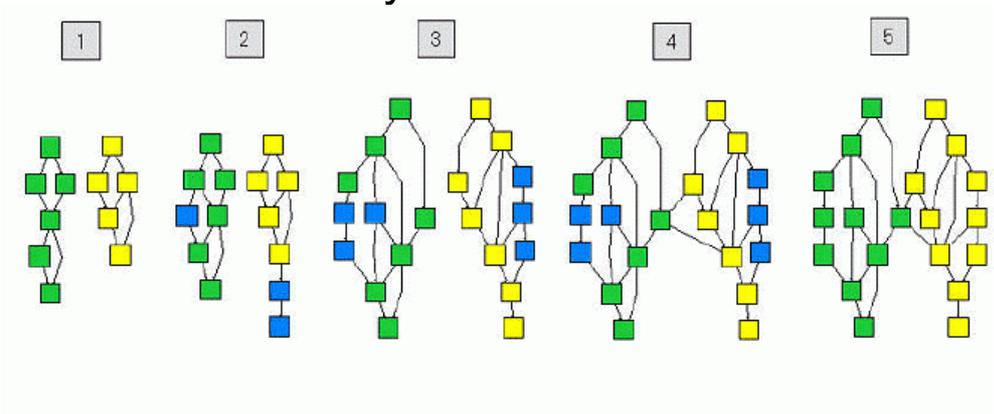
"Spring embedder" Layout (kräftebasiert):



kräftebasiertes Layout, weiteres Beispiel:

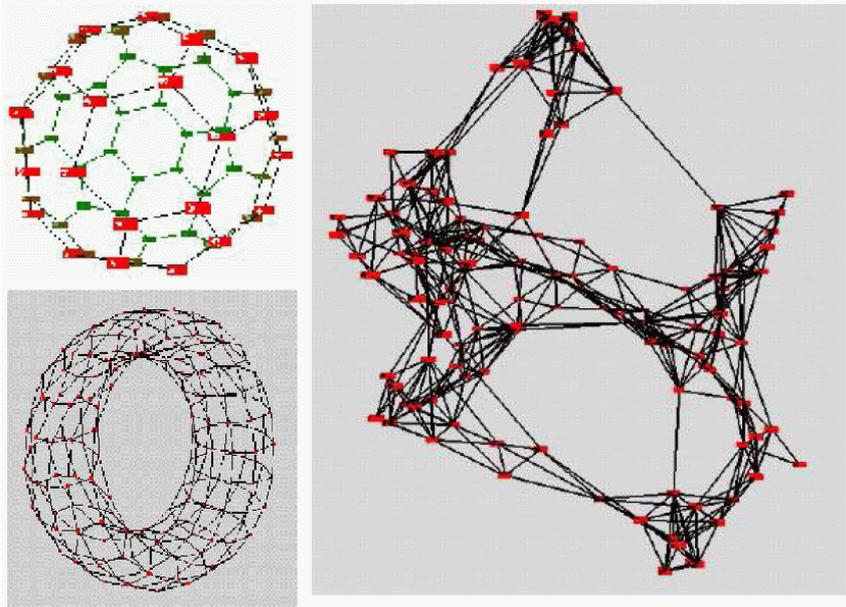


Inkrementelles Layout:

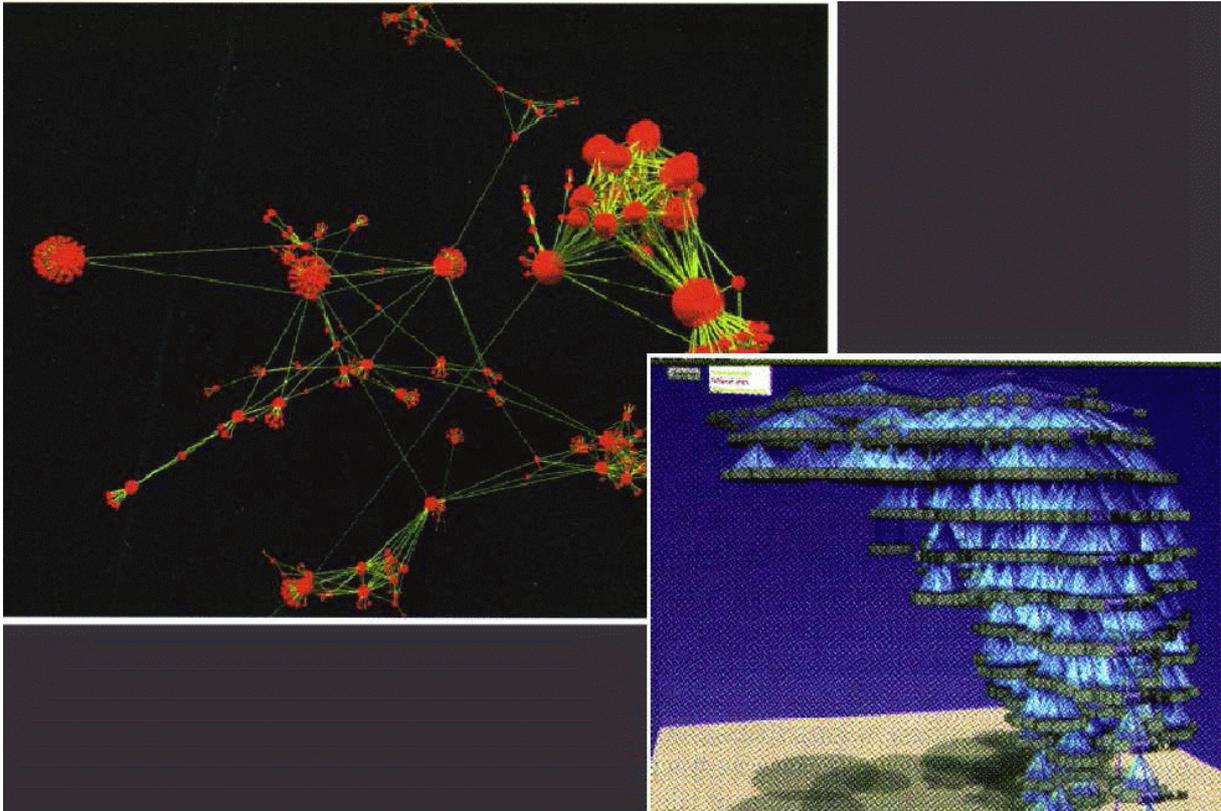


Relationen zwischen existierenden Elementen bleiben erhalten.

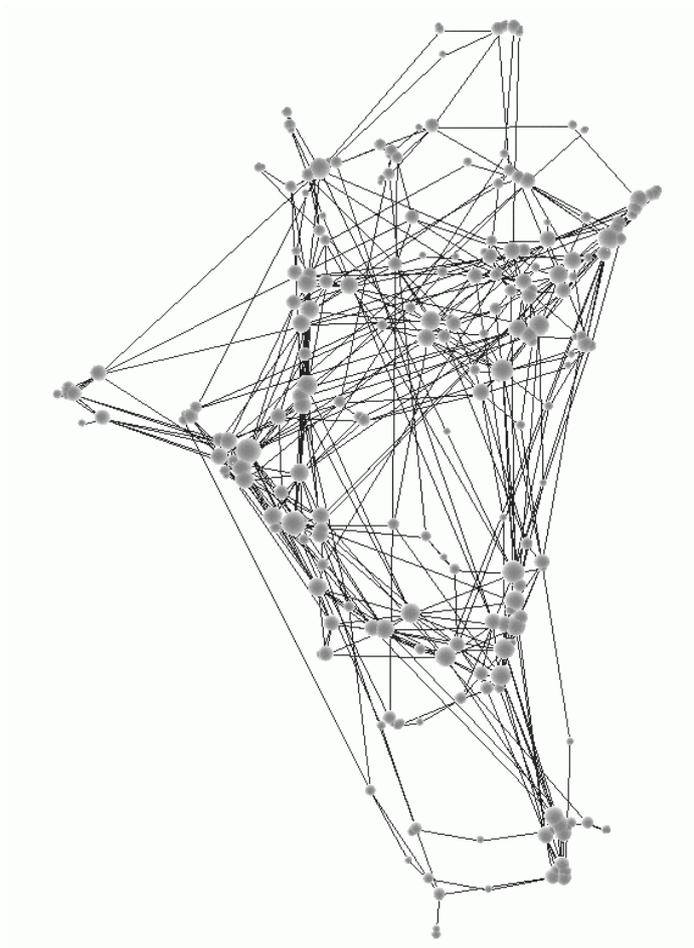
3D-Graph-Layouts:



weitere 3D-Beispiele:



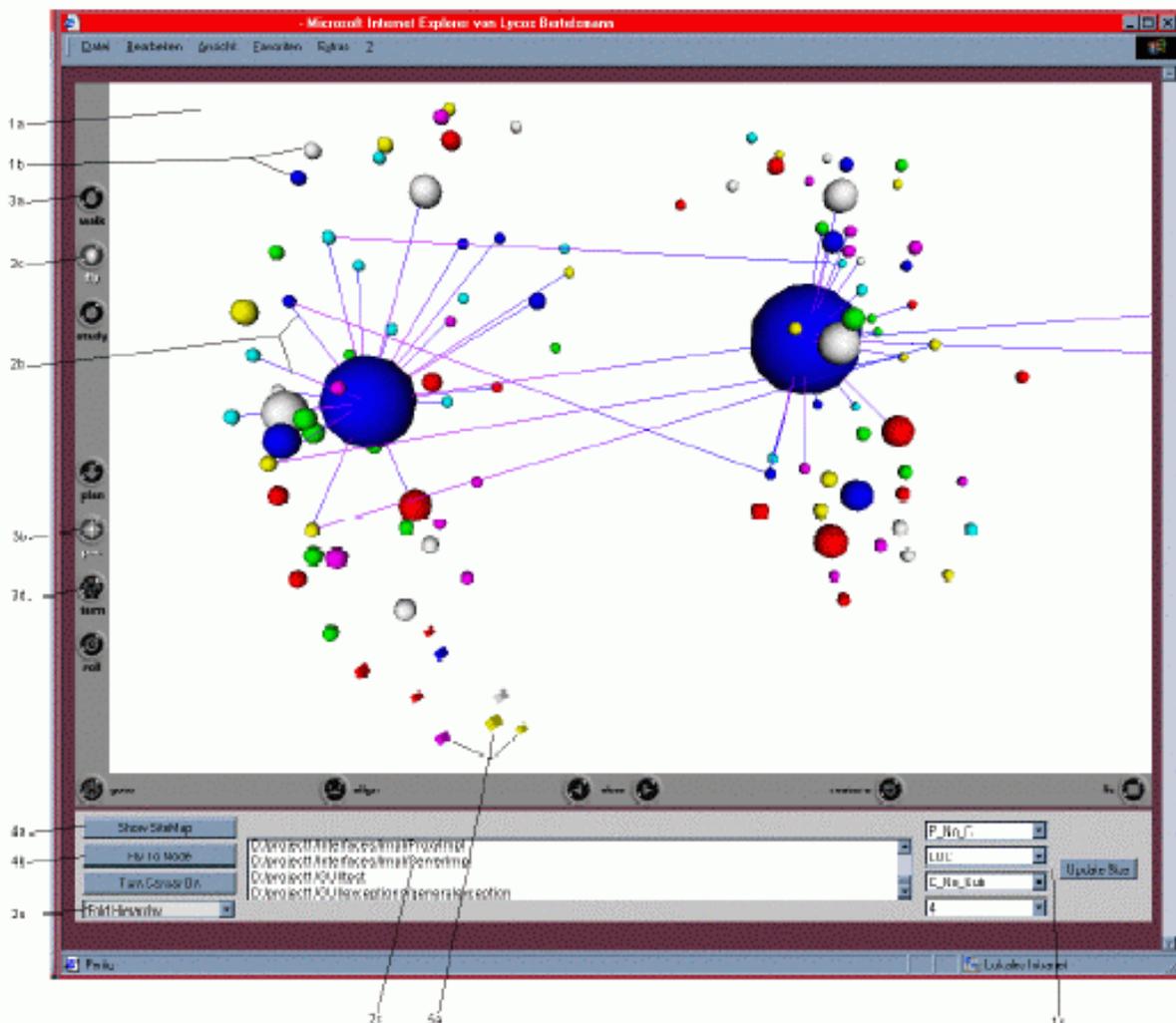
(Weimar 2005)



(Lewerentz et al.)

# Software zur graphbasierten, visuellen Analyse von großen Softwaresystemen (Lehrstuhl Softwaresystemtechnik, BTU Cottbus):

Anhang A: Annotierter Screenshot von CrocoCosmos



Dabei bedeuten:

- 1.x Kohäsionsdaten: Dargestellt wird die Objektübersicht (1a) zusammen mit der Objektdarstellung (1b). Die Parameter für die letzteren Daten werden für die jeweilige Abstraktionsstufe (von oben nach unten: Subsystem, Datei, Klasse) unter Verwendung eines Skalierungsfaktors in (1c) eingestellt.
- 2.x Strukturdaten: Mittels des Pull-Down-Menüs (2a) kann zwischen verschiedenen Modi gewählt werden, die jeweils unterschiedliches Verhalten für die Selektion eines Objekts repräsentieren (s.u.). Diese umfassen u.a. die explizite Darstellung von Strukturdaten wie z.B. Vererbungsbeziehungen (2b) und nähere Angaben zur Identifikation (2c).
- 3.x Navigation: Hier können die verschiedenen Navigationstechniken ausgewählt werden.
- 4.x Die zusätzlich mittels (4a) darzustellende Projekthierarchie kann durch Auswahl von (4b) für die Navigation innerhalb der Welt verwendet werden.
- 5.x Abstraktionsniveauselektion: Das Pull-Down-Menü (2a) beinhaltet ebenfalls beide Möglichkeiten der Abstraktionsniveauselektion (global und für einzelne Objekte). Im Screenshot sind unterschiedliche Abstraktionsniveaus durch unterschiedliche Formen symbolisiert (vgl. 5a und 1b). Das Pull-Down-Menü ermöglicht darüber hinaus die Generierung unterschiedlicher Visualisierungen knotenorientierter Meßdaten.

(siehe Webseite des LS SST)