High-Dimensional Data Analysis

Michael Gruber

Overview

- Dimstiller
  - DimStiller: Workflows for dimensional analysis and reduction, Stephen Ingram, Tamara Munzner, Veronika Irvine, Melanie Tory, Steven Bergner, Torsten Möller, Proceedings of the 5th IEEE Conference on Visual Analytics in Science and Technology (VAST) 2010, pp. ???--???, October 2010

- DOSFA
  - Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets. Jing Yang, Wei Peng, Matthew O. Ward and Elke A. Rundensteiner .Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609

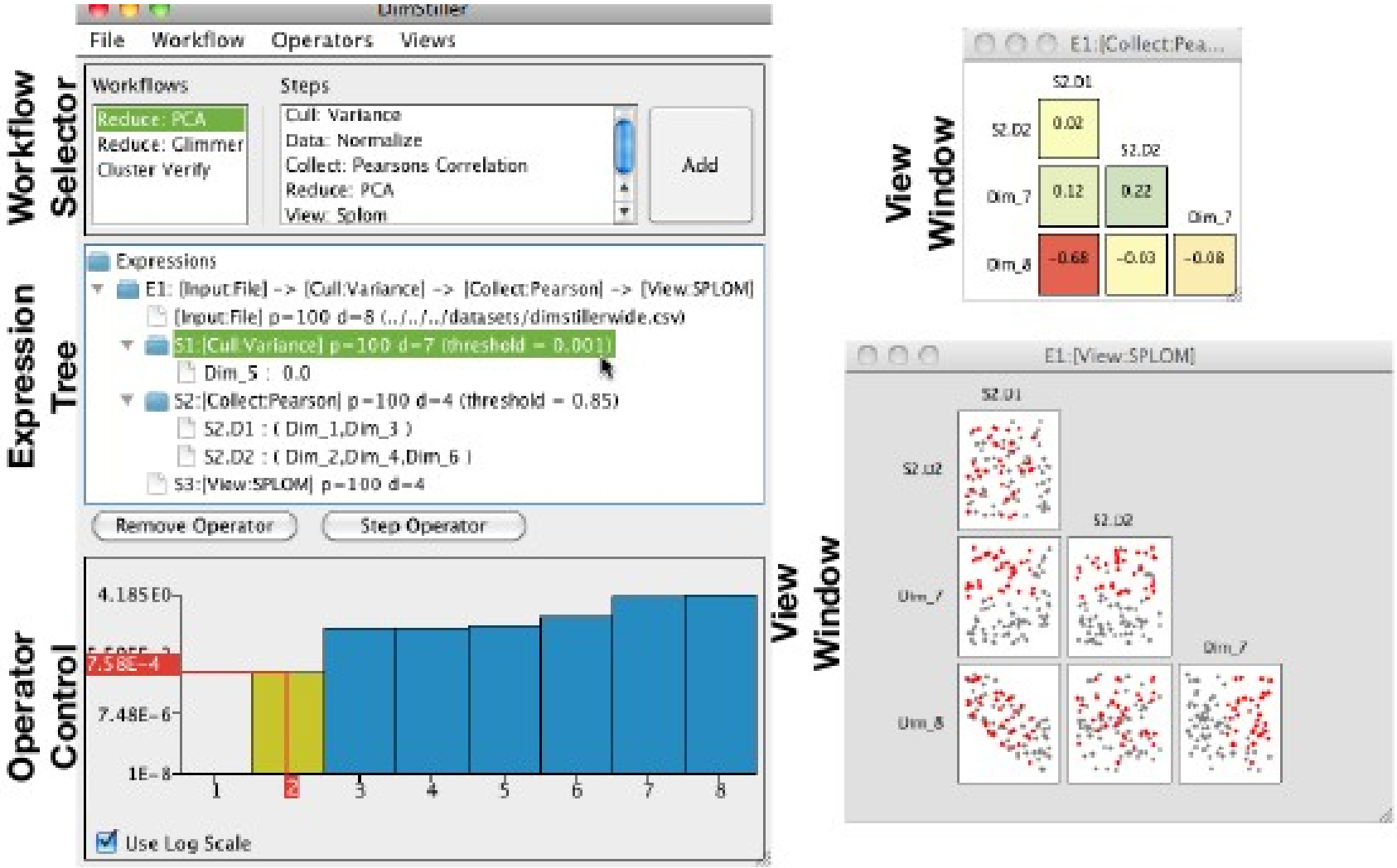
Dimstiller: Intro

- System for dimensionality reduction and visualization
- Works on data tables
- Provides large amount of possible transformations and visualization options
- Workflows by chaining operators together
- Guides user to choose most useful operations

Dimstiller: Architecture

- Expression
  - Pipeline of Operators
- Operator
  - Data Transformation
  - Visual display
- Workflow
  - Predefined or saved data-independent expression

Dimstiller: UI Overview



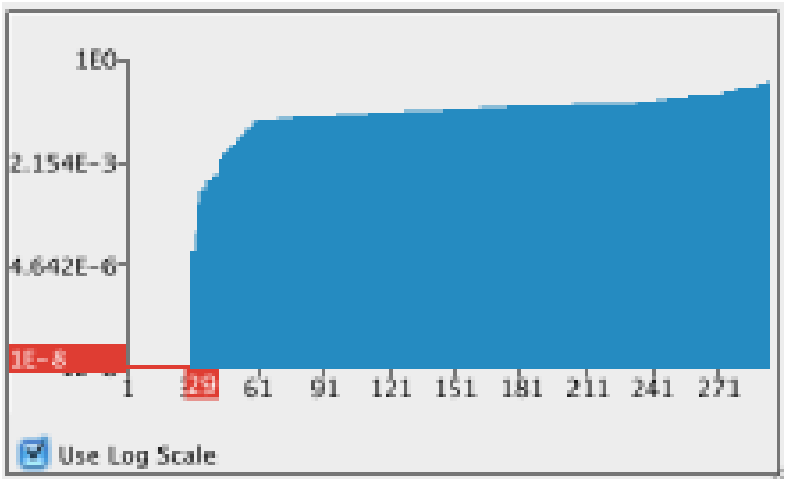
Dimstiller: Operators

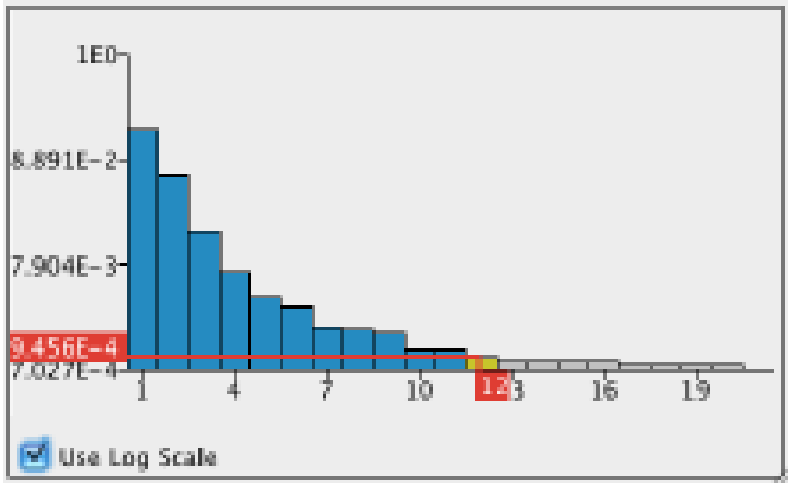
- Operators
  - Cull:(Var|Name)
  - Collect:pearson
  - Data:Norm
  - Reduce:(PCA|MDS)
  - View:(SPLOM|Histo)
  - Input,Attrib
- Easily extensible by user

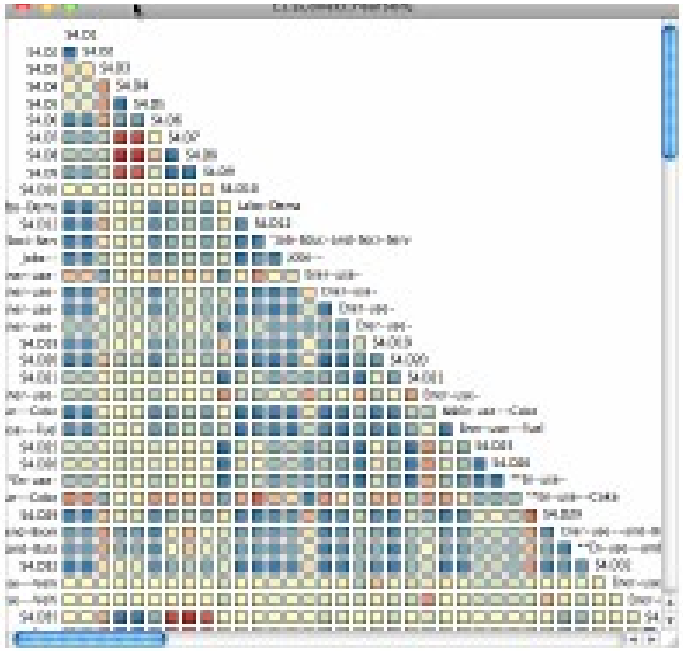
Dimstiller: Architecture

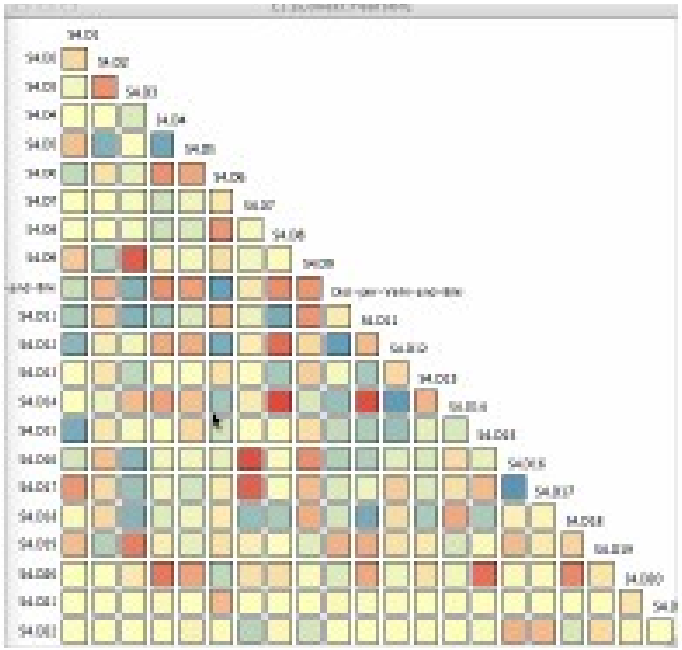
- Works on Tables
  - Input as cvs
- Dimensions
  - Data: quantitative vs categorical
  - Attributes: colormaping, selection
    - Ignored for data transformations
- View
  - Visual representation of associated table
  - Fully interlinked
  - Selection propagation

Dimstiller: View examples



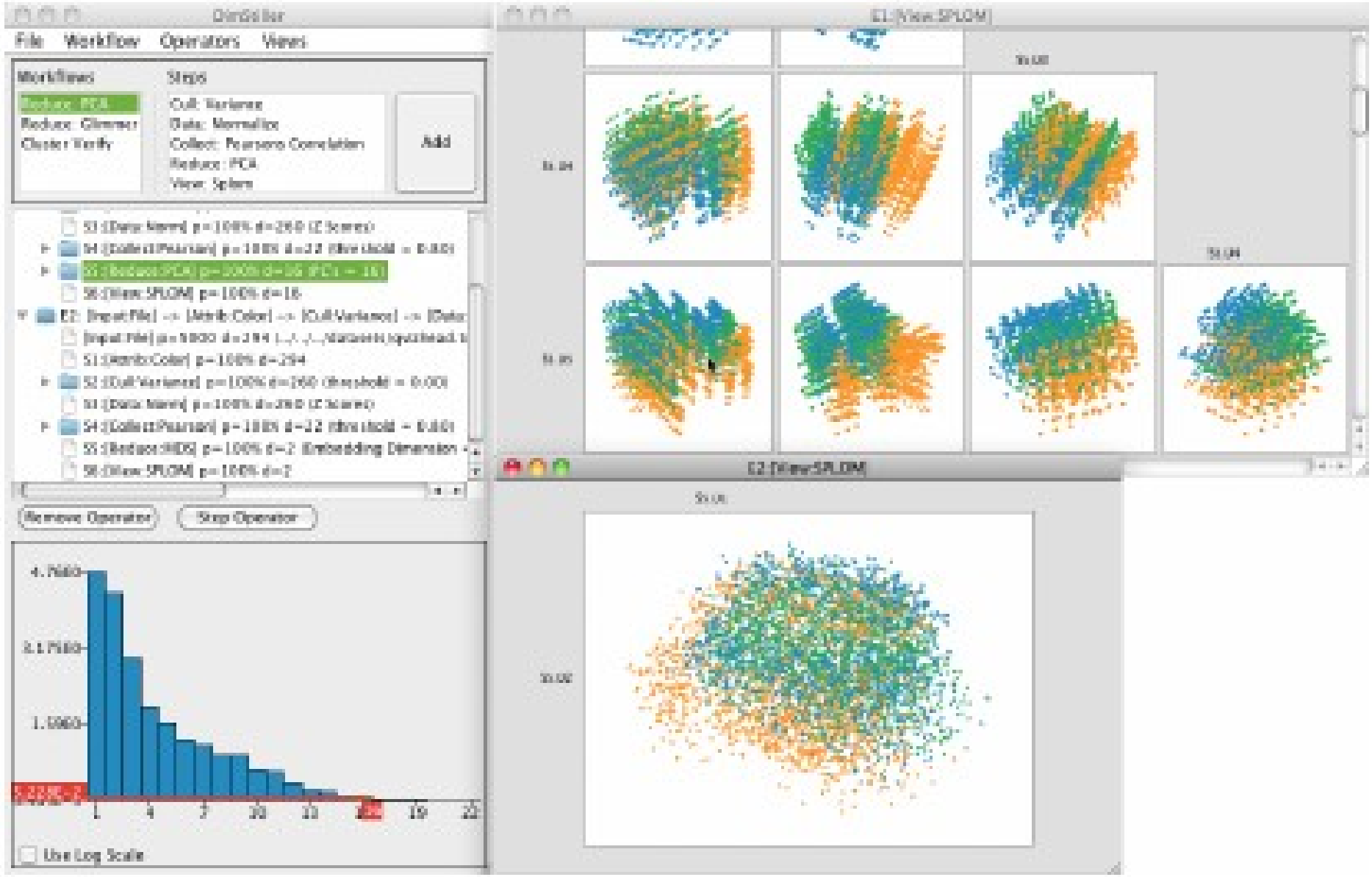






Dimstiller: Case Studies

- Sustainability Simulation
  - Social/environmental indicators affected by simulated government policy
  - 294 dimensions
    - Cull:variance – reduced to 260
    - Collect:pearson
      - Threshold 1.0 – 147
      - Threshold 0.8 - 22


DimStalker
File
Workflow
Operators
Views
Workflows
Steps
Reduce: PCA
Reduce: Gimmer
Cluster Verify
Cull: Variance
Data: Normalize
Collect: Pearsons Correlation
Reduce: PCA
View: Splom
Add
Remove Operator
Stop Operator
Use Log Scale
E1: View SPLOM
E2: View SPLOM

DOSFA: Intro

- Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets

- Alter arrangement of dimensions
- Similarity-based spacing
- Similarity/variance-based filtering

DOSFA: Visualization Techniques

- Scatter-plots
- Parallel Coordinates
- Star Glyphs
- Other arrangement-sensitive techniques

DOSFA: Dimensional Hierarchy

- Clustered by similarity
  - Iterative clustering
  - Representative dimensions
  - Data clusters

- drill-down/roll-up, pan, zoom, and rotation, scale, orientation, distortion

DOSFA: Dimensional Ordering

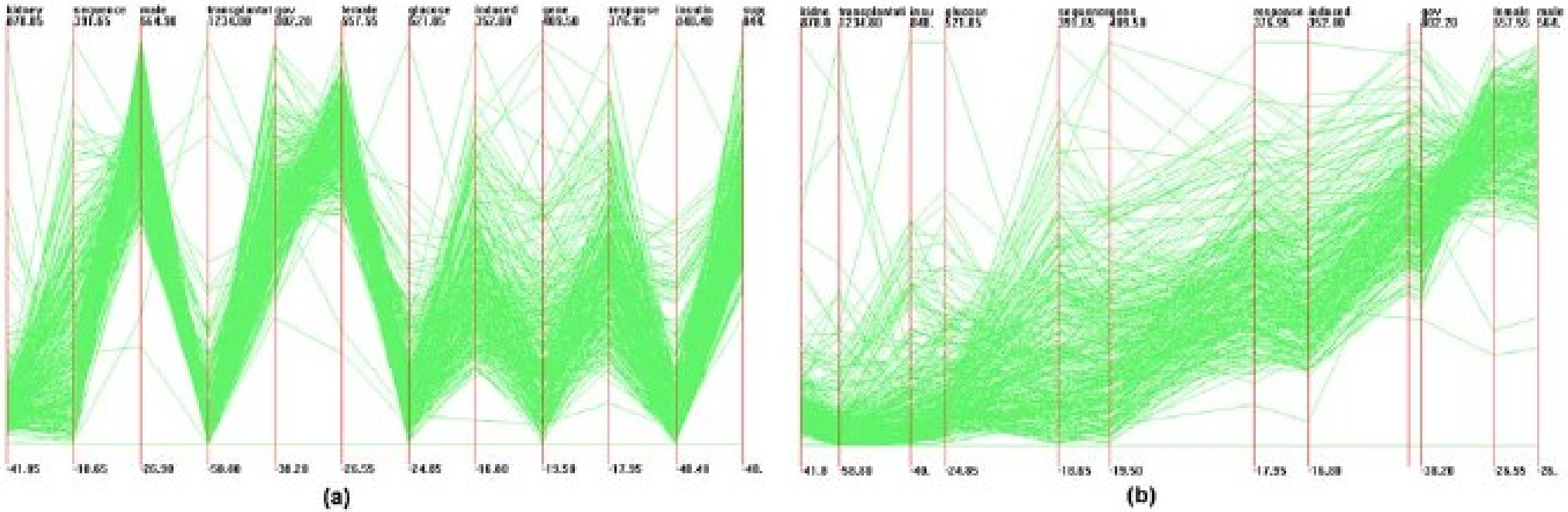
- Make similarities obvious
- More prevalent placing
- Similarity-oriented
  - Optimal (high complexity)
  - Heuristics
- Importance-oriented
  - Based on variance
- Interactive

DOSFA: Dimension Spacing

- Especially useful for parallel coordinates
- Places less similar dimensions further apart
- Implemented using pairwise correlation-factor
- Distance also influenced by clusters

- Interactive
  - Zooming, panning, distortion

DOSFA: Ordering/Spacing



DOSFA: Dimensional Filtering

- Based on hierarchy
- Automatic
  - Resursive descent into hierarchy
  - Merge very similar dimensions
  - Ignore “unimportant” dimensions
- Interactive
  - Adjust threshold-values
  - Selectively only on certain clusters
  - Manually add/remove dimensions

DOSFA: Interaction

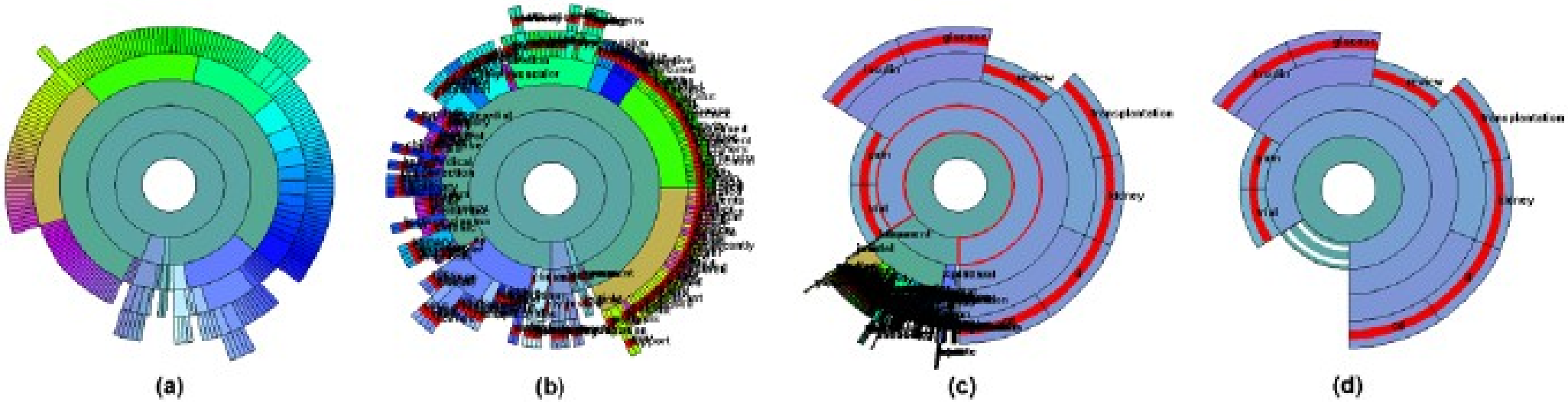
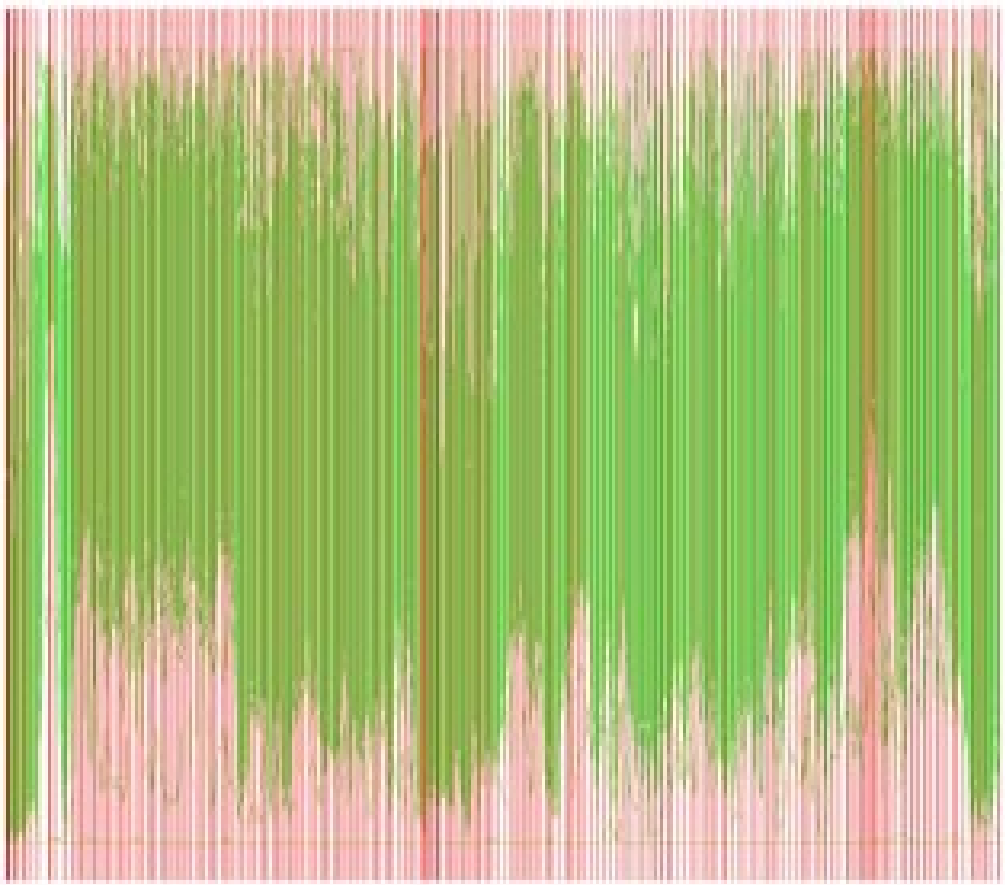
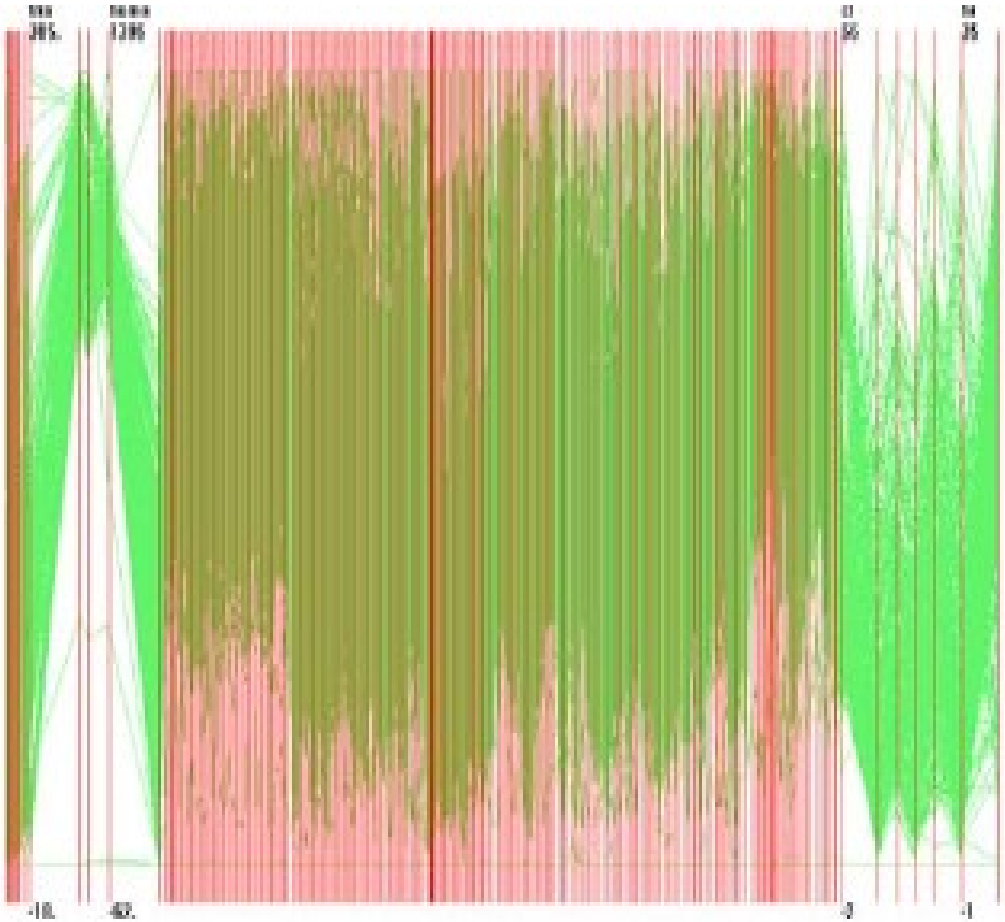


Figure 1: InterRing. (a): Dimension hierarchy of OHSUMED dataset in InterRing. (b): after reordering. (c): after distortion. (d): after roll-up operation.

DOSFA: UI



(c)



(d)

DOSFA: UI

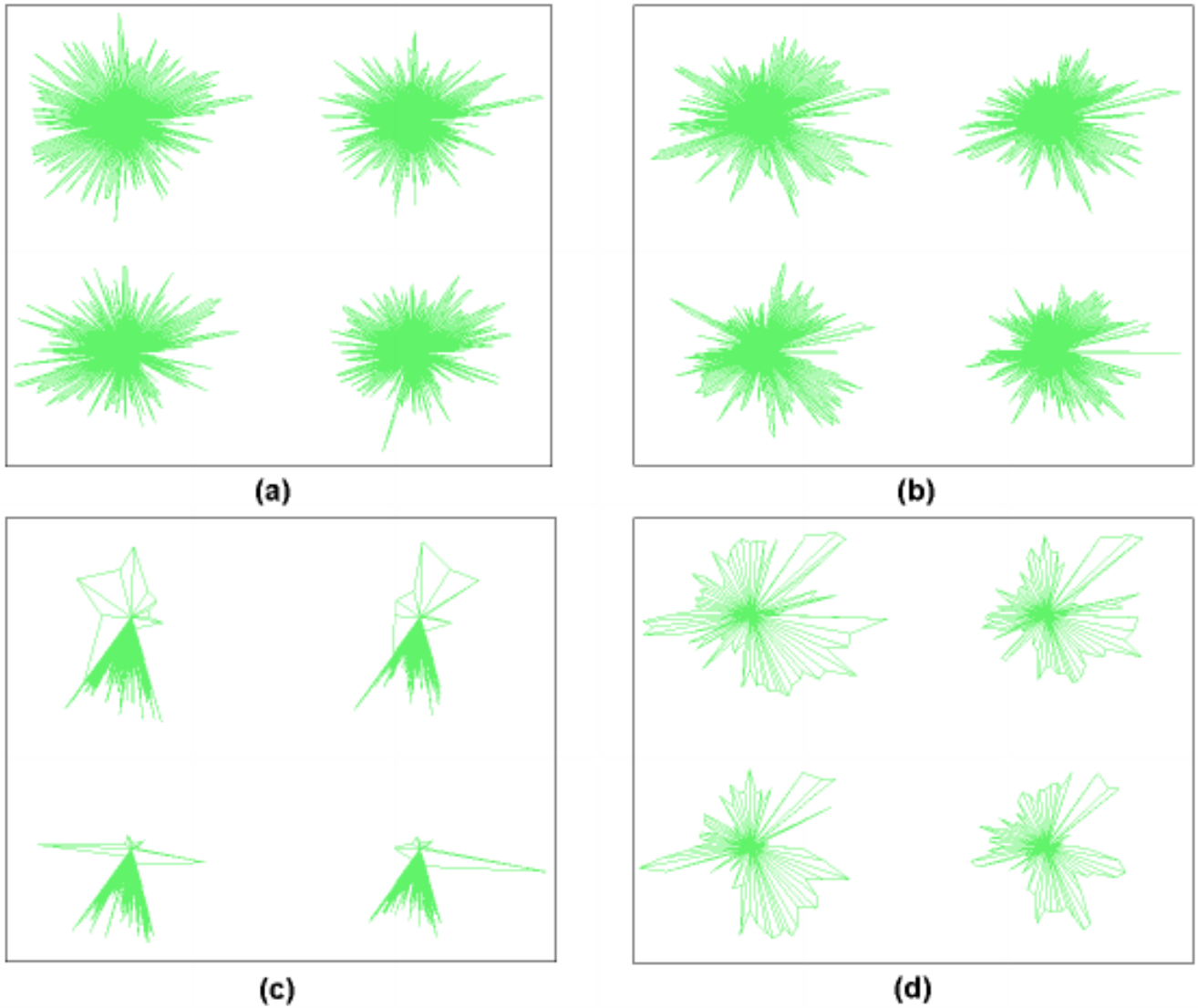
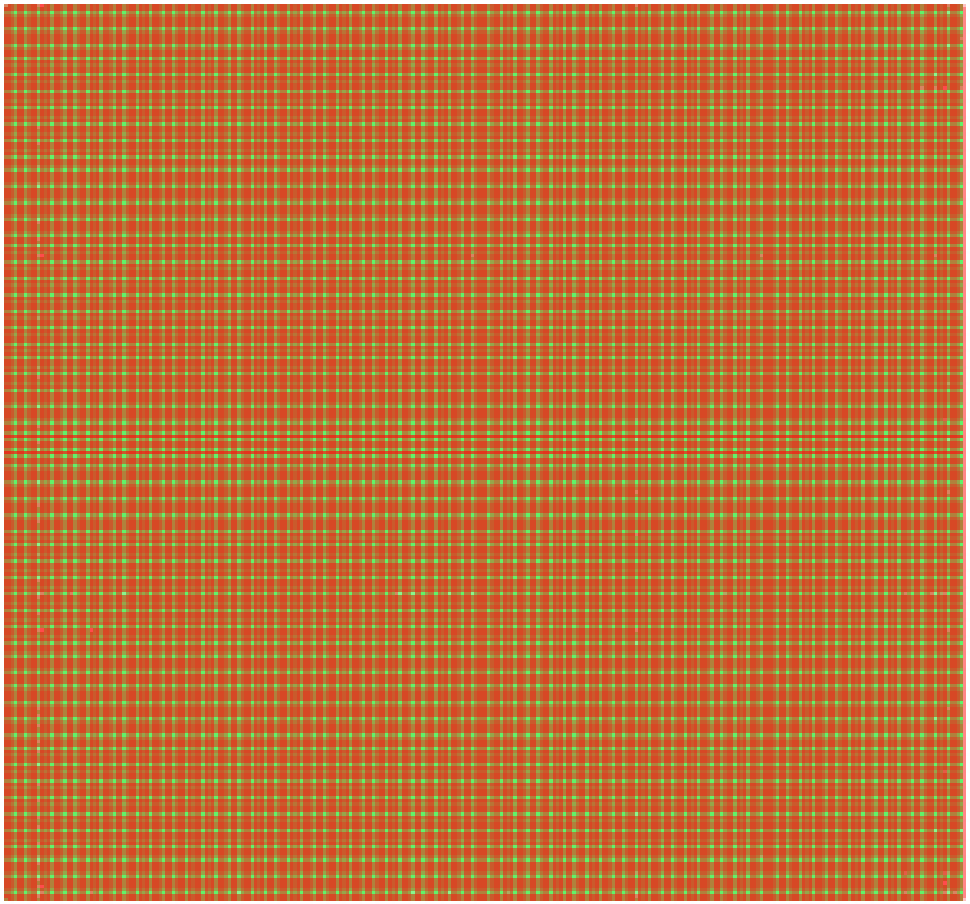
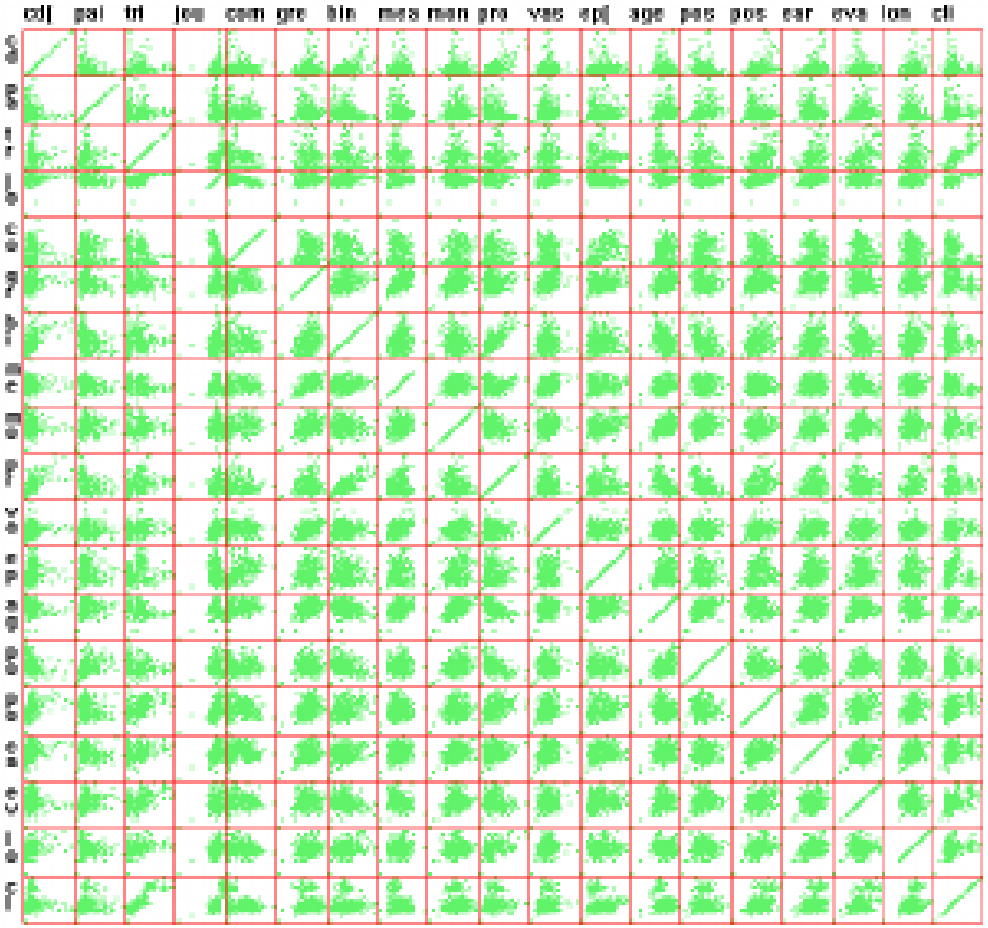


Figure 3: Star Glyphs. (a): OHSUMED dataset without DOSFA. (b): after ordering. (c): distorted star glyphs. (d): after filtering.

DOSFA: UI



(a)


odj pal fri jou com gre bin mea man pro vac epi age pas pos ear eva lon cli

(b)