Mündliche Doktorprüfung (Kolloquium)
10. February 2017, Konstanz

Universität Konstanz

Visual Analytics Methods for Exploring Large Amounts of Relational Data with Matrix-based Representations.

Michael Behrisch | Data Analysis and Visualization Group
Universität Konstanz

Mündliche Doktorprüfung in Form eines Kolloquiums über die Dissertation von Michael Behrisch.

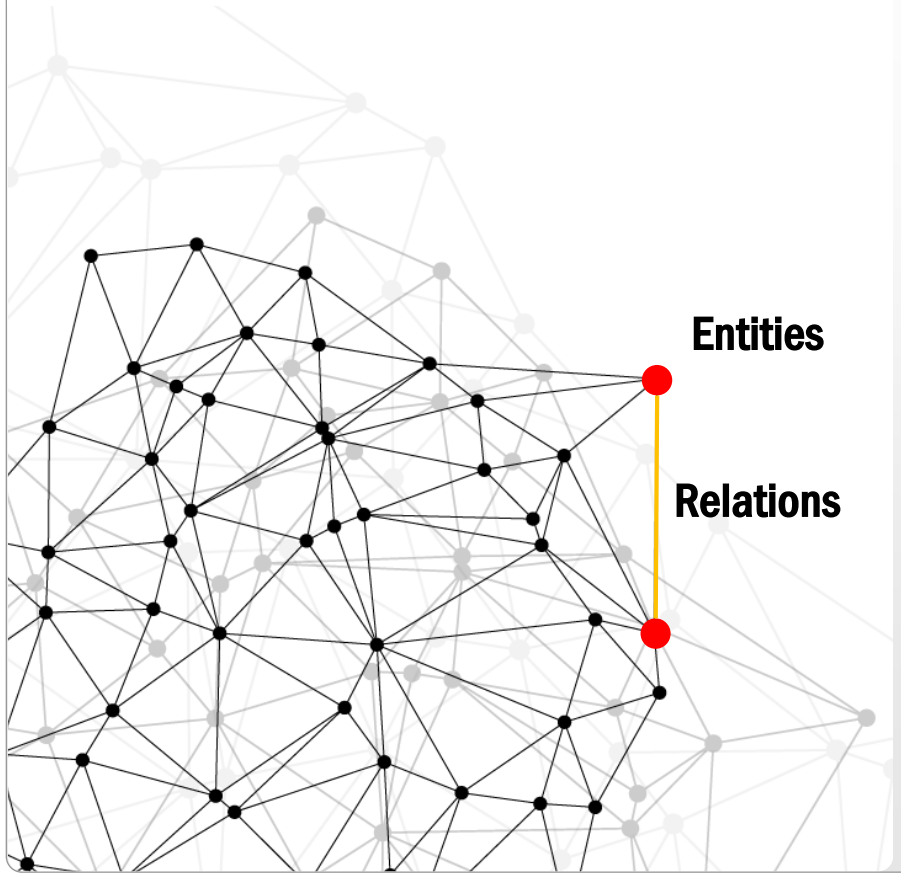
Prüfungskommission

Prof. Dr. Ulrik Brandes	(Vorsitzender und mündlicher Prüfer)
Prof. Dr. Tobias Schreck	(Gutachter und mündlicher Prüfer)
Prof. Dr. Jean-Daniel Fekete	(Gutachter und mündlicher Prüfer)
Prof. Dr. Falk Schreiber	(Gutachter)

Motivation.

Why is Relational Data so powerful?

Application Examples



Entities

Attributes, properties, features describing objects

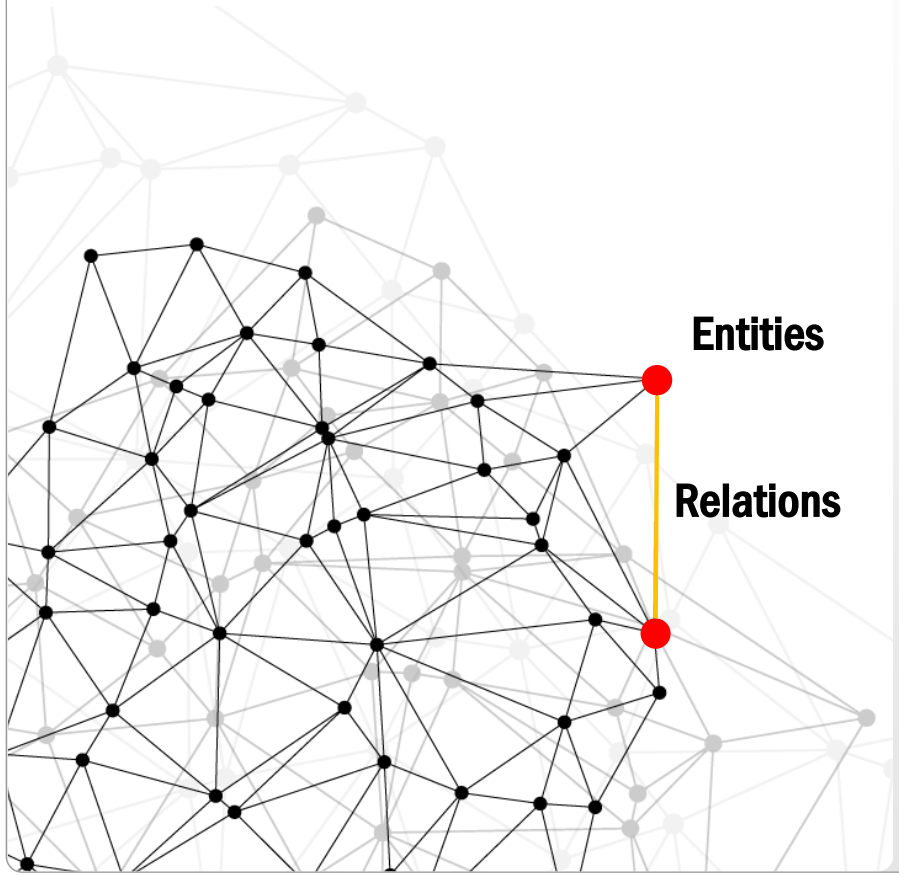
Relations

Attributes, properties, features describing the context

Motivation.

Why is Relational Data so powerful?

Application Examples



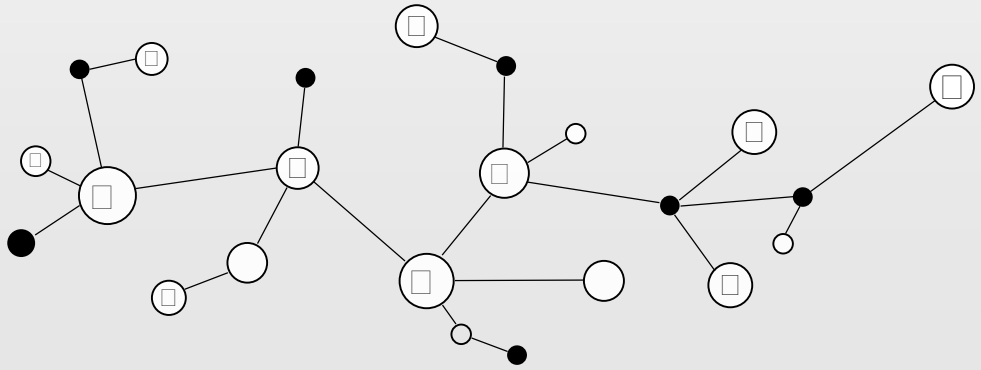
Entities

Attributes, properties, features describing objects

Relations

Attributes, properties, features describing the context

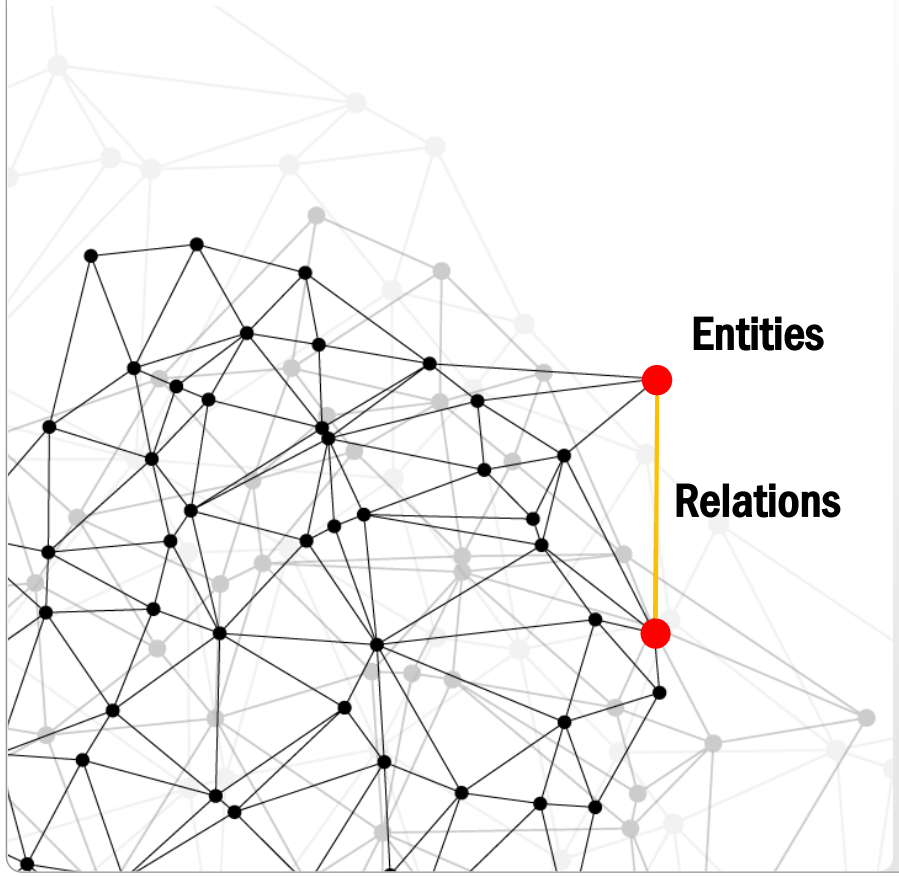
Social Network Analysis



Motivation.

Why is Relational Data so powerful?

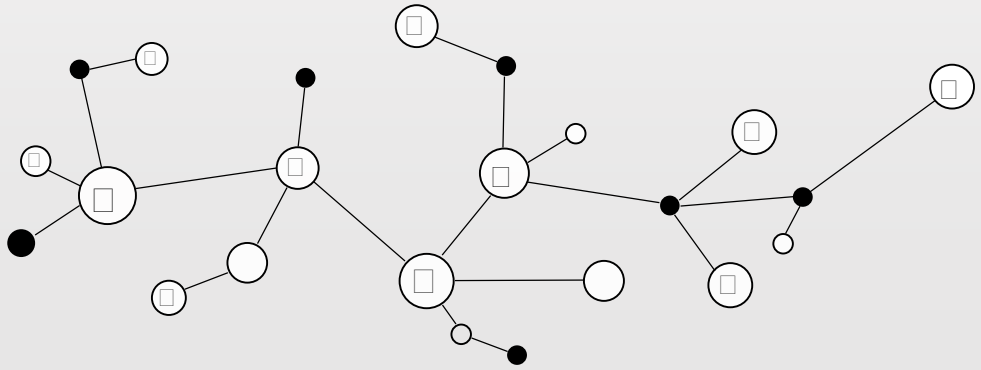
Application Examples



Entities
Attributes, properties, features describing objects

Relations
Attributes, properties, features describing the context

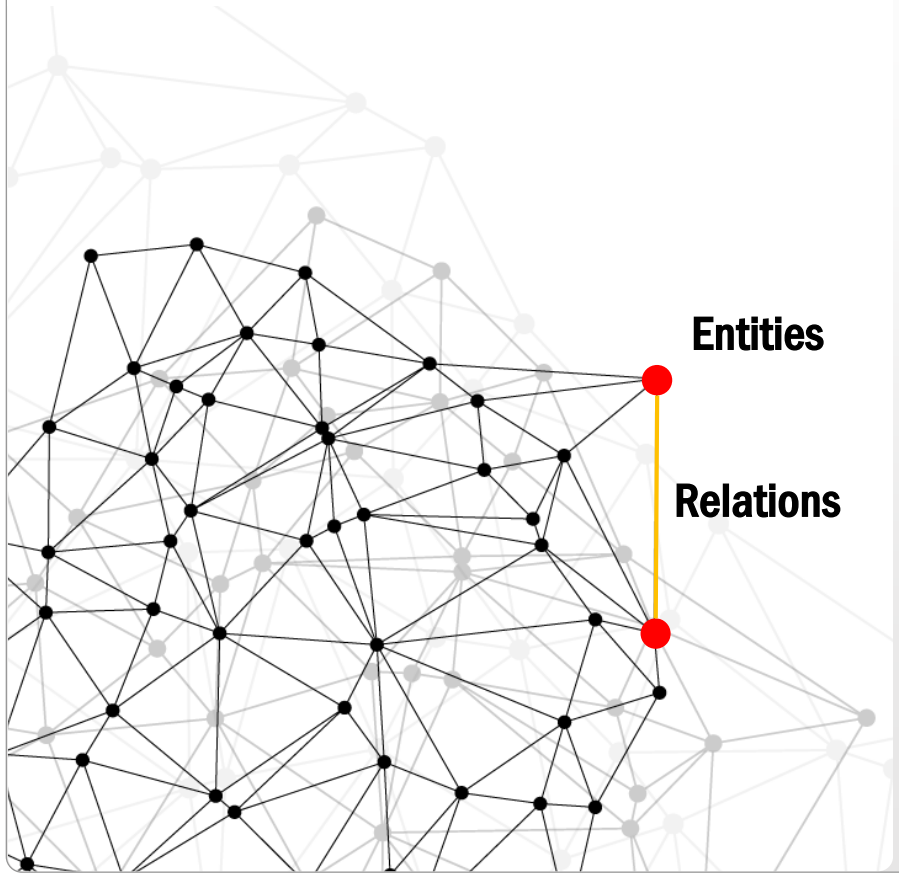
Biological Experiments
Health Data



Motivation.

Why is Relational Data so powerful?

Application Examples



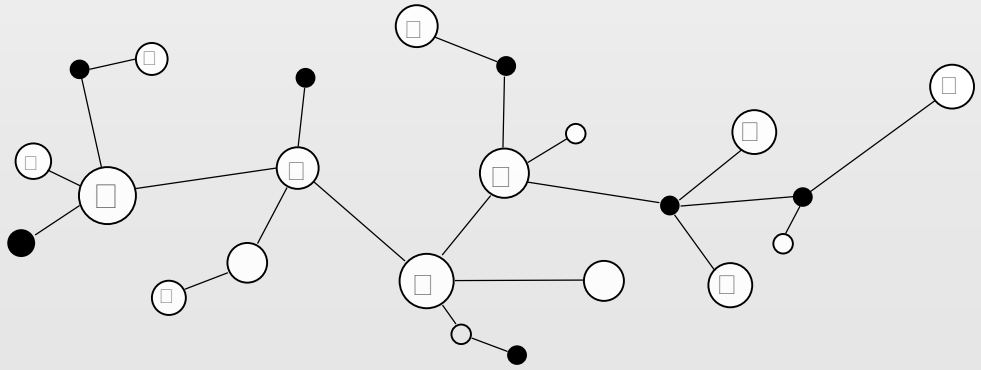
Entities

Attributes, properties, features describing objects

Relations

Attributes, properties, features describing the context

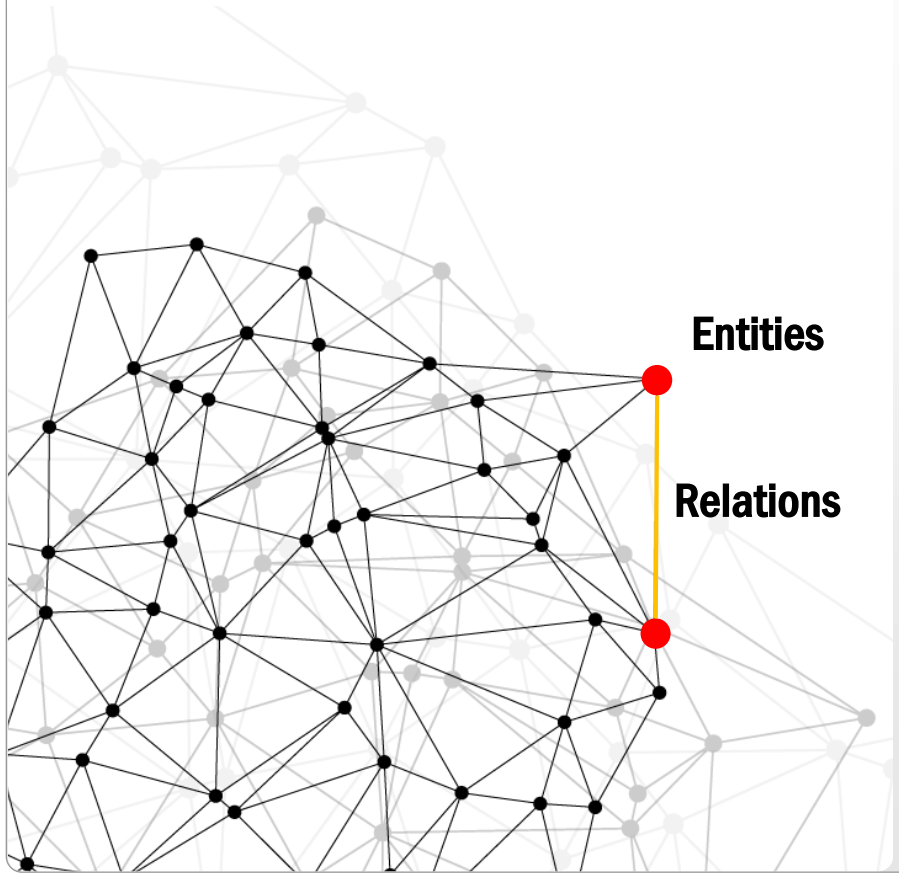
Computer Networks



Motivation.

Why is Relational Data so powerful?

Application Examples



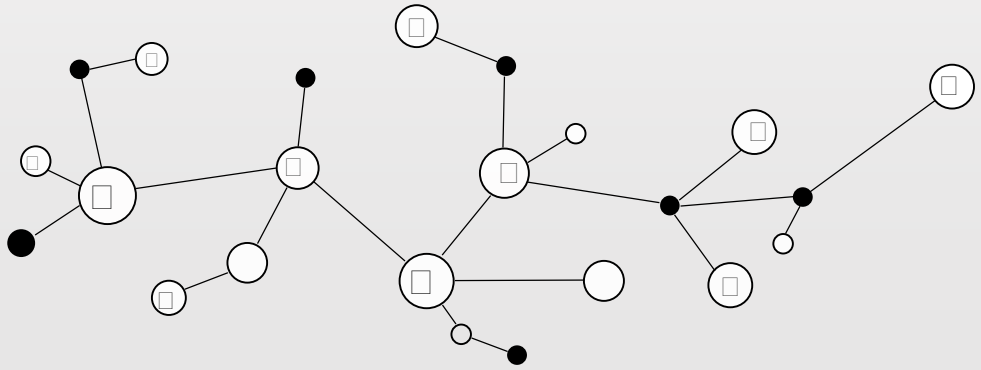
Entities

Attributes, properties, features describing objects

Relations

Attributes, properties, features describing the context

Infrastructure Networks

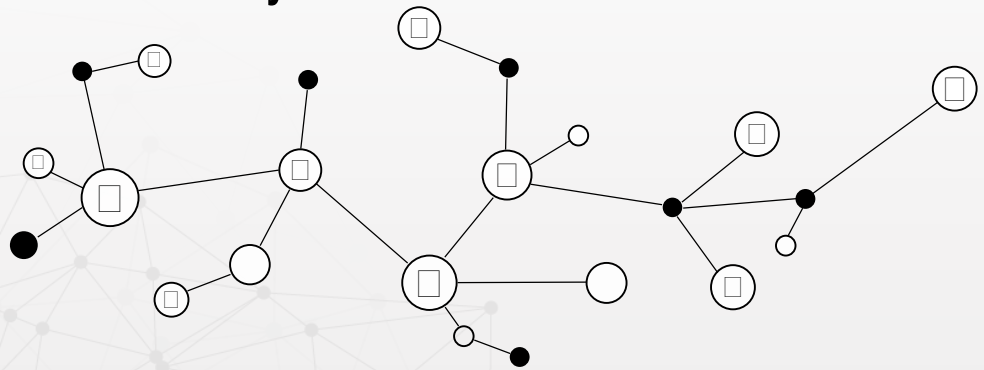


Motivation.

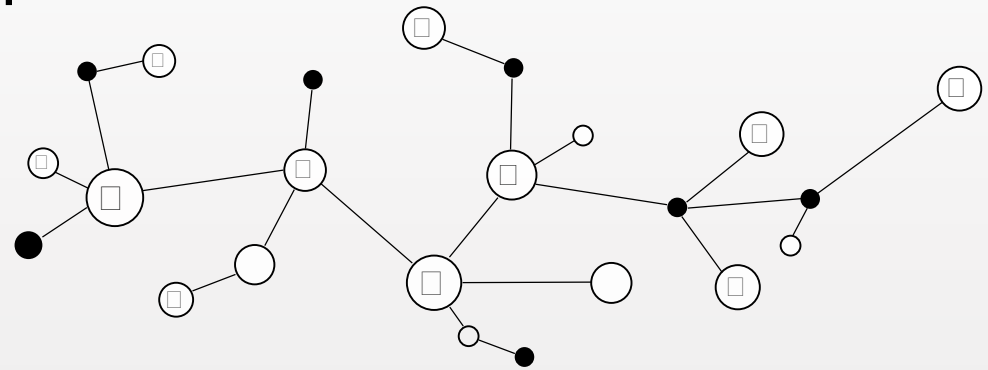
Why is Relational Data so powerful?

Application Examples.

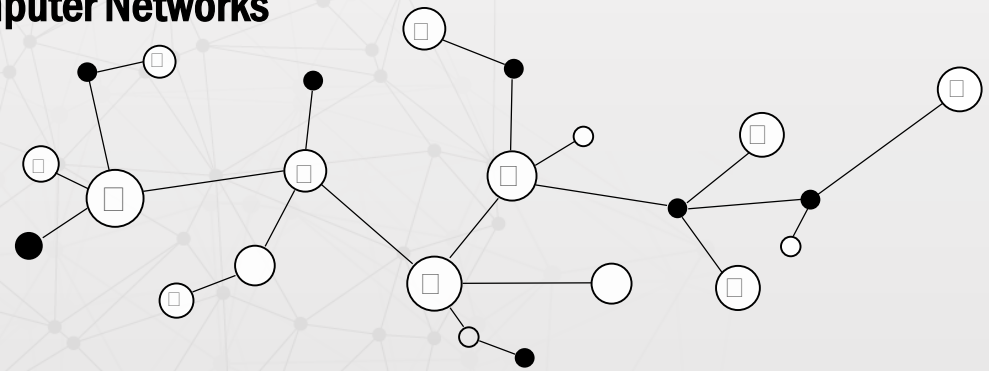
Social Network Analysis



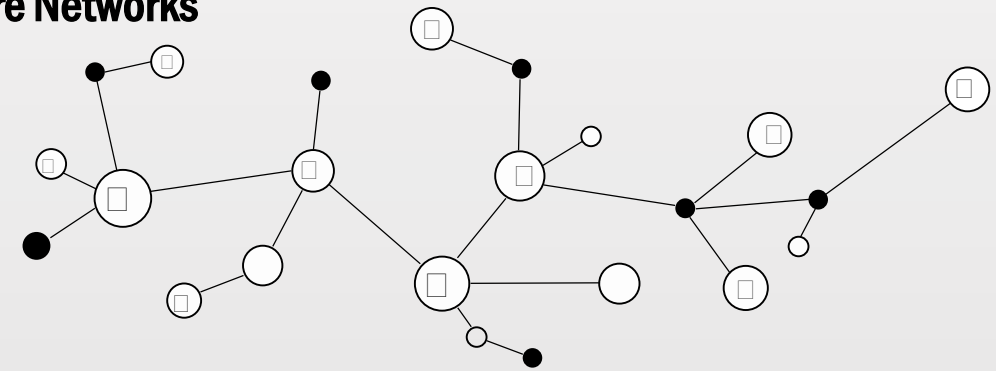
Biological Experiments
Health Data



Computer Networks



Infrastructure Networks



Motivation.
Relational Data.

Application Examples

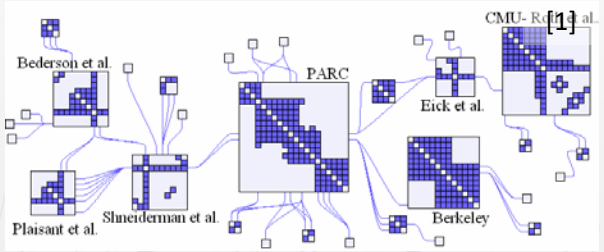
Every Graph Structure can be Represented as a Matrix

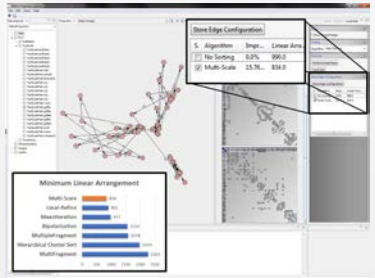
Motivation.

Why is Relational Data so powerful?

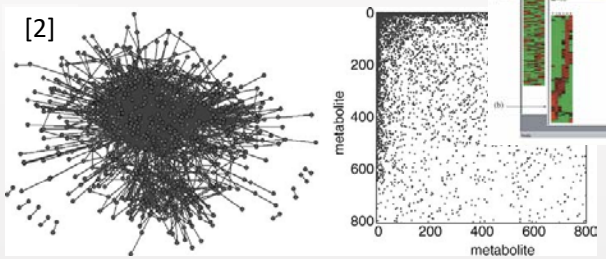
Application Examples.

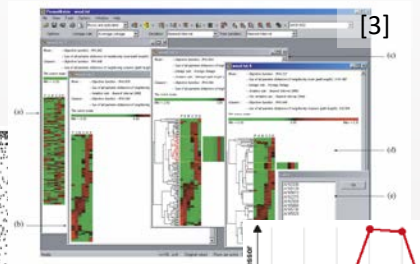
Social Network Analysis

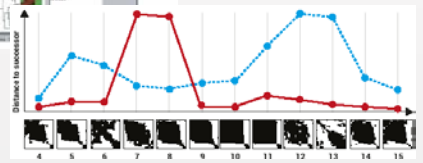




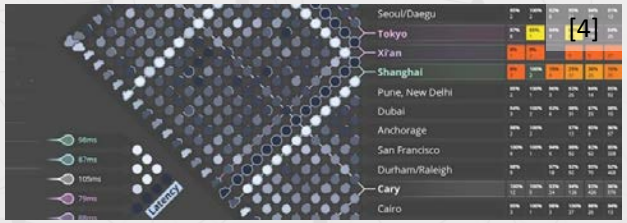
Biological Experiments Health Data

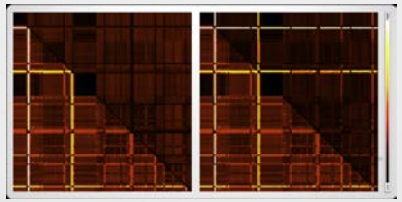




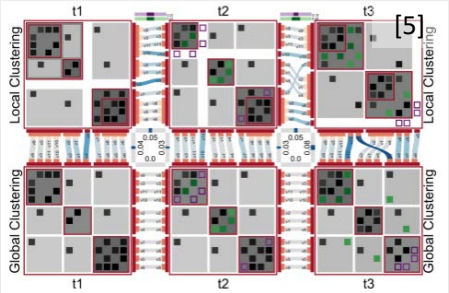


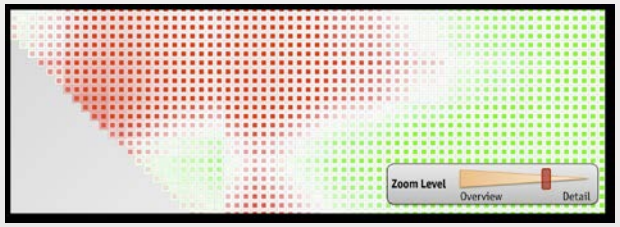
Computer Networks





Infrastructure Networks





Pictures courtesy: 1) Henry, Nathalie. Exploring social networks with matrix-based representations. Diss. PhD thesis, Université Paris Sud, France, and University of Sydney, Australia, 2008. 2) Junker, Björn H., and Falk Schreiber. Analysis of biological networks. Vol. 2. John Wiley & Sons, 2011.
3) Eisen, M., Spellman, P., Brown, P., and Botstein, D. (1998), "Cluster Analysis and Display of Genome-Wide Expression Patterns," Proceedings of the National Academy of Sciences, 95, 14863–14868 4) Benson, Jordan Riley, and Rajiv Ramarajan. "NOCturne: A Scalable Large Format Visualization for Network Operations." IEEE VIS Conference, Atlanta. 2013. 5) Vehlow, Corinna, Fabian Beck, and Daniel Weiskopf. "Visualizing Dynamic Hierarchies in Graph Sequences." (2015).

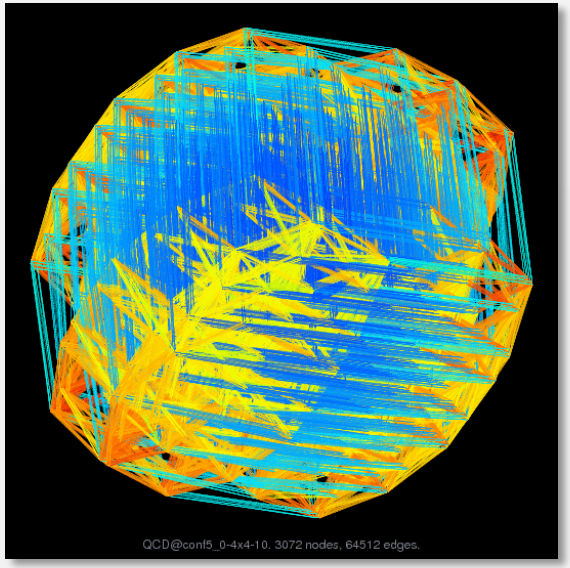
Motivation.

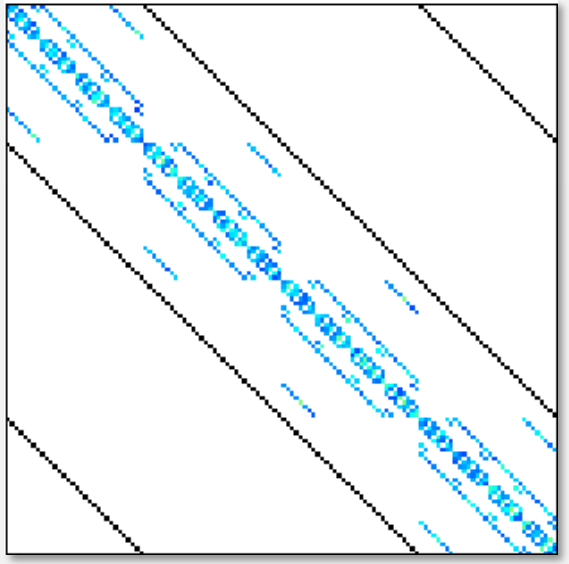
Why Matrices?

Properties and Benefits.

Matrix Properties:

- Compact representation
- No Overplotting Issues





Quantum chemistry problem: Quantum chromodynamics conf 5.0-00l 4x4-1000:
3072 Nodes; **119,808** Edges

Images courtesy: The SuiteSparse Matrix Collection (formerly known as the University of Florida Sparse Matrix Collection) https://www.cise.ufl.edu/research/sparse/matrices/QCD/conf5_0-4x4-10.html

Motivation.

Why Matrices?

Properties and Benefits.

Matrix Properties:

- Compact representation
- No Overplotting Issues
- Order of rows / columns can be freely changed

Motivation.

Why Matrices?

Properties and Benefits.

Matrix Properties:

- Compact representation
- No Overplotting Issues
- Order of rows / columns can be freely changed
- Simultaneous Overview and Detail

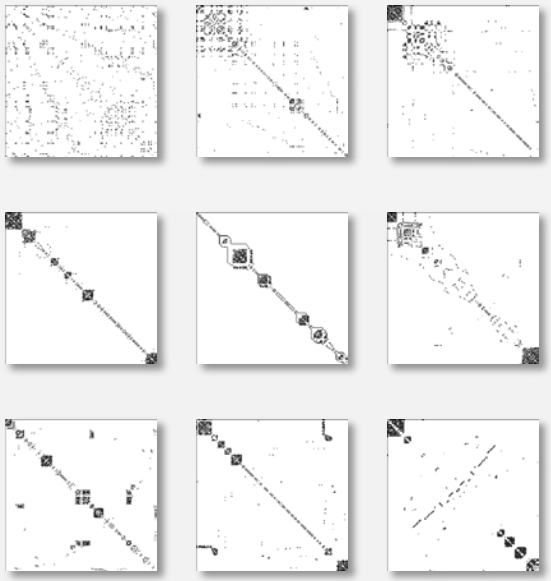
Matrix Analysis Task:

Overview tasks can be answered **independent** of matrix **ordering**.

Higher level tasks **require** an **reordering** of rows and columns.

Other tasks only if **patterns** are **visible.**

All matrices same graph (Brain Connectivity)



Motivation.

Why Matrices?

Central Question

Central Question of the Dissertation:

What is a good Matrix Representation?

Motivation.

Why Matrix Research?

Central Questions and Topics of the Work.

Affects

Visualization

Data Analysis

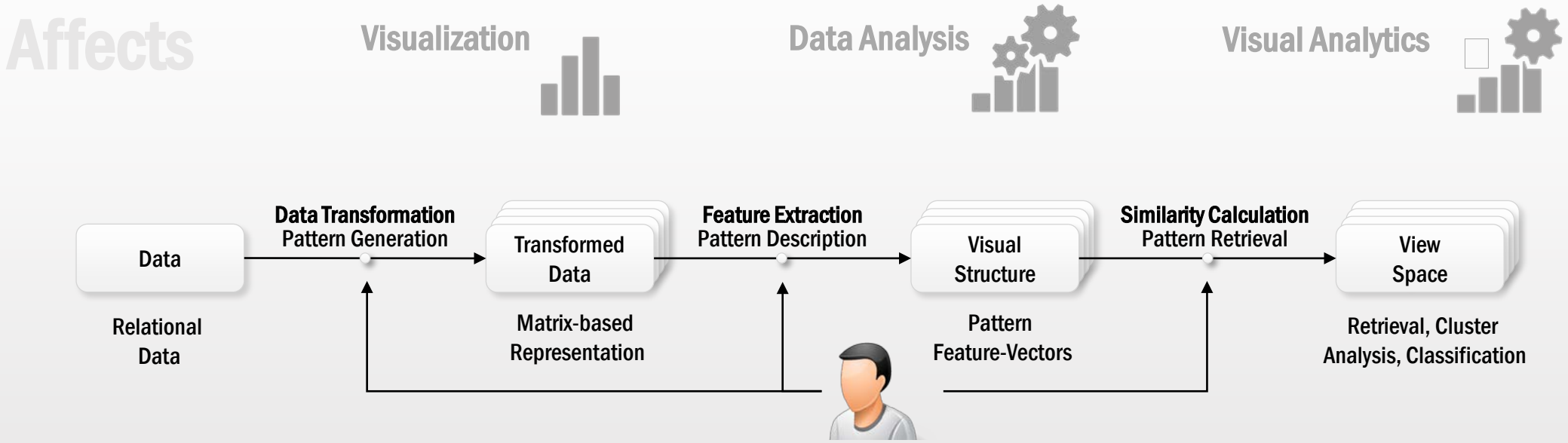
Visual Analytics

- How can we visually support the **effectiveness** of matrices?
- How can we automatically assess the **quality** of a matrix?
- How can we use **interestingness** for exploring large matrix spaces?

Research Framework.

Pattern-Driven Analysis of Large Matrix-Spaces.

What is a good (Matrix) Representation?

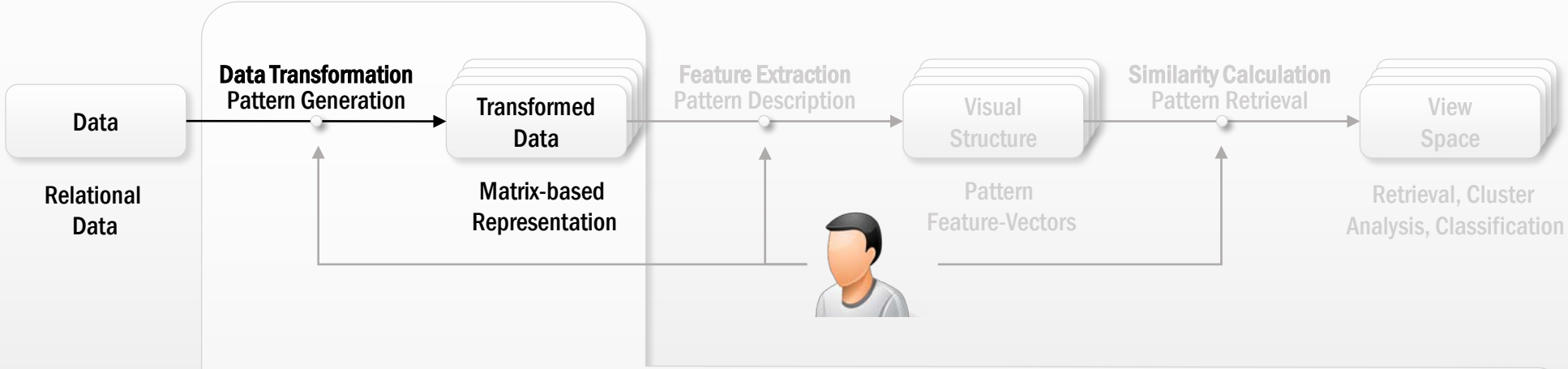


Research Framework.

Pattern-Driven Analysis of Large Matrix-Spaces.

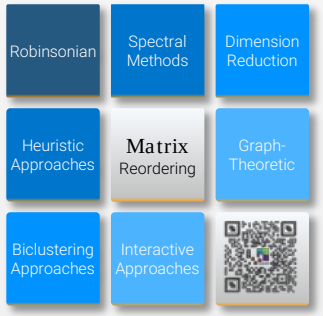
Matrix Reordering

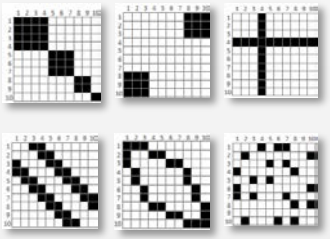
What is a good (Matrix) Representation?



"Matrix Reordering Methods for Table and Network Visualization"

Behrisch, Bach, Riche, Schreck, and Fekete.
Computer Graphics Forum, The Eurographics Association and John Wiley & Sons Ltd., 2016, 35, 693-716. [Beh+16b]



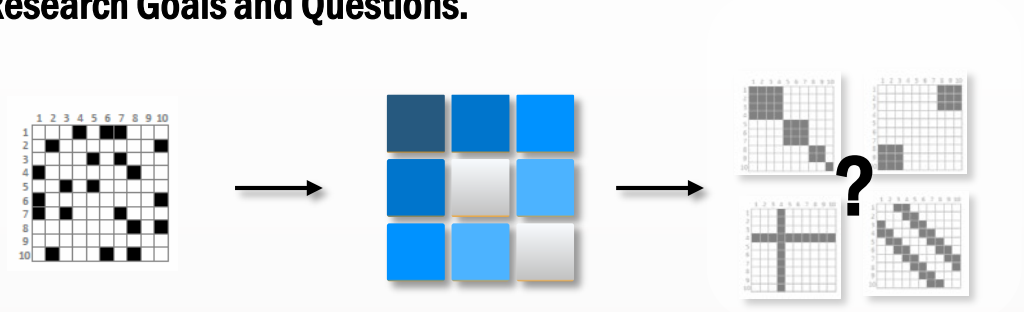


Matrix Reordering State-of-the-Art.
Research Objectives and Contributions.

Matrix Reordering

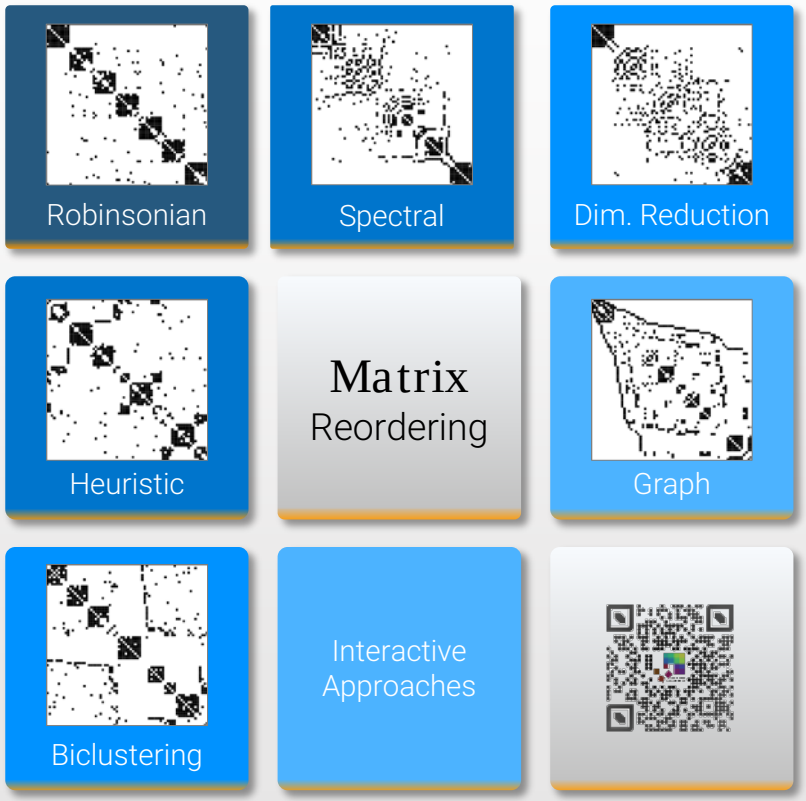
Structure and Goals of the Survey.

Research Goals and Questions.



Core Question: Which algorithms tend to produce which patterns?

- **Unified presentation** and discussion of the (design) impacts.
- **Practical Guidance** on algorithm selection.
- **Open Challenges** and **Future Research Opportunities**



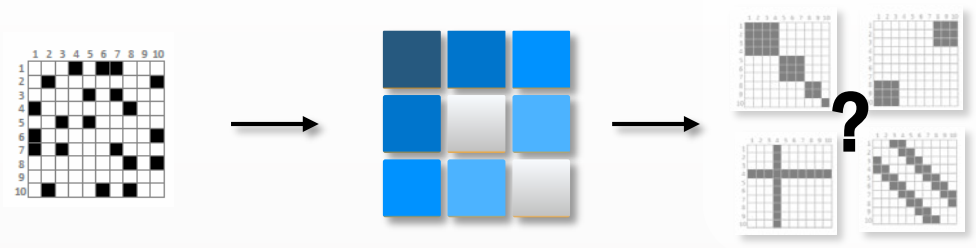
Same Graph: Clustered Nodes: 70; Edges 312; Density 0.12; Clustering Coefficient: 8.77

Matrix Reordering State-of-the-Art.
Research Objectives and Contributions.

Matrix Reordering

Structure and Goals of the Survey.

Research Goals and Questions.



Core Question: Which algorithms tend to produce which patterns?

- **35** Algorithms and **150** Graphs.
- **14** performance measures for each trial.
- Reviewed ~**159** Papers

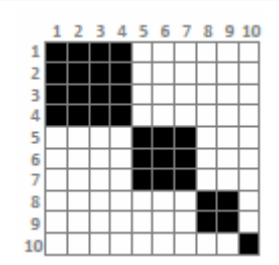
Visit our Website
matrixreordering.dbvis.de

Background.

Patterns in Matrices.

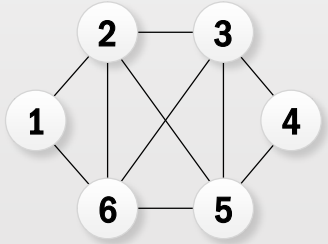
Matrix Reordering

Block-Diagonal Form



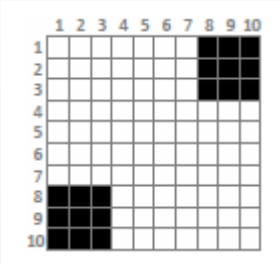
Visual Description
Coherent rectangular areas around the main diagonal

Topology Interpretation
Strongly connected components or cliques
Single missing connections (“holes”)
“Off-diagonal dots” link to other clusters



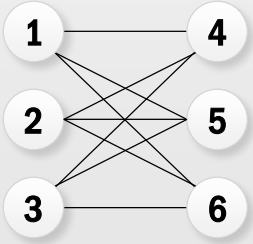
Analysis Purpose / Task
Central/expected pattern
Partitioning and Grouping (Vertices in blocks share similar connection characteristics)

Off-Diagonal Blocks



Visual Description
Coherent areas apart from the main diagonal

Topology Interpretation
Bi-partite Subgraph (Bi-Graph)
May also contain missing links “holes”



Analysis Purpose / Task
Adjacency Assessment and Connection Overview
Understand *how* groups are related

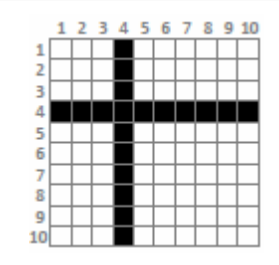
See also: Shen, Zeqian, and Kwan-Liu Ma. "Path visualization for adjacency matrices." Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on Visualization. Eurographics Association, 2007.

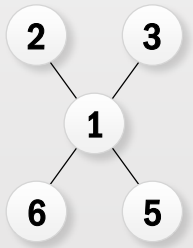
Background.

Patterns in Matrices.

Matrix Reordering

Star/Line



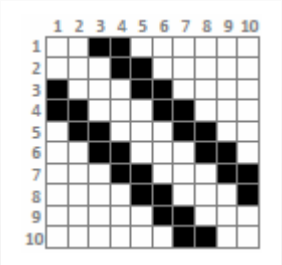


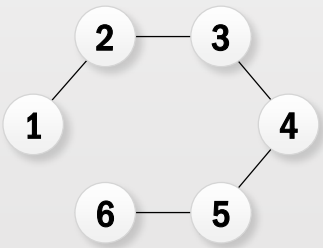
Visual Description
Continuous horizontal and vertical lines

Topology Interpretation
Vertex is strongly connected to several distinct other vertices; Line/Star differentiates uni-directional, resp. bi-directional, connection

Analysis Purpose / Task
Connectivity aspects

Bands





Visual Description
Continuous lines apart from the main diagonal

Topology Interpretation
Paths, cycles or meshes, “Few” connectors; Width relates to alternative routes, not strength of the path

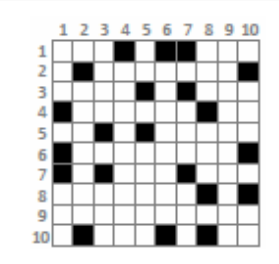
Analysis Purpose / Task
Adjacency Assessment

See also: Shen, Zeqian, and Kwan-Liu Ma. "Path visualization for adjacency matrices." Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on Visualization. Eurographics Association, 2007.

Background.

Anti-Patterns in Matrices.

Noise

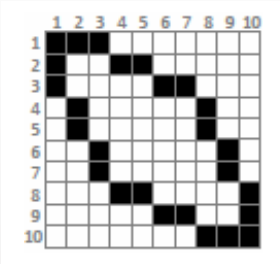


Visual Description
Noise, also called salt-and-pepper
No structure can be perceived

Topology Interpretation
Potentially no global (entire matrix) or local (submatrix) topology or reordering algorithm was not able to grasp it

Analysis Purpose / Task
(Overview)

Bandwidth



Visual Description
Grouping of non-zero elements within an enclosure around the main diagonal

Topology Interpretation
Normally just an artifact of the algorithm's permutation enumeration (breath-first search)

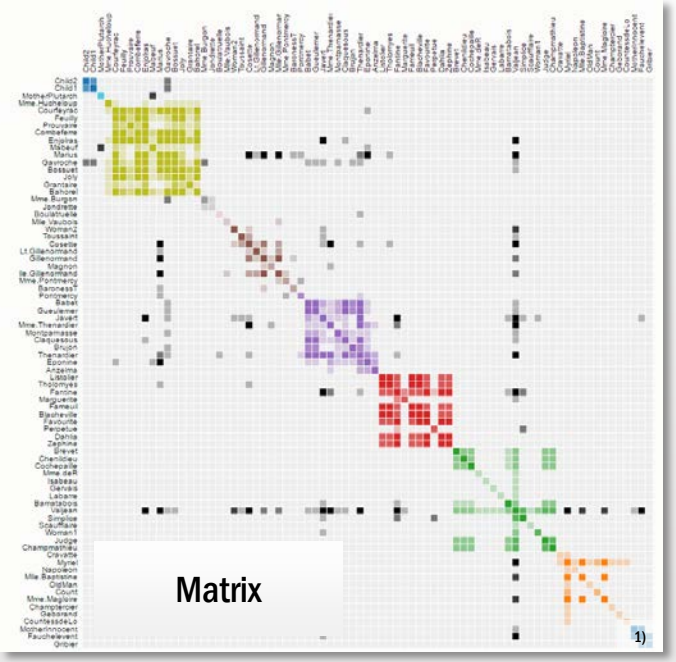
Analysis Purpose / Task
(Overview)

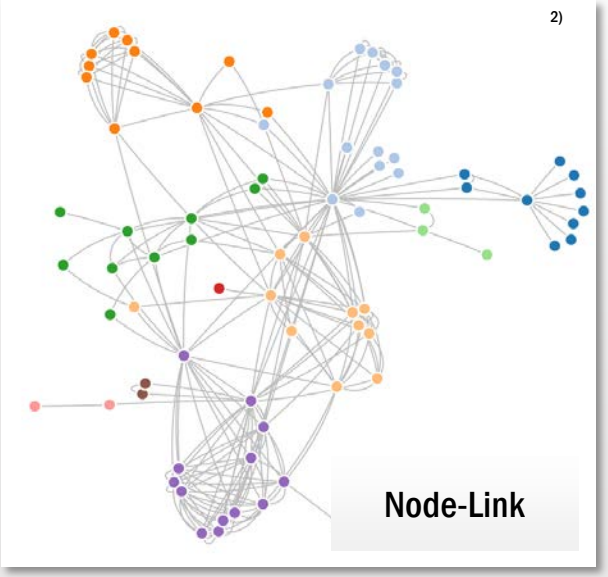
See also: Shen, Zeqian, and Kwan-Liu Ma. "Path visualization for adjacency matrices." Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on Visualization. Eurographics Association, 2007.

Background.

Patterns in Matrices.

Combination of Patterns.





Problem Statements.

Problem 1:

Computationally $\mathcal{NP}$-Hard Problem

For a n×m table, there are n!×m! permutations

Problem 2:

No consensual objective function

Which pattern should be favored/emphasized?

Pictures courtesy: 1) https://bost.ocks.org/mike/miserables/; 2) https://bl.ocks.org/mbostock/4600693

Matrix Reordering Approaches.

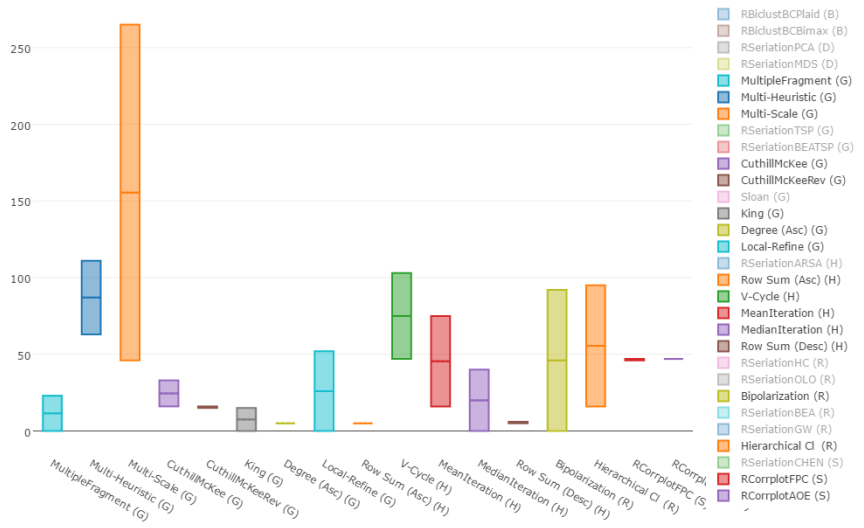
Algorithm Selection Guidelines

Matrix Reordering

Findings and Lessons Learned.

- **Fast algorithms first.**
- Heuristic algorithms offer runtime and quality tradeoffs.
- Optimizing for cluster identification.
- Quality takes time.

Sub-second runtime is possible!



- **Cuthill-McKee or RCorrplotSortingAOE**

Matrix Reordering Approaches.

Algorithm Selection Guidelines

Findings and Lessons Learned.

- Fast algorithms first.
- **Heuristic algorithms offer runtime and quality tradeoffs.**
- Optimizing for cluster identification.
- Quality takes time.

Often produce **more patterns** than very fast algorithms, and **improve** them via multiple iterations.

Often tend to reveal **data partitioning first**, but patterns within these partitions require more iterations.

- Barycenter or RSeriationARSA

Matrix Reordering Approaches.

Algorithm Selection Guidelines

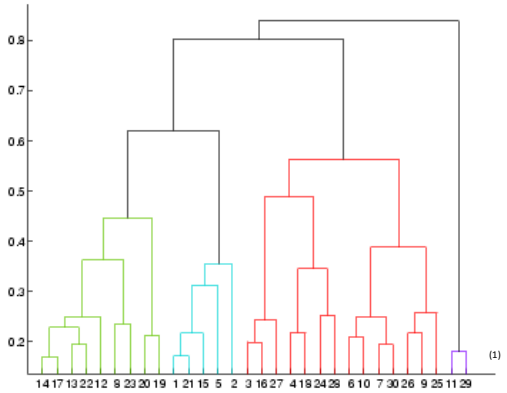
Matrix Reordering

Findings and Lessons Learned.

- Fast algorithms first.
- Heuristic algorithms offer runtime and quality tradeoffs.
- **Optimizing for cluster identification.**
- Quality takes time.

Some algorithms explicitly **detect clusters** and order each one individually, placing them at the matrix diagonal.

Alternatively spectral algorithms also highlight clusters.



- (Hierarchical) clustering or RSeriationChen

1) http://www-rohan.sdsu.edu/doc/matlab/toolbox/stats/dendrogr.gif

Matrix Reordering Approaches.

Algorithm Selection Guidelines

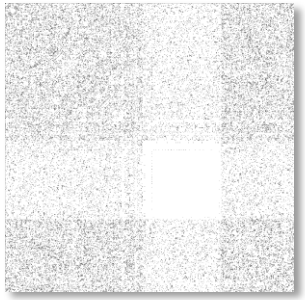
Findings and Lessons Learned.

- Fast algorithms first.
- Heuristic algorithms offer runtime and quality tradeoffs.
- Optimizing for cluster identification.
- **Quality takes time.**

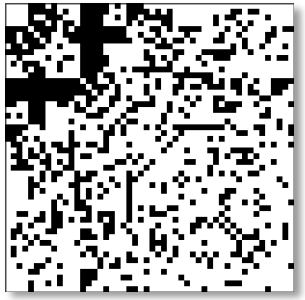
Local (sub-)structures and patterns might become discernable

As a direct consequence, algorithms attempting to optimize functions on the entire data may not be able to grasp these phenomena.

OLO
on Brain_0



TSP
on Petit_randomA2



Biclustering
on Barabasi_s50_d10

- Optimal-Leaf-Ordering, Biclustering or TSP

Matrix Reordering Approaches.
Conclusion.

Matrix Reordering

Take Home Messages.

- **Matrix Reordering is Powerful.**
  Let's us focus on the data topology and task at hand.
- **Data Topology can be interpreted.**
  Structure visible on a Global and Local level.
- **Visual Analytics will change the field (once again).**
  Opening the Black-Box will lead to novel algorithms.

matrixreordering.dbvis.de

Matrix Reordering Approaches.
Conclusion.

Take Home Messages.

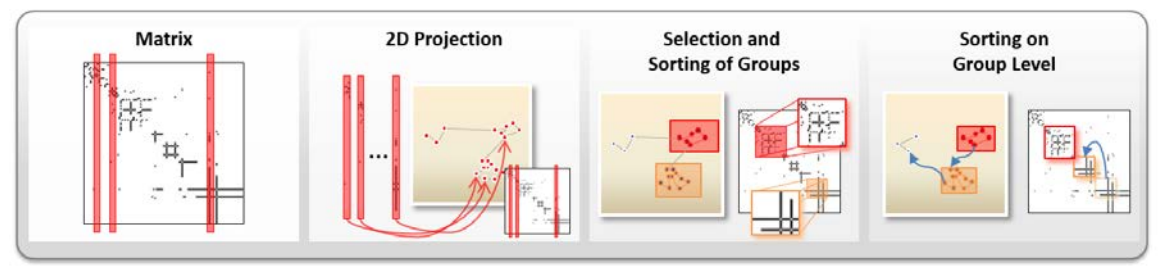
Matrix Reordering is Powerful.
Lets us focus on the data topology and task at hand.

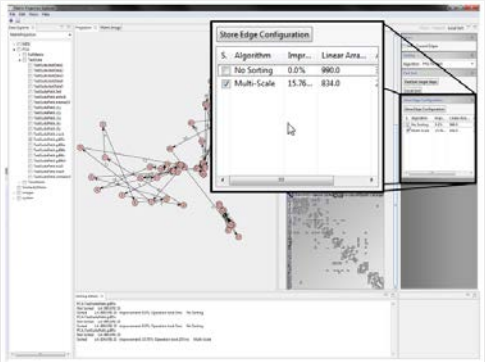
Data Topology can be interpreted.
Structure visible on a Global and Local level.

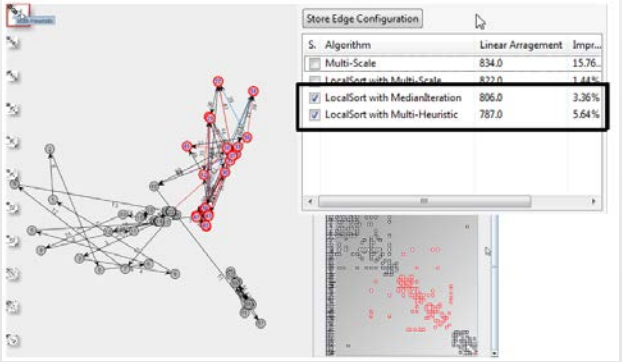
Visual Analytics will change the field (once again).
Opening the Black-Box will lead to novel algorithms.

User-Steerable Iterative Matrix Reordering.

Make matrix sorting an interactively steerable process.





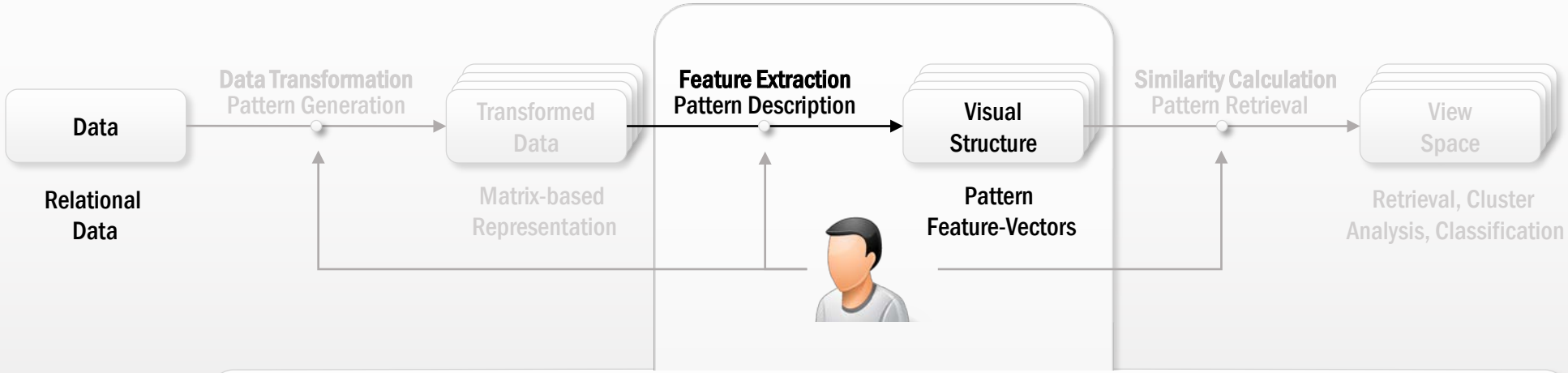


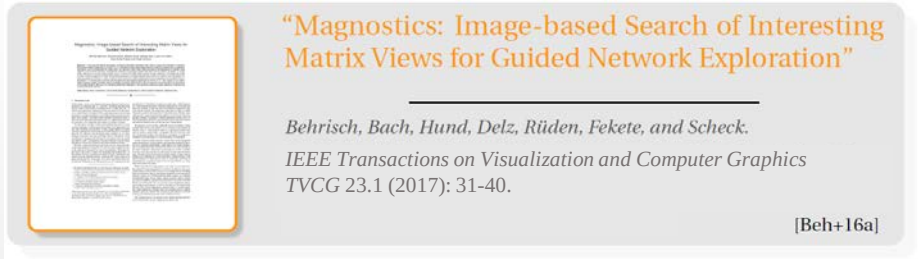
Research Framework.

Pattern-Driven Analysis of Large Matrix-Spaces.

Magnostics

What is a good (Matrix) Representation?



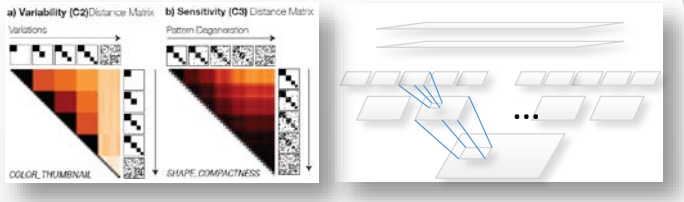


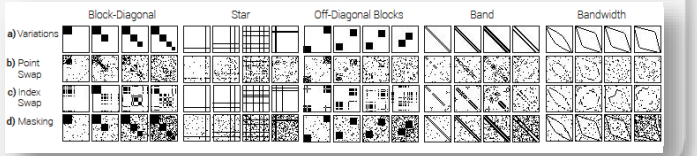
"Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration"

Behrisch, Bach, Hund, Delz, Rüden, Fekete, and Scheck.

IEEE Transactions on Visualization and Computer Graphics TVCG 23.1 (2017): 31-40.

[Beh+16a]

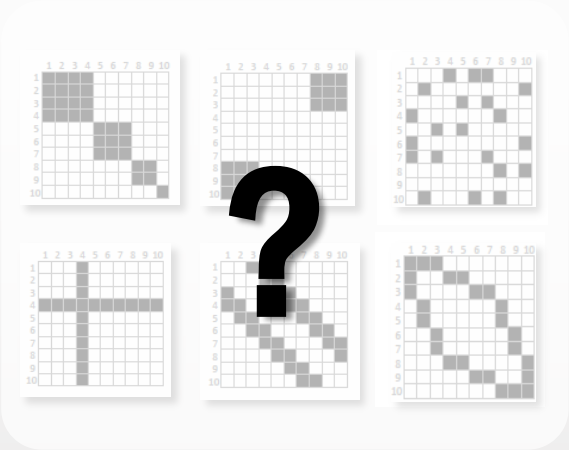




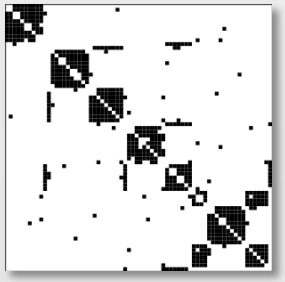
Magnostics.
Research Objectives and Contributions.

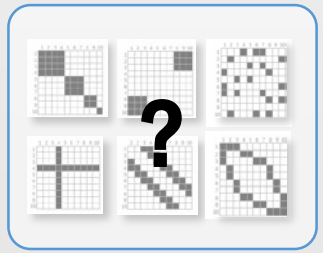
Magnostics

Research Goals and Questions.

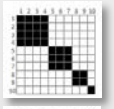


Imaginary Example 1

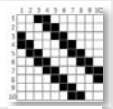




MAGNOSTICS



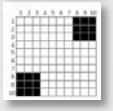
≈ 89%



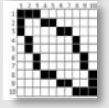
≈ 1%

≈ 6%

≈ 1%



≈ 3%



≈ 0%

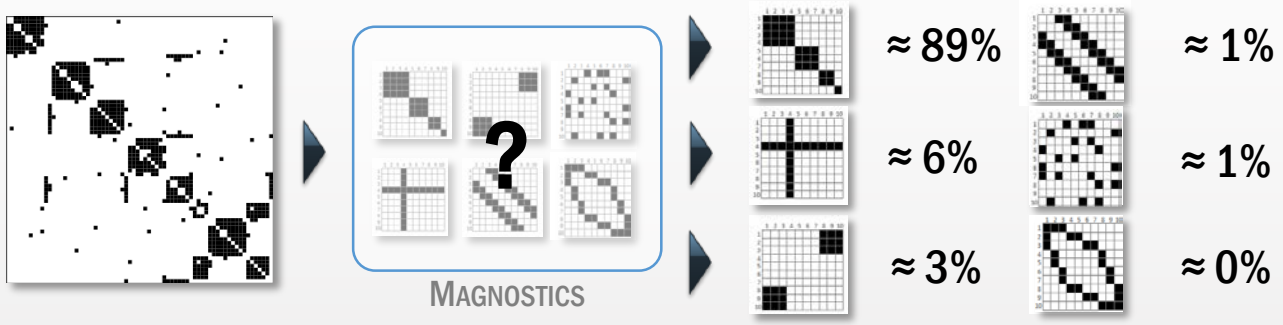
Can we **quantify** the **existence** of **visual patterns** in matrices?

- **C1** Pattern Response
- **C2** Pattern Variability
- **C3** Pattern Sensitivity
- **C4** Pattern Discrimination

Magnostics.
What is it (not)?

Magnostics

Contributions.



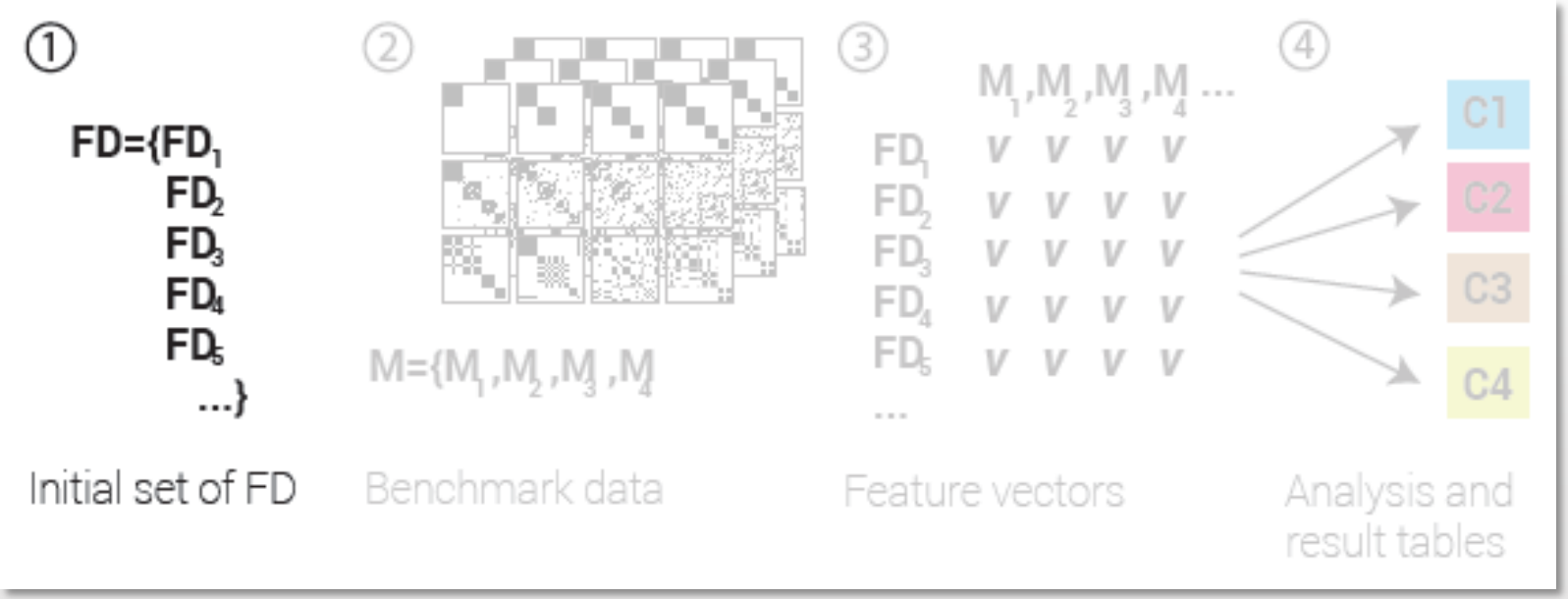
- **Frontend:** Best Performing Image-based Feature Descriptors (FDs).
- **Backend:** Evaluation Framework and Matrix Pattern Benchmark Dataset.

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

Automatic Analysis of Matrix Patterns.



Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

Automatic Analysis of Matrix Patterns.

Feature Descriptor	Group	Reference
GLOBAL_COLOR_HISTOGRAM	Color	[30]
AUTO_COLOR_CORRELOGRAM		[18]
FUZZY_HISTOGRAM		[13]
FUZZY_OPPONENT_HISTOGRAM		[41]
COLOR_OPPONENT_HISTOGRAM		
THUMBNAIL		
MPEG7_COLOR_LAYOUT	Color Layout	
LUMINANCE_LAYOUT		
CEDD		
FCTH		
JCD		
EDGEHIST	Edge	
MPEG7_EDGE_HISTOGRAM		
HOUGH		

Feature Descriptor	Group	Reference
SURF	Point of Interest	[3]
FAST		[29]
BLOCKS	Shape	Java
COMPACTNESS		[25]
ECCENTRICITY		[47]
ADAPTIVE_GRID_RESOLUTION		
JPEG_COEFFICIENT_HISTOGRAM	Structure	
PROFILES		
FRACTAL_BOX_COUNTER		
PHOG		

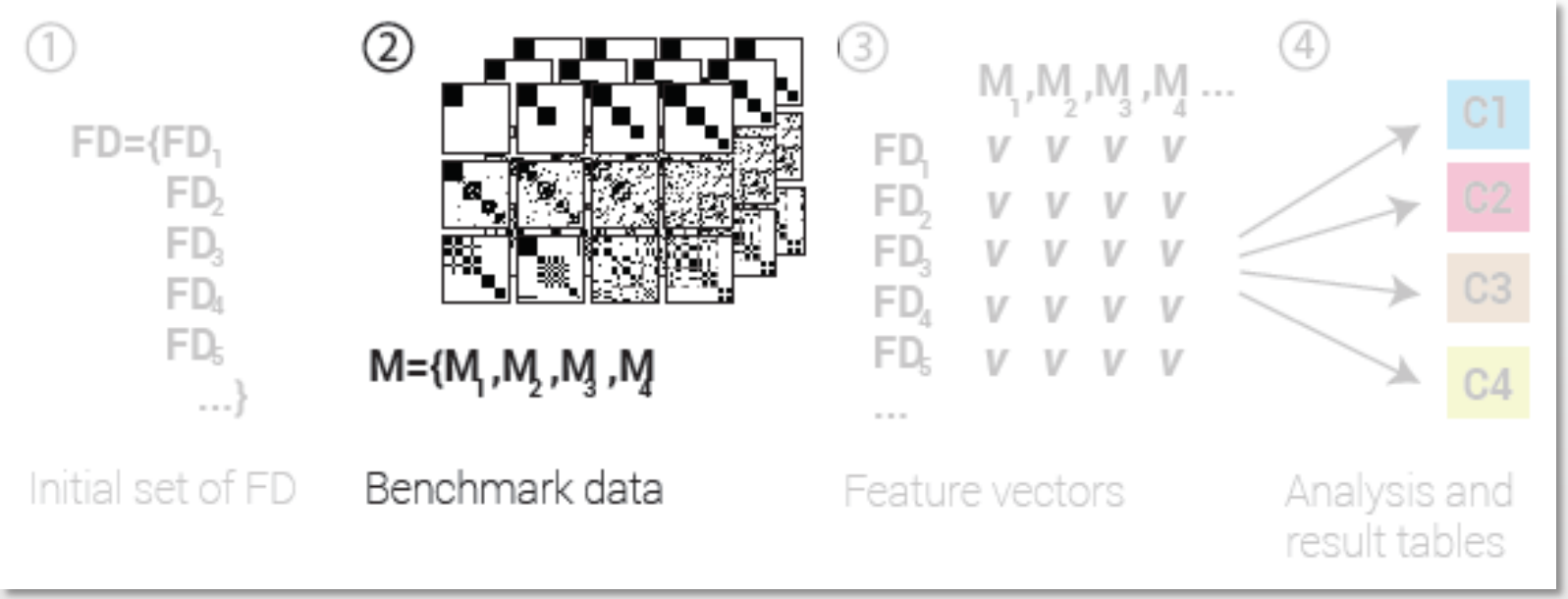
Feature Descriptor	Group	Reference
HARALICK	Texture	[14]
GABOR		[23]
TAMURA		[38]
LOCAL_BINARY_PATTERN		[15]
NOISE_STATISTICAL_SLIDINGWINDOW		Java
NOISE_DISSIMILARITY		[1]
GRADIENT		[42]

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

Automatic Analysis of Matrix Patterns.

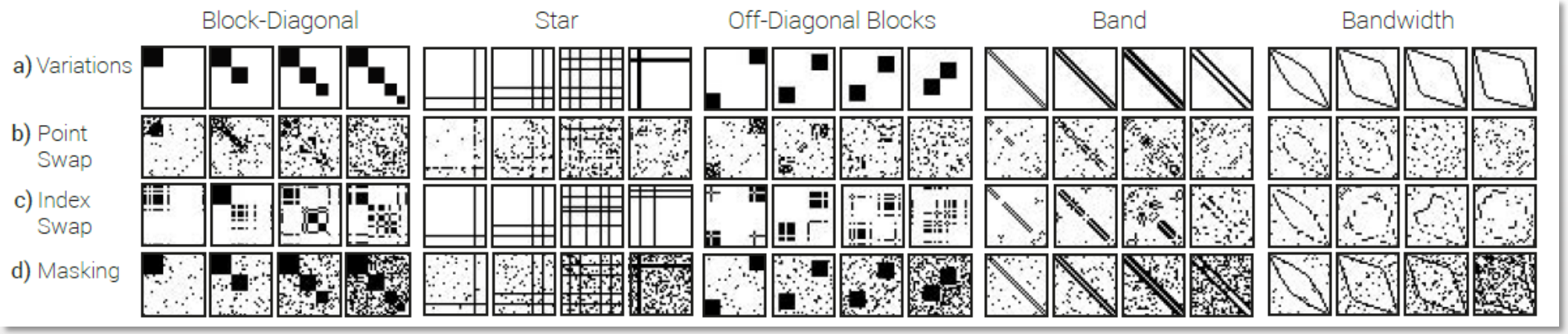


Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

Automatic Analysis of Matrix Patterns.



5570 Artificially Generated Benchmark Matrix Plots

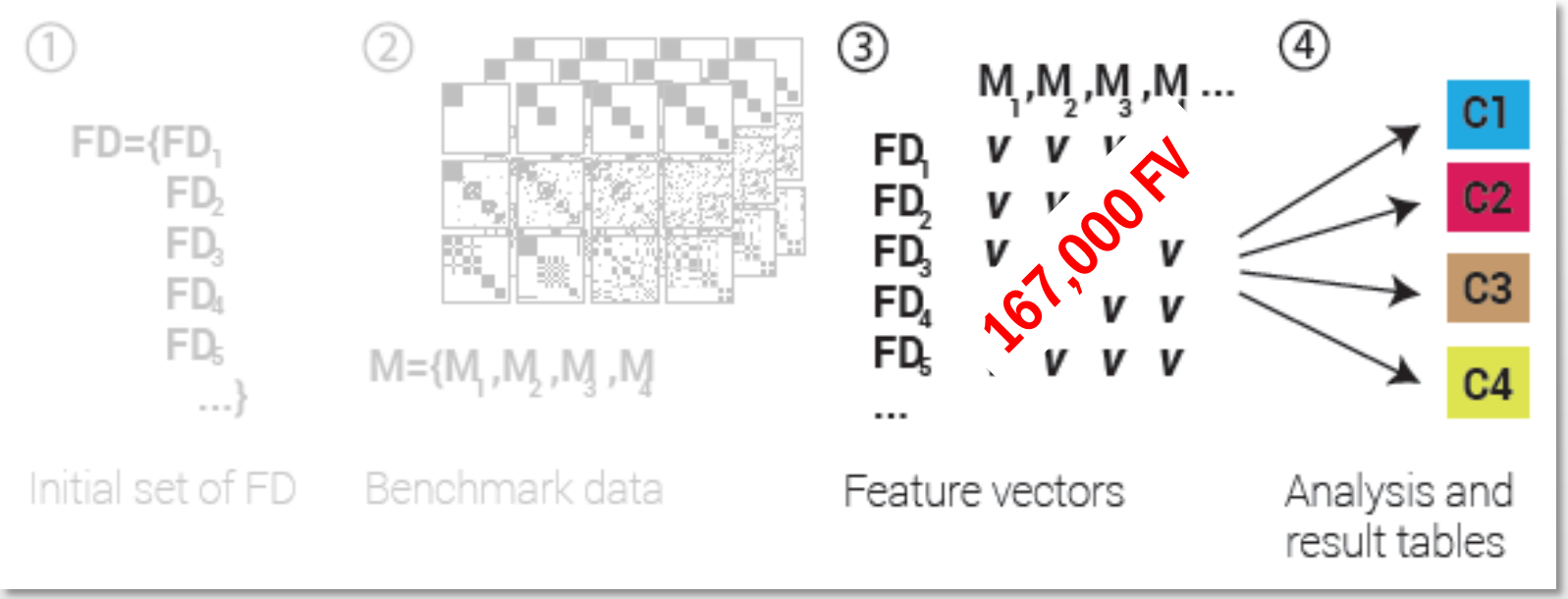
- Five base patterns + one noise antipattern
- Three “degeneration methods”
- Up to ten base pattern variations

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

Automatic Analysis of Matrix Patterns.



Magnostics.

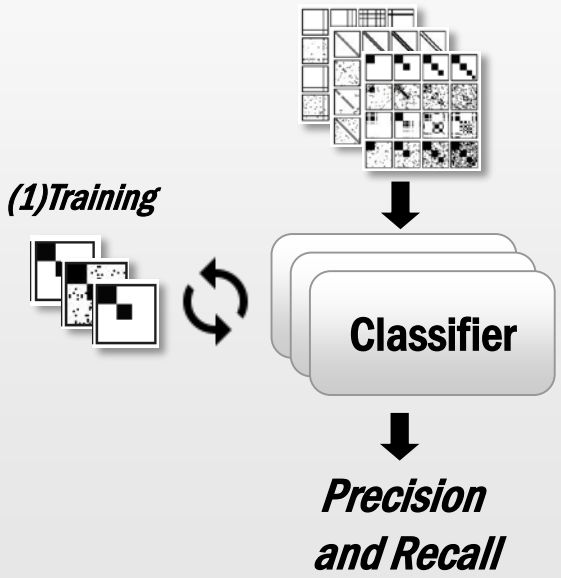
Pattern-Driven Analysis of Matrices.

Magnostics

Evaluation Criteria

C1 **Pattern Response** *To which specific matrix pattern(s) does the FD respond?*

Experiment Setup



Metric Target/Goal

Favoring High Recall over **High Precision**
retrieve a large fraction of all correct answers.

Quality Metric

$$F_\beta = (1+\beta^2)\cdot\frac{\text{precision}\cdot\text{recall}}{(\beta^2\cdot\text{precision})+\text{recall}}$$

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

Evaluation Criteria

C2	**Pattern Variability**	*How much variation in the pattern can the FD detect?*

Magnostics.

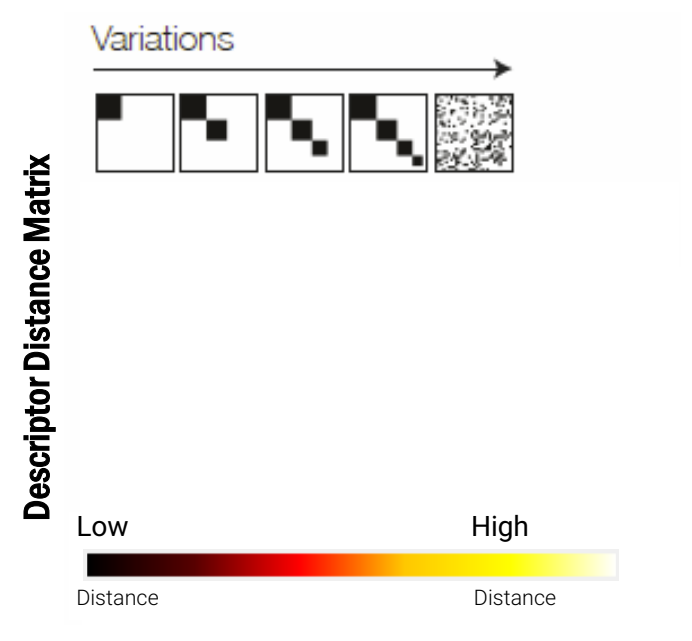
Pattern-Driven Analysis of Matrices.

Magnostics

Evaluation Criteria

C2 **Pattern Variability** *How much variation in the pattern can the FD detect?*

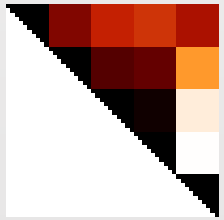
Experiment Setup



Metric Target/Goal

High Standard Deviation represents FDs ability to discriminate the base patterns.

Negative Example



FUZZY_
OPPONENT_
HISTOGRAM

Quality Metric

$$\sigma_{FD_i}(PV) = \sqrt{\frac{1}{|PV|} \sum_{1 \le i,j \le |PV|} \left(ndist(PV_i, PV_j) - \bar{x}\right)^2}$$

Magnostics.

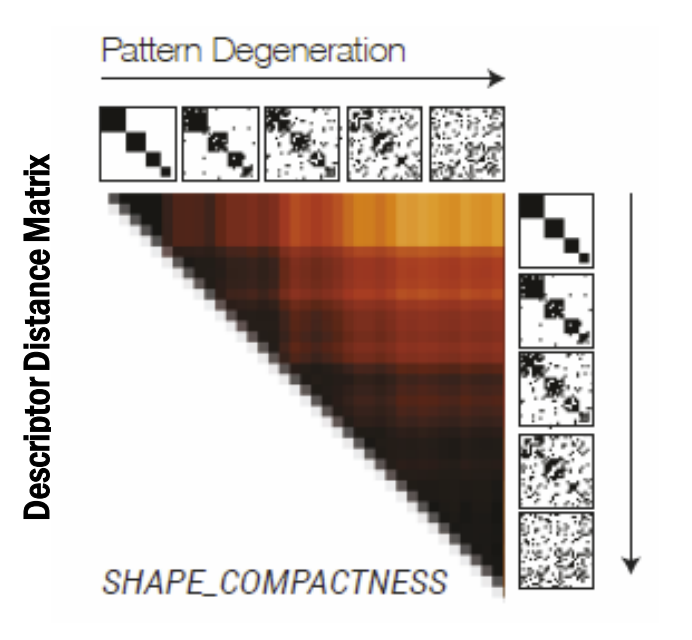
Pattern-Driven Analysis of Matrices.

Magnostics

Evaluation Criteria

C3 **Pattern Sensitivity** *How sensitive is the FD to pattern degeneration (noise)?*

Experiment Setup

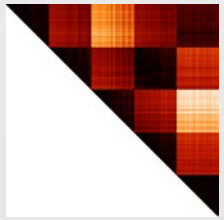


Metric Target/Goal

(1) Monotony degeneration degree reflected by the FD distances

(2) Sensitivity highly sensitive (fast increase of distances) versus tolerant (slow increase of distances).

Negative Example



TEXTURE
GABOR

Quality Metric

(1)
$$monotony(FD_i, P_j) = \begin{cases} 1, & (mean_{l+1} - mean_l + 0.05) > 0; \\ & \forall l \in noise_levels. \\ 0, & otherwise \end{cases}$$

(2)
$$sensitivity(FD_i, P_j) = \sum_{l=0}^{|noise_levels|-1} \left((mean_k - (\frac{l}{|noise_levels|})) \right)$$

Magnostics.

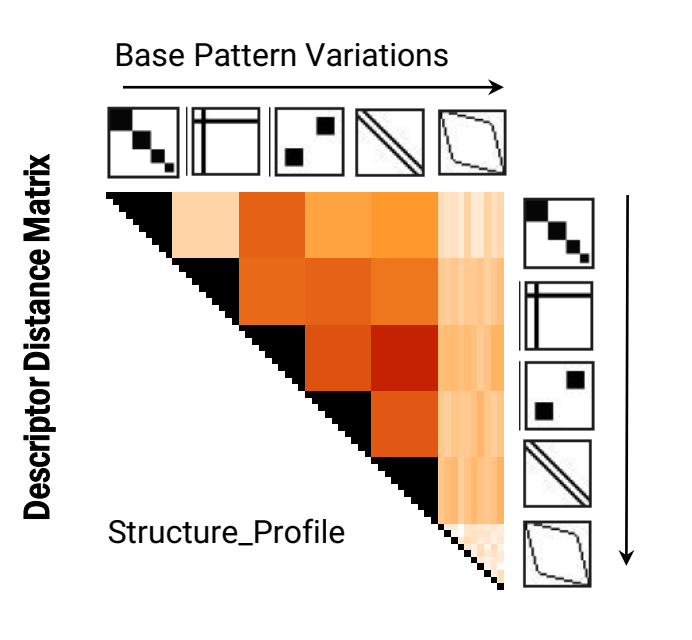
Pattern-Driven Analysis of Matrices.

Magnostics

Evaluation Criteria

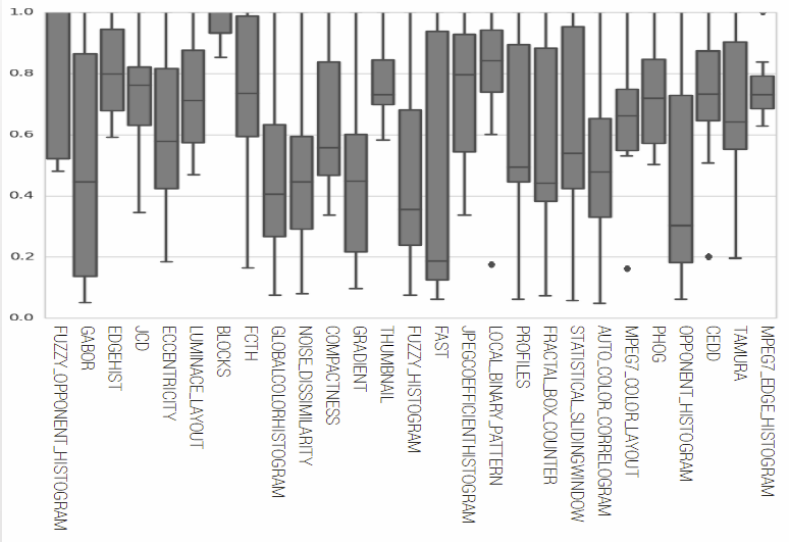
C4 **Pattern Discrimination** *How different does the FD respond to different patterns?*

Experiment Setup



Metric Target/Goal

High Median and **Low Variance** reflects strong discrimination



Negative Example

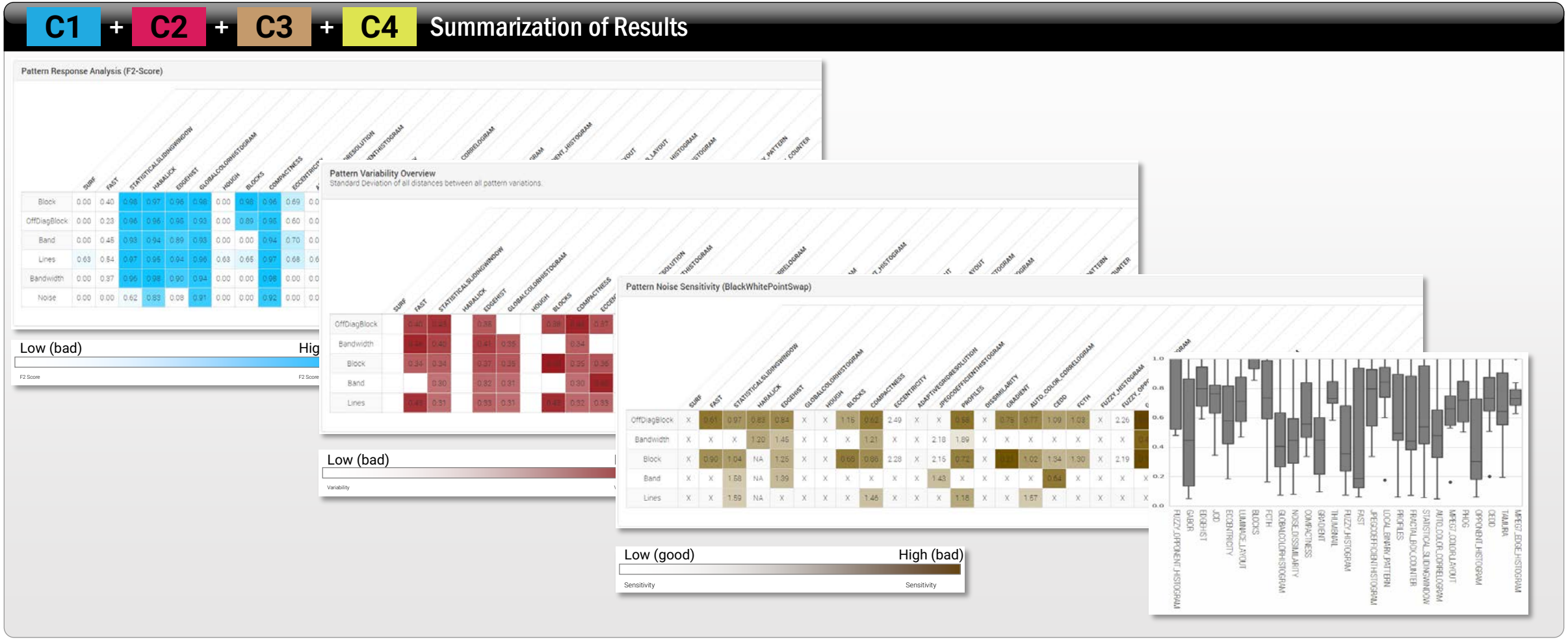
COLOR
FuzzyOpponent
Histogram

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

C1 + C2 + C3 + C4 Summarization of Results

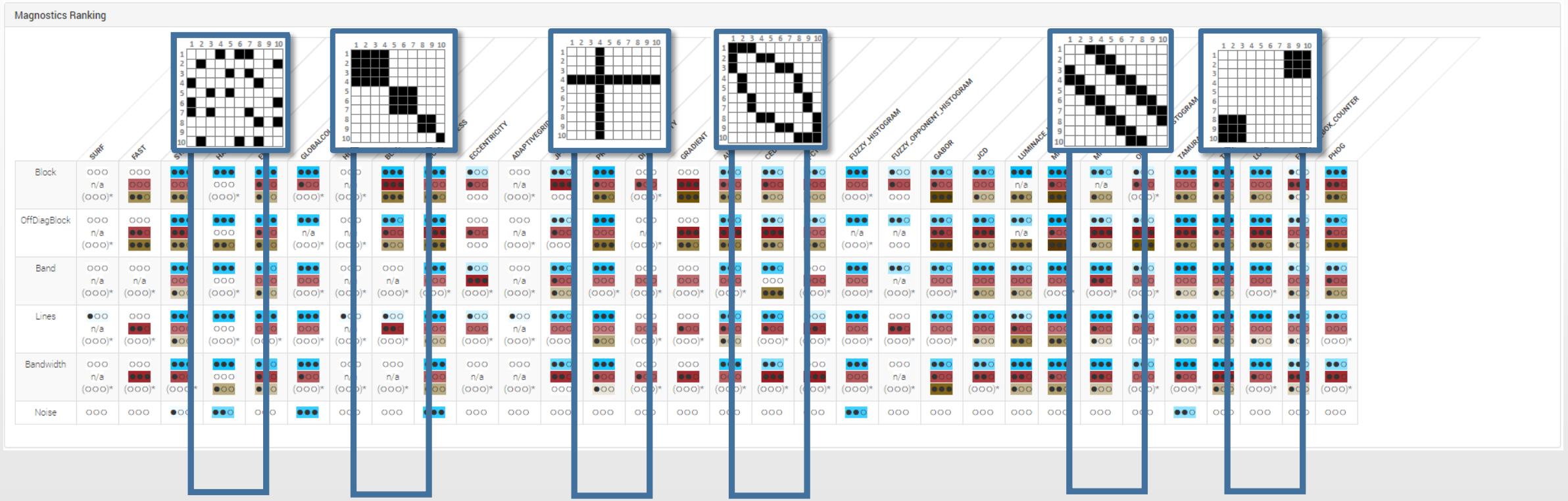


Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

C1 + C2 + C3 + C4 Summarization of Results



Haralick

Blocks

Profiles

CEDD

MPEG7 Edge Histogram

Local Binary Patterns

Magnostics.

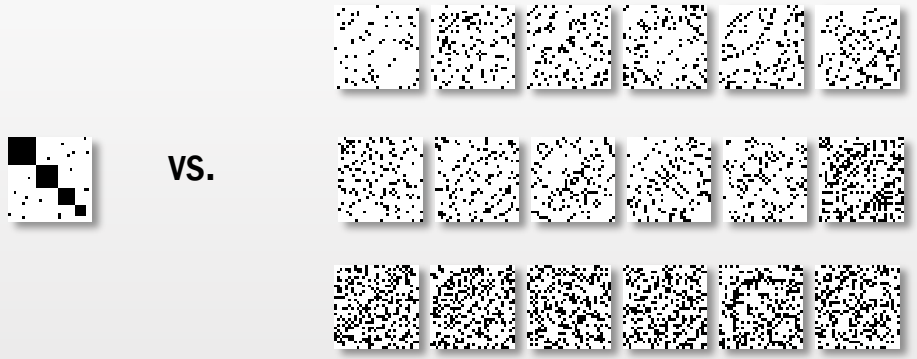
Pattern-Driven Analysis of Matrices.

Magnostics

C1 + C2 + C3 + C4 Visual Inspection of Results

Distance to Noise

Independent from benchmark



Noise more and more visible

Pattern more and more visible

Distance to Base Pattern

More natural understanding, but benchmark dependent

Pattern more and more visible

Noise more and more visible

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

C1 + **C2** + **C3** + **C4** Visual Inspection of Results Discrimination

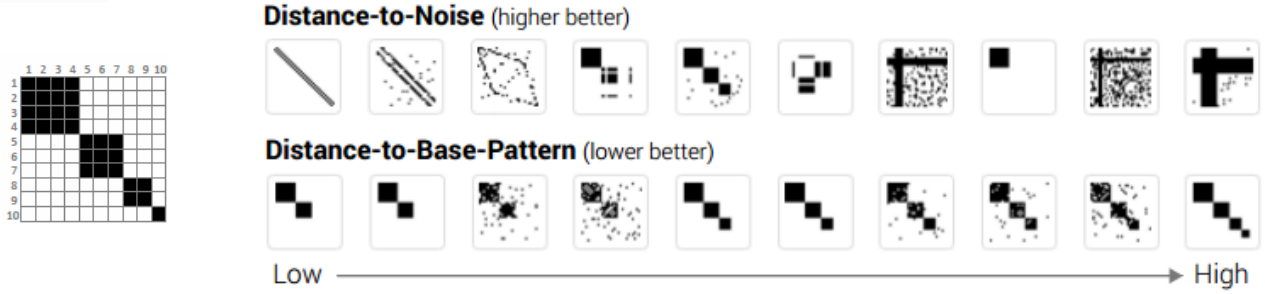


Figure 1.10 Examples for our Block Descriptor, specifically engineered to retrieve blocks around the matrix diagonal.

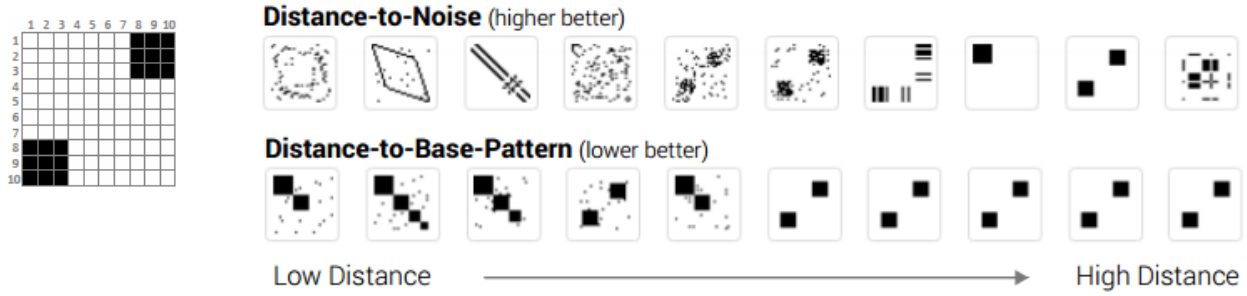


Figure 1.11 Examples for the Local Binary Pattern Descriptor.

Magnostics.

Pattern-Driven Analysis of Matrices.

Magnostics

C1 + **C2** + **C3** + **C4** Visual Inspection of Results Discrimination

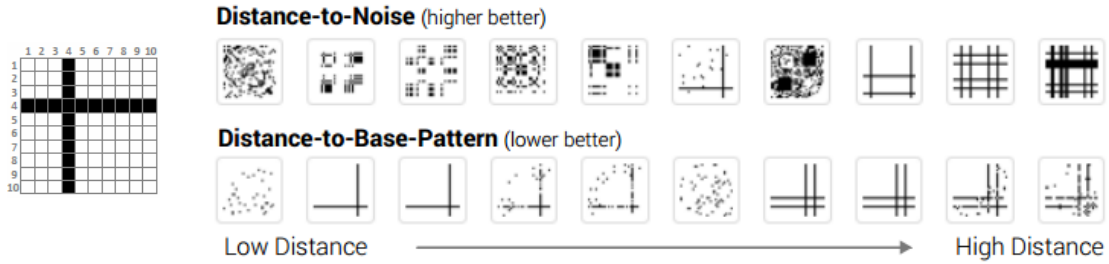


Figure 1.12 Examples for our Profile Descriptor.

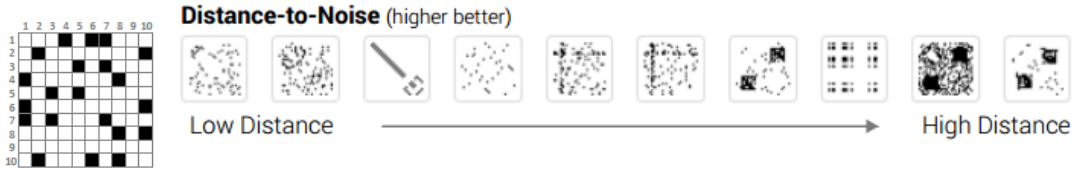


Figure 1.15 Examples for our STATISTICAL SLIDING WINDOW Descriptor.

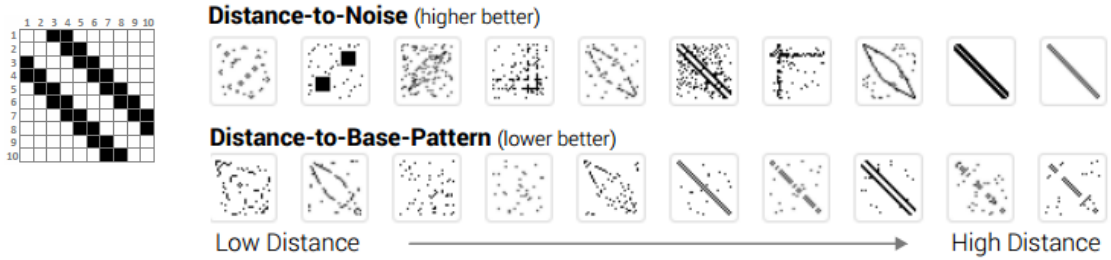


Figure 1.13 Examples for the MPEG7 Edge Histogram Descriptor.

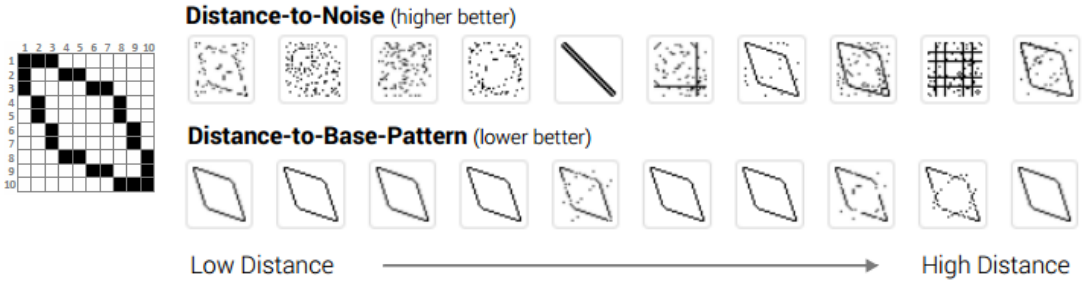


Figure 1.14 Examples for the CEDD Descriptor.

Magnostics.
Conclusion.

Magnostics

Take Home Messages.

- Visual **Matrix Pattern Analysis** is possible.
- Allows for a totally **New Approach** to **Matrix Analysis and Exploration.**
- **Visual Analytics** allows to relate **Pattern** to **User Tasks.**

magnostics.dbvis.de

Magnostics.
Conclusion.

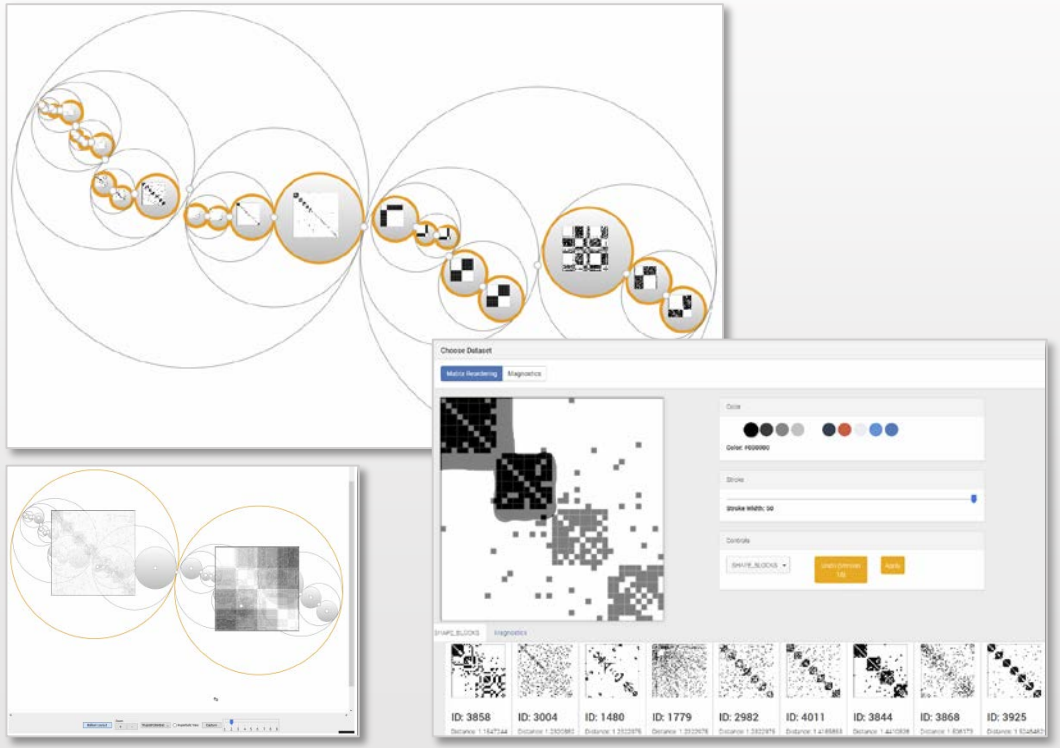
Magnostics

Take Home Messages.

- Visual **Matrix Pattern Analysis** is possible.
- Allows for a totally **New Approach** to **Matrix Analysis and Exploration.**
- **Visual Analytics** allows to relate **Pattern** to **User Tasks.**

Pattern-Driven Exploration Interfaces.

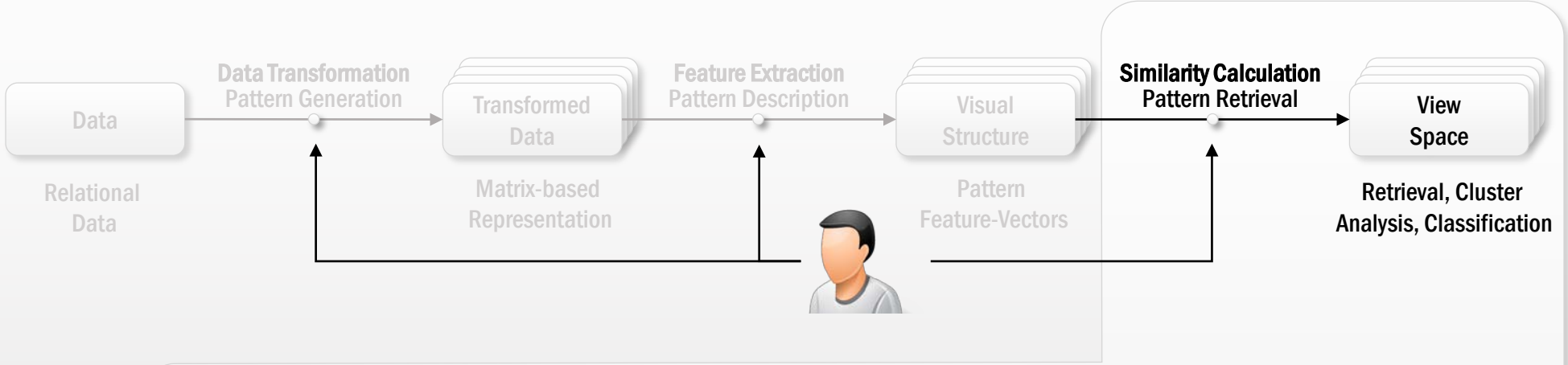
Cluster-by-Visual Patterns or Query-by-Sketch.



Research Framework.

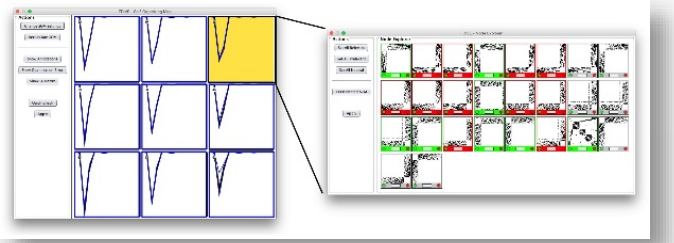
Pattern-Driven Analysis of Large Matrix-Spaces.

What is a good (Matrix) Representation?



"Feedback-Driven Interactive Exploration of Large Multidimensional Data Supported by Visual Classifier"

Behrisch, Korkmaz, Shao, and Schreck.
Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on, IEEE CS Press, 2014, 43-52. [Beh+14a]

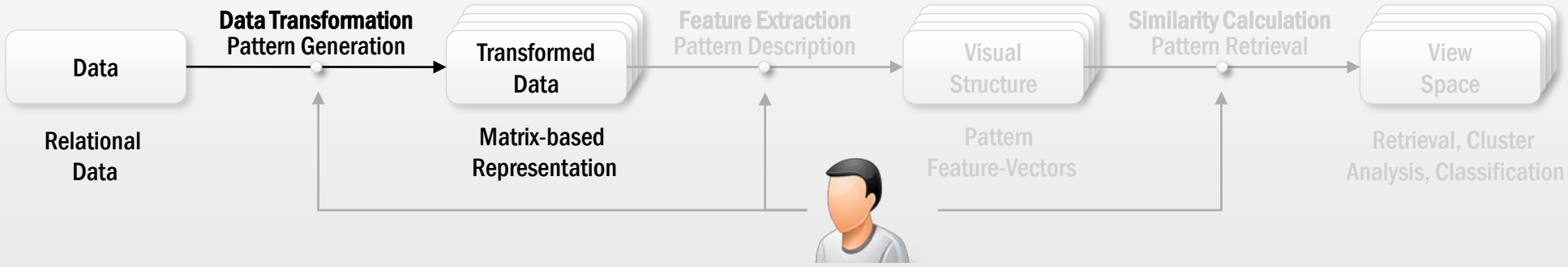


Feedback-Driven Interactive Exploration.

Motivation and Background.

Why Feedback-Driven Interactive Exploration?

Problem Characterization

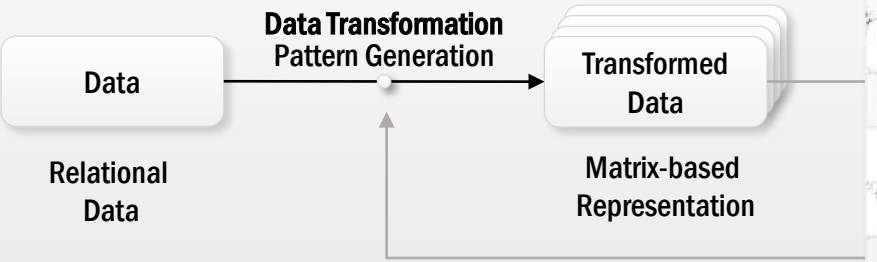


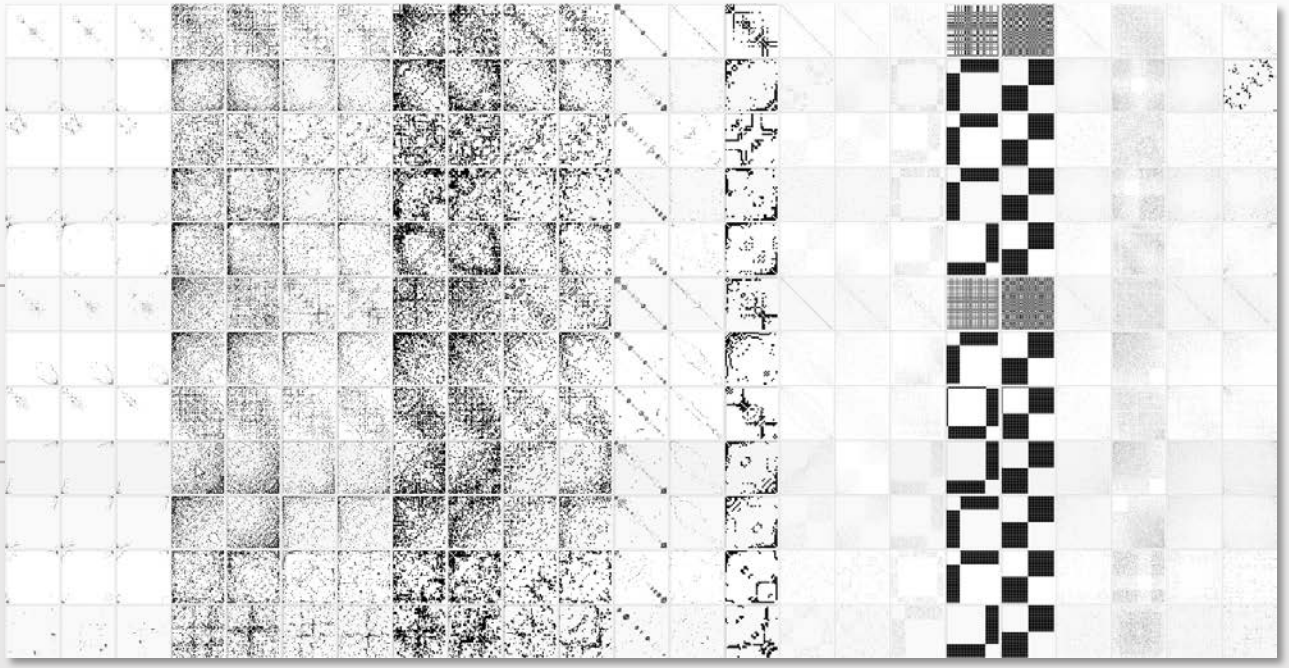
Feedback-Driven Interactive Exploration.

Motivation and Background.

Why Feedback-Driven Interactive Exploration?

Problem Characterization

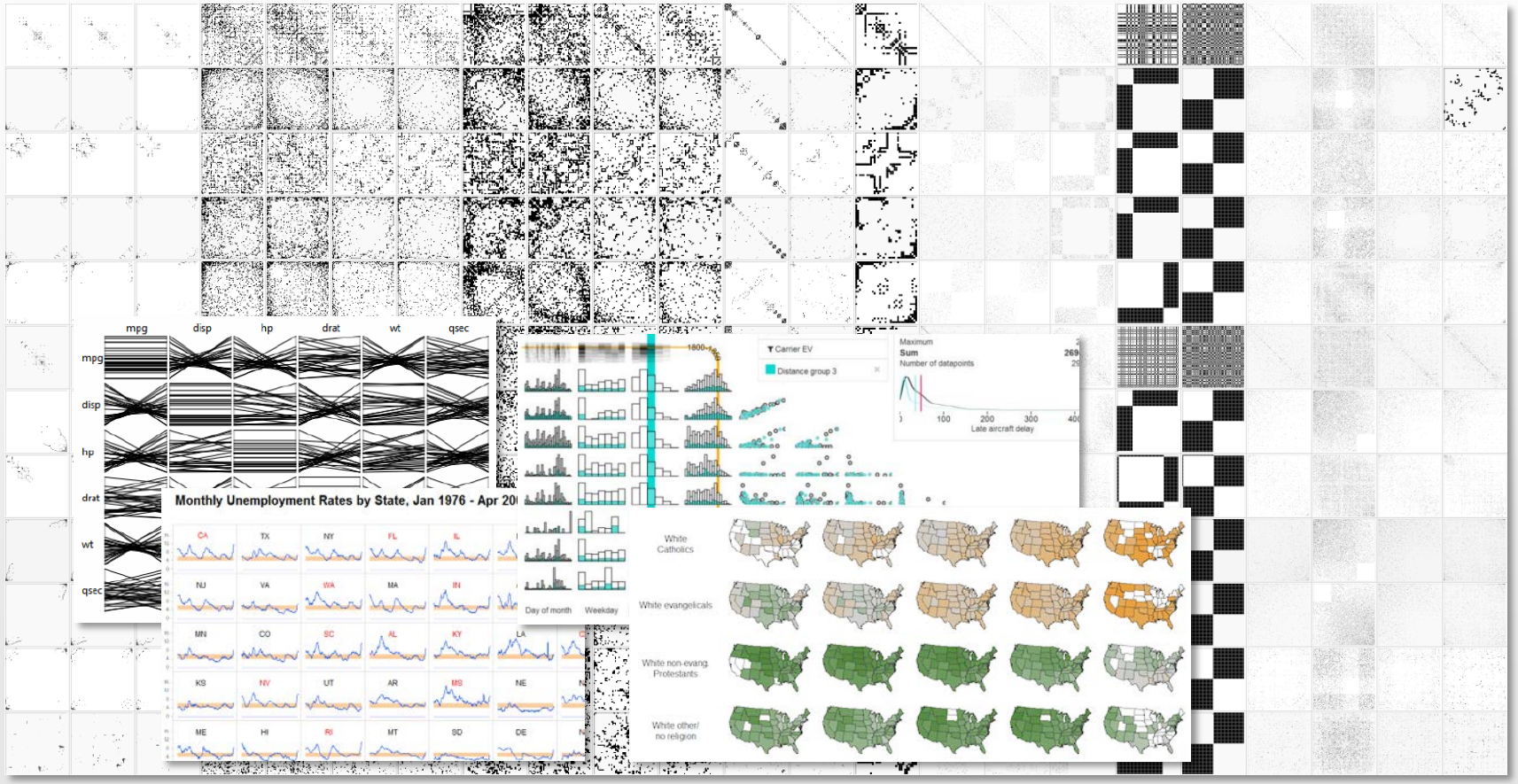




Feedback-Driven Interactive Exploration.

Motivation and Background.

Motivation and Use-Cases.



Images courtesy: 1) Dasgupta, Aritra, and Robert Kosara. "Pargnostics: Screen-space metrics for parallel coordinates." IEEE Transactions on Visualization and Computer Graphics 16.6 (2010): 1017-1026. 2) Im, Jean-François, Michael J. McGuffin, and Rock Leung. "GPLOM: the generalized plot matrix for visualizing multidimensional multivariate data." IEEE Transactions on Visualization and Computer Graphics 19.12 (2013): 2606-2614. 3) Better Know a Visualization: Small Multiples Zach Gemignani: http://www.juiceanalytics.com/writing/better-know-visualization-small-multiples

Feedback-Driven Interactive Exploration.

Motivation and Background.

VA

Problem Statement.

In large View Spaces only a Few Views are Interesting

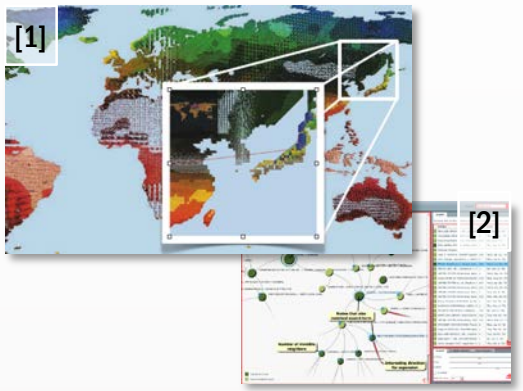
Images courtesy: 1) Dasgupta, Aritra, and Robert Kosara. "Pargnostics: Screen-space metrics for parallel coordinates." IEEE Transactions on Visualization and Computer Graphics 16.6 (2010): 1017-1026. 2) Im, Jean-François, Michael J. McGuffin, and Rock Leung. "GPLOM: the generalized plot matrix for visualizing multidimensional multivariate data." IEEE Transactions on Visualization and Computer Graphics 19.12 (2013): 2606-2614. 3) Better Know a Visualization: Small Multiples Zach Gemignani: http://www.juiceanalytics.com/writing/better-know-visualization-small-multiples

Feedback-Driven Interactive Exploration.

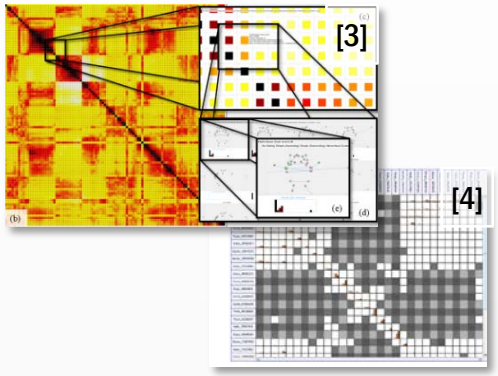
Motivation and Background.

VA

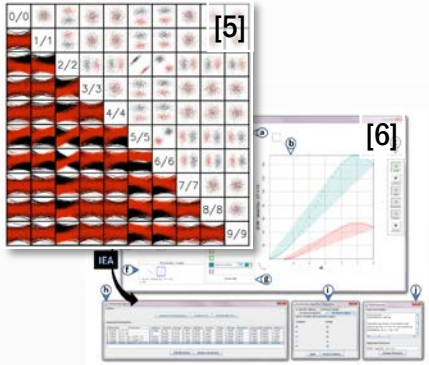
Why Feedback-Driven Interactive Exploration?



Focus+Context



Interaction Concepts



Quality Metrics / Subspace Search

- **Misleading abstractions** can ultimately lead to wrong exploration paths.
- **Query formulation problematic**, initial understanding required.
- Heuristic approaches **can not grasp user intention** and/or non-interactive.

[1] Healey, Christopher G., and Brent M. Dennis. "Interest driven navigation in visualization." TVCG 18.10 (2012)
[2] Van Ham, Frank, Adam Perer. "“Search, show context, expand on demand”: supporting large graph exploration with degree-of-interest." TVCG 15.6 (2009).
[3] Behrisch, Michael, et al. “VA of Sets of Heterogeneous Matrices Using Projection-Based Distance Functions and Semantic Zoom." EuroVis 2014.
[4] Elmqvist, Niklas, et al. "ZAME: Interactive large-scale graph visualization." Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific. IEEE, 2008.
[5] Tatu, Andrada, et al. "Automated Visual Analysis Methods for an Effective Exploration of High-Dimensional Data.." TVCG 17.5 (2011): 584-597.
[6] Cancino, Waldo, et al. "Evolutionary visual exploration: experimental analysis of algorithm behaviour." Proceeding of the 5th Genetic and evolutionary computation conference. ACM, 2013.

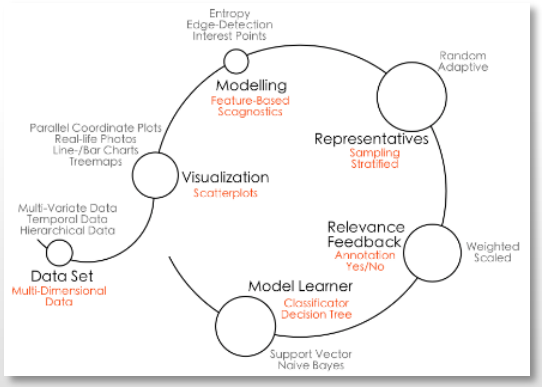
Feedback-Driven Interactive Exploration.

Approach.

VA

Novel Approach towards the Interesting View Problem.

FB-Driven Interactive Expl. of Large Multidimensional Data Supported



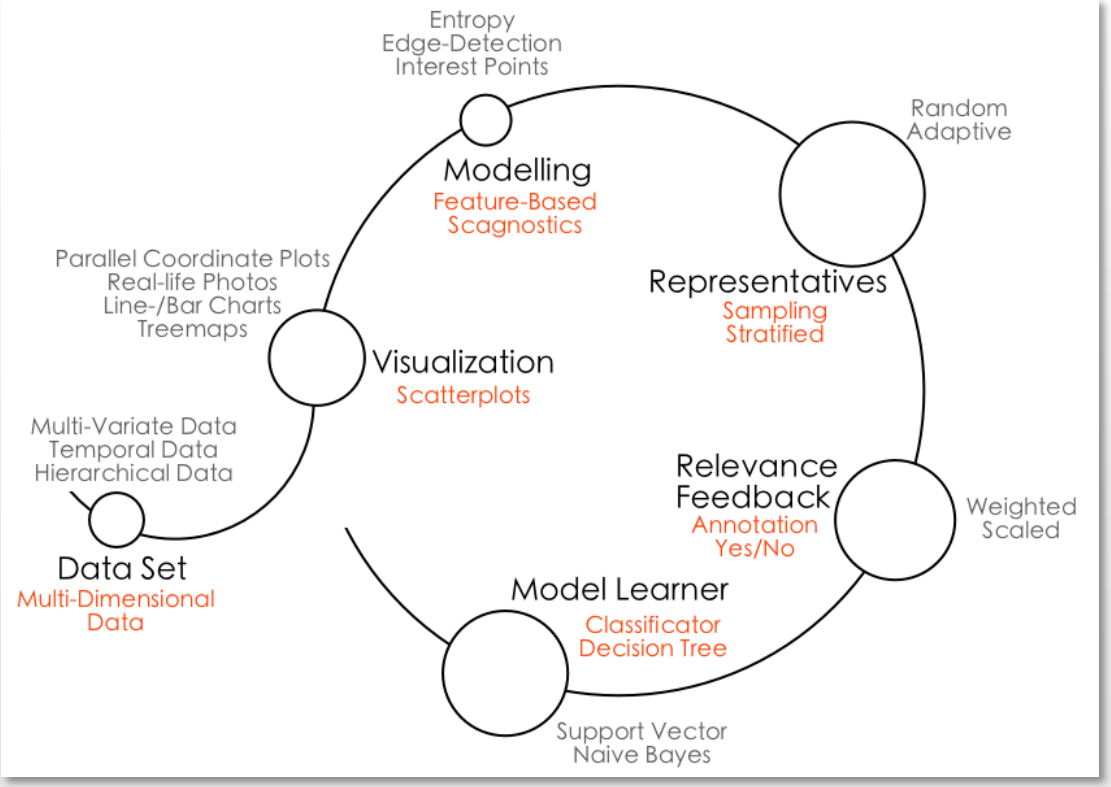
- Framework for an interactive RF driven data exploration.
- Inspired by Active Learning Approaches

Interesting Reference: Van Ham, Frank, and Adam Perer. "**Search, show context, expand on demand**": supporting large graph exploration with degree-of-interest. IEEE Transactions on Visualization and Computer Graphics 15.6 (2009).

Feedback-Driven Interactive Exploration.

Approach.

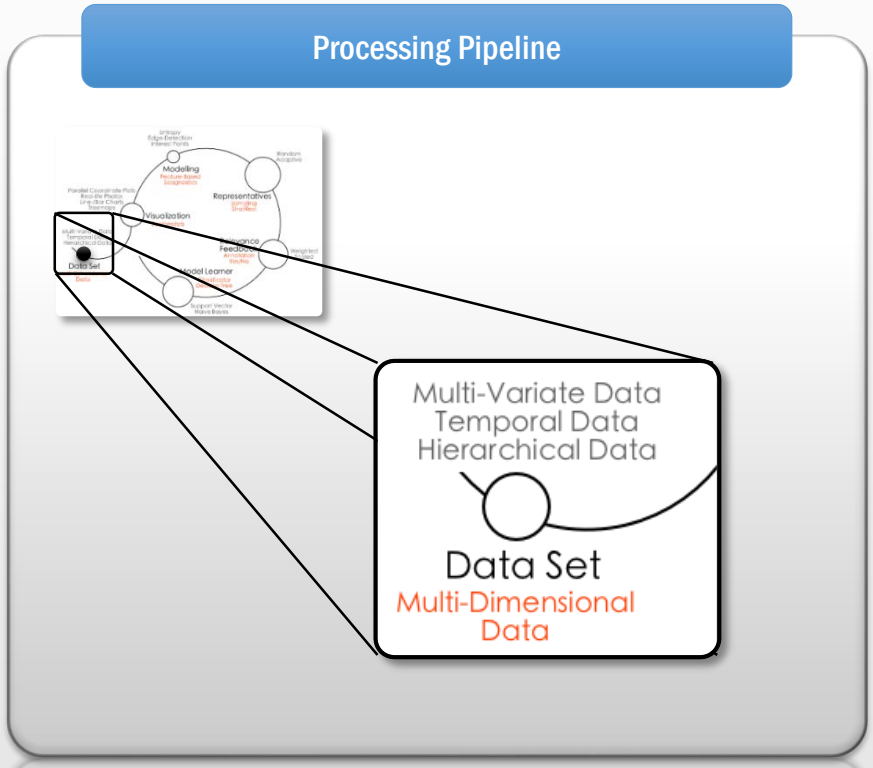
Processing Pipeline

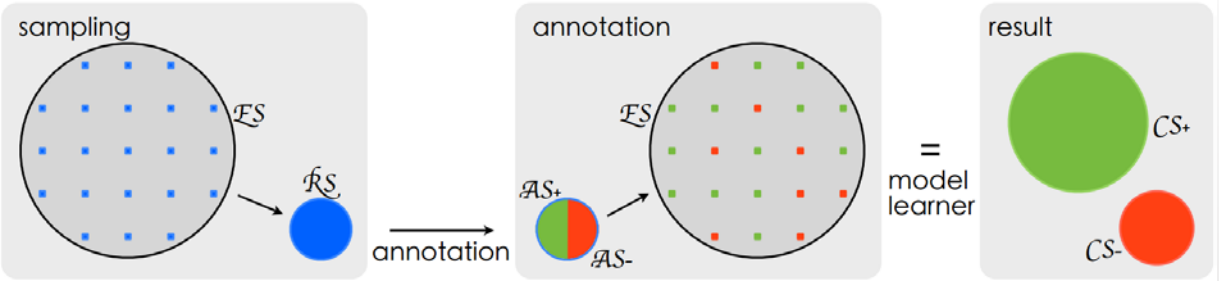


Feedback-Driven Interactive Exploration.

Processing Pipeline.

Data Set – General Idea – Definitions





RS = Representation Set

ES = Exploration Set

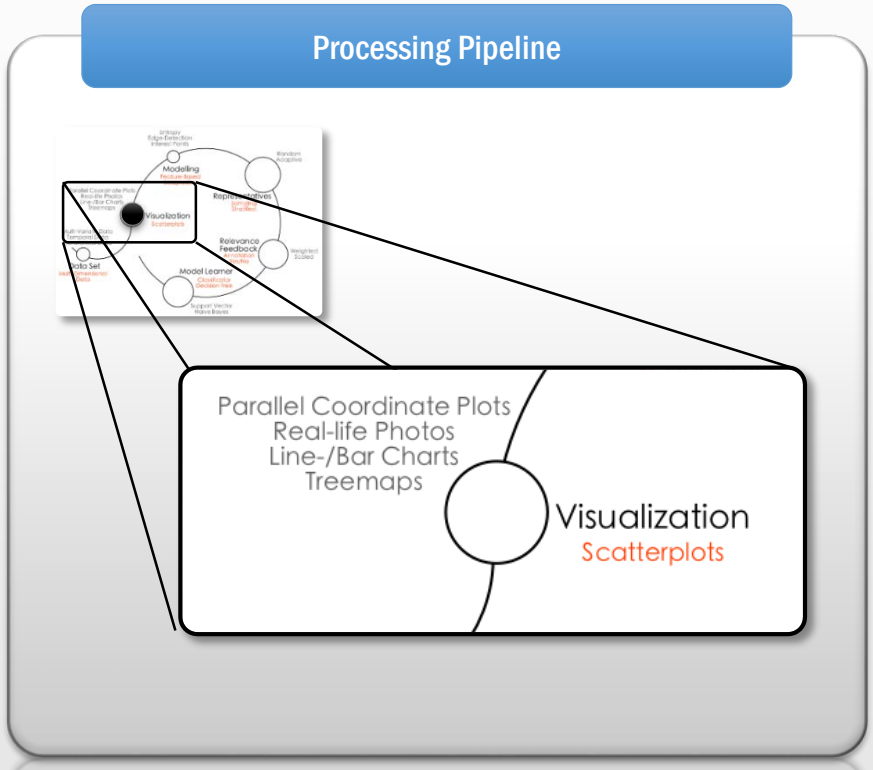
AS^+/AS^- = Positively/Negatively Annotated Items

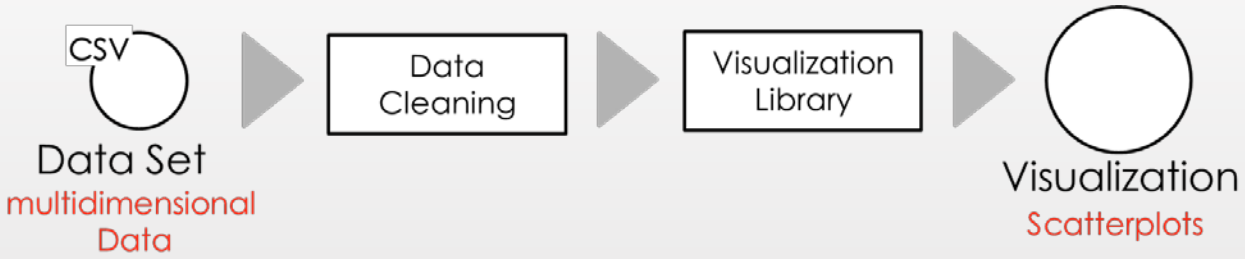
CS^+/CS^- = Positively/Negatively Classified Items

Feedback-Driven Interactive Exploration.

Processing Pipeline.

Visualization – View Space



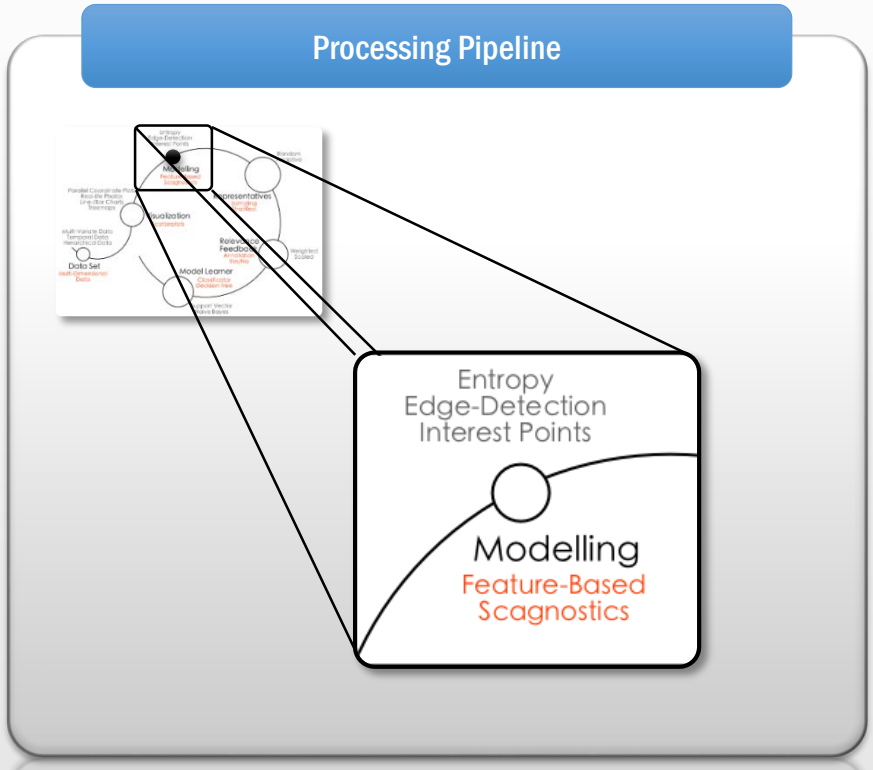


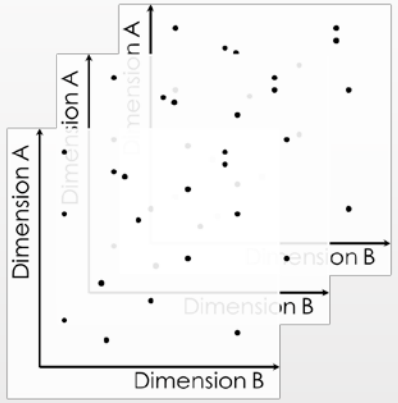
Feedback-Driven Interactive Exploration.

Processing Pipeline.

VA

Modelling of the View Space





Feature Extraction
(Scagnostics)

(Quasi-) Semantic properties [1]
Regressional Features [2].

$FV_1 = \langle s_0, s_1, \ldots, s_8 \rangle$
$FV_2 = \langle s_0, s_1, \ldots, s_8 \rangle$
$\vdots$
$FV_n = \langle s_0, s_1, \ldots, s_8 \rangle$

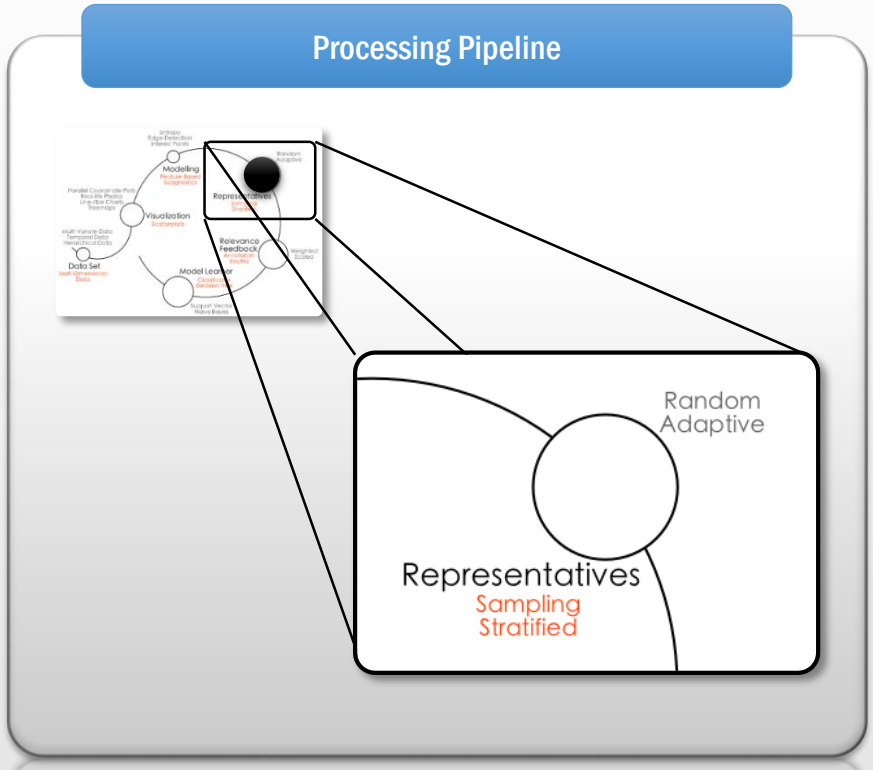
Feature Vectors

[1] Visual Document Analysis: Towards a Semantic Analysis of Large Document Collections, Daniela Oelke, PhD thesis 2010.
[2] Scherer, Maximilian, Jürgen Bernard, and Tobias Schreck. "Retrieval and exploratory search in multivariate research data repositories using regressional features." Proc. 11th ACM/IEEE Digital libraries. 2011.

Feedback-Driven Interactive Exploration.

Processing Pipeline.

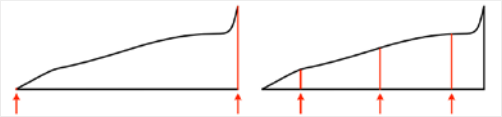
Finding Representatives



$FV_1 = \langle s_0, s_1, \ldots, s_8 \rangle$
$FV_2 = \langle s_0, s_1, \ldots, s_8 \rangle$
$\vdots$
$FV_n = \langle s_0, s_1, \ldots, s_8 \rangle$

All Feature Vectors
(exploration set)

Sampling method



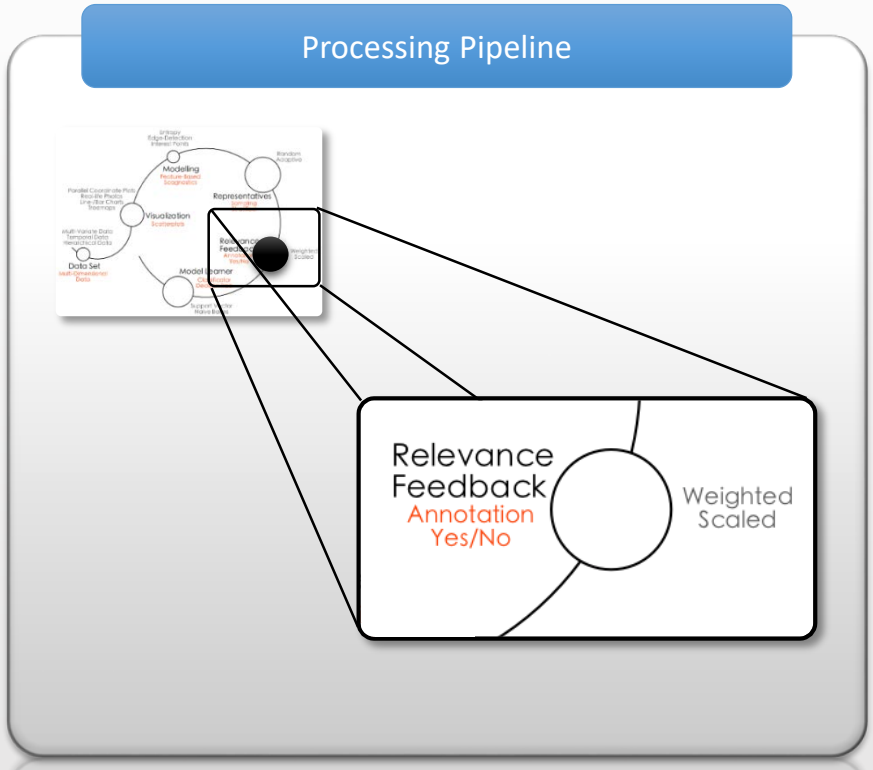
$FV_1 = \langle s_0, s_1, \ldots, s_8 \rangle$
$FV_{23} = \langle s_0, s_1, \ldots, s_8 \rangle$
$FV_{59} = \langle s_0, s_1, \ldots, s_8 \rangle$

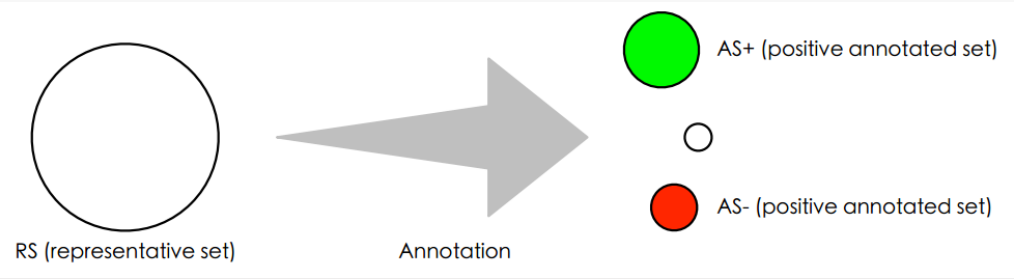
Selection of Feature Vectors
(representative set)

Feedback-Driven Interactive Exploration.

Processing Pipeline.

Receiving User Feedback.

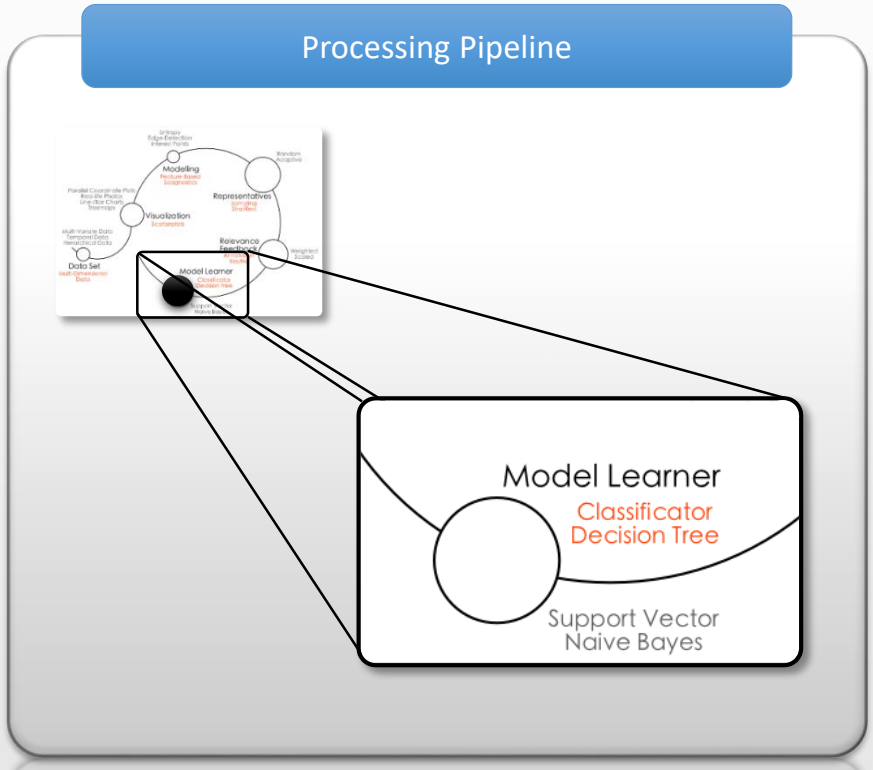


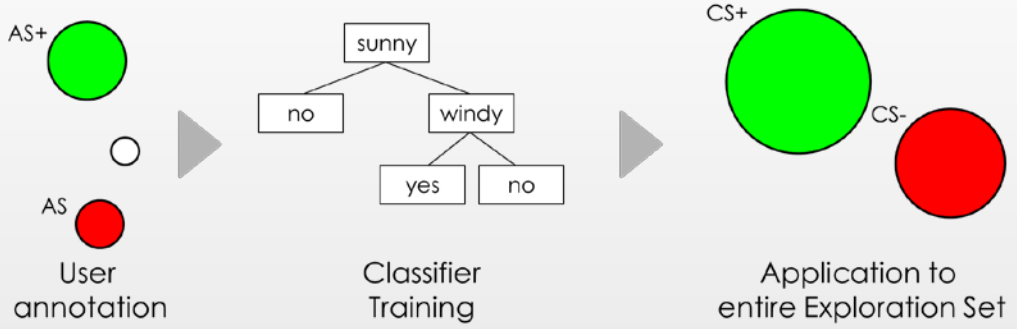


Feedback-Driven Interactive Exploration.

Processing Pipeline.

Model Learning Approach to Reflect the User's Preferences



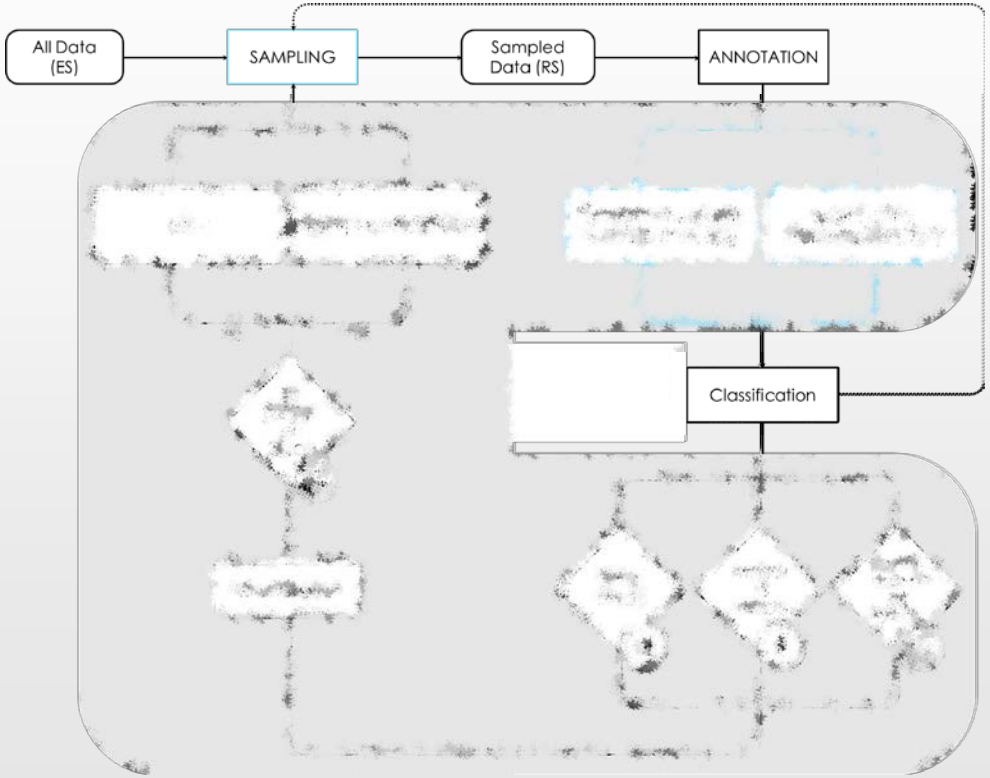


[1] H. Yang and S. Fong. Incrementally optimized decision tree for noisy big data. In Proceedings of the 1st International Workshop on Big Data,Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, BigMine '12, pages 36–44, New York, NY, USA, 2012. ACM

Feedback-Driven Interactive Exploration.

Decision Support.

Decision Support System View on the Framework



Decision Support System

- Contribute to the **Efficiency** of the Exploration
- Contribute to the **Effectiveness** of the Exploration
- Contribute to the **Trust** in the System

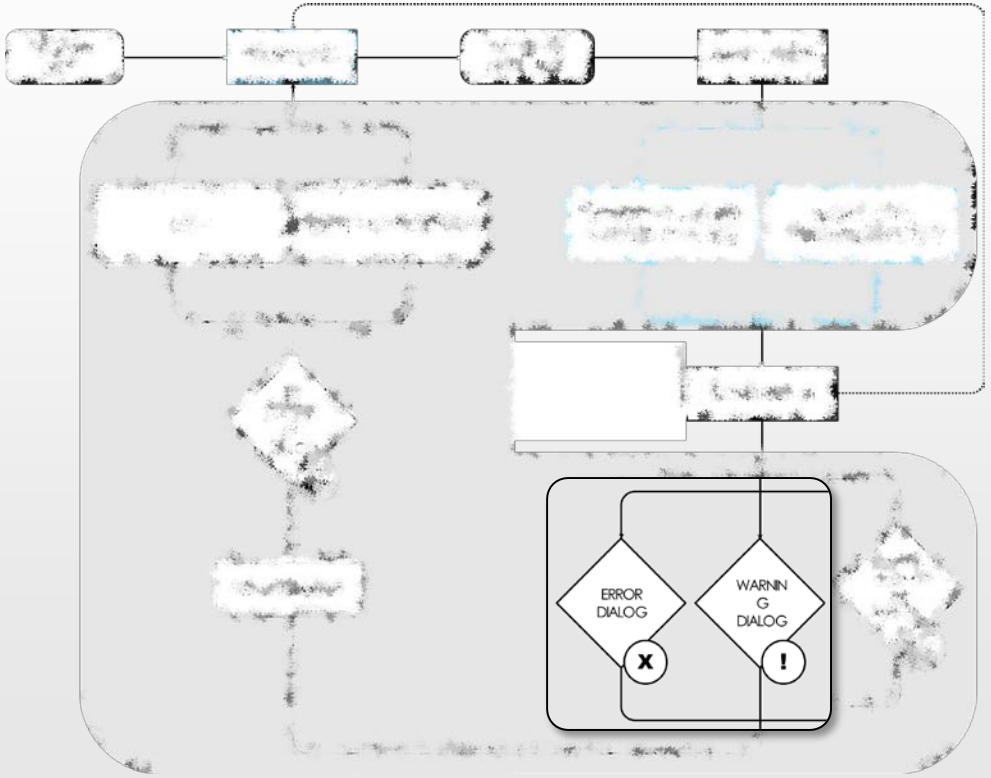
- **Transparency** important! Forecast potential intervention outcome.

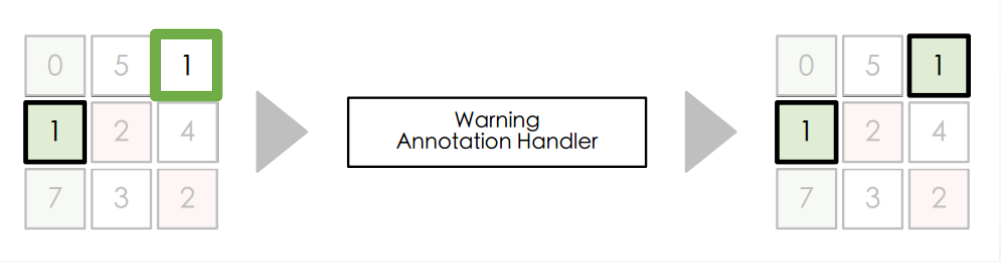
Feedback-Driven Interactive Exploration.

Decision Support.

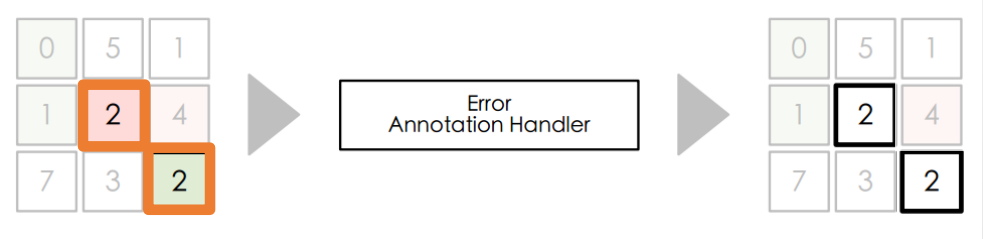
VA

Decision Support System - Dialogs





- Incomplete annotations are tracked and a warning is raised for the user
- In case of multiple occurrences Decision Support enables auto-highlighting



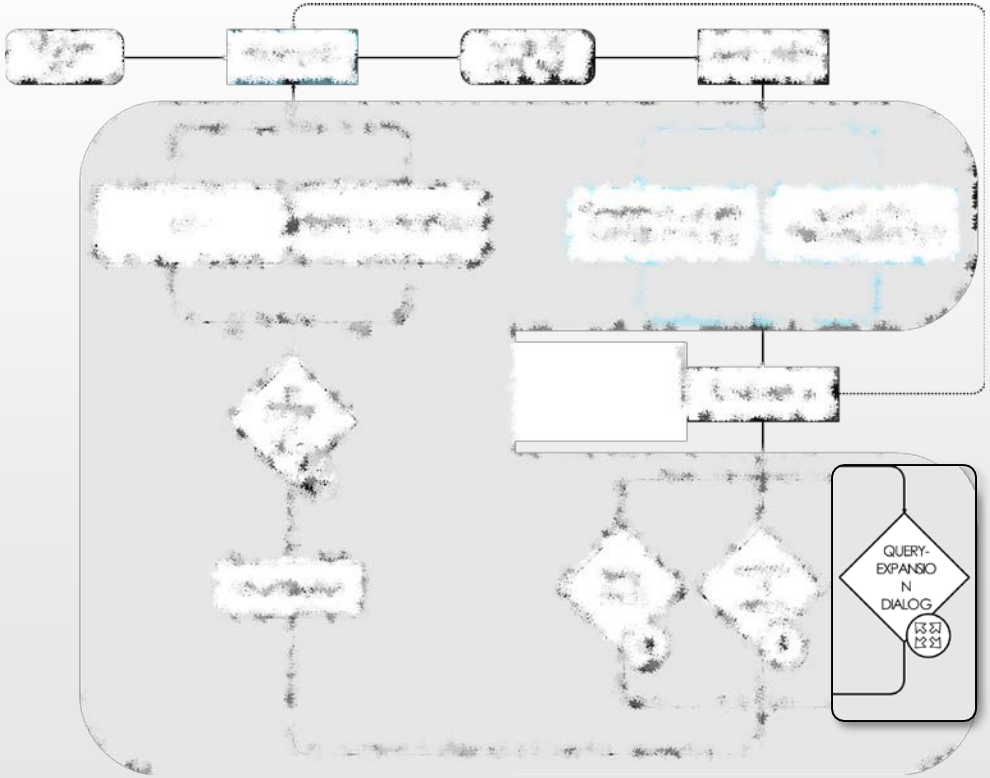
- User has to annotate related scatterplots again
- A quick-check functionality allows to calculate the implications on the fly

Feedback-Driven Interactive Exploration.

Decision Support.

VA

Decision **Support** System – Reducing the work load.



- Potentially promising solutions can be recommended

  Positive/Negative recommendations:

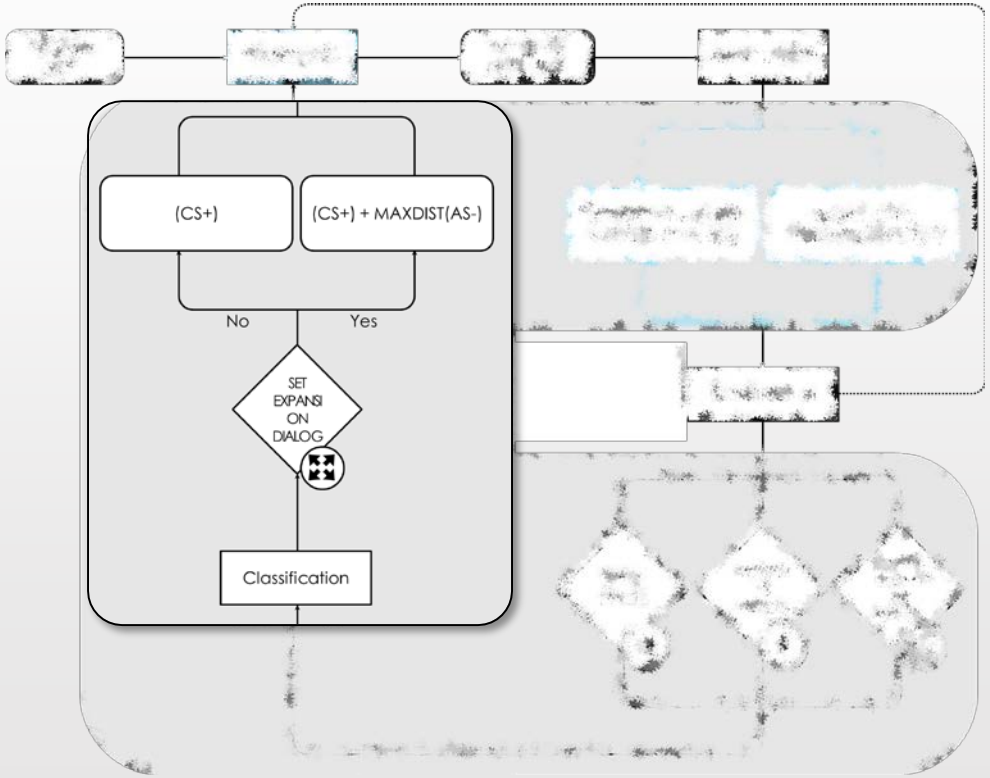
$$Candidate^{\pm} = \{x \mid \forall x \in ES, y \in AS^{\pm} \wedge \min\big(dist(x,y)\big) \wedge dist(x,y) < \alpha\}$$

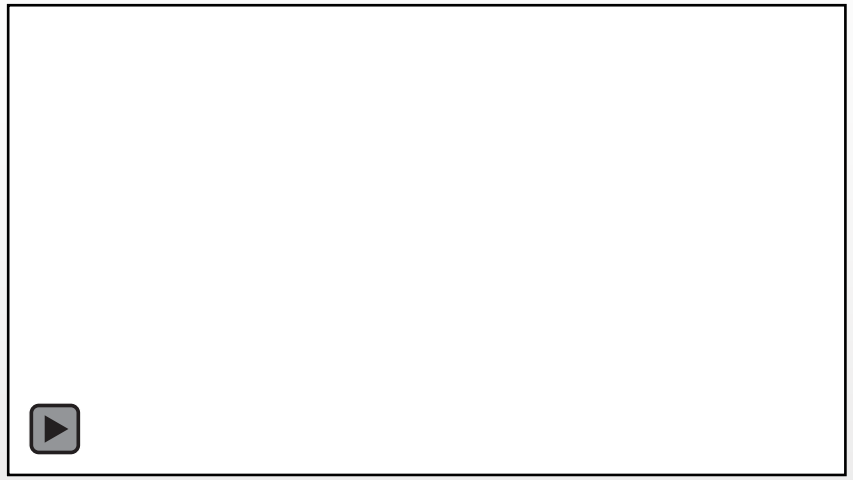
Feedback-Driven Interactive Exploration.

Decision Support.

VA

Decision Support System – Search stability and convergence





- Expansion of the exploration set due to classifier uncertainty.
- Set will with expanded all uncertain decisions MAXDIST(AS-).
- The further the exploration the less we interrupt the user.

Uncertain decisions:

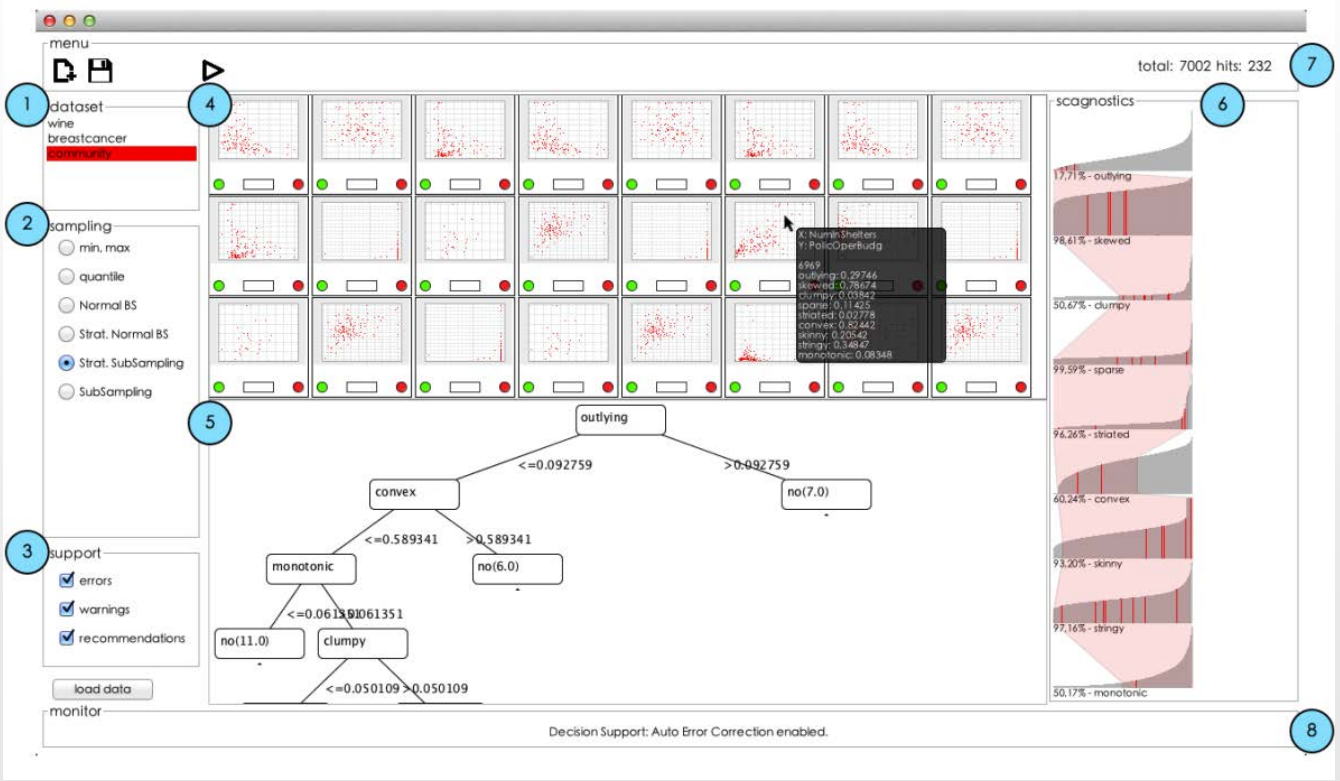
$$MaxDist(AS^{-}) = \left\{ \begin{matrix} x \mid \exists x \in (CS^{-} \setminus AS^{-}), \forall y \in AS^{-} \wedge \\ \min(dist(x,y)) \wedge dist(x,y) > \beta \end{matrix} \right\}$$

Feedback-Driven Interactive Exploration.

Prototypical Framework Instantiation.

VA

Visualization – First Design Alternatives Implementation



VIDEO

Feedback-Driven Interactive Exploration.

Prototypical Framework Instantiation.

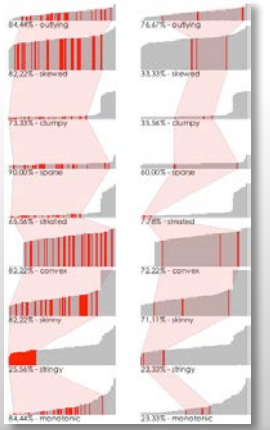
Feedback-Driven Interactive Exploration.

Prototypical Framework Instantiation.

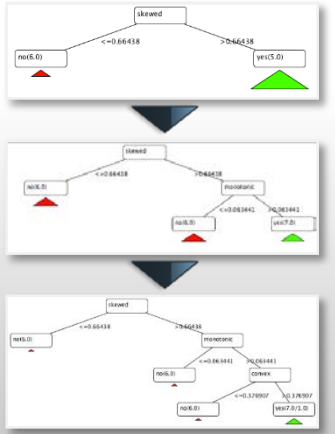
VA

Exploration Path Convergence

Feature Tube Appearance



Decision Tree Appearance



Annotation Set Divergence

$$dist(AS_1^+, AS_2^+)$$

$$dist(AS_1^-, AS_2^-)$$

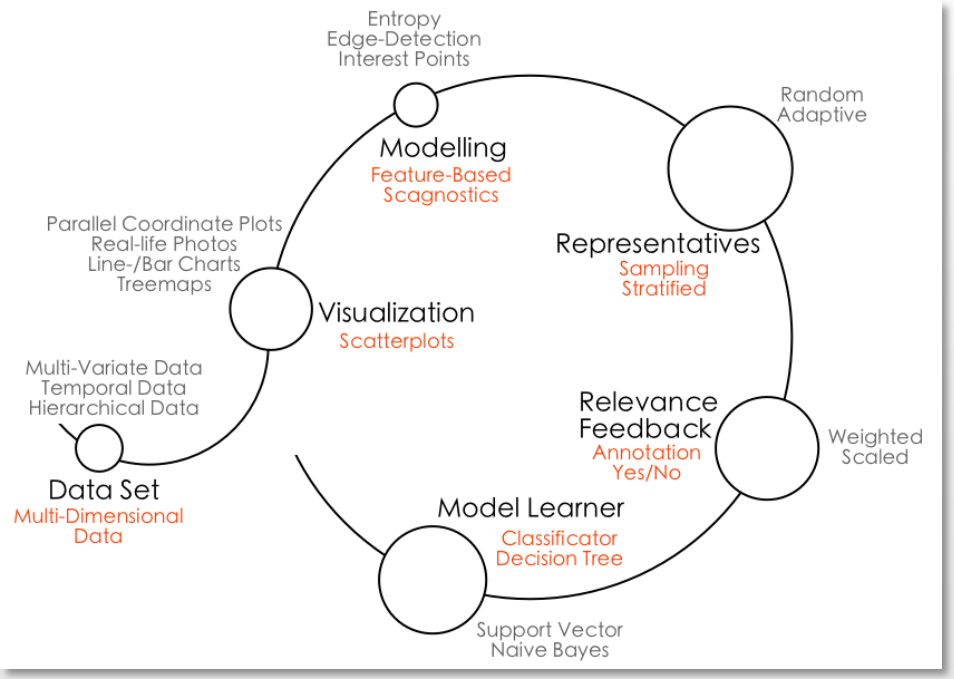
Decision Support Intervention Points	Change Visualization Type	Change Modeling Scheme	Change Sampling Method	Ask for Confidence Score	Assess Classification Certainty
Default Parameter Setting	Visualization Scatter plots	Modelling Scagnostics	Representatives Sampling	Feedback Yes/No	Model Learner Decision tree

Feedback-Driven Interactive Exploration.

Conclusion.

Take Home Messages.

- **General framework** for an **Interest-driven data exploration.**
- Even simple heuristics can help – But are **hard to evaluate.**
- First approach with **many extension possibilities.**



Feedback-Driven Interactive Exploration.

Conclusion.

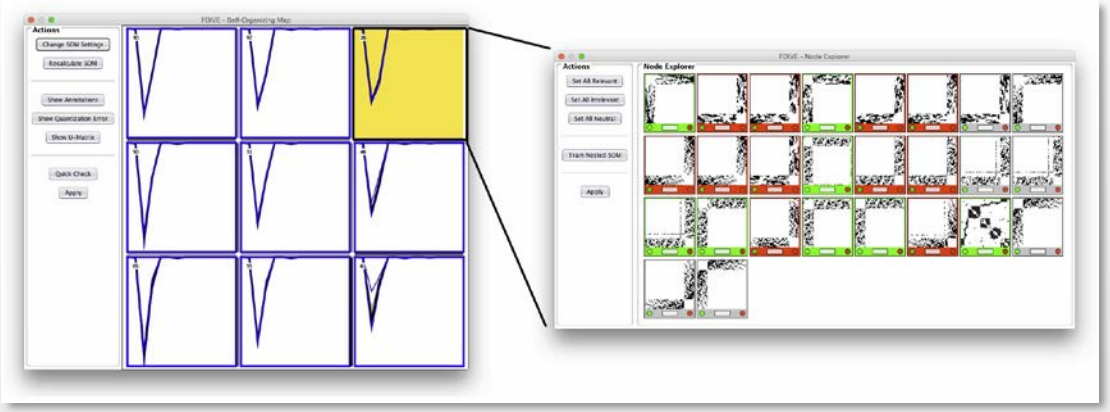
VA

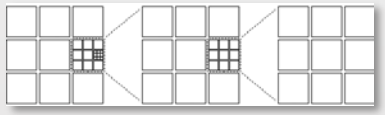
Take Home Messages.

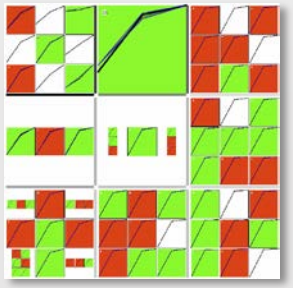
- **General framework** for an **Interest-driven data exploration.**
- Even simple heuristics can help – But are hard to evaluate.
- First approach with many extension possibilities.

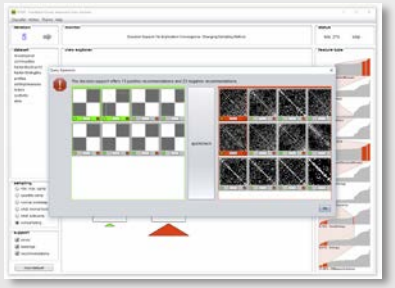
FDIVE 2.0

Application/Generalization to Matrix Data and Nested SOM.









Summary and Future Work.

Pattern-Driven Analysis of Large Matrix-Spaces.

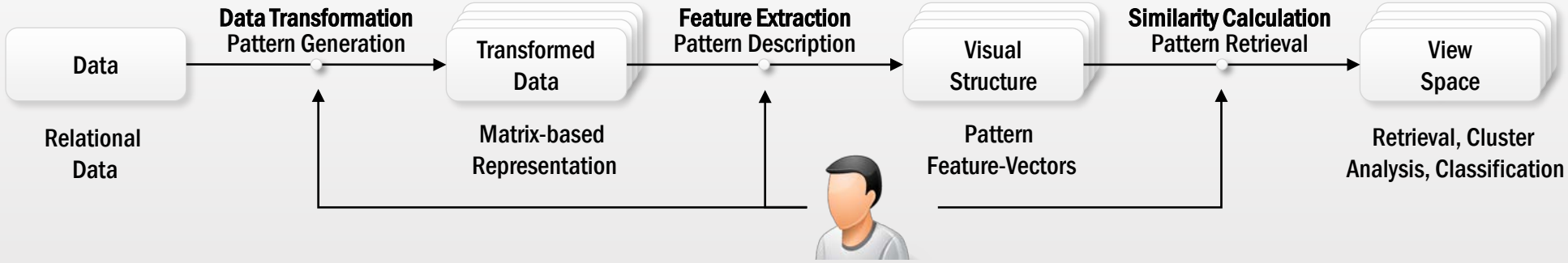
What is a good (Matrix) Representation?

Affects

Visualization

Data Analysis

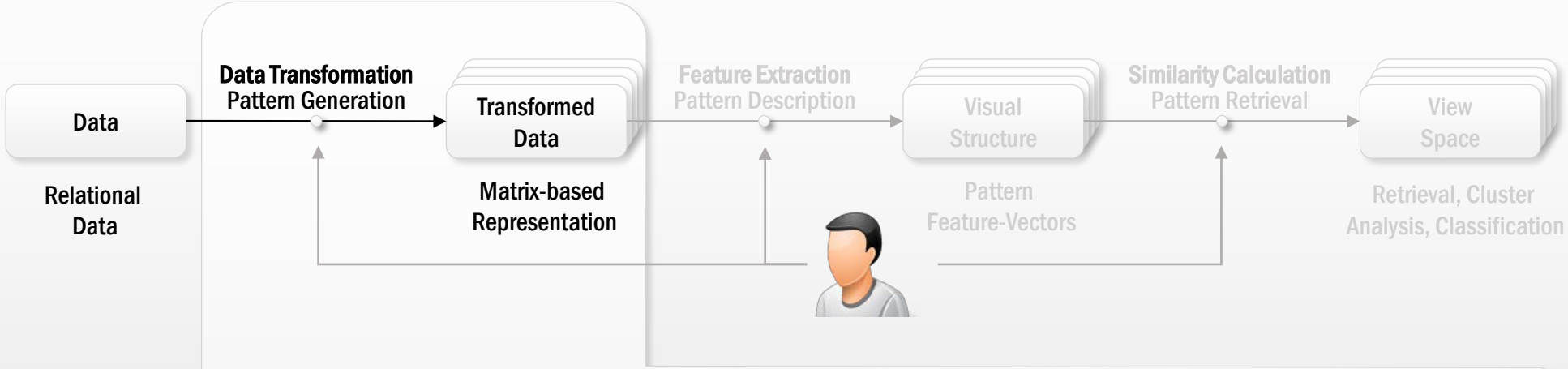
Visual Analytics



Summary and Future Work.

Pattern-Driven Analysis of Large Matrix-Spaces.

Summary of Contributions

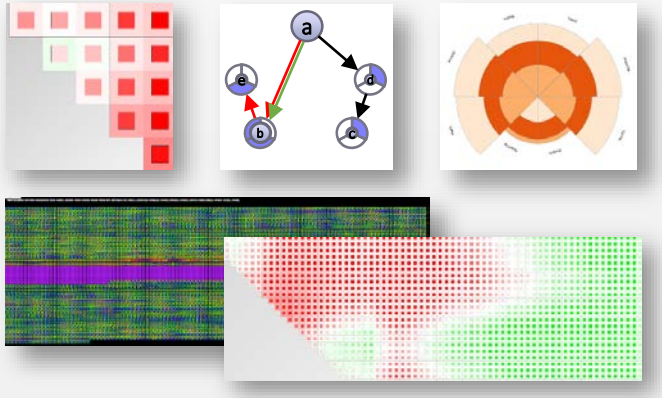


Research Questions

How can we (visually) **support** the **exploration process**?
How can **enhance** the **expressiveness** and **effectiveness**?

Contributions

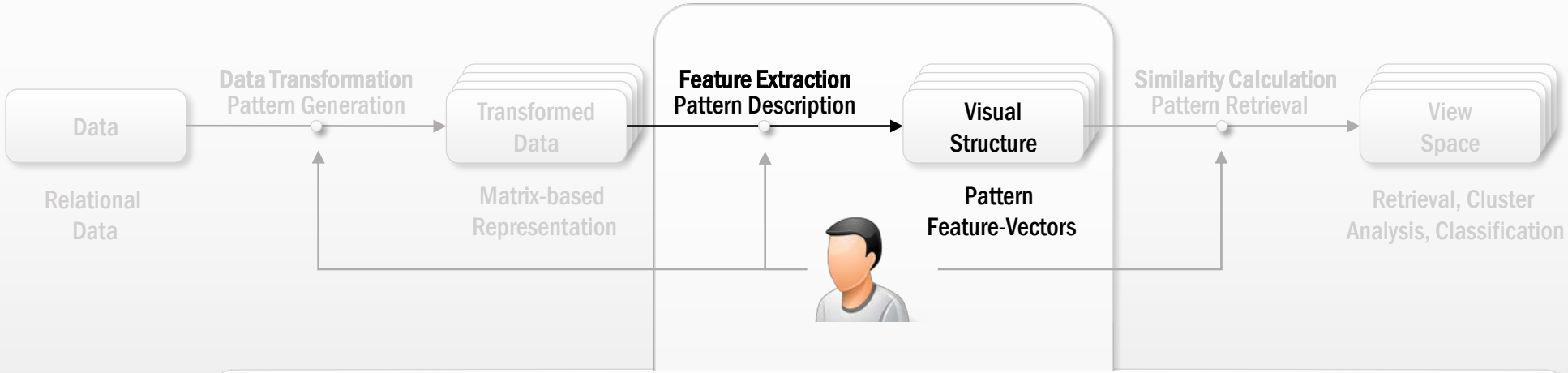
Matrix Reordering State-of-the-Art.
Interactive matrix glyphs-cell concept.
Semantic zoom navigation and interaction.



Summary and Future Work.

Pattern-Driven Analysis of Large Matrix-Spaces.

Summary of Contributions

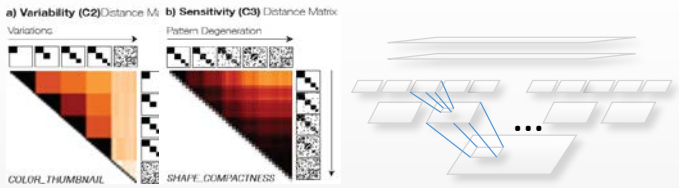


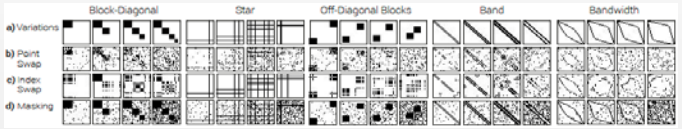
Research Questions

How can **quantify** the **interestingness** of matrices?
How can we automatically **measure** visual matrix **patterns**?

Contributions

Magnostics Feature Descriptor Evaluation Approach.

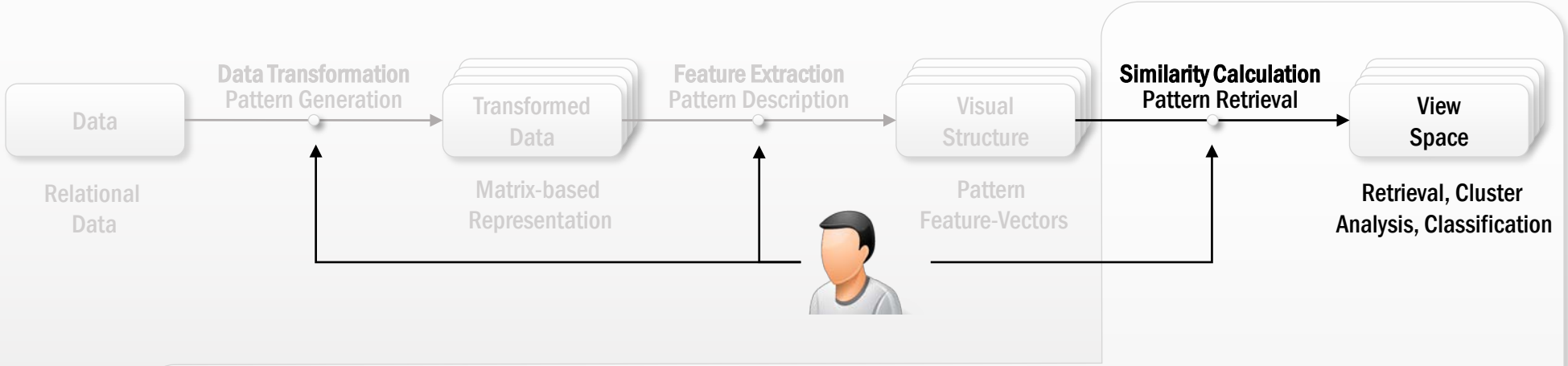




Summary and Future Work.

Pattern-Driven Analysis of Large Matrix-Spaces.

Summary of Contributions

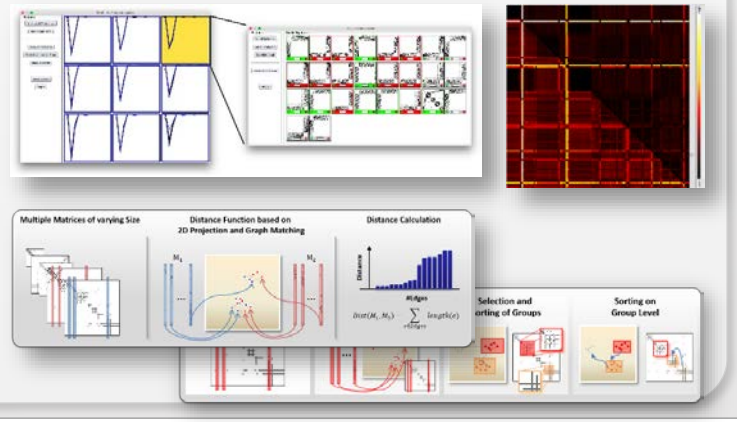


Research Questions

How can we **navigate** and **query** in large matrix spaces?
How can we **compare** matrices?
How can we train computers to **learn** interestingness?

Contributions

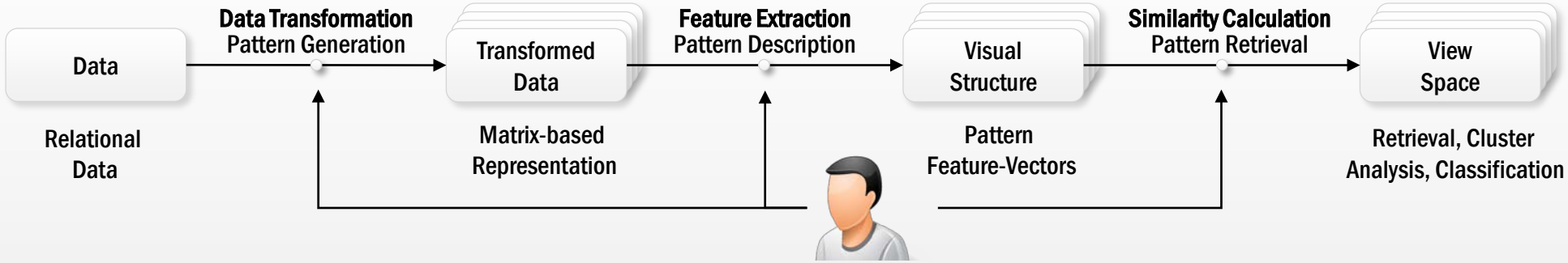
Relevance Feedback-Driven Exploration.
VA Matrix Reordering and Similarity Definition.



Summary and Future Work.

Pattern-Driven Analysis of Large Matrix-Spaces.

What is a good (Matrix) Representation?



Visualization Literacy

Matrix Simplification

Matrix/NodeLink Attribute filtering

Matrix Pattern Glyphs

Pattern Description

Hybrid Image-/Data Space Measures

Novel Image Based Measures

Patterns Diagnostics in other VIS

Pattern-Driven Analysis

Interactive Model Learning

Exploration and Navigation in Pattern Spaces

Thank You.

Pattern-Driven Analysis of Large Matrix-Spaces.

Collaboration Partners

Universität Konstanz

TU Graz

Fraunhofer IGD

Inria informatics mathematics

RWTH AACHEN UNIVERSITY

TECHNISCHE UNIVERSITÄT DARMSTADT

VIS-SENSE
www.vis-sense.eu

vasa VISUAL ANALYTICS FOR SECURITY APPLICATIONS

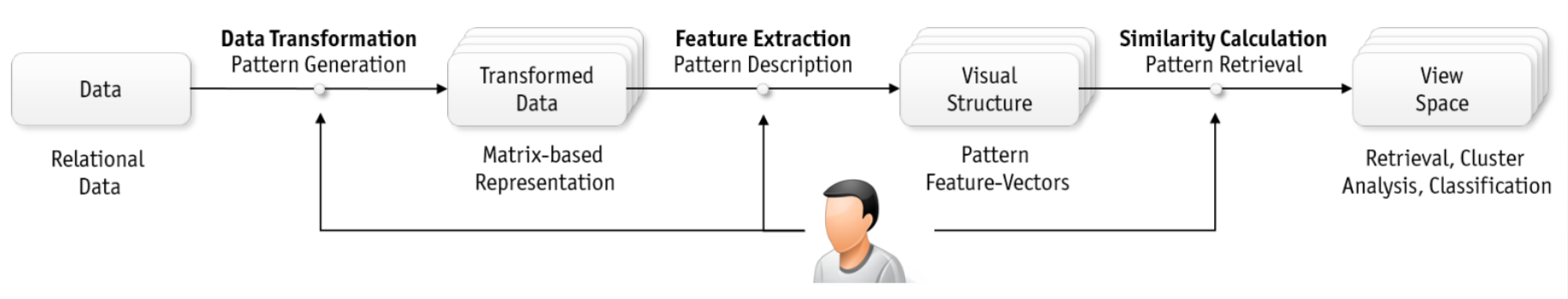
Symantec.

Research Framework.

Pattern-Driven Analysis of Large Matrix-Spaces.

Embedding in the General Pattern-Driven Analysis Vision

Questions?



Publications

2016

M. Behrisch, B. Bach, M. Hund, M. Delz, L. von Rüden, J.-D. Fekete and T. Schreck.
Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration.
IEEE Transactions on Visualization and Computer Graphics, PP(99):1-1, DOI: 10.1109/TVCG.2016.2598467, 2016.

L. Shao, T. Schleicher, M. Behrisch, T. Schreck, I. Sipiran and D. A. Keim.
Guiding the exploration of scatter plot data using motif-based interest measures (Journal).
Journal of Visual Languages & Computing , 36():1 - 12, DOI: https://dx.doi.org/10.1016/j.jvlc.2016.07.003, 2016.

M. Behrisch, B. Bach, N. Henry Riche, T. Schreck and J.-D. Fekete.
Matrix Reordering Methods for Table and Network Visualization.
Computer Graphics Forum, The Eurographics Association and John Wiley & Sons Ltd., DOI: 10.1111/cgf.12935, 2016.

2015

L. von Rüden, M.-A. Hermanns, M. Behrisch, D. A. Keim, B. Mohr and F. Wolf.
Separating the Wheat from the Chaff: Identifying Relevant and Similar Performance Data with Visual Analytics.
Proceedings of the 2nd Workshop on Visual Performance Analysis, ACM, pages 4:1--4:8, DOI: 10.1145/2835238.2835242, 2015.

L. Shao, T. Schleicher, M. Behrisch, T. Schreck, I. Sipiran and D. A. Keim.
Guiding the Exploration of Scatter Plot Data Using Motif-based Interest Measures.
IEEE Int. Symposium on Big Data Visual Analytics (Best Paper Award), DOI: 10.1109/BDVA.2015.7314294, 2015.

M. Hund, M. Behrisch, I. Färber, M. Sedlmair, T. Schreck, T. Seidl and D. A. Keim.
Subspace Nearest Neighbor Search - Problem Statement, Approaches, and Discussion.
Similarity Search and Applications, Springer International Publishing, pages 307-313, DOI: 10.1007/978-3-319-25087-8_29, 2015.

M. Behrisch, L. Shao, J. Buchmüller and T. Schreck.
Quality Metrics Driven Approach to Visualize Multidimensional Data in Scatterplot Matrix.
Eurographics Conference on Visualization (poster paper), 2015.

J. Twellmeyer, M. Hutter, M. Behrisch, J. Kohlhammer and T. Schreck.
The Visual Exploration of Aggregate Similarity for Multi-Dimensional Clustering.
Proc. Int. Conference on Information Visualization Theory and Applications, 2015.

2014

M. Behrisch, L. Shao, B. C. Kwon, T. Schreck, I. Sipiran and D. A. Keim.
Quality Metrics Driven Approach to Visualize Multidimensional Data in Scatterplot Matrix.
Poster Presentation at GI Workshop Big Data Visual Computing – Quantitative Perspectives for Visual Computing, September 22, 2014 Stuttgart, Germany, 2014.

J. Bernard, D. Sessler, M. Behrisch, M. Hutter, T. Schreck and J. Kohlhammer.
Towards a User-Defined Visual-Interactive Definition of Similarity Functions for Mixed Data.
Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on, pages 227-228, DOI: 10.1109/VAST.2014.7042503, 2014.

L. Shao, M. Behrisch, T. Schreck, I. Sipiran, B. C. Kwon and D. A. Keim.
Identifying Locally Interesting Motifs for Exploration of Scatter Plot Matrices.
Poster Presentation at GI Workshop Big Data Visual Computing – Quantitative Perspectives for Visual Computing, September 22, 2014 Stuttgart, Germany, 2014.

M. Behrisch, F. Korkmaz, L. Shao and T. Schreck.
Feedback-Driven Interactive Exploration of Large Multidimensional Data Supported by Visual Classifier.
Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on, IEEE CS Press, pages 43-52, DOI: 10.1109/VAST.2014.7042480, 2014.

L. Shao, M. Behrisch, T. Schreck, T. von Landesberger, M. Scherer, S. Bremm and D. A. Keim.
Guided Sketching for Visual Search and Exploration in Large Scatter Plot Spaces.
Proc. EuroVA International Workshop on Visual Analytics, The Eurographics Association, DOI: 10.2312/eurova.20141140, 2014.

M. Behrisch, J. Davey, F. Fischer, O. Thonnard, T. Schreck, D. A. Keim and J. Kohlhammer.
Visual Analysis of Sets of Heterogeneous Matrices Using Projection-Based Distance Functions and Semantic Zoom.
Eurographics Conference on Visualization (EuroVis 2014), The Eurographics Association and John Wiley & Sons Ltd. Published by John Wiley & Sons Ltd., pages 411–420, DOI: 10.1111/cgf.12397, 2014.

2013

M. Behrisch, J. Davey, S. Simon, T. Schreck, D. A. Keim and J. Kohlhammer.
Visual Comparison of Orderings and Rankings.
EuroVis Workshop on Visual Analytics, The Eurographics Association, DOI: 10.2312/PE.EuroVAST.EuroVA13.007-011, 2013.

2012

M. Behrisch, J. Davey, T. Schreck, J. Kohlhammer and D. A. Keim.
Matrix-Based Visual Correlation Analysis on Large Timeseries Data.
Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on, pages 209-210, DOI: 10.1109/VAST.2012.6400549, 2012.

L. Zhang, A. Stoffel, M. Behrisch, S. Mittelstädt, T. Schreck, R. Pompl, S. Weber, H. Last and D. A. Keim.
Visual analytics for the big data era - a comparative review of state-of-the-art commercial systems.
Proceedings of IEEE Symposium on Visual Analytics Science and Technology, DOI: 10.1109/VAST.2012.6400554, 2012.

M. Behrisch, M. Krstajic, T. Schreck and D. A. Keim.
The News Auditor: Visual Exploration of Clusters of Stories.
Proc. EuroVA International Workshop on Visual Analytics, Eurographics, pages 61-65, DOI: 10.2312/PE/EuroVAST/EuroVA12/061-065, 2012.