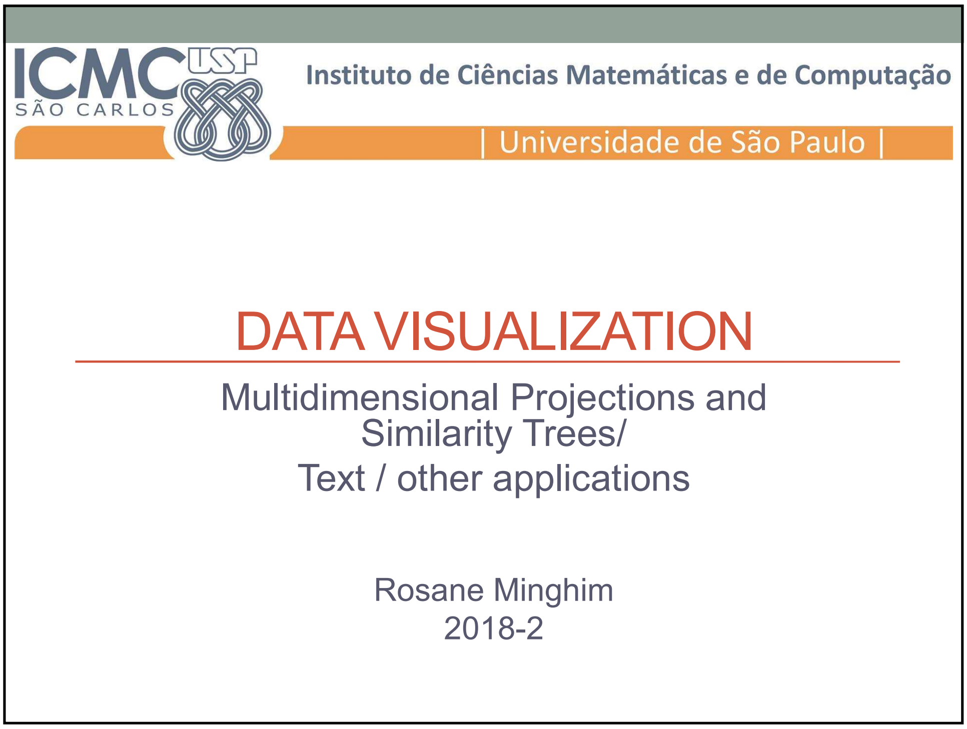


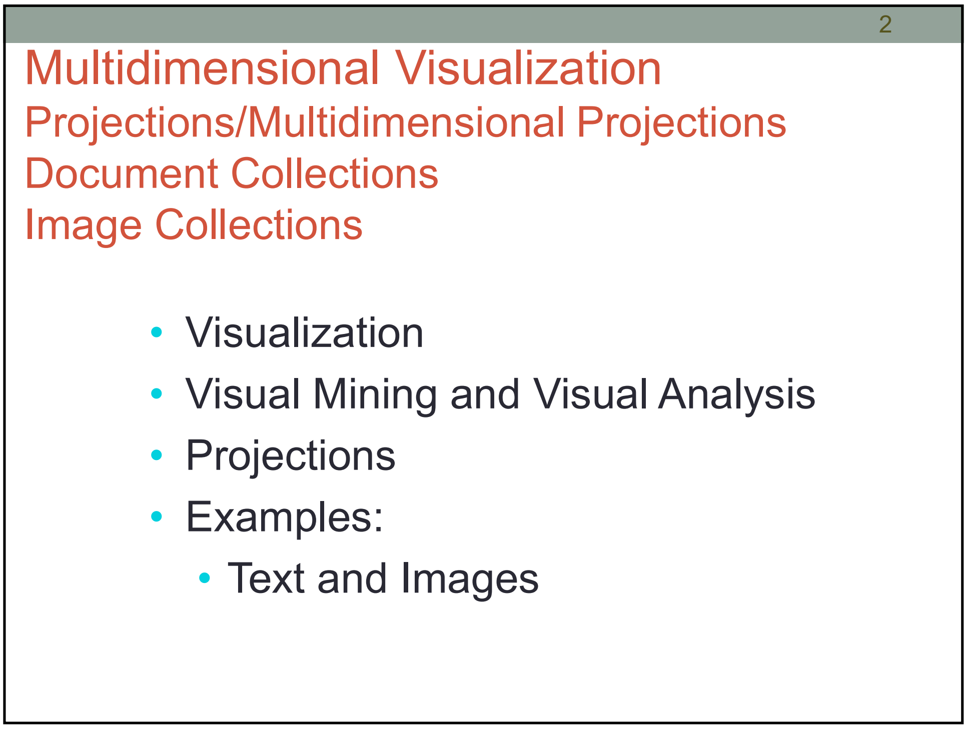
ICMC SÃO CARLOS USP Instituto de Ciências Matemáticas e de Computação

| Universidade de São Paulo |

DATA VISUALIZATION

Multidimensional Projections and Similarity Trees/
Text / other applications

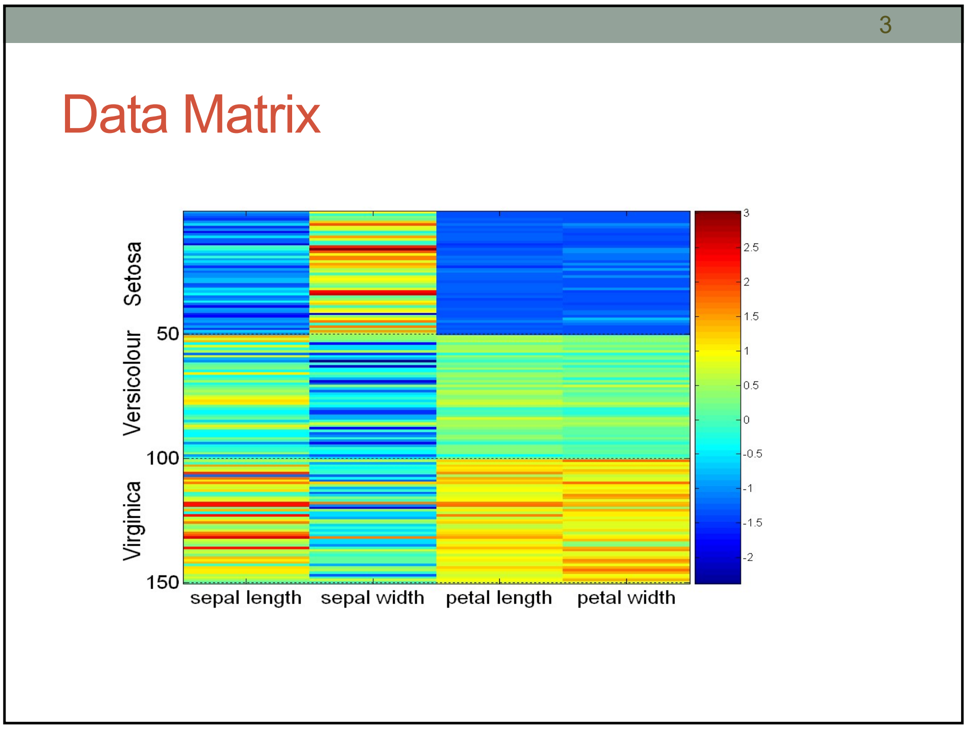
Rosane Minghim
2018-2



2

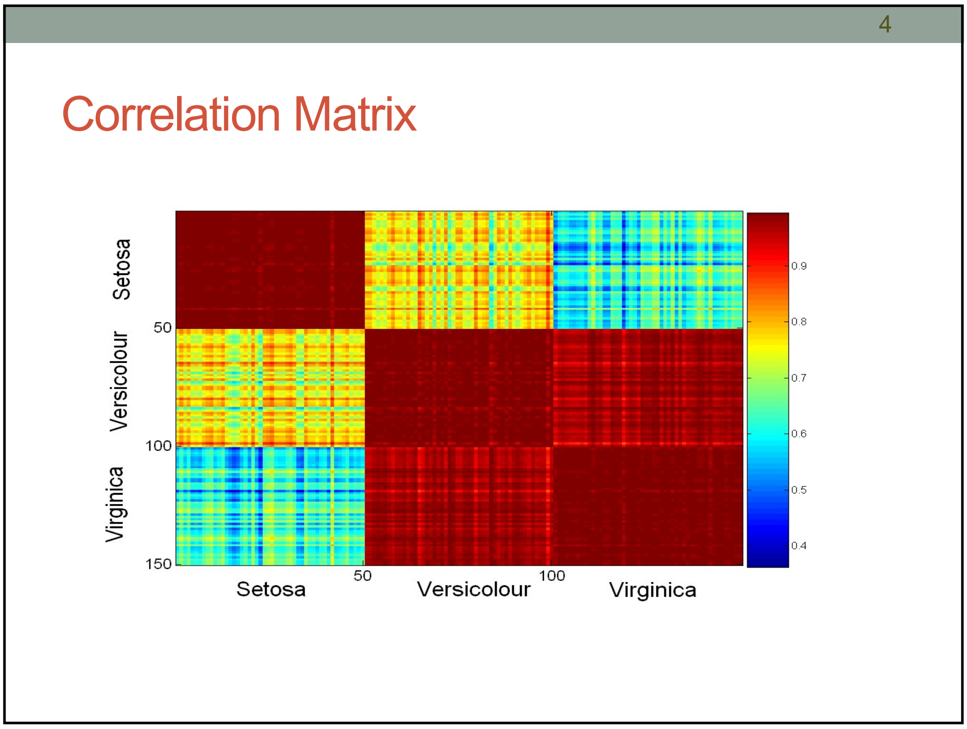
Multidimensional Visualization
Projections/Multidimensional Projections
Document Collections
Image Collections

- Visualization
- Visual Mining and Visual Analysis
- Projections
- Examples:
  - Text and Images



3

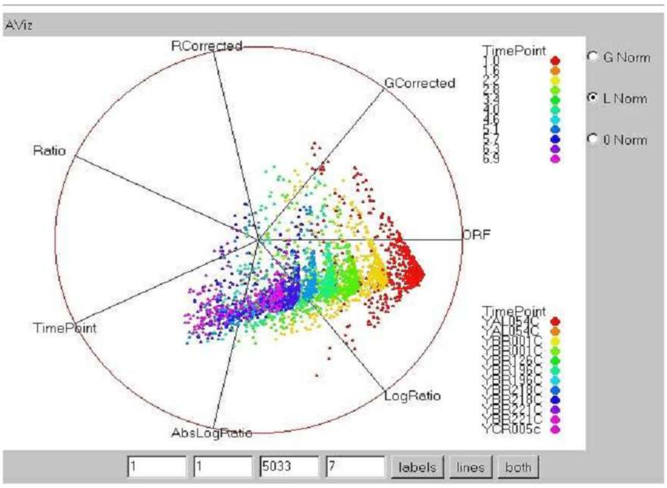
Data Matrix



4

Correlation Matrix

RadViz

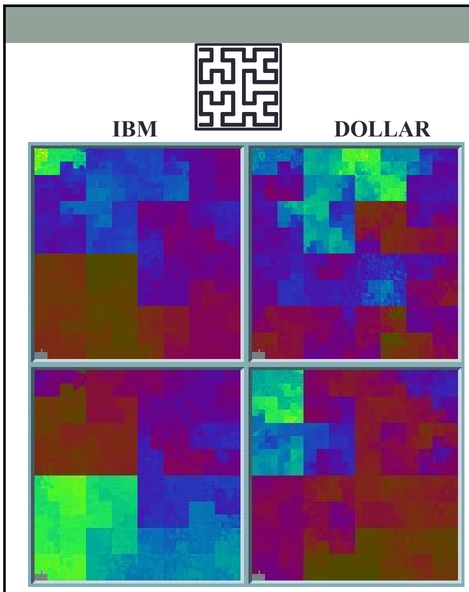


○ G Norm

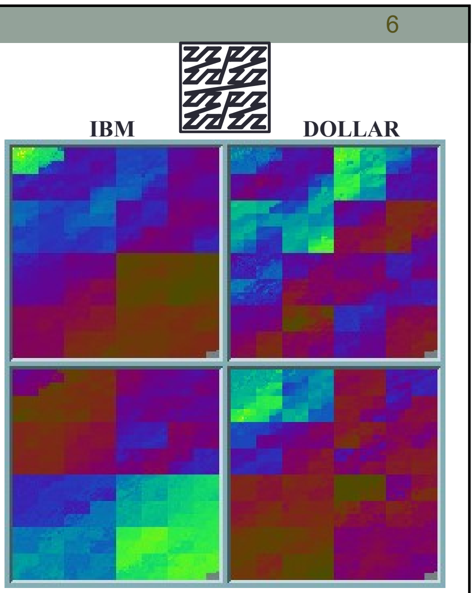
● L Norm

○ 0 Norm

1 | 1 | 5033 | 7



Peano-Hilbert

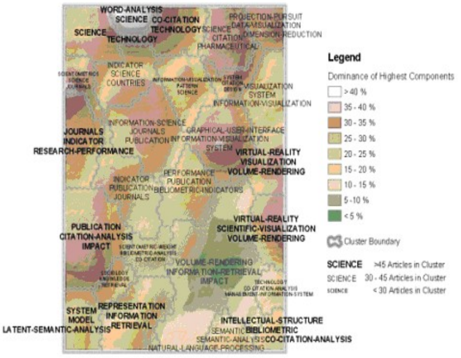


Morton (Z-Curve)

Space-Filling Curves

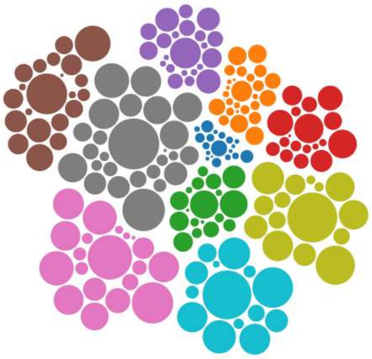
SOM based

- **Self-Organization Maps (SOMs) cartográficos (ex. Skurpin 2002)**



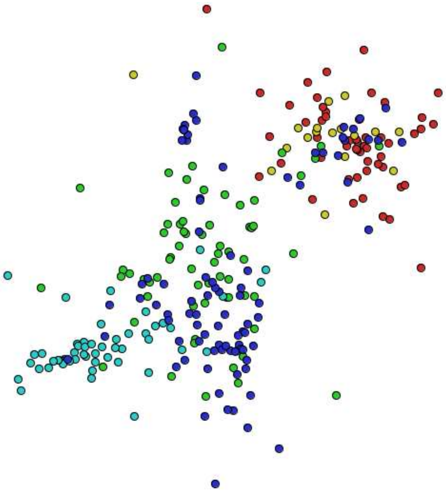
Clustered Force Layout

- http://bl.ocks.org/mbostock/1747543

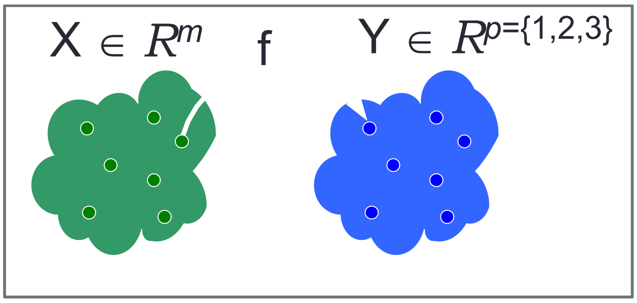


Mapeamento para o plano permitindo a exploração.

Ex: Patents surgery, drugs, molecular bio

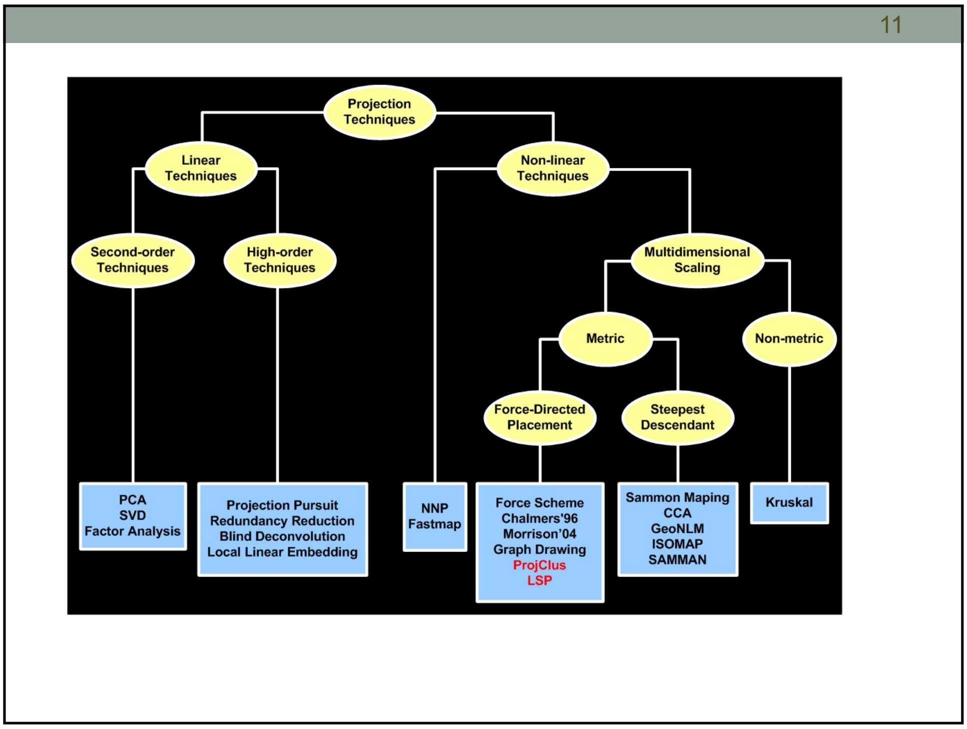


Projection Techniques



$X \in \mathcal{R}^m$ f $Y \in \mathcal{R}^{p=\{1,2,3\}}$

- δ: x_i, $x_j \rightarrow \mathcal{R}$, $x_i, x_j \in X$
- d: y_i, $y_j \rightarrow \mathcal{R}$, $y_i, y_j \in Y$
- f: $X \rightarrow Y, |\delta(x_i, x_j) - d(f(x_i), f(x_j))| \approx 0, \forall\ x_i, x_j \in X$



11

Projection Techniques

Linear Techniques

Non-linear Techniques

Second-order Techniques

High-order Techniques

Multidimensional Scaling

Metric

Non-metric

Force-Directed Placement

Steepest Descendant

PCA
SVD
Factor Analysis

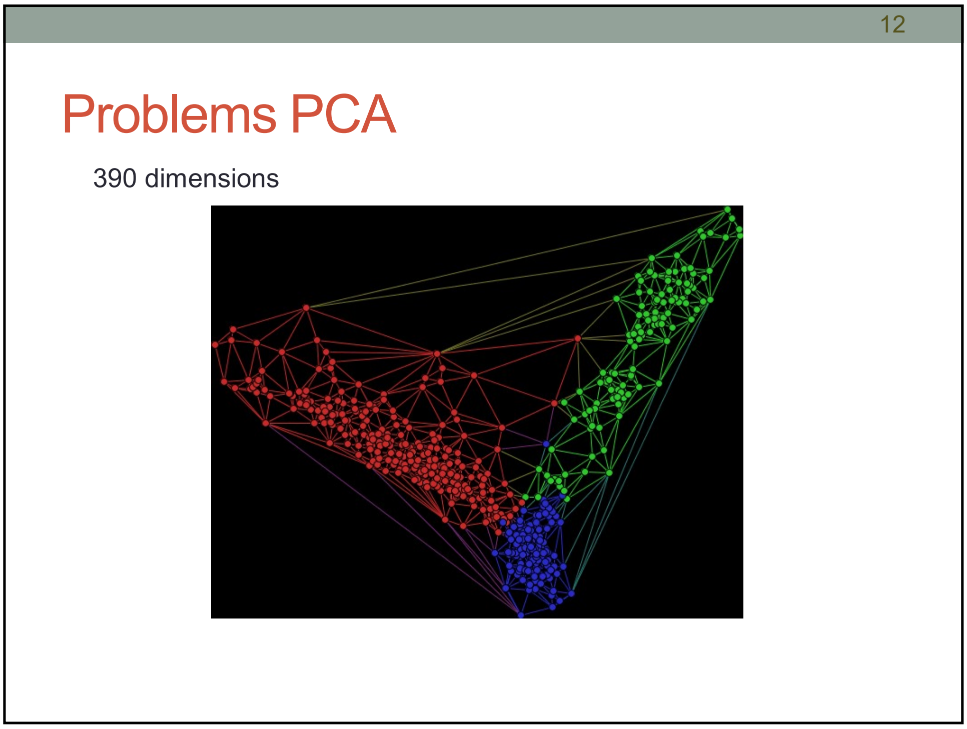
Projection Pursuit
Redundancy Reduction
Blind Deconvolution
Local Linear Embedding

NNP
Fastmap

Force Scheme
Chalmers'96
Morrison'04
Graph Drawing
ProjClus
LSP

Sammon Maping
CCA
GeoNLM
ISOMAP
SAMMAN

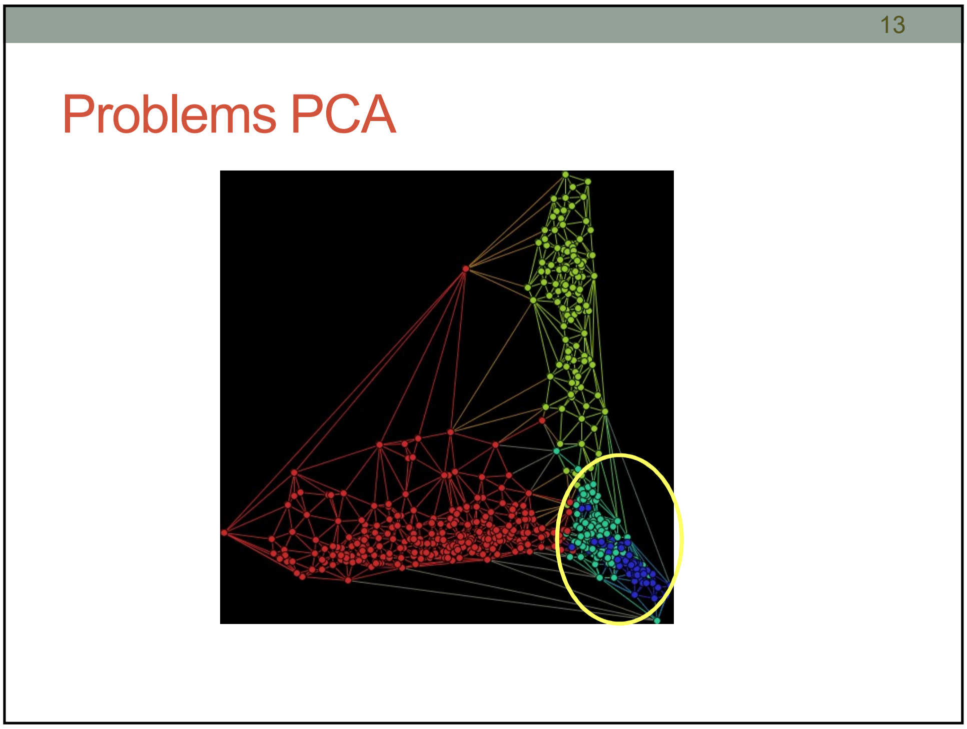
Kruskal



12

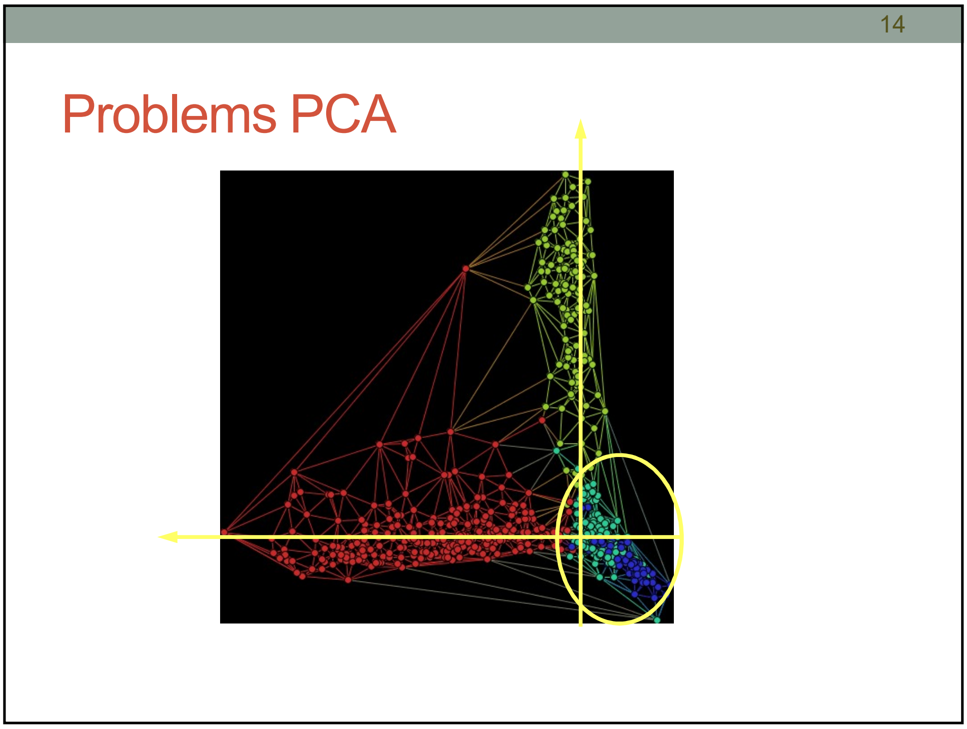
Problems PCA

390 dimensions



13

Problems PCA

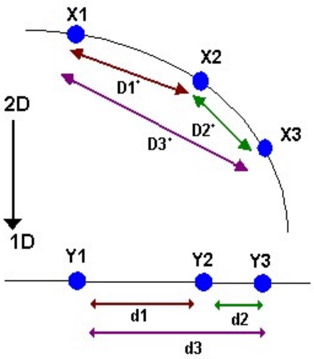


14

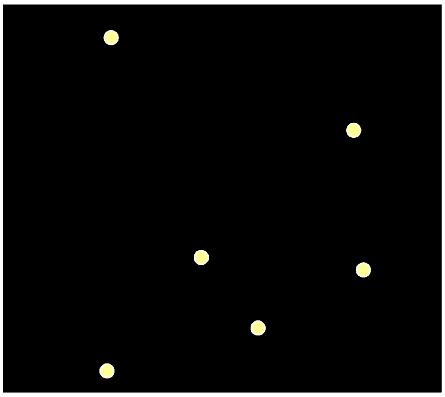
Problems PCA

Ex: Sammon Mapping

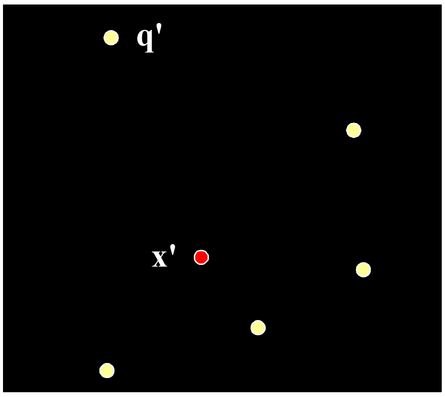
- Let **X** be the points in the original space R^n, we apply a distance measure d_{ij}* between Xi an Xj., and find **Y, the projected point**, ex. R^2 and d_{ij} the Euclidean distance between them.
- Sammon's method applies an error function to measure the target.



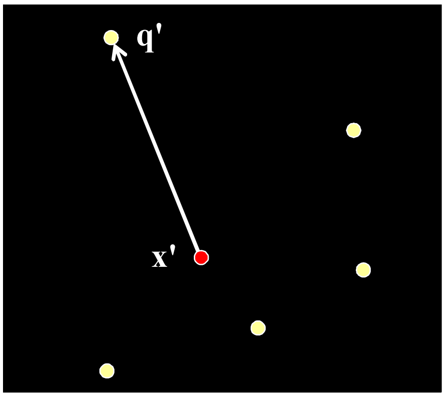
Force Based Point Placement



Force Scheme [Tejada et al., 2003]

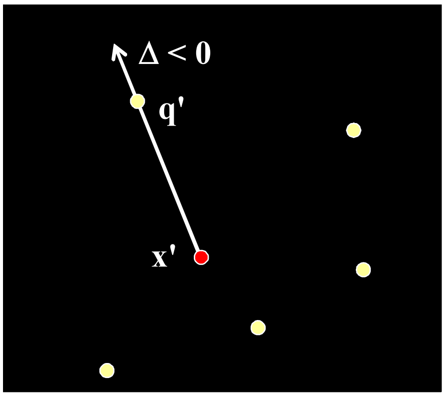


Force Scheme [Tejada et al., 2003]



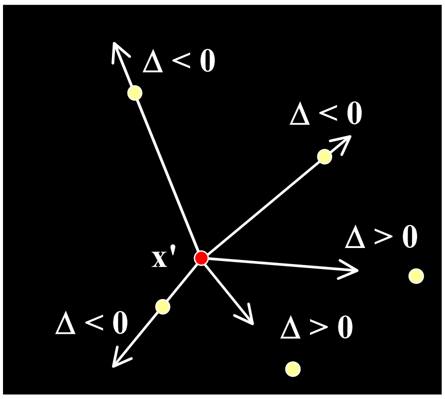
19

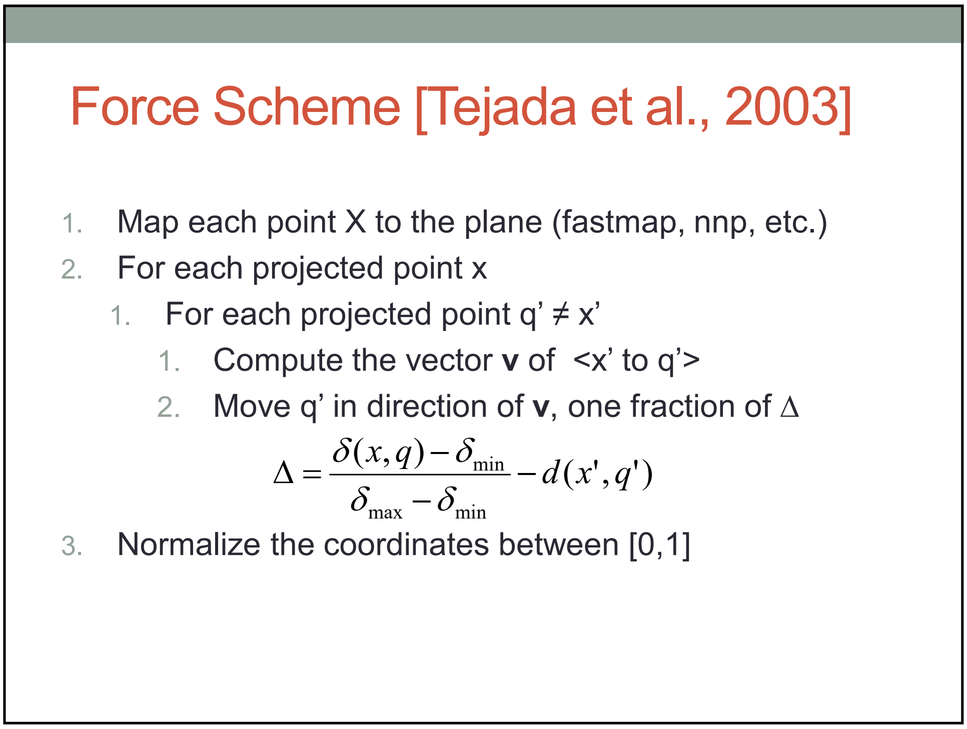
Force Scheme [Tejada et al., 2003]



20

Force Scheme [Tejada et al., 2003]



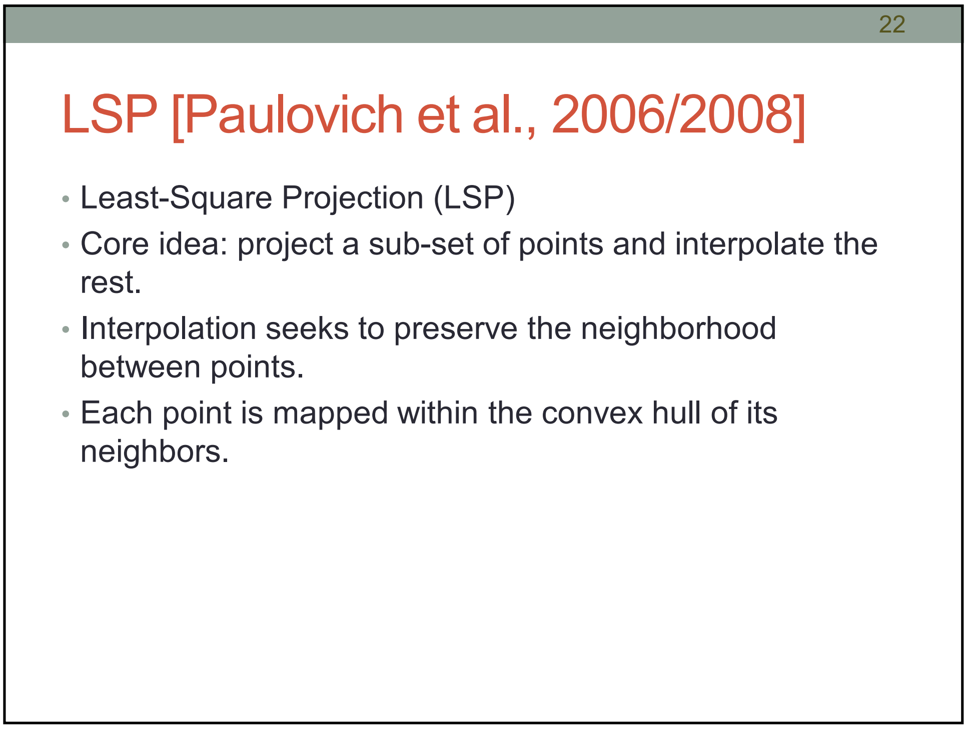


Force Scheme [Tejada et al., 2003]

1. Map each point X to the plane (fastmap, nnp, etc.)
2. For each projected point x
    1. For each projected point q’ ≠ x’
        1. Compute the vector **v** of <x’ to q’>
        2. Move q’ in direction of **v**, one fraction of Δ

$$\Delta = \frac{\delta(x,q)-\delta_{\min}}{\delta_{\max}-\delta_{\min}} - d(x',q')$$

3. Normalize the coordinates between [0,1]



22

LSP [Paulovich et al., 2006/2008]

- Least-Square Projection (LSP)
- Core idea: project a sub-set of points and interpolate the rest.
- Interpolation seeks to preserve the neighborhood between points.
- Each point is mapped within the convex hull of its neighbors.

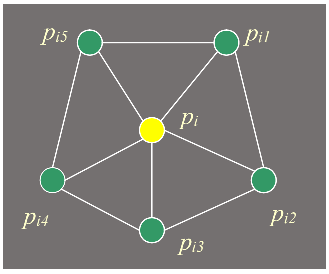
LSP [Paulovich et al., 2006/2008]

- Three main steps:
  1. Select a subset of points(control points) and Project these in *R^p*
  2. Determine the neighborhood of points
  3. Create a linear system whose answers are the Cartesian coordinates of points p_i in R^p

LSP: Laplacian Matrix

- Let V_i = {p_{i1},…,p_{iki}} *be the neighborhood of a point* p_i *and* c_i the coodinates of p_i in R^p

$$c_i - \frac{1}{ki}\sum_{p_j \in V_i} c_j = 0$$



- Each p_i will be the centroid of points in V_i

25

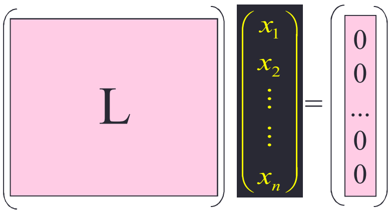
LSP: Laplacian Matrix

L**x**$_1$=0, L**x**$_2$=0, …, L**x**$_p$=0

where x_1, x_2,…, x_p are vectors containing the Cartesian coordinates of the points

and L is the matrix defined by*:*

$$Lij = \begin{cases} 1 & i = j \\ -\dfrac{1}{ki} & p_j \in V_i \\ 0 & otherwise \end{cases}$$



26

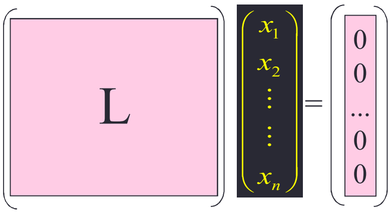
LSP: Matriz Laplaciana

L**x**$_1$=0, L**x**$_2$=0, …, L**x**$_p$=0

onde x_1, x_2,…, x_p são vetores contendo as coordenadas cartesianas dos pontos e *L*

- é a matriz dada por

$$Lij = \begin{cases} 1 & i = j \\ -\dfrac{1}{ki} & p_j \in V_i \\ 0 & \text{caso contrário} \end{cases}$$



27

LSP: Adding control points

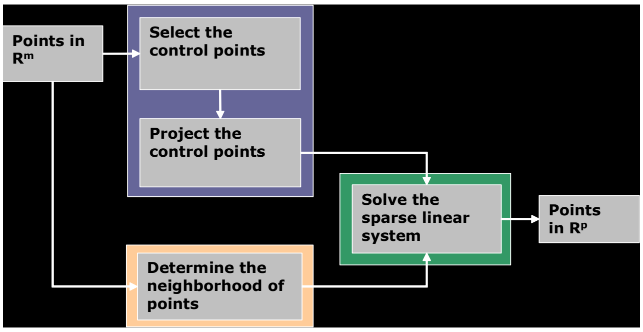
$$A = \begin{pmatrix} L \\ C \end{pmatrix} \qquad Cij = \begin{cases} 1 & p_j \text{ is a control point} \\ 0 & otherwise \end{cases}$$

$$b_i = \begin{cases} 0 & i \leq n \\ x_{p_{c_i}} & n < i \leq n + nc \end{cases}$$

$$\begin{pmatrix} & & L & & \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ c_1 \\ c_2 \end{pmatrix}$$

28

LSP: Overview


Points in R^m
Select the control points
Project the control points
Solve the sparse linear system
Points in R^p
Determine the neighborhood of points

29

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$$L = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \end{bmatrix}$$

30

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$$L = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \end{bmatrix}$$

31

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$$L = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \end{bmatrix}$$

32

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

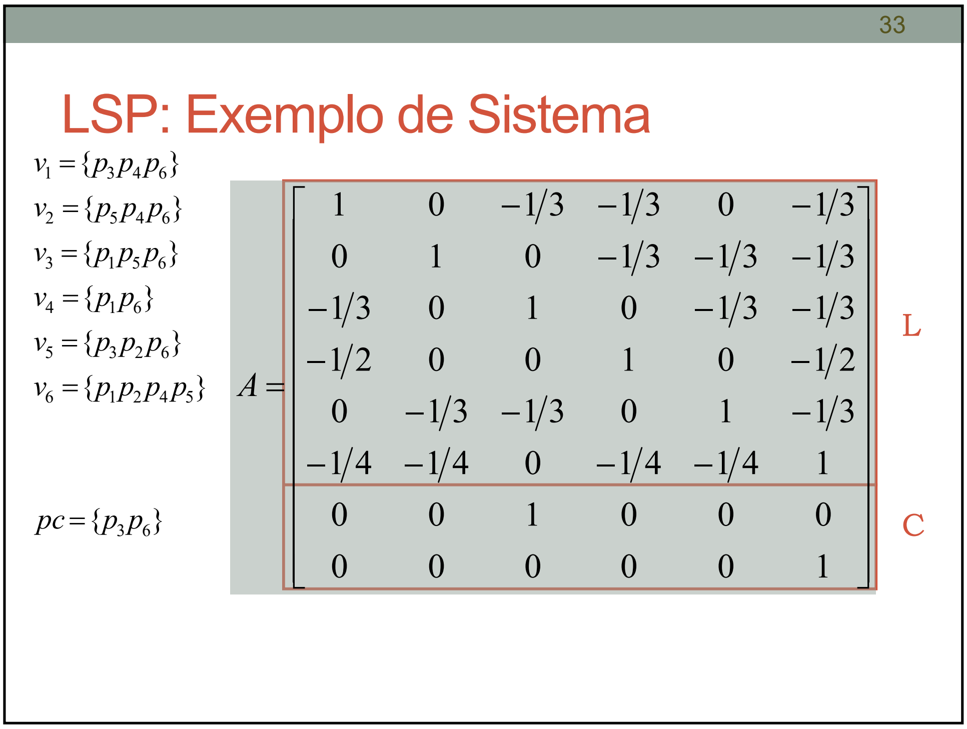
$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

L



33

LSP: Exemplo de Sistema

$v_1=\{p_3p_4p_6\}$

$v_2=\{p_5p_4p_6\}$

$v_3=\{p_1p_5p_6\}$

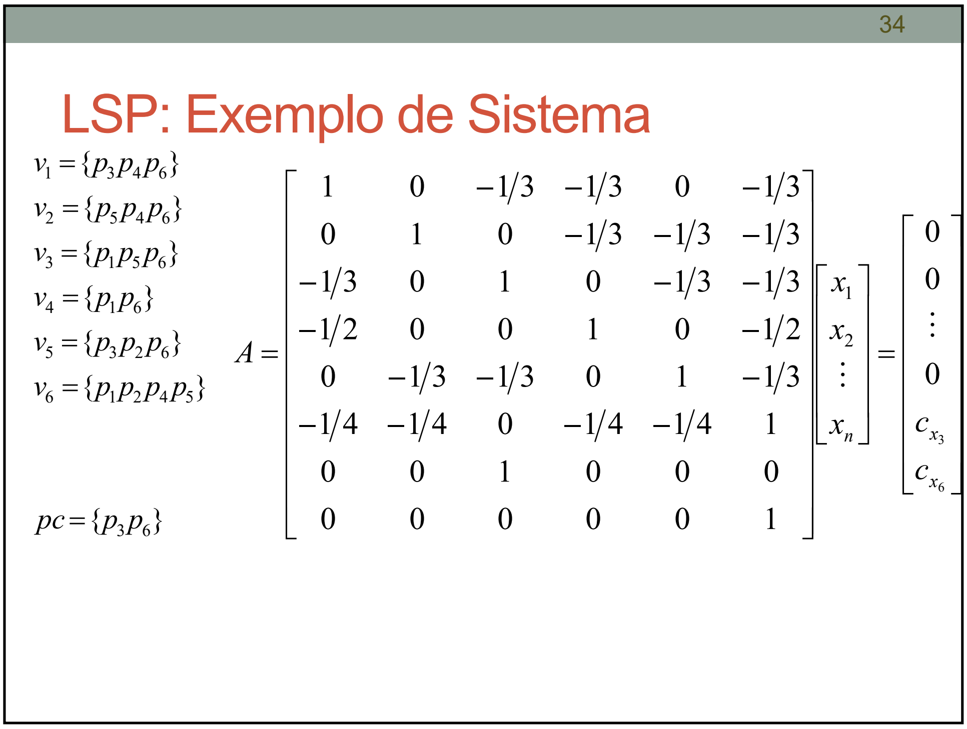
$v_4=\{p_1p_6\}$

$v_5=\{p_3p_2p_6\}$

$v_6=\{p_1p_2p_4p_5\}$

$pc=\{p_3p_6\}$

$$A=\begin{array}{c}\left[\begin{array}{cccccc} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]\end{array}\begin{array}{l} \\ \\ \mathrm{L} \\ \\ \\ \\ \mathrm{C} \\ \\ \end{array}$$



34

LSP: Exemplo de Sistema

$v_1=\{p_3p_4p_6\}$

$v_2=\{p_5p_4p_6\}$

$v_3=\{p_1p_5p_6\}$

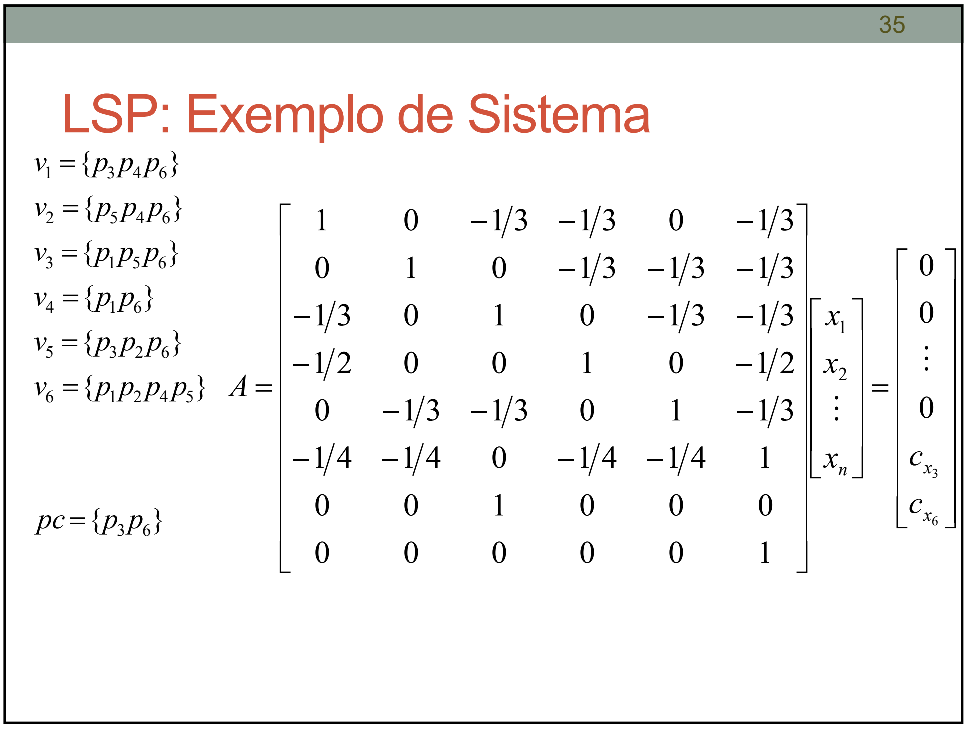
$v_4=\{p_1p_6\}$

$v_5=\{p_3p_2p_6\}$

$v_6=\{p_1p_2p_4p_5\}$

$pc=\{p_3p_6\}$

$$A=\begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{x_3} \\ c_{x_6} \end{bmatrix}$$



35

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

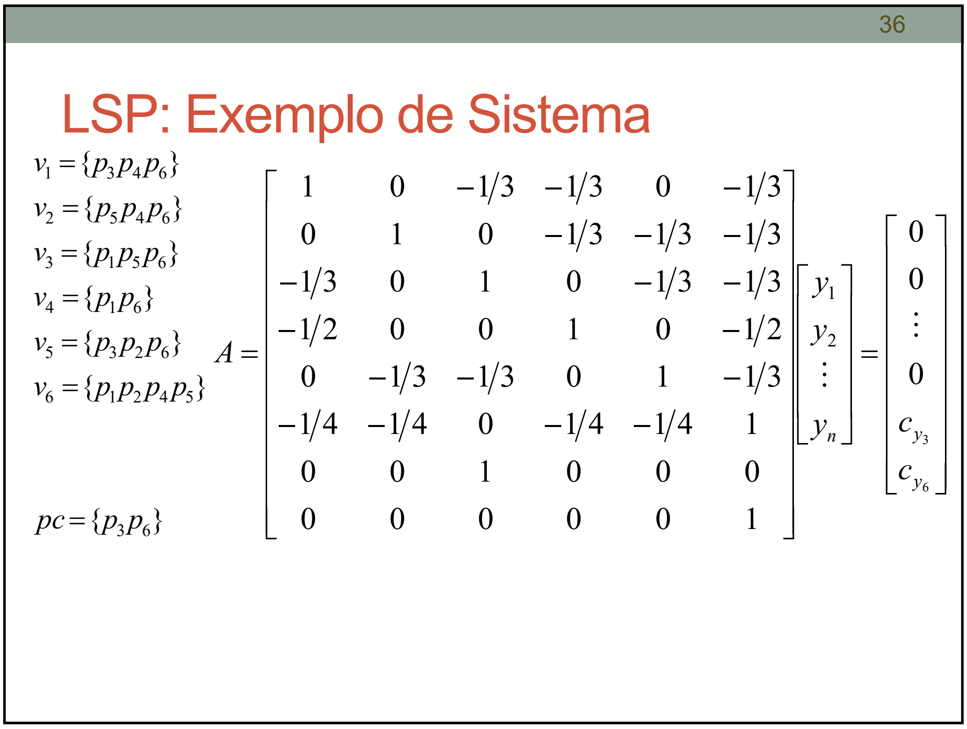
$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{x_3} \\ c_{x_6} \end{bmatrix}$$

$pc = \{p_3 p_6\}$



36

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

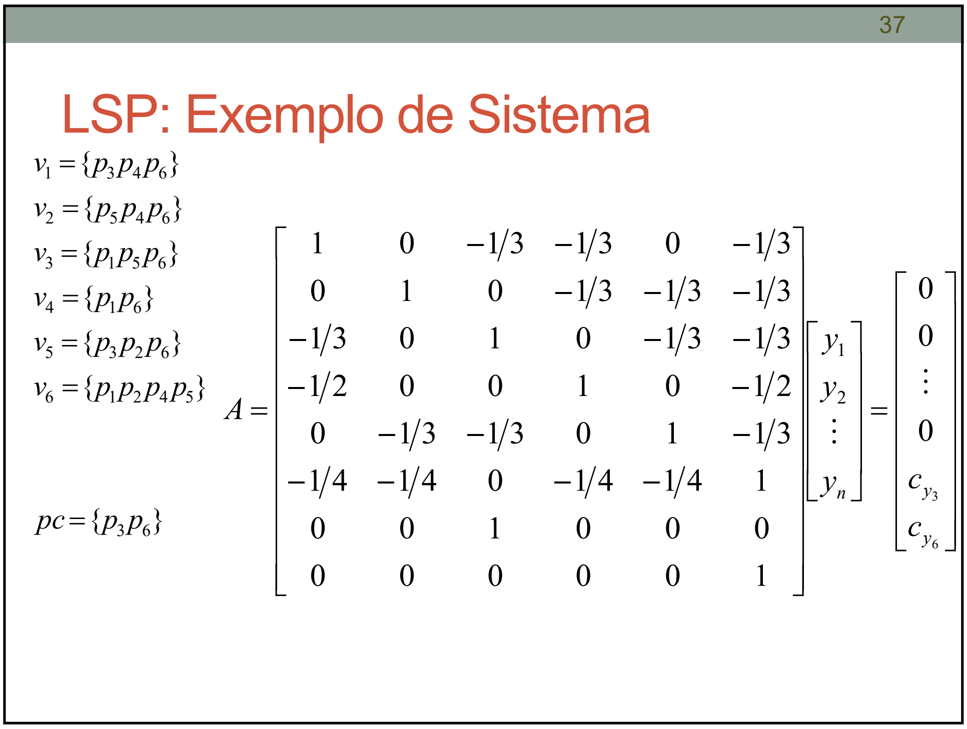
$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{y_3} \\ c_{y_6} \end{bmatrix}$$

$pc = \{p_3 p_6\}$



LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

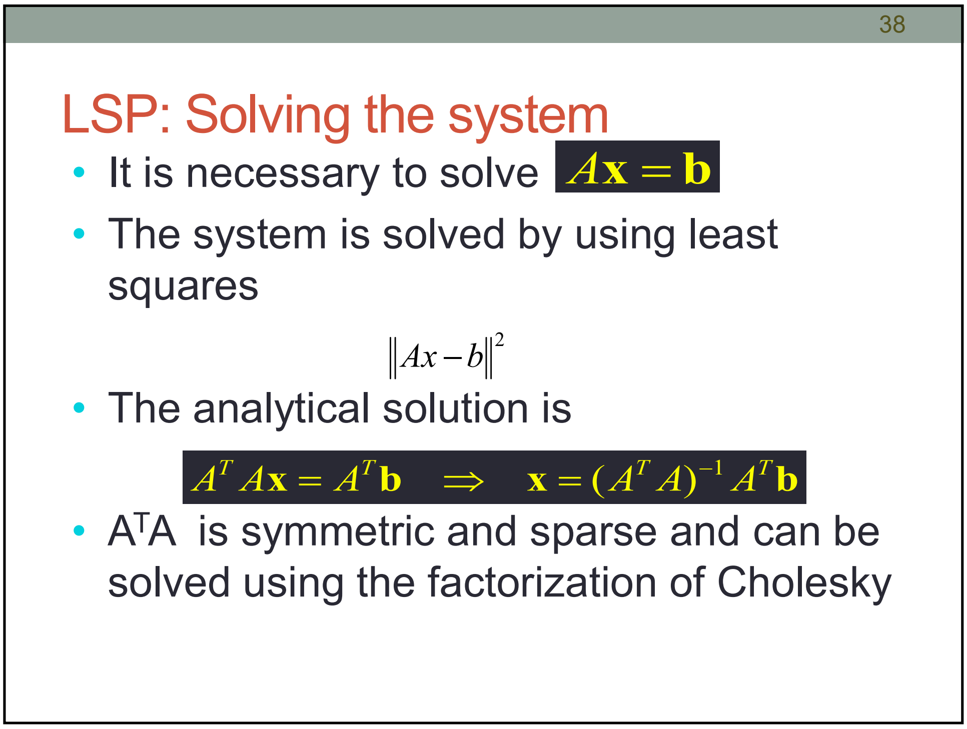
$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{y_3} \\ c_{y_6} \end{bmatrix}$$



LSP: Solving the system

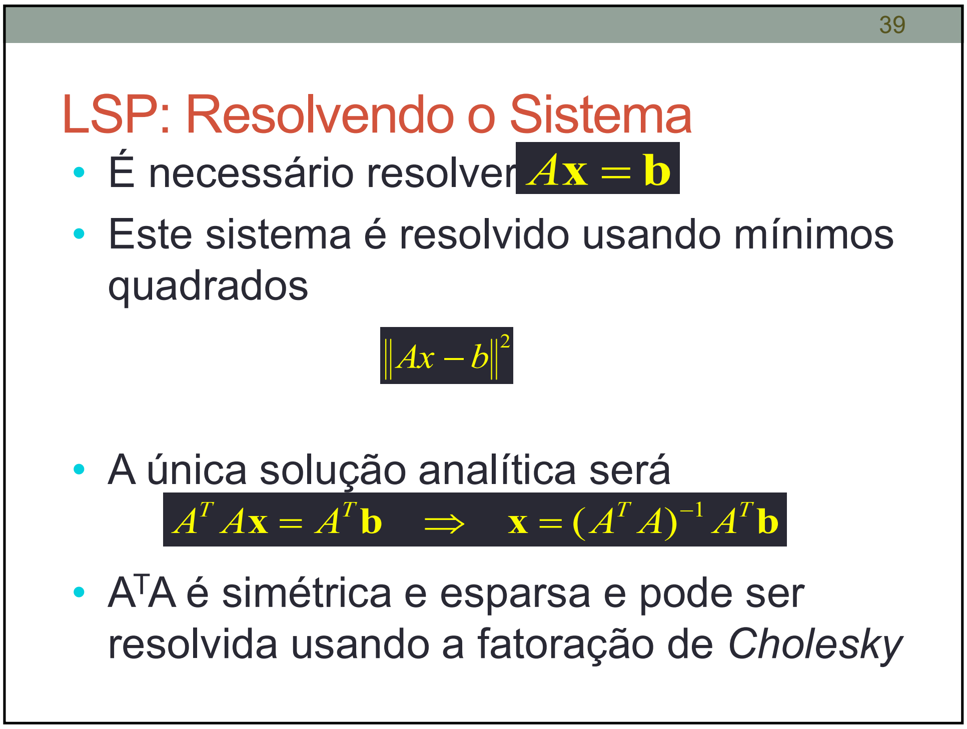
- It is necessary to solve $\mathbf{A}\mathbf{x} = \mathbf{b}$
- The system is solved by using least squares

$$\|Ax - b\|^2$$

- The analytical solution is

$$A^T A\mathbf{x} = A^T\mathbf{b} \quad \Rightarrow \quad \mathbf{x} = (A^T A)^{-1} A^T\mathbf{b}$$

- A^TA is symmetric and sparse and can be solved using the factorization of Cholesky



39

LSP: Resolvendo o Sistema

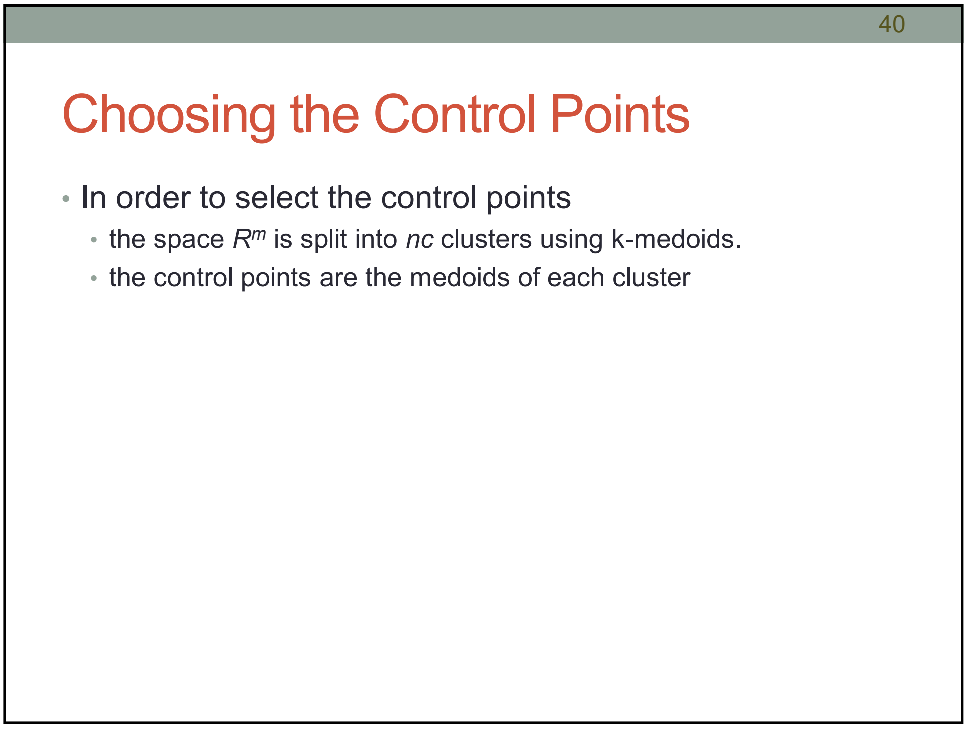
- É necessário resolver $A\mathbf{x} = \mathbf{b}$
- Este sistema é resolvido usando mínimos quadrados

$$\|Ax - b\|^2$$

- A única solução analítica será

$$A^T A\mathbf{x} = A^T\mathbf{b} \quad \Rightarrow \quad \mathbf{x} = (A^T A)^{-1} A^T\mathbf{b}$$

- $A^T A$ é simétrica e esparsa e pode ser resolvida usando a fatoração de *Cholesky*



40

Choosing the Control Points

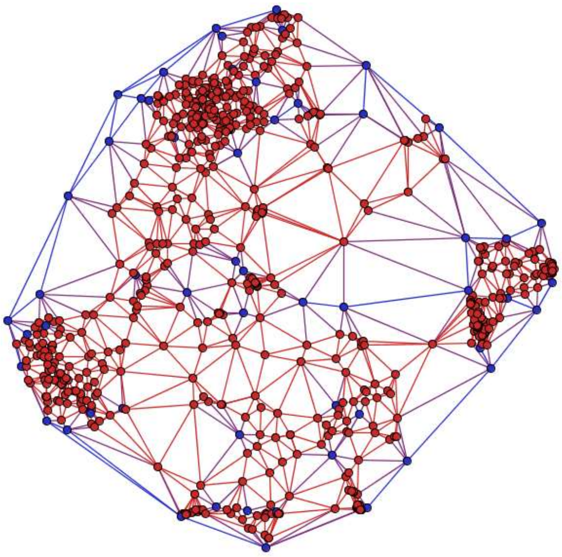
- In order to select the control points
  - the space R^m is split into *nc* clusters using k-medoids.
  - the control points are the medoids of each cluster

41

Choosing the Control Points

- Once the control points are chosen, these points are projected onto R^d through a fast dimensionality reduction method
  - Fast Projection (Fastmap or NNP)
  - Force Placement

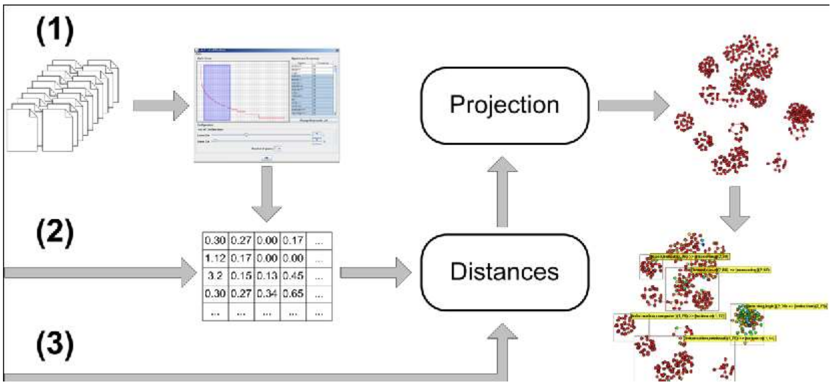
42



Control points in blue

43

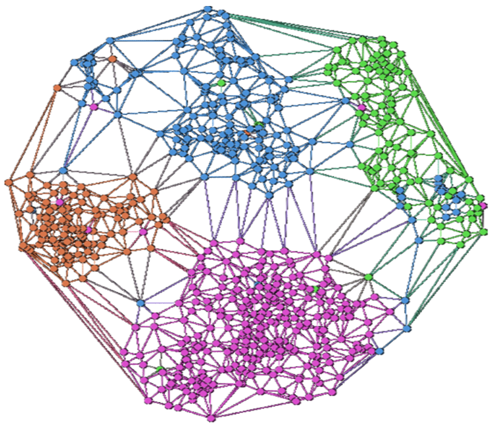
Content – based by Projections



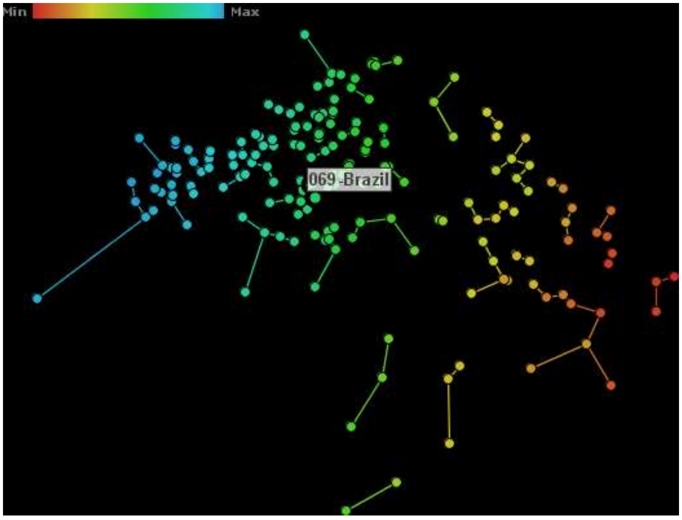
0.30	0.27	0.00	0.17	...
1.12	0.17	0.00	0.00	...
3.2	0.15	0.13	0.45	...
0.30	0.27	0.34	0.65	...
...	...	...	...	...

44

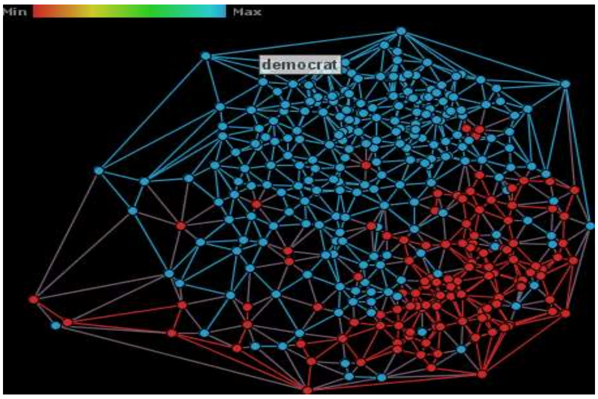
Projection



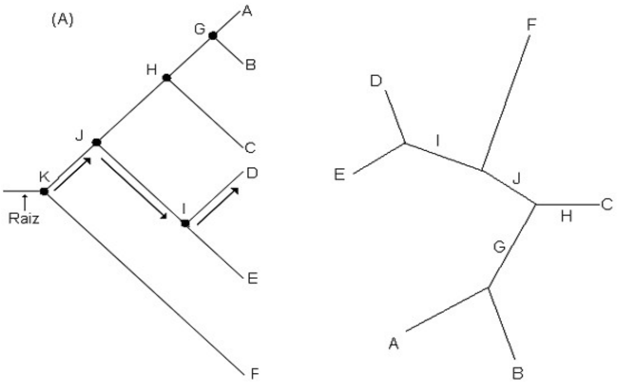
Projection: HDI



Projection: Voting

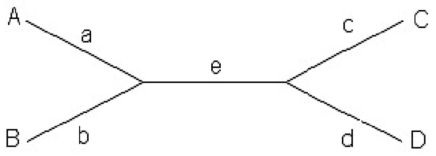


Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)



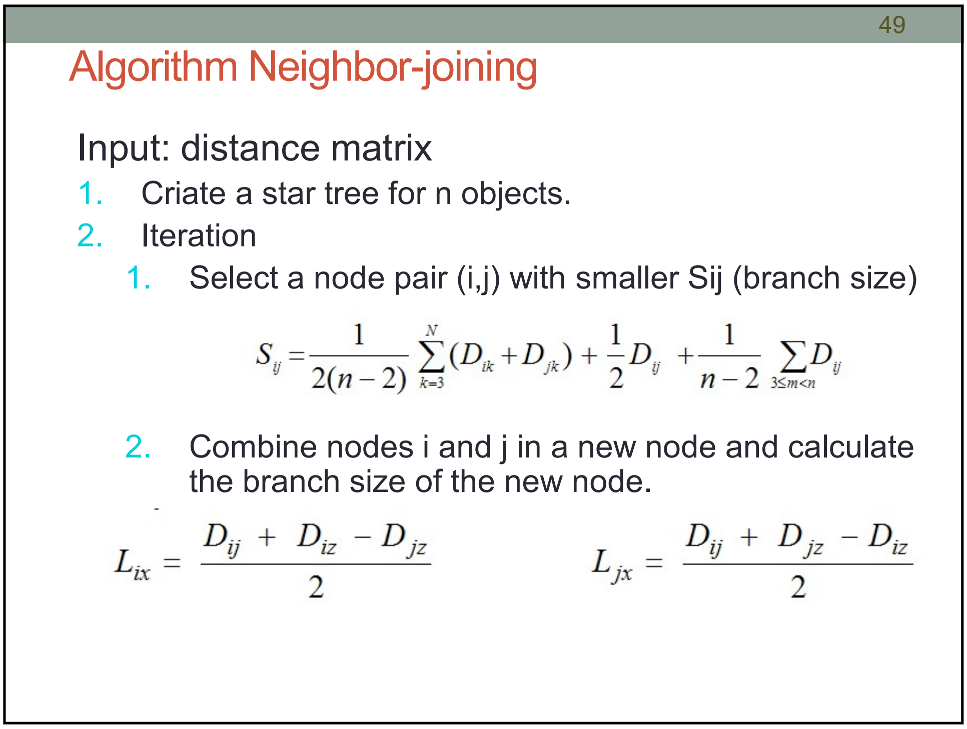
Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)

$$d_{AB} + d_{CD} \le \max\,(d_{AC} + d_{BD}\,,\, d_{AD} + d_{BC})$$



$d_{AB} = a + b$ $d_{AC} = a + e + c$ $d_{AD} = a + e + d$

$d_{CD} = c + d$ $d_{BD} = b + e + d$ $d_{BC} = b + e + a$



Algorithm Neighbor-joining

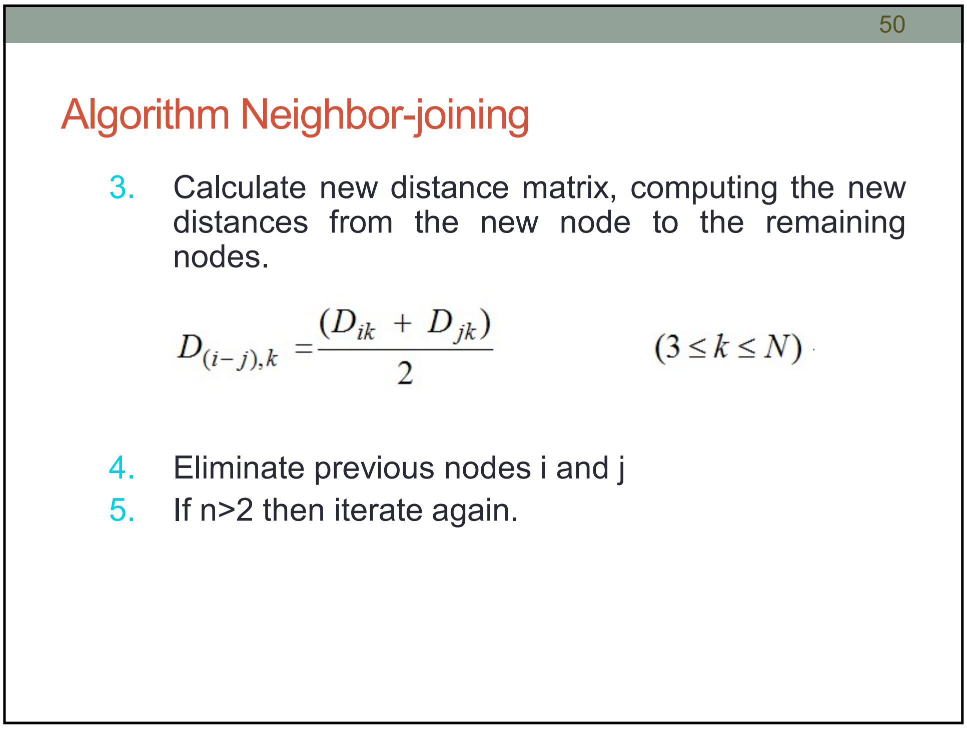
Input: distance matrix

1. Criate a star tree for n objects.
2. Iteration
   1. Select a node pair (i,j) with smaller Sij (branch size)

$$S_{ij} = \frac{1}{2(n-2)} \sum_{k=3}^{N} (D_{ik} + D_{jk}) + \frac{1}{2} D_{ij} + \frac{1}{n-2} \sum_{3 \le m < n} D_{ij}$$

   2. Combine nodes i and j in a new node and calculate the branch size of the new node.

$$L_{ix} = \frac{D_{ij} + D_{iz} - D_{jz}}{2} \qquad L_{jx} = \frac{D_{ij} + D_{jz} - D_{iz}}{2}$$

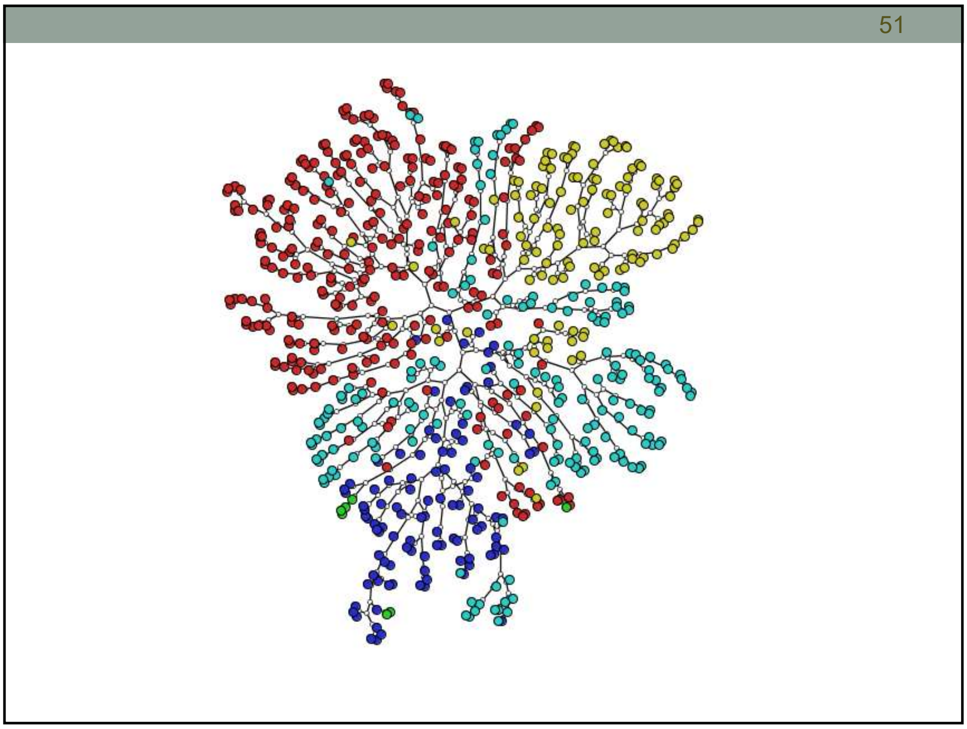


Algorithm Neighbor-joining

3. Calculate new distance matrix, computing the new distances from the new node to the remaining nodes.

$$D_{(i-j),k} = \frac{(D_{ik} + D_{jk})}{2} \qquad (3 \le k \le N)$$

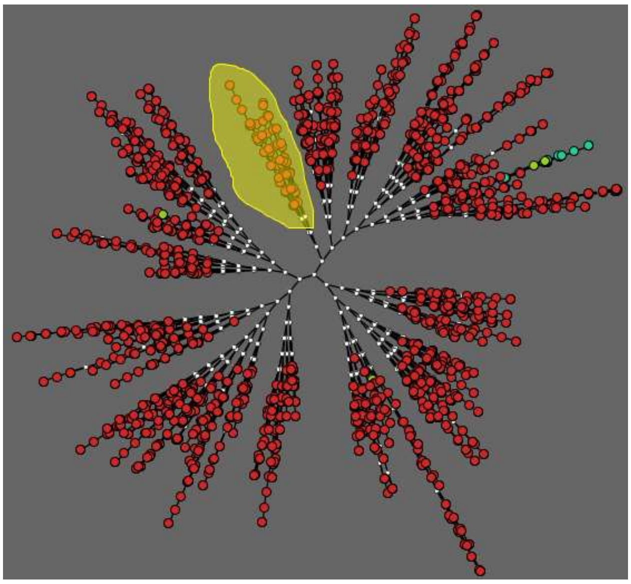
4. Eliminate previous nodes i and j
5. If n>2 then iterate again.



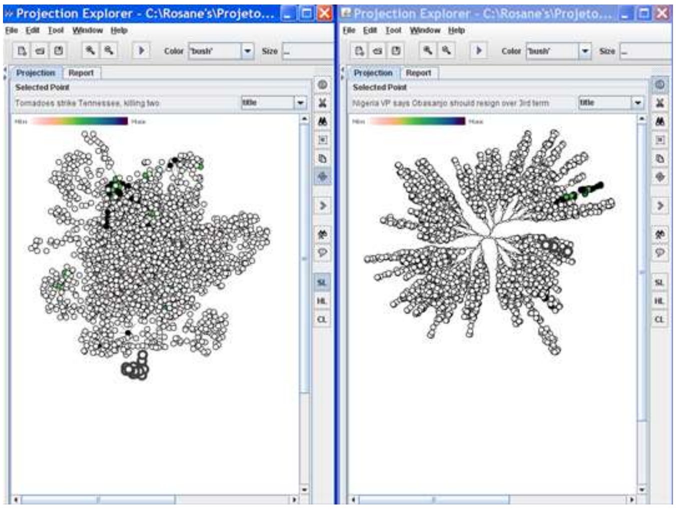
51

52

- Alternate view (N-J Tree)



Projections & Trees



Projection Explorer - C:\Rosane's\Projeto...

Color 'bush'

Selected Point

Tornadoes strike Tennessee, killing two

title

Projection Explorer - C:\Rosane's\Projeto...

Color 'bush'

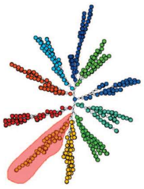
Selected Point

Nigeria VP says Obasanjo should resign over 3rd term

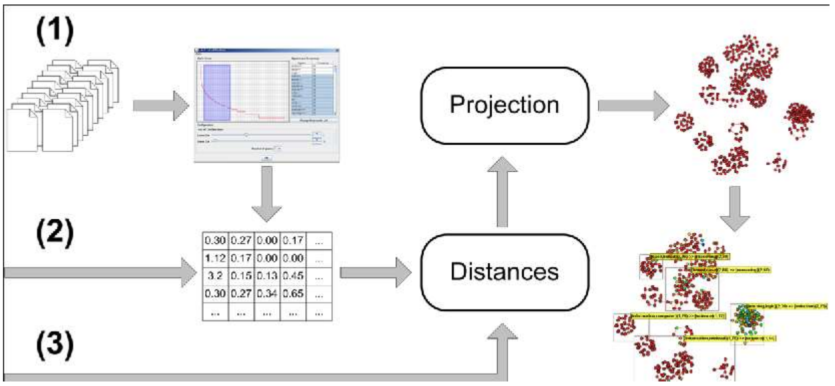
title

Application to Visual Data Classification

- Sample selection
- Classification
- Evolution of models
- Cooperation IC/UNICAMP
  - Helio Pedrini
  - William Schwartz (now UFMG)
- *Aplications: GPS data, Systems biology Data, data on quality of text*
  - *Cooperations with UNESP / Presidente Prudente, LNBio/CNPEM, NILC/ICMC*



Visualization by Projections



0.30	0.27	0.00	0.17	...
1.12	0.17	0.00	0.00	...
3.2	0.15	0.13	0.45	...
0.30	0.27	0.34	0.65	...
...	...	...	...	...

57

The case of document collections

- Applications
  - Teaching/Research
  - Search
  - Investigation

  - Patents
  - Medical reports
  - News

58

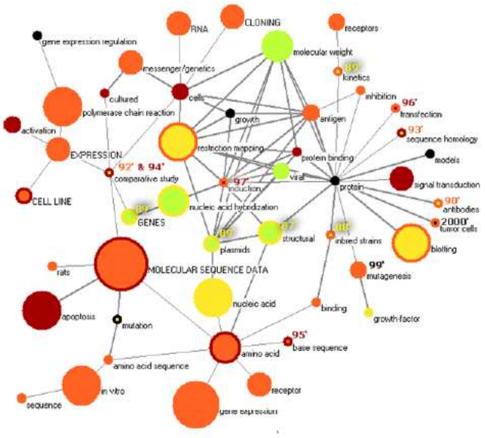
O caso de coleções de documentos

- Aplicações
  - Ensino/Pesquisa
  - Busca
  - Investigação

  - Patentes
  - Laudos médicos
  - Notícias

- Maps of text Collections
  - Based on Relationships (Borner & Chen)
    - Co-authorship, co-citation
  - Based on Content
    - Similarity and Grouping
    - Common underlying subject
    - ➔Topics

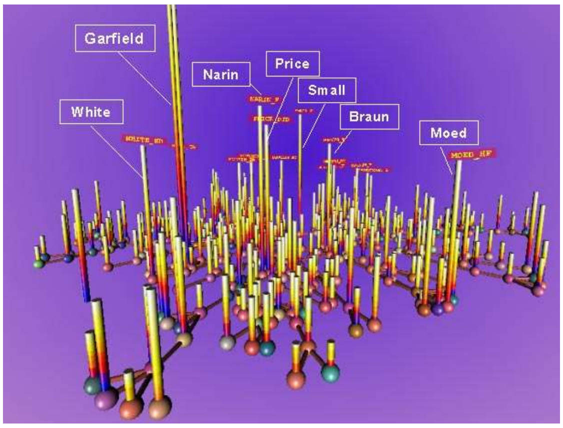
Relationships : Topic Busts and co-word



(Mane and Borner)
2004

61

Relationships : Citation and Co-citation



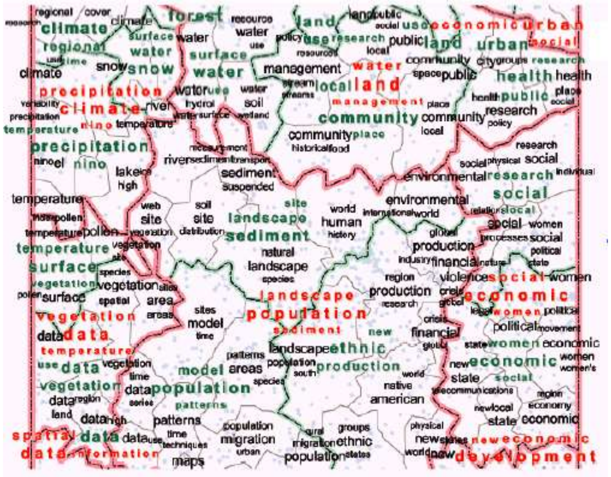
(Borner)
(2003)

62

Content-based Text Mapping

- Approach 1: Dimension reduction
  ex. MSD, SVD, PCA
- Approach 2: Point Placement (PP)
- Approach 3: Clustering
- Approach 4: Projections
  ex. FASPMAP, NNP, LSP

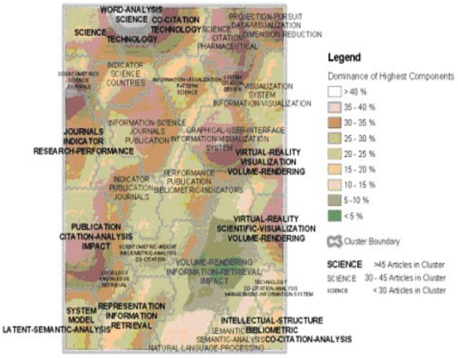
Content - based



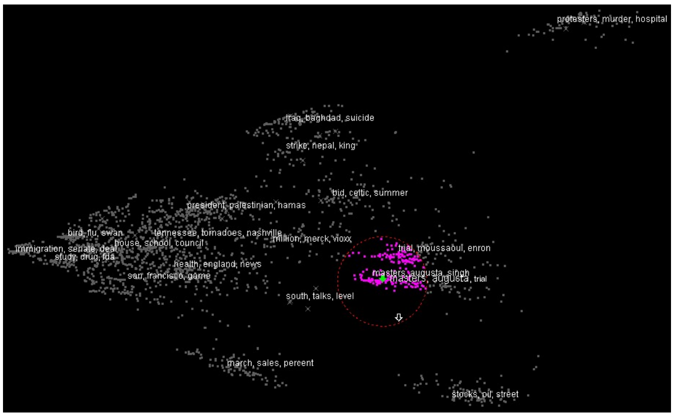
(Skupin)
(2002)
(abstracts)
SOM

SOM based

- **Self-Organization Maps (SOMs) cartográficos (ex. Skurpin 2002)**



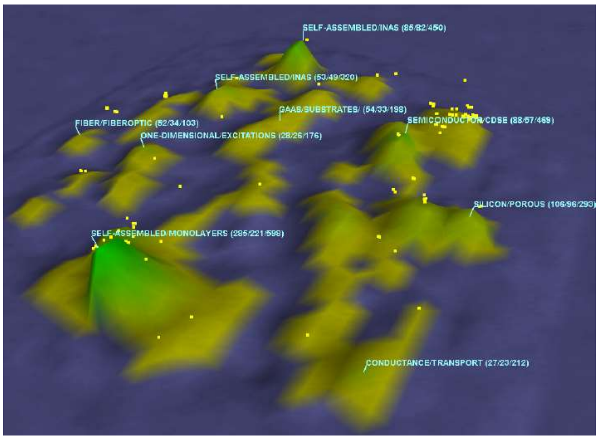
Content - based



(Dimensional Reduction)
News flash
IN-SPIRE
(PNL)

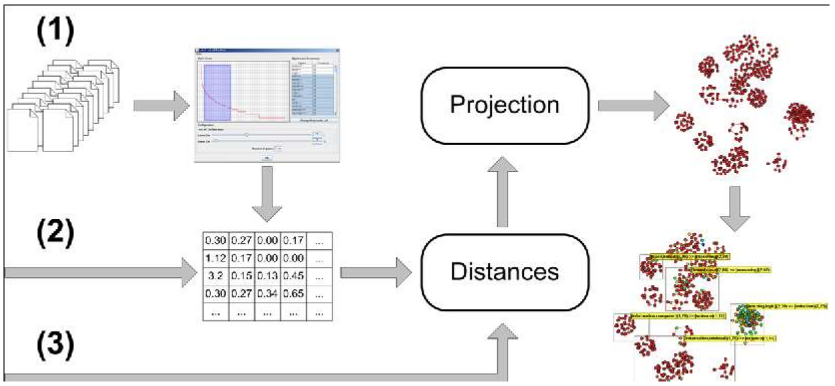
VxInsight

- Sandia National Laboratories, mountain metaphor (Boyack et al., 2002).


SELF-ASSEMBLED/INAS (85/82/450)
SELF-ASSEMBLED/INAS (53/49/320)
GAAS/SUBSTRATES/ (54/33/198)
FIBER/FIBEROPTIC (52/34/103)
SEMICONDUCTOR/CDSE (88/57/469)
ONE-DIMENSIONAL/EXCITATIONS (28/26/176)
SILICON/POROUS (106/96/293)
SELF-ASSEMBLED/MONOLAYERS (285/221/598)
CONDUCTANCE/TRANSPORT (27/23/212)

67

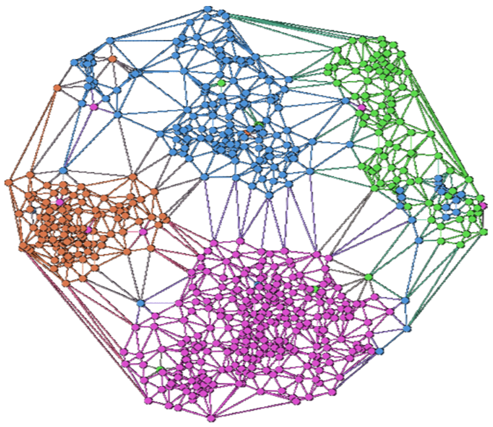
Content – based by Projections



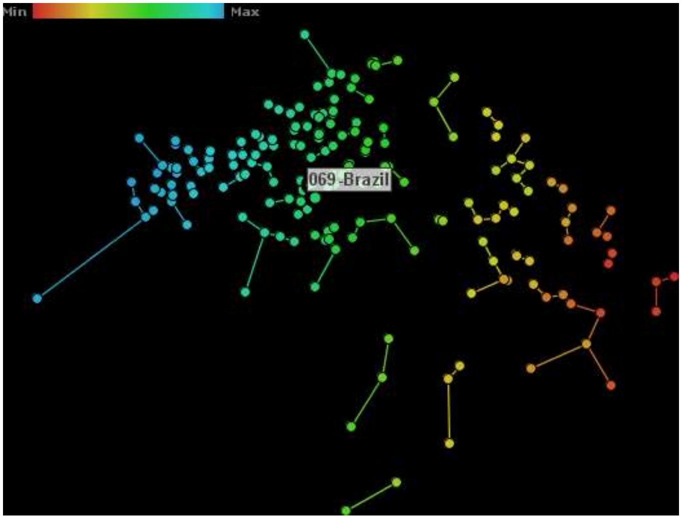
0.30	0.27	0.00	0.17	...
1.12	0.17	0.00	0.00	...
3.2	0.15	0.13	0.45	...
0.30	0.27	0.34	0.65	...
...	...	...	...	...

68

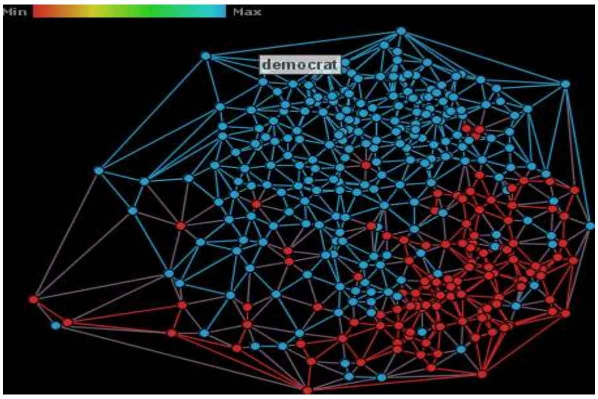
Exemplo de Projeção

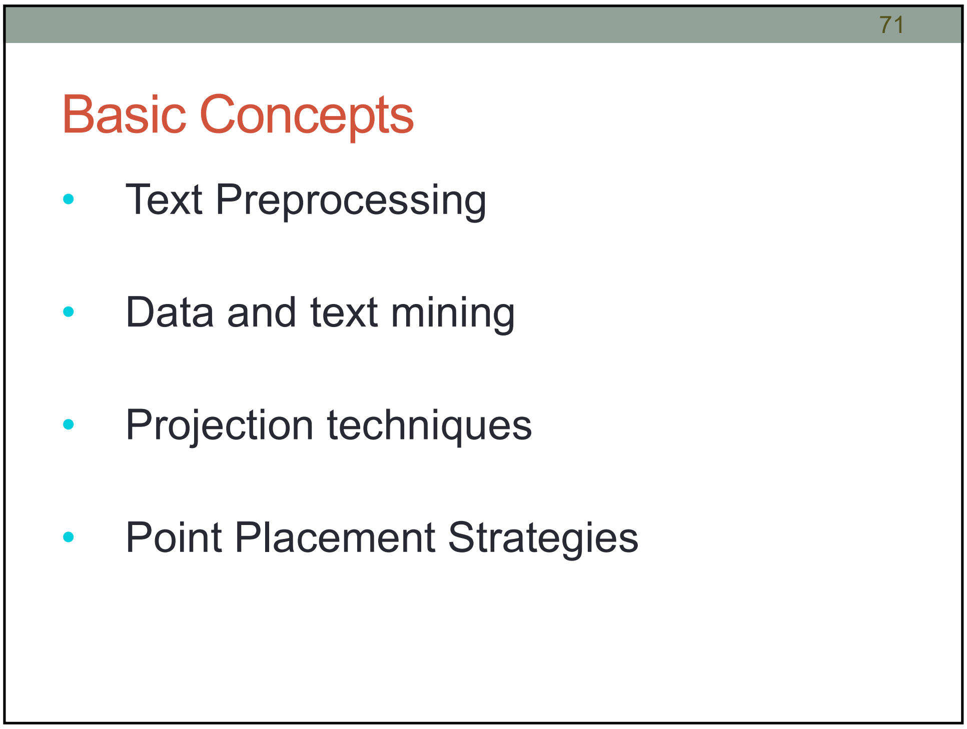


Exemplo de Projeção: IDH



Exemplo de Projeção: Votação

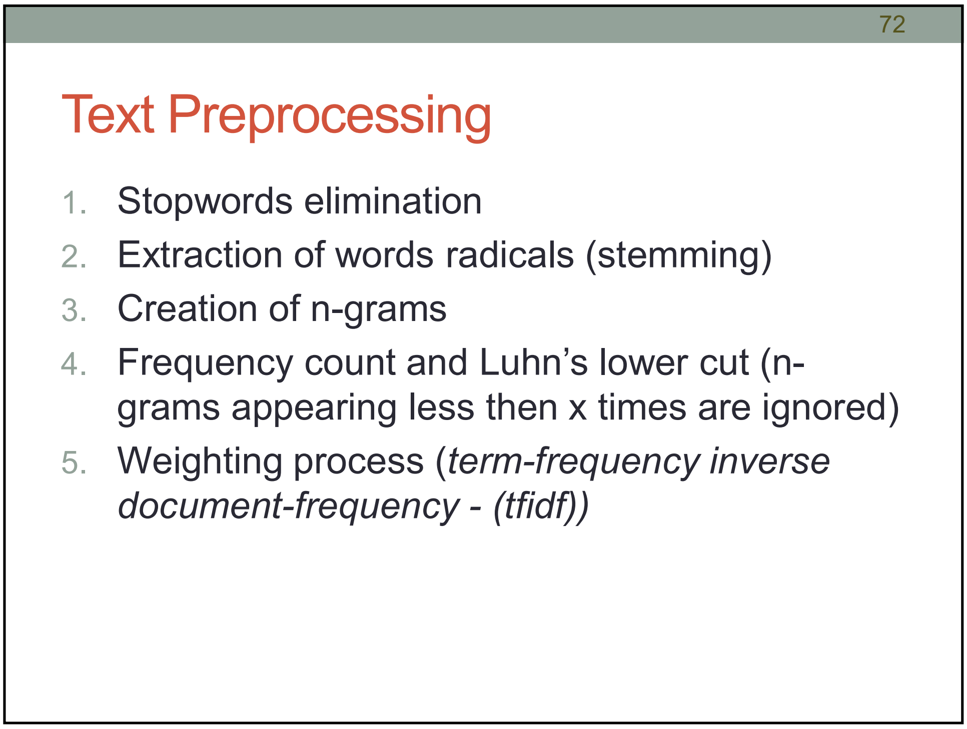




71

Basic Concepts

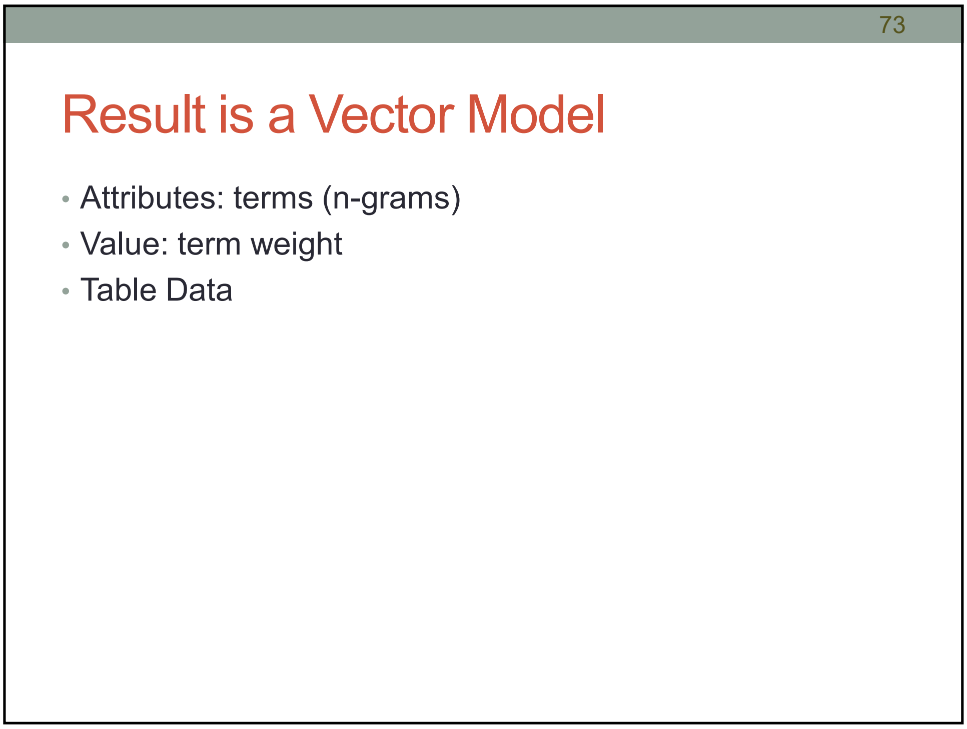
- Text Preprocessing
- Data and text mining
- Projection techniques
- Point Placement Strategies



72

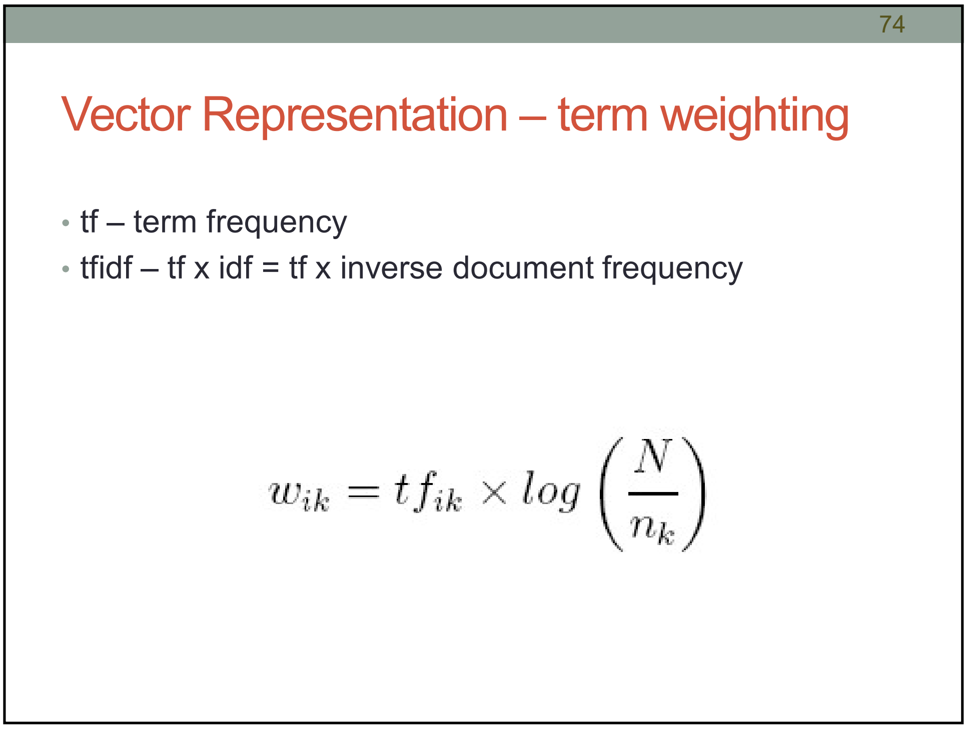
Text Preprocessing

1. Stopwords elimination
2. Extraction of words radicals (stemming)
3. Creation of n-grams
4. Frequency count and Luhn’s lower cut (n-grams appearing less then x times are ignored)
5. Weighting process (*term-frequency inverse document-frequency - (tfidf))*



Result is a Vector Model

- Attributes: terms (n-grams)
- Value: term weight
- Table Data



Vector Representation – term weighting

- tf – term frequency
- tfidf – tf x idf = tf x inverse document frequency

$$w_{ik} = tf_{ik} \times log\left(\frac{N}{n_k}\right)$$

Vector Representation

	term_1	term_2	term_3	term_4	...	term_m
Doc_1	0.92	0.62	0.92	0.10	...	0.67
Doc_2	0.13	0.11	1.00	0.34	...	0.33
Doc_3	0.52	0.00	0.00	0.44	...	0.77
...	...	...	...	...	...	...
Doc_n	0.02	0.12	0.22	0.92	...	0.00

Vector Representation – Similarity calculatin

EUCLIDEAN

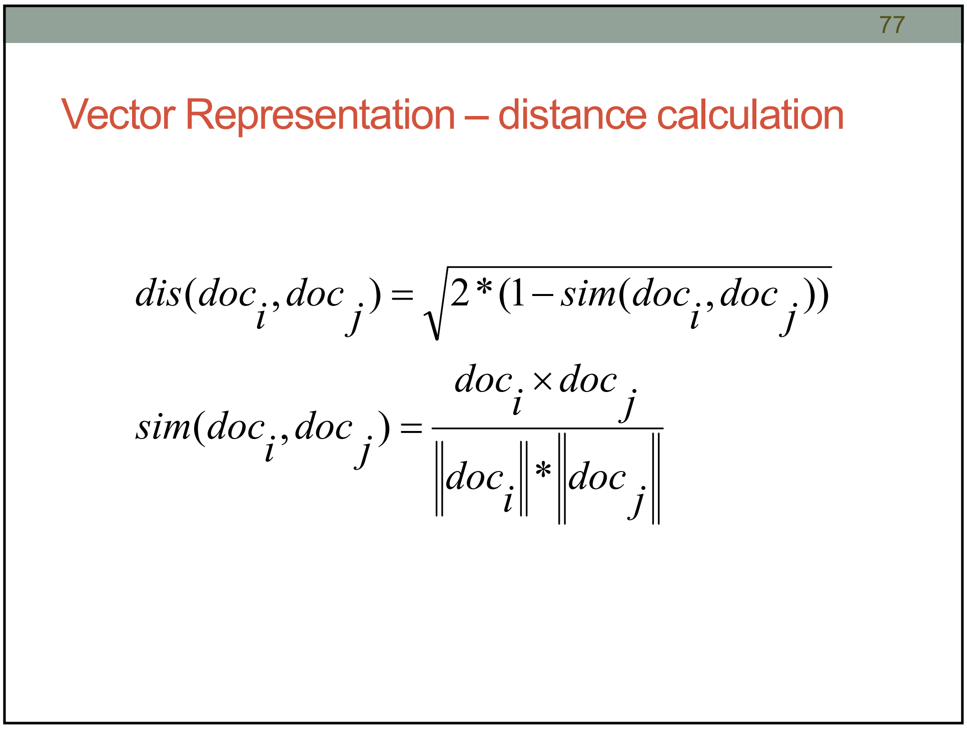
$$sim_{i,j} = \sqrt{(w_{i,1} - w_{j,1})^2 + \ldots + (w_{i,k} - w_{j,k})^2}$$

MANHATAN

$$sim_{i,j} = |w_{i,1} - w_{j,1}| + \ldots + |w_{i,1} - w_{j,1}|$$

COSINE

$$sim_{i,j} = \frac{(w_{i,1} \times w_{j,1}) + \ldots + (w_{i,k} \times w_{j,k})}{(w_{i,1}^2 + \ldots + w_{i,k}^2) \times (w_{j,1}^2 + \ldots + w_{j,k}^2)}$$

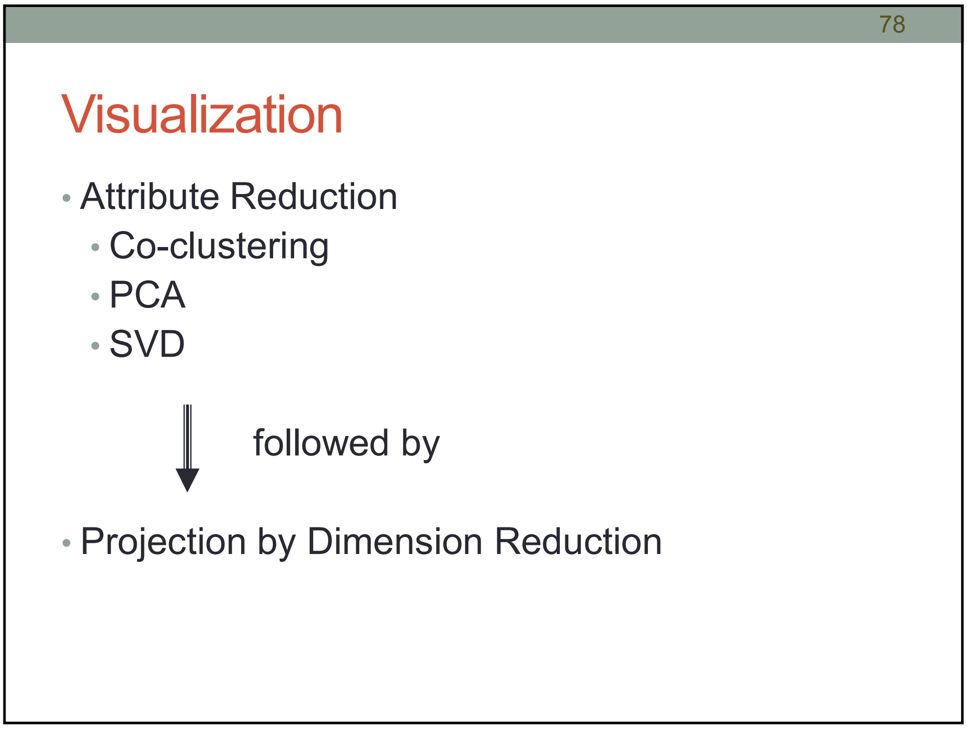


77

Vector Representation – distance calculation

$$dis(doc_i, doc_j) = \sqrt{2 * (1 - sim(doc_i, doc_j))}$$

$$sim(doc_i, doc_j) = \frac{doc_i \times doc_j}{\left\| doc_i \right\| * \left\| doc_j \right\|}$$



78

Visualization

- Attribute Reduction
  - Co-clustering
  - PCA
  - SVD

↓ followed by

- Projection by Dimension Reduction

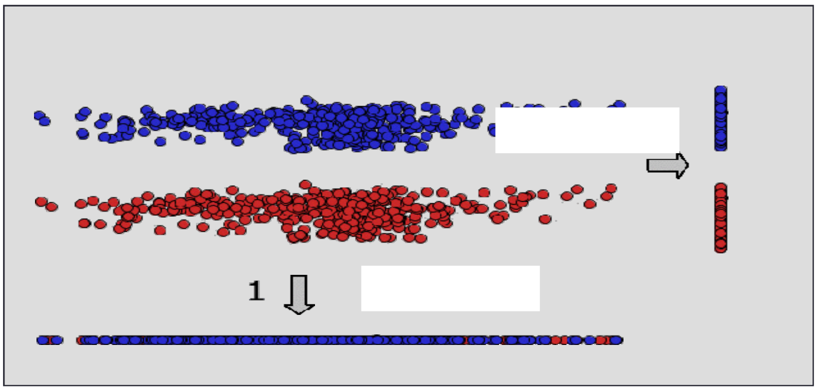
79

Alternative to Vector Representation

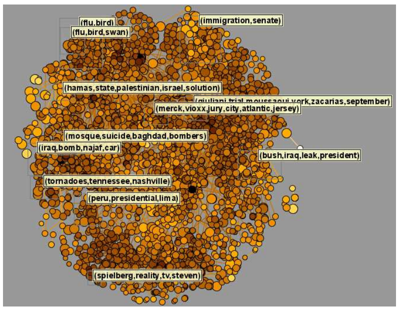
- Similarity Calculation text against text
  - Ex: NCD Normalized Compression Distance
    - Approximation of Kolmogorov Complexity
    - Ver: G. P. Telles, R. Minghim, and F. V. Paulovich. 2007. Visual Analytics: Normalized compression distance for visual analysis of document collections. *Comput. Graph.* 31, 3 (June 2007), 327-337. DOI=http://dx.doi.org/10.1016/j.cag.2007.01.024
  - Editing distance
    - Dice's coefficient
    - Matching's coefficient
    - Overlap's coefficient
    - Qgram Distance
    - Ver: Frizzi San Roman Salazar. Um estudo sobre o papel de medidas de similaridade na visualização de coleções de documentos. 2012. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Fundação de Amparo à Pesquisa do Estado de São Paulo. Orientador: Maria Cristina Ferreira de Oliveira.

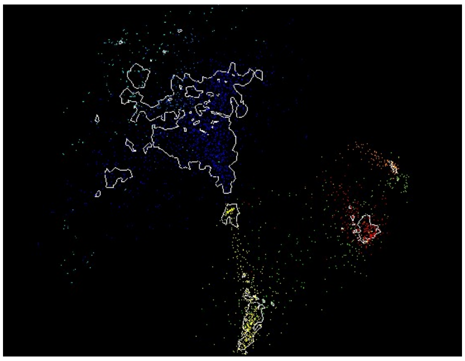
80

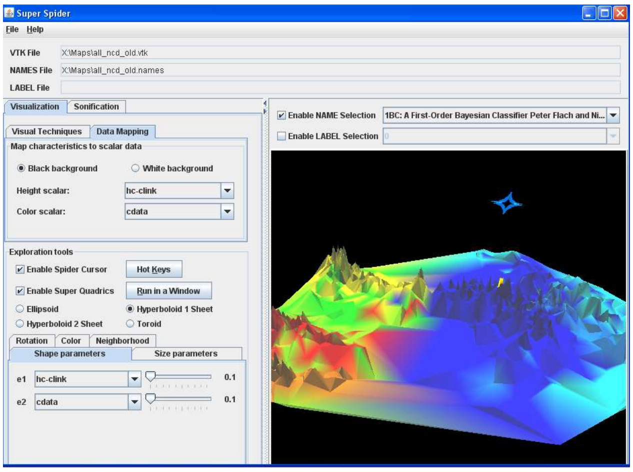
Problems with Stress measurements



Visual representations: graphs, surfaces, volumes, triangulations


(flu,bird)
(flu,bird,swan)
(immigration,senate)
(hamas,state,palestinian,israel,solution)
(giuliani,trial,moussaoui,york,zacarias,september)
(merck,vioxx,jury,city,atlantic,jersey)
(mosque,suicide,baghdad,bombers)
(iraq,bomb,najaf,car)
(bush,iraq,leak,president)
(tornadoes,tennessee,nashville)
(peru,presidential,lima)
(spielberg,reality,tv,steven)





VTK File X:\Maps\all_ncd_old.vtk

NAMES File X:\Maps\all_ncd_old.names

LABEL File

Map characteristics to scalar data

● Black background ○ White background

Height scalar: hc-clink

Color scalar: cdata

Exploration tools

☑ Enable Spider Cursor

☑ Enable Super Quadrics

○ Ellipsoid ● Hyperboloid 1 Sheet

○ Hyperboloid 2 Sheet ○ Toroid

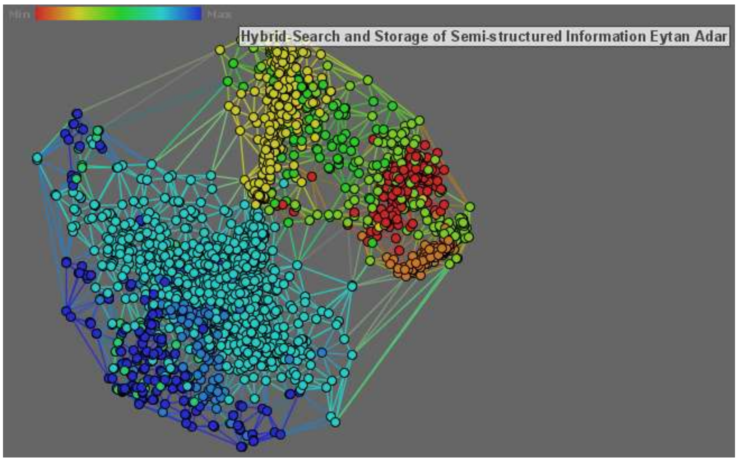
e1 hc-clink 0.1

e2 cdata 0.1

☑ Enable NAME Selection 1BC: A First-Order Bayesian Classifier Peter Flach and Ni...

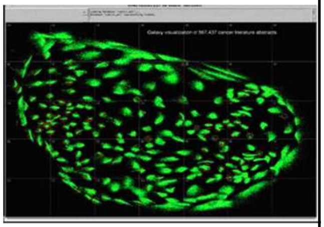
☐ Enable LABEL Selection 0

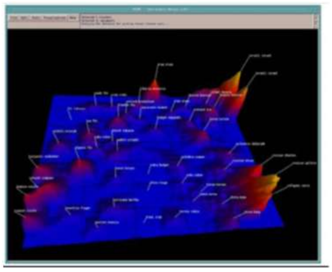
Exploration



IN-SPIRE

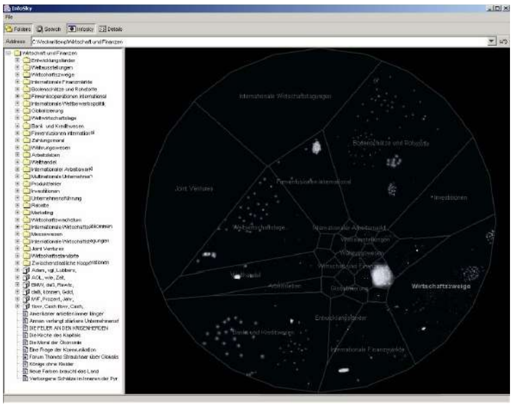
- Spatial Paradigm for Information Retrieval - Pacific Nortwest National Laboratories
- Two Visualization Metaphors:
  - Galaxies – dimensional reduction
  - Themescape

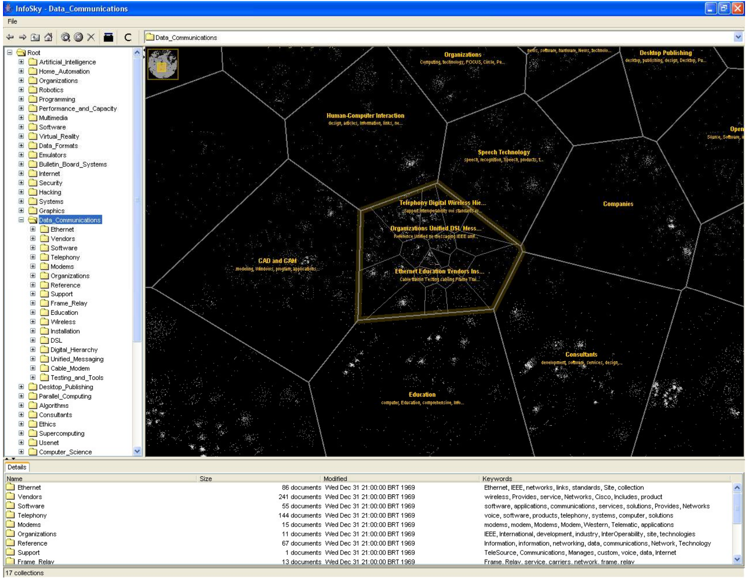




InfoSky

Granitzer (Granitzer et al., 2004) also employs galaxy metaphor





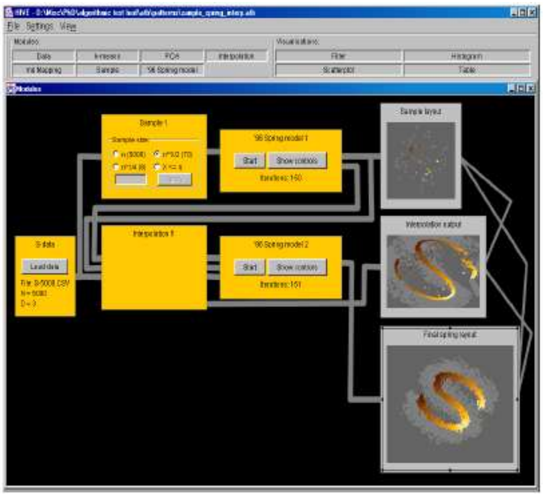
Name	Size	Modified	Keywords
Ethernet	86 documents	Wed Dec 31 21:00:00 BRT 1969	Ethernet, IEEE, networks, links, standards, Site, collection
Vendors	241 documents	Wed Dec 31 21:00:00 BRT 1969	wireless, Provides, service, Networks, Cisco, Includes, product
Software	55 documents	Wed Dec 31 21:00:00 BRT 1969	software, applications, communications, services, solutions, Provides, Networks
Telephony	144 documents	Wed Dec 31 21:00:00 BRT 1969	voice, software, products, telephony, systems, computer, solutions
Modems	15 documents	Wed Dec 31 21:00:00 BRT 1969	modems, modem, Modems, Modem, Western, Telematic, applications
Organizations	11 documents	Wed Dec 31 21:00:00 BRT 1969	IEEE, International, development, industry, InterOperability, site, technologies
Reference	67 documents	Wed Dec 31 21:00:00 BRT 1969	Information, information, networking, data, communications, Network, Technology
Support	1 documents	Wed Dec 31 21:00:00 BRT 1969	TeleSource, Communications, Manages, custom, voice, data, Internet
Frame_Relay	13 documents	Wed Dec 31 21:00:00 BRT 1969	Frame, Relay, service, carriers, network, frame, relay

17 collections

http://en.know-center.at/forschung/wissenserschliessung/downloads_demos/infosky_demo

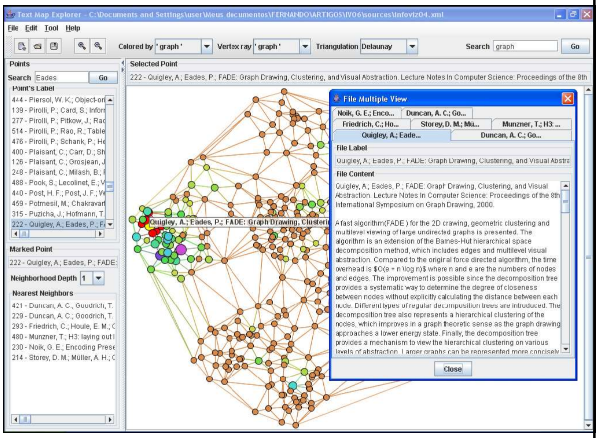
HIVE (Ross and Chalmers 2003)

- Interconnected components:
  - Import
  - Transform
  - Render multi-dim data



Projection Explorer (PEx)

- Projection and Point placement
- Precision
- Graphs and surfaces (Super Spider)



Colored by ' graph ' Vertex ray ' graph ' Triangulation Delaunay Search graph

Search Eades

444 - Piersol, W. K.; Object-ori
139 - Pirolli, P.; Card, S.; Inforr
277 - Pirolli, P.; Pitkow, J.; Rac
514 - Pirolli, P.; Rao, R.; Table
476 - Pirolli, P.; Schank, P.; He
400 - Plaisant, C.; Carr, D.; Sh
126 - Plaisant, C.; Grosjean, J
248 - Plaisant, C.; Milash, B.; F
488 - Pook, S.; Lecolinet, E.; V
440 - Post, H. F.; Post, J. F.; W
459 - Potmesil, M.; Chakravart
315 - Puzicha, J.; Hofmann, T.
222 - Quigley, A.; Eades, P.; F/

Selected Point: 222 - Quigley, A.; Eades, P.; FADE: Graph Drawing, Clustering, and Visual Abstraction. Lecture Notes In Computer Science: Proceedings of the 8th

Marked Point: 222 - Quigley, A.; Eades, P.; FADE:

Neighborhood Depth 1

421 - Duncan, A. C., Goodrich, T.
229 - Duncan, A. C.; Goodrich, T.
293 - Friedrich, C.; Houle, E. M.; C
480 - Munzner, T.; H3: laying out l
230 - Noik, G. E.; Encoding Prese
214 - Storey, D. M.; Müller, A. H.; C

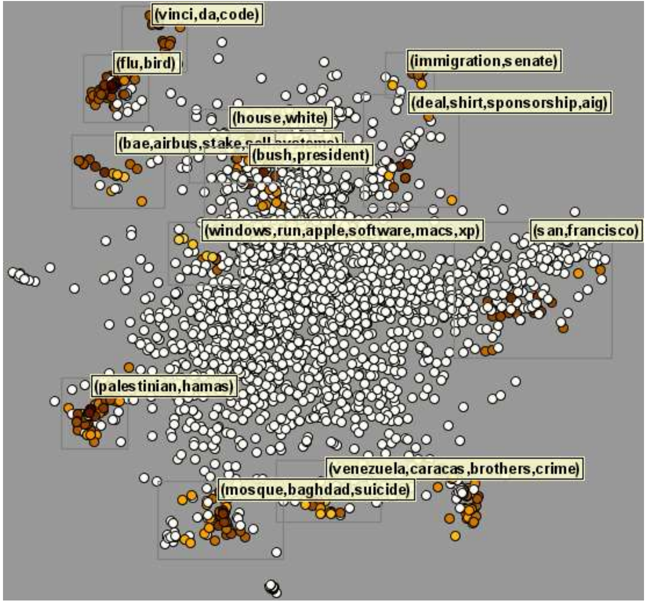
File Label: Quigley, A.; Eades, P.; FADE: Graph Drawing, Clustering, and Visual Abstra

Quigley, A.; Eades, P.; FADE: Grapł Drawing, Clustering, and Visual Abstraction. Lecture Notes In Computer Science: Proceedings of the 8th International Symposium on Graph Drawing, 2000.

A fast algorithm(FADE) for the 2D crawing, geometric clustering and multilevel viewing of large undirected graphs is presented. The algorithm is an extension of the Barnes-Hut hierarchical space decomposition method, which includes edges and multilevel visual abstraction. Compared to the origir al force directed algorithm, the time overhead is $O(e + n \log n)$ where n and e are the numbers of nodes and edges. The improvement is possible since the decomposition tree provides a systematic way to determine the degree of closeness between nodes without explicitly calculating the distance between each node. Different types of regular decomposition trees are introduced. The decomposition tree also represents a hierarchical clustering of the nodes, which improves in a graph theoretic sense as the graph drawing approaches a lower energy state. Finally, the decomposition tree provides a mechanism to view the hierarchical clustering on various levels of abstraction. Larger graphs can be represented more concisely

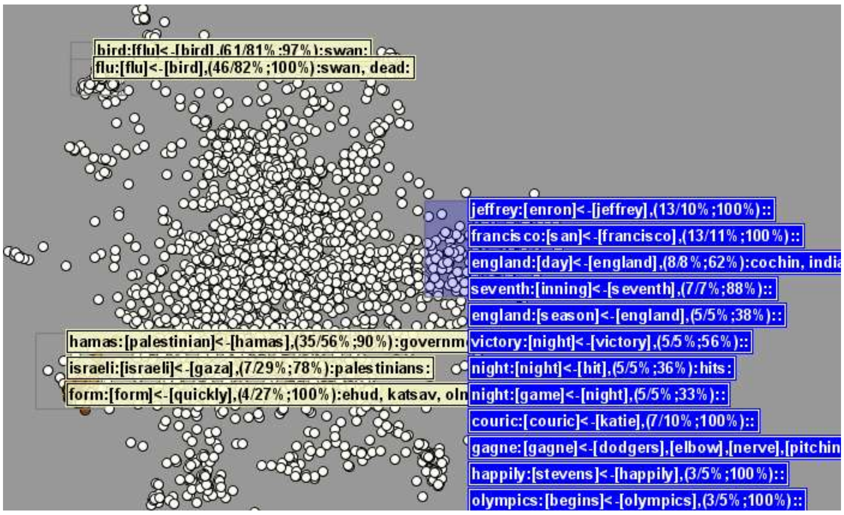
Mapping Text Collections via Projections and Point Placement

- Positioning and labeling



- Detailing topics

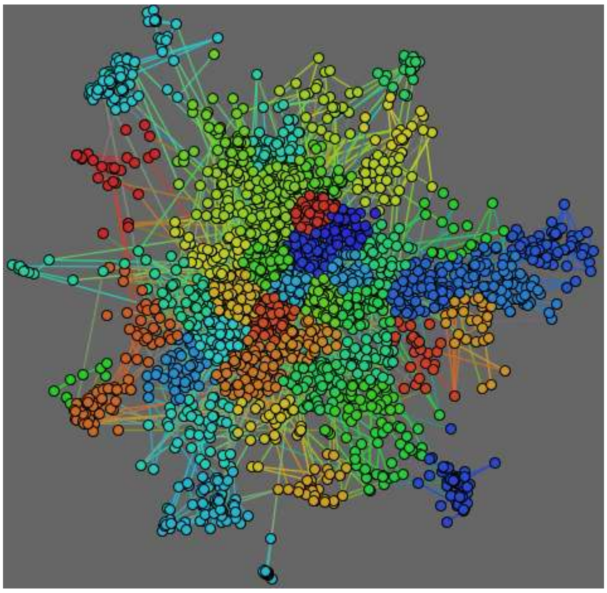
→


bird:[flu]<-[bird],(61/81%;97%):swan:
flu:[flu]<-[bird],(46/82%;100%):swan, dead:
jeffrey:[enron]<-[jeffrey],(13/10%;100%)::
francisco:[san]<-[francisco],(13/11%;100%)::
england:[day]<-[england],(8/8%;62%):cochin, india
seventh:[inning]<-[seventh],(7/7%;88%)::
england:[season]<-[england],(5/5%;38%)::
hamas:[palestinian]<-[hamas],(35/56%;90%):governm
victory:[night]<-[victory],(5/5%;56%)::
israeli:[israeli]<-[gaza],(7/29%;78%):palestinians:
night:[night]<-[hit],(5/5%;36%):hits:
form:[form]<-[quickly],(4/27%;100%):ehud, katsav, oln
night:[game]<-[night],(5/5%;33%)::
couric:[couric]<-[katie],(7/10%;100%)::
gagne:[gagne]<-[dodgers],[elbow],[nerve],[pitchin
happily:[stevens]<-[happily],(3/5%;100%)::
olympics:[begins]<-[olympics],(3/5%;100%)::

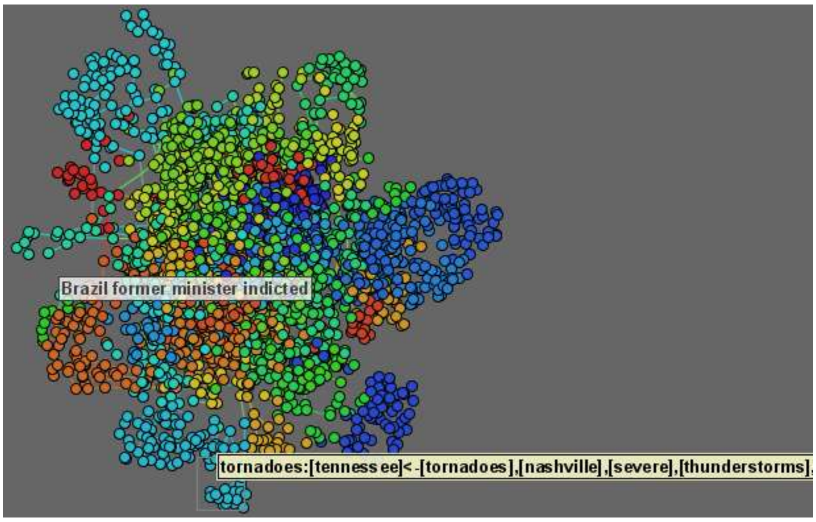
93

- Finding Relationships

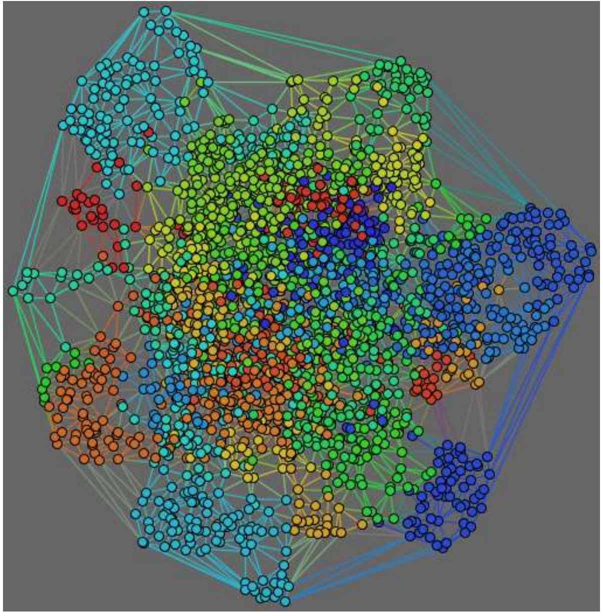
94



- Untangling


Brazil former minister indicted
tornadoes:[tennessee]<-[tornadoes],[nashville],[severe],[thunderstorms],

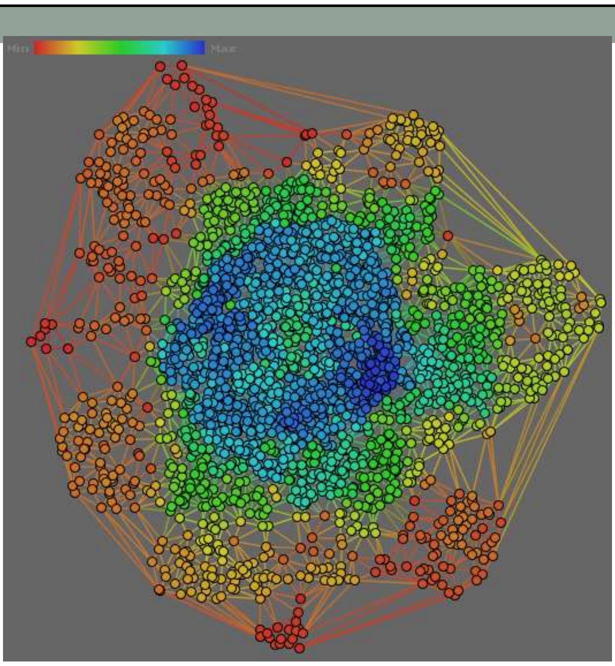
- Building a mesh



99

- Coloring by degree of proximity

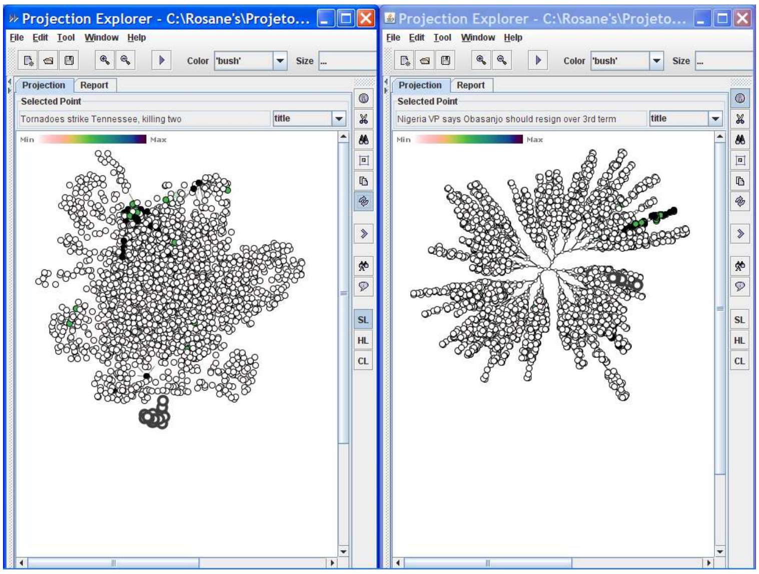
10
0



101

- Coordinating

102



Projection Explorer - C:\Rosane's\Projeto...

Color 'bush'

Selected Point: Tornadoes strike Tennessee, killing two

title

Projection Explorer - C:\Rosane's\Projeto...

Color 'bush'

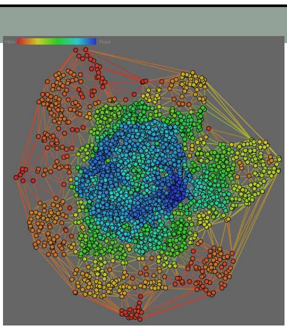
Selected Point: Nigeria VP says Obasanjo should resign over 3rd term

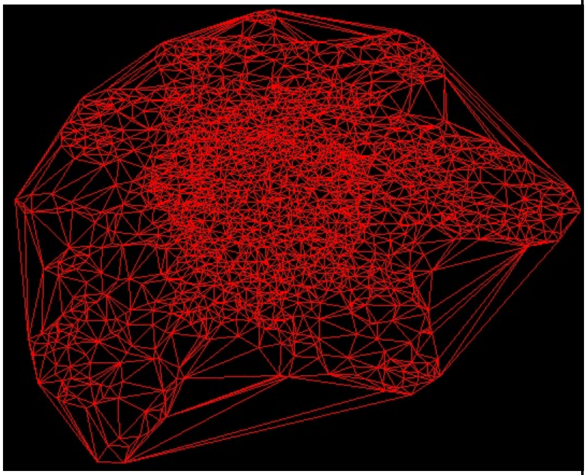
title

10
3

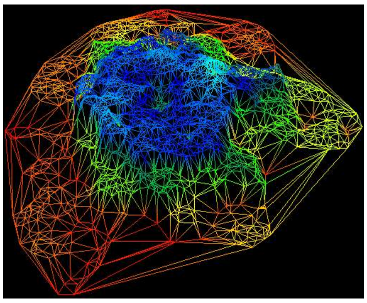
- Building a Surface

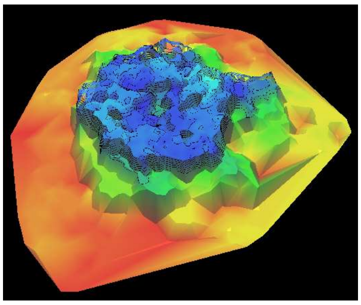
10
4

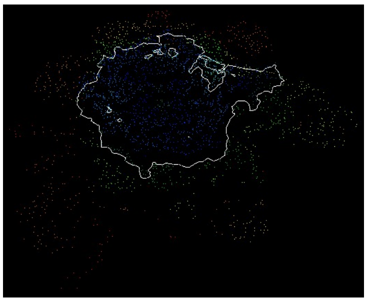




10
5

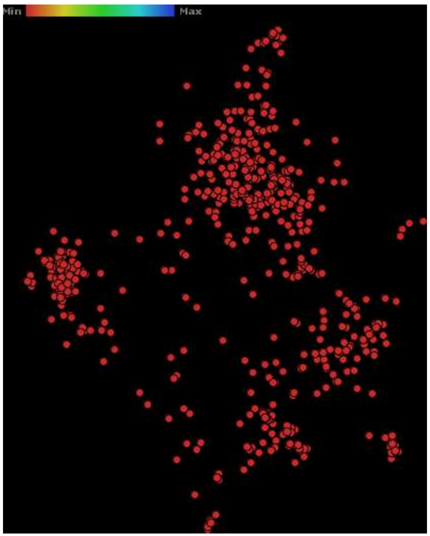






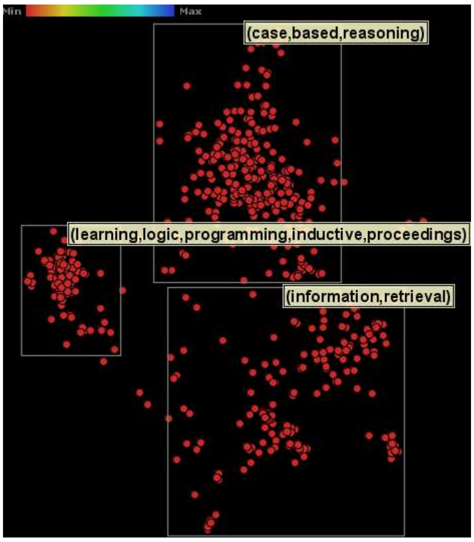
10
6

Explorando

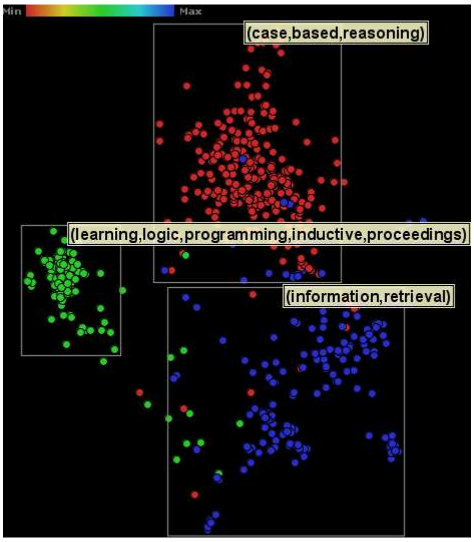


- Case-Base Reasoning
- Information Retrieval
- Inductive Logic Programming

Exemplos de Mapas

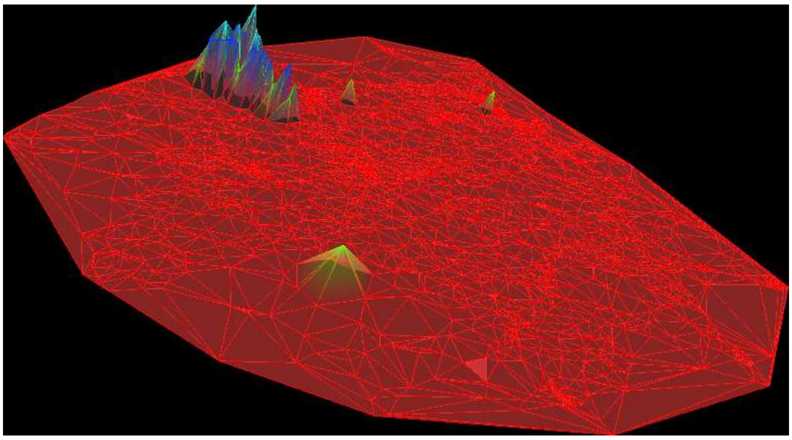


Exemplos de Mapas



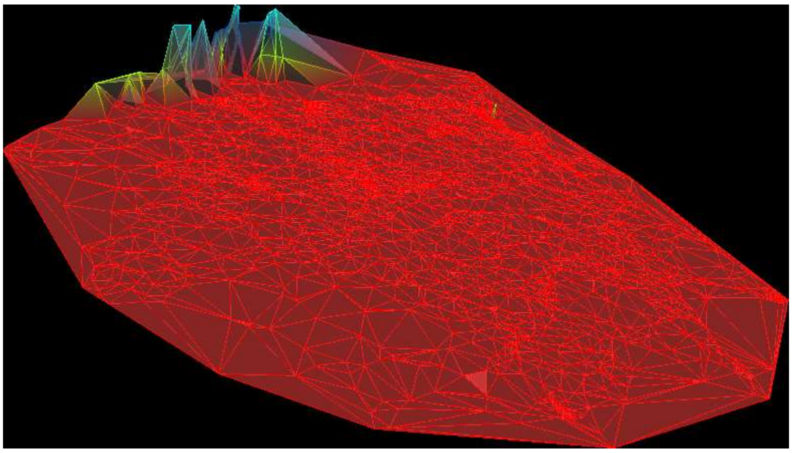
109

RSS News Flash

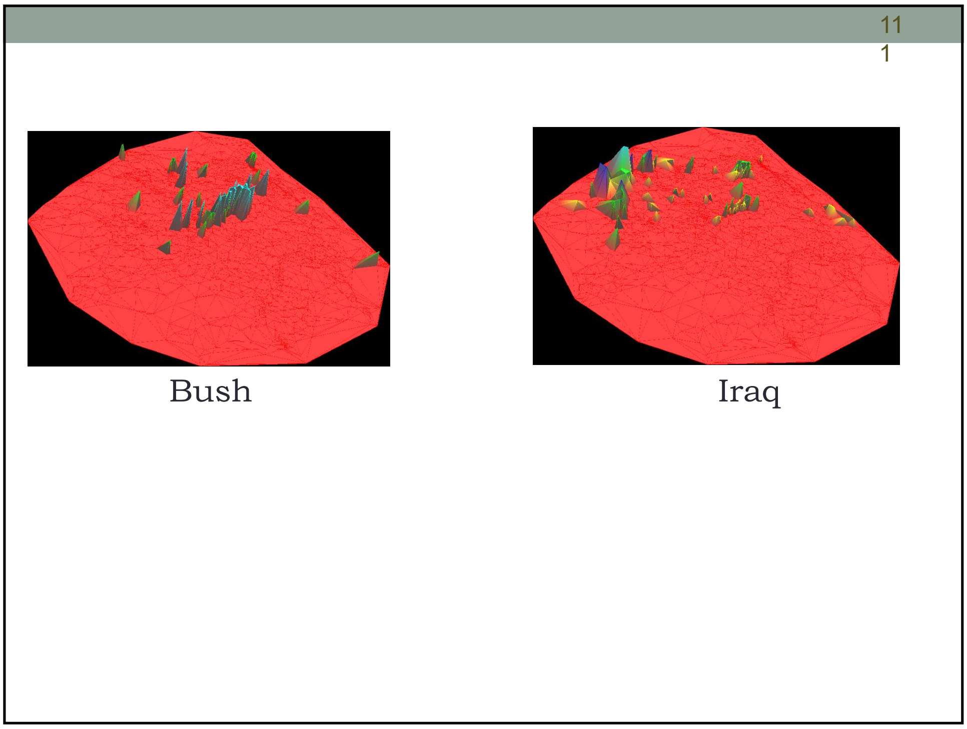


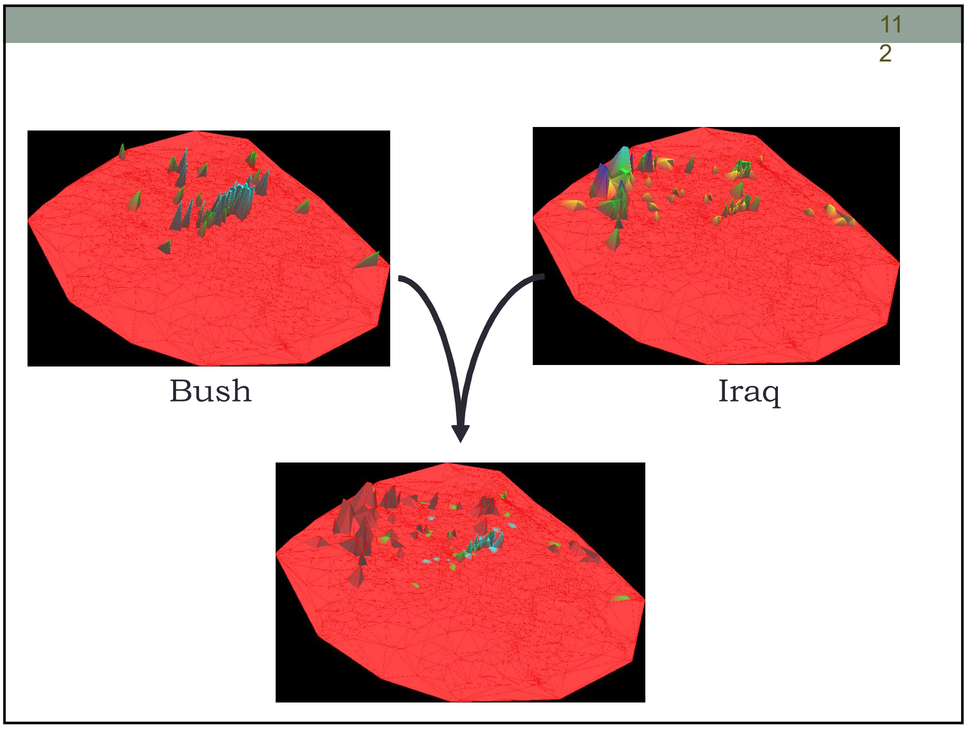
Bird and Flu

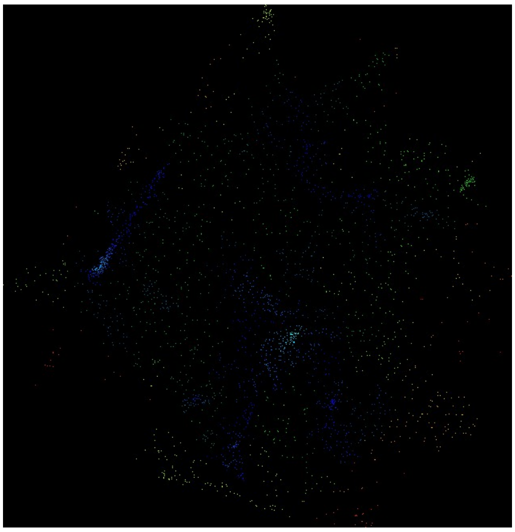
110



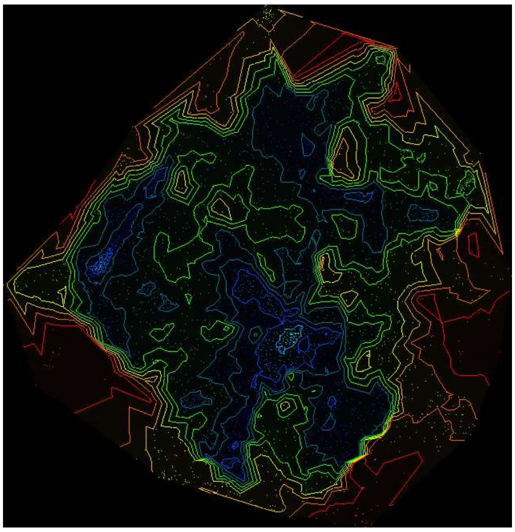
Palestinian


111
Bush
Iraq


112
Bush
Iraq



Curvas de Nível



Time Series - Flow in Hydro. P. P.

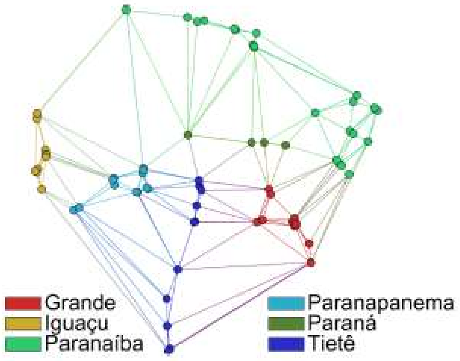
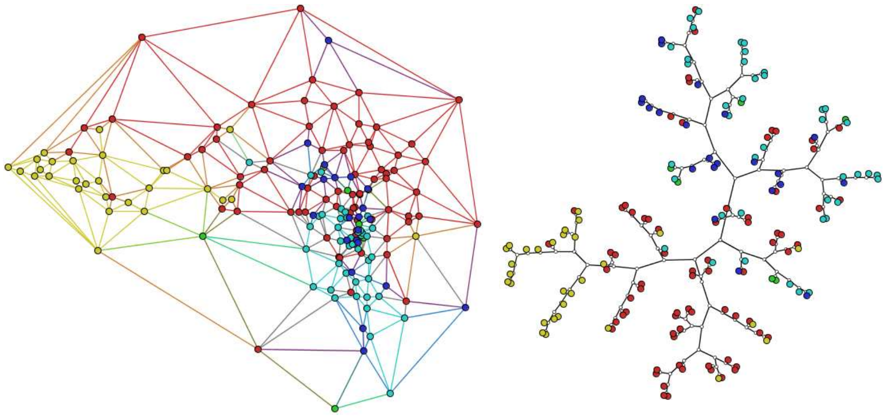


Figure 2. Power plants of the basin Paraná

Further Example - patents

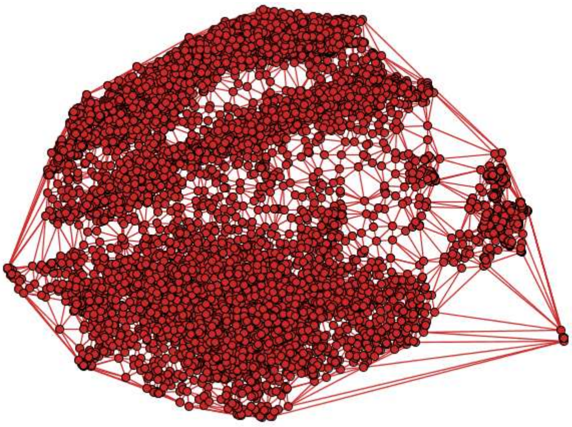


Further Example

- Cattle performance data
  - Translated to text from categorical information, e.g.,
    - Ranges of weight to words such as:

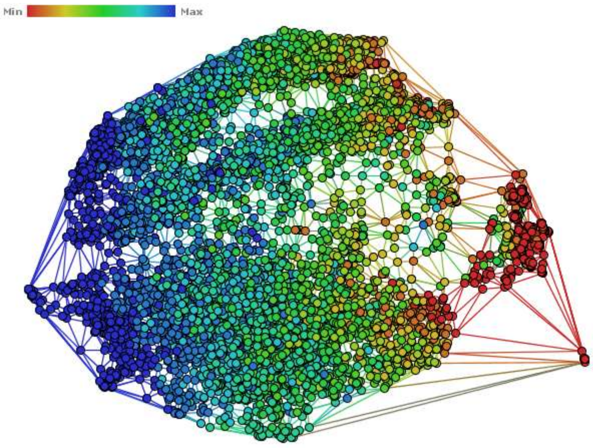
      {weight_below_fifty_percent; weight_between_fifty_seventy_five; etc..}
  - 9135 individuals

Cattle performance data



11
9

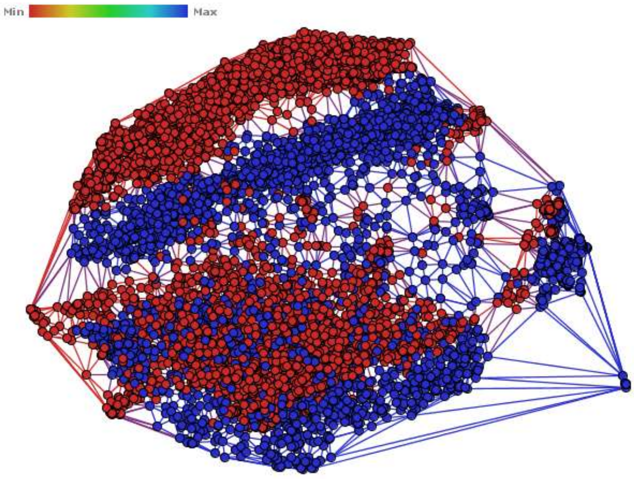
Cattle performance data



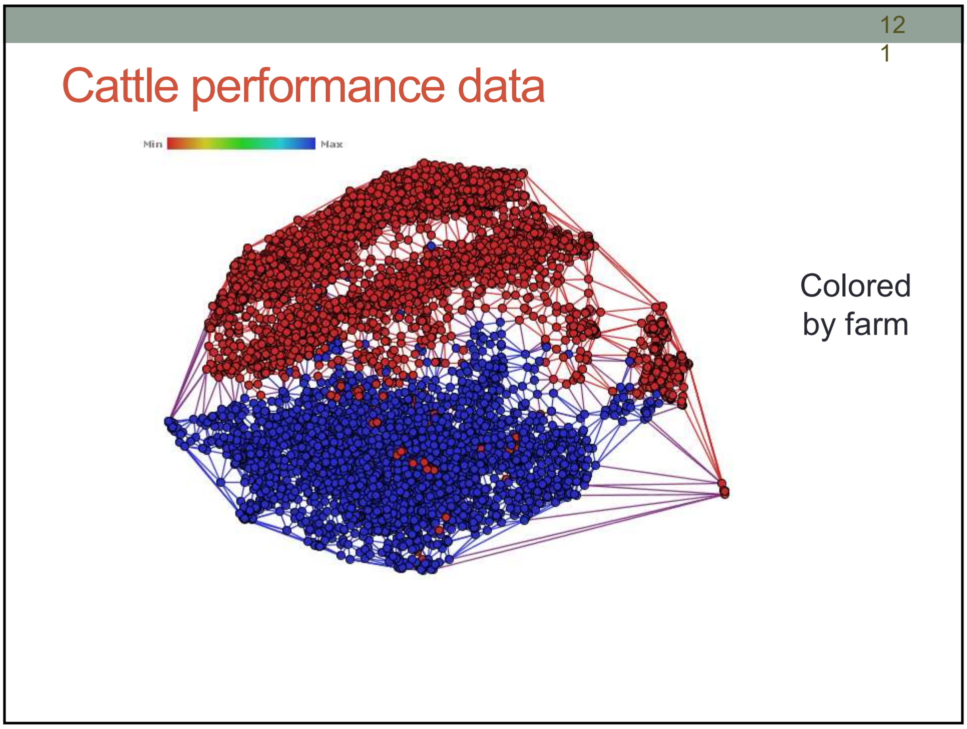
Colored
by word
‘top’

12
0

Cattle performance data



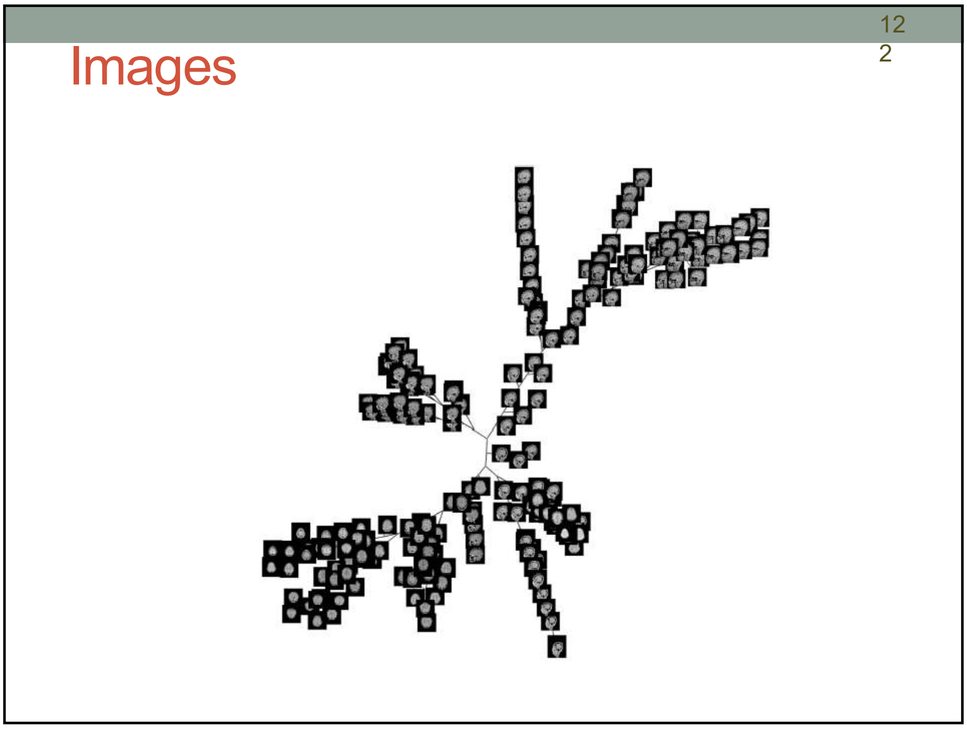
Colored
by
female



121

Cattle performance data

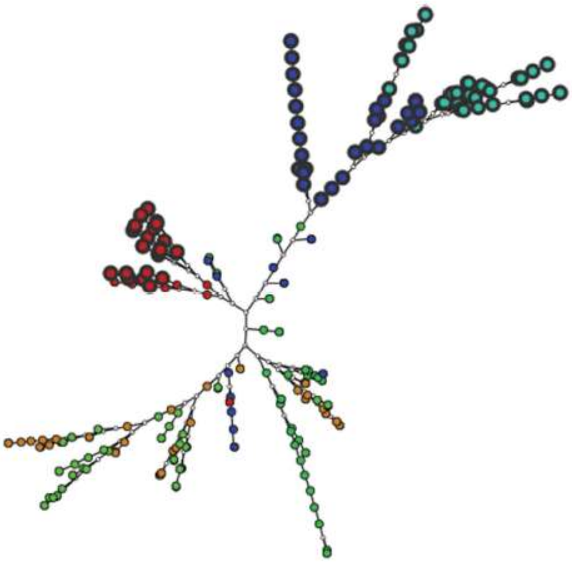
Colored
by farm



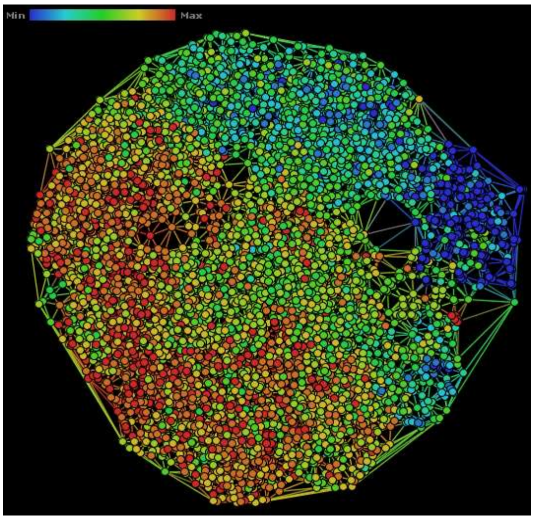
122

Images

Images

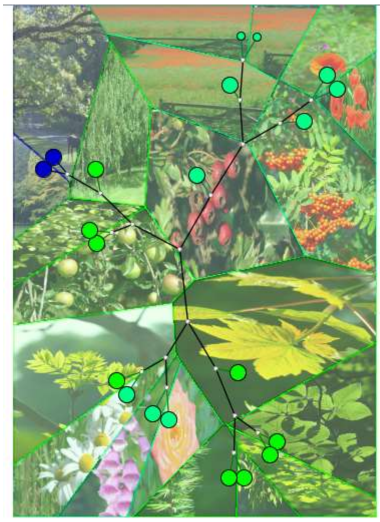


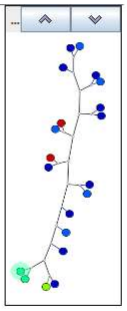
Cattle performance data



Colored by word 'top'

Scalability
The Visual Super Tree



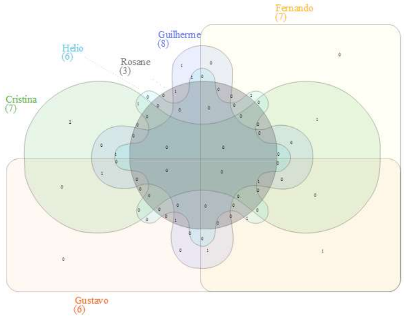


Cooperation:

Guilherme Pimentel Telles
IC/UNICAMP

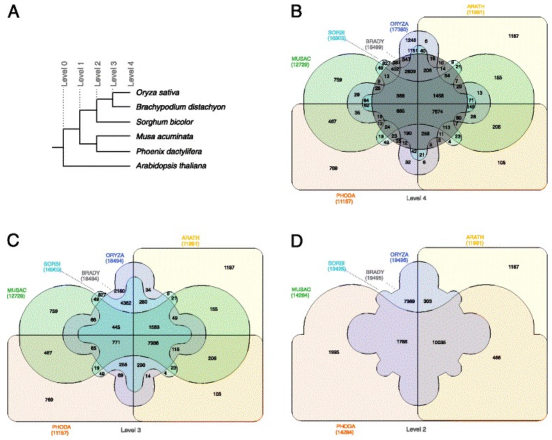
Application: Gene Expression and Systems biology data

- Cooperation ICMC/UNICAMP/LNBio (Campinas)/ Embrapa (Campinas)
- Complex Networks for Biological Data.
- Venn diagrams for Biological Data:
    - interactivenn.net, interactivenn.org



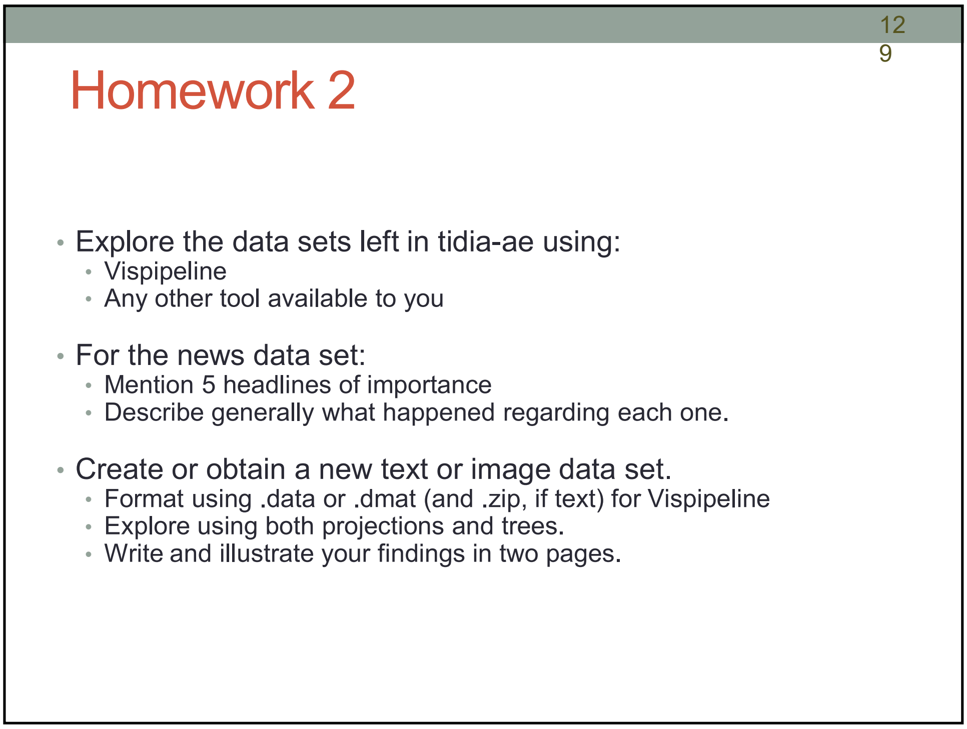
Application: comparison of sets

- Cooperation ICMC/UNICAMP/LNBio (Campinas)/ Embrapa (Campinas)
- Fig.: Comparison of gene lists from different species
- www.interactiVenn.net



Context: Visual Data Mining

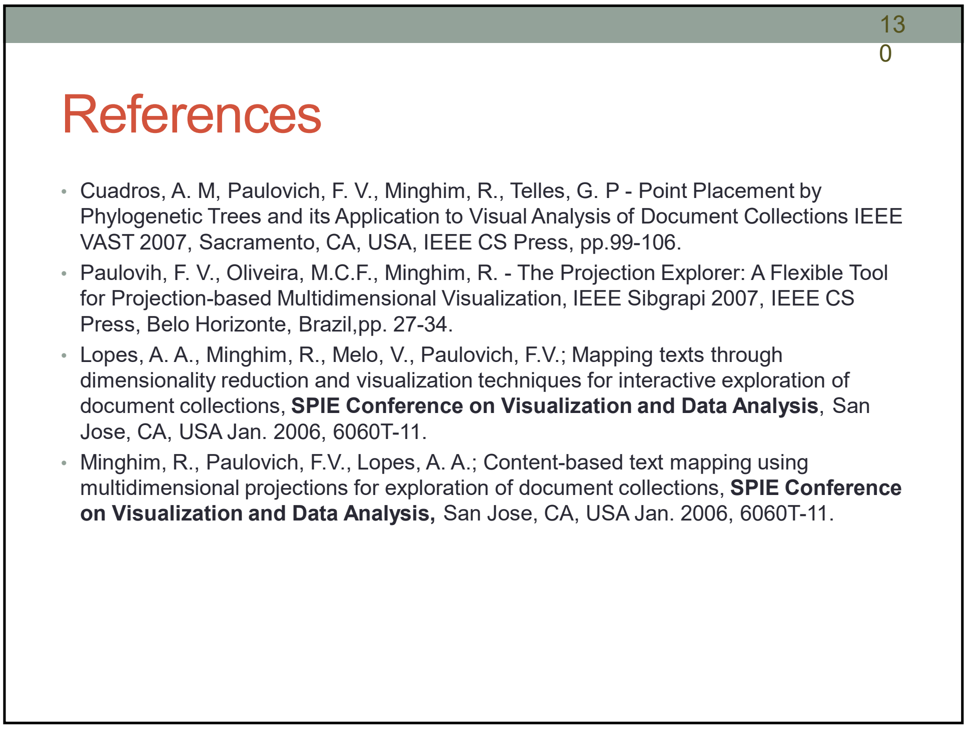
- Definition [Ankerst 2000]
  - step in process of knowledge discovery / extraction (KDD)
  - utilizes visualization as communication channel between computer and user
  - to support identification of new and interpretable patterns



129

Homework 2

- Explore the data sets left in tidia-ae using:
  - Vispipeline
  - Any other tool available to you
- For the news data set:
  - Mention 5 headlines of importance
  - Describe generally what happened regarding each one.
- Create or obtain a new text or image data set.
  - Format using .data or .dmat (and .zip, if text) for Vispipeline
  - Explore using both projections and trees.
  - Write and illustrate your findings in two pages.



130

References

- Cuadros, A. M, Paulovich, F. V., Minghim, R., Telles, G. P - Point Placement by Phylogenetic Trees and its Application to Visual Analysis of Document Collections IEEE VAST 2007, Sacramento, CA, USA, IEEE CS Press, pp.99-106.
- Paulovih, F. V., Oliveira, M.C.F., Minghim, R. - The Projection Explorer: A Flexible Tool for Projection-based Multidimensional Visualization, IEEE Sibgrapi 2007, IEEE CS Press, Belo Horizonte, Brazil,pp. 27-34.
- Lopes, A. A., Minghim, R., Melo, V., Paulovich, F.V.; Mapping texts through dimensionality reduction and visualization techniques for interactive exploration of document collections, **SPIE Conference on Visualization and Data Analysis**, San Jose, CA, USA Jan. 2006, 6060T-11.
- Minghim, R., Paulovich, F.V., Lopes, A. A.; Content-based text mapping using multidimensional projections for exploration of document collections, **SPIE Conference on Visualization and Data Analysis,** San Jose, CA, USA Jan. 2006, 6060T-11.

References

- Pinho, R. D. ; Oliveira, M. C. F. ; Minghim, R. ; Andrade, M. G. . Voromap: A Voronoi-based Tool for Visual Exploration of Multidimensional Data. In: **10th International Conference on Information Visualization**, 2006, Londres. Proceedings of Information Visualisation 2006, 2006. v. 1. p. 39-44
- Paulovich, F. V. ; Minghim, R. . Text Map Explorer: a Tool to Create and Explore Document Maps. In: Information Visualisation 2006 (IV06) **10th International Conference on Information Visualisation**, 2006, Londres. Proceedings of Information Visualisation 2006, 2006. v. 1. p. 245-251.
- Paulovich, F. V. ; Nonato, L. G. ; MINGHIM, R. ; Levkowitz, H. . Least Square Projection: a fast high precision multidimensional projection technique and its application to document mapping. IEEE Transactions on Visualization and Computer Graphics, 2008.
- Minghim, R. ; Levkowitz, H. ; Nonato, L. G. ; Watanabe, L. S. ; Salvador, V. C. L. ; Lopes, H. ; Pesco, S. ; Tavares, G. . Spider Cursor: A simple versatile interaction tool for data visualization and exploration. In: **ACM GRAPHITE'05** - 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, 2005, Dunedin. Proceedings of Graphite 2005, 2005. p. 307-314.
- Heberle, H.; Meirelles, G. V.; da Silva, F. R.; Telles, G. P.; Minghim, R. ***InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams***. BMC Bioinformatics 16:169 (2015).

5.3 Topic Extraction and Visualization

- Topic Definition by Covariance
- Topic Extraction by Seeded Generation of Association Rules (pruning by relevant terms)
- Labeling and Viewing

Topic Extraction and Visualization

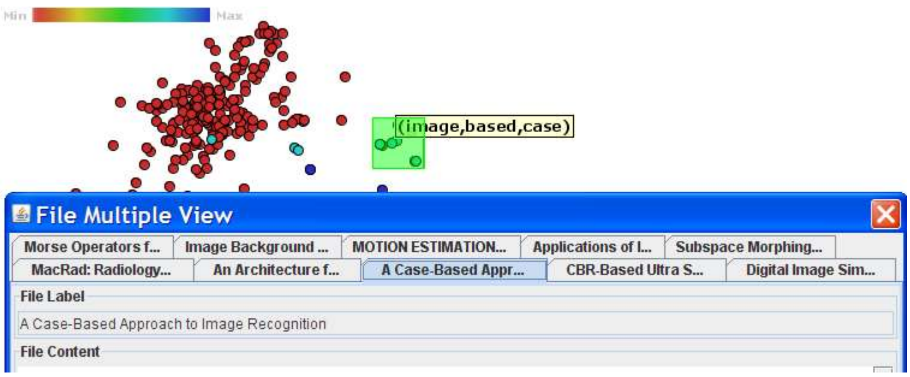
Topic Definition by Covariance

- Pair of words with highest covariance

$$cov(t_i,t_j) = \frac{1}{n-1}\sum_{k=1}^{n}(t_{ki}-\bar{t}_i)(t_{kj}-\bar{t}_j)$$

- For all the other words, highest mean covariance compared to first two.
- Add to label if above threshold.

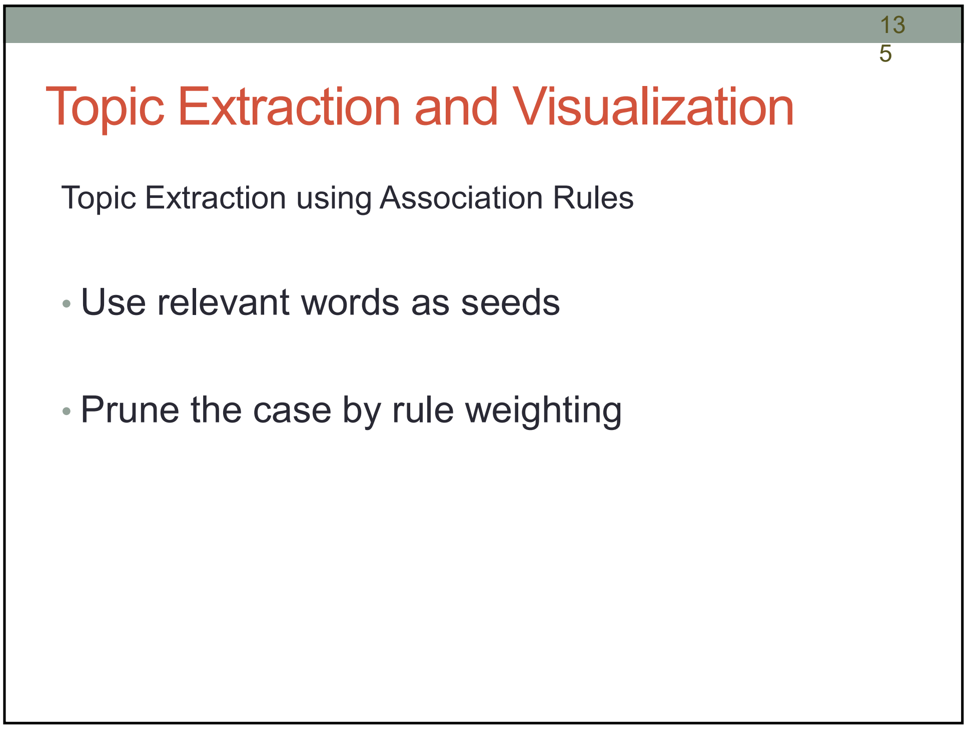
Topic Definition by Covariance



File Label

A Case-Based Approach to Image Recognition

File Content

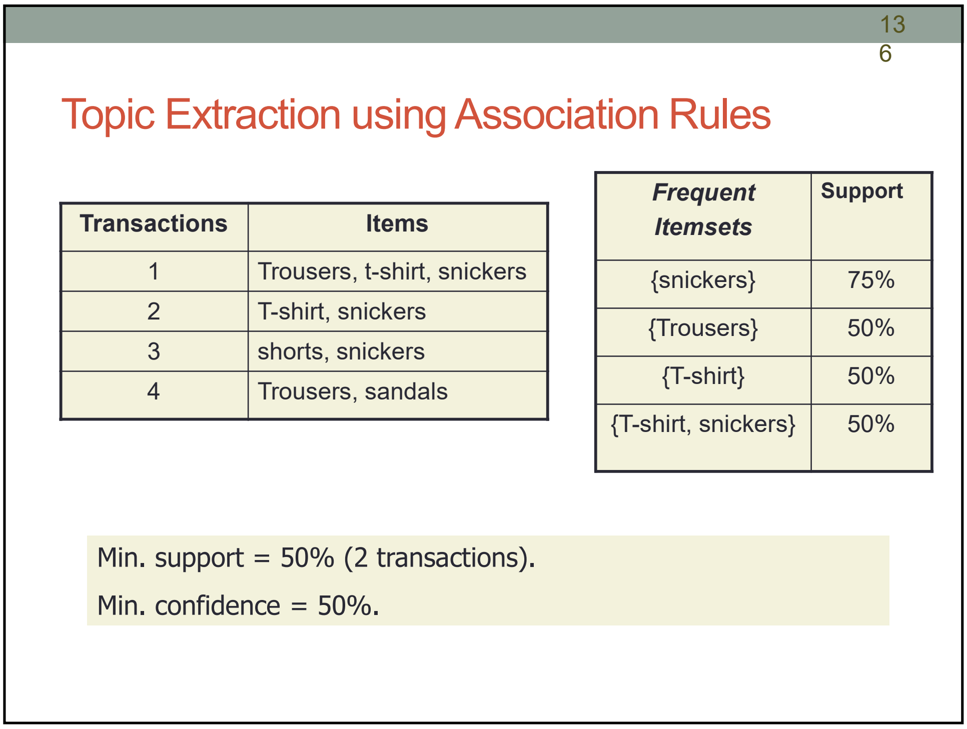


13
5

Topic Extraction and Visualization

Topic Extraction using Association Rules

- Use relevant words as seeds
- Prune the case by rule weighting



13
6

Topic Extraction using Association Rules

Transactions	Items
1	Trousers, t-shirt, snickers
2	T-shirt, snickers
3	shorts, snickers
4	Trousers, sandals

Frequent Itemsets	**Support**
{snickers}	75%
{Trousers}	50%
{T-shirt}	50%
{T-shirt, snickers}	50%

Min. support = 50% (2 transactions).

Min. confidence = 50%.

137

Topic Extraction using Association Rules

tenis → t-shirt

$$\text{support} = \text{support}(\{\text{snickers}, \text{t - shirt}\}) = 50\%$$

$$confidence = \frac{\text{support}(\{\text{snickers}, t - shirt\})}{\text{support}(\{t - shirt\})} = \frac{50}{50} = 100\%$$

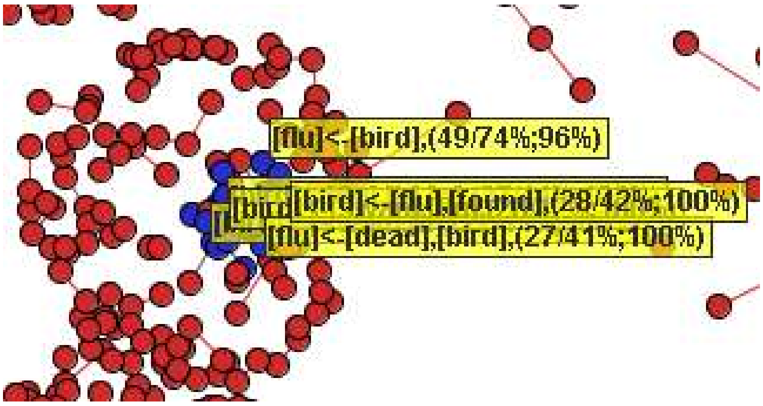
T-shirt → snickers

$$\text{support} = \text{support}(\{\text{snickers}, \text{ t} - \text{shirt}\}) = 50\%$$

$$confidence = \frac{\text{support}(\{\text{snickers}, \text{t} - \text{shirt}\})}{\text{support}(\{snic\,\text{ker}\,s\})} = \frac{50}{75} = 66{,}6\%$$

138

Topic Extraction using Association Rules (example)


[flu]<-[bird],(49/74%;96%)
[bird
[bird]<-[flu],[found],(28/42%;100%)
[flu]<-[dead],[bird],(27/41%;100%)

Topic Extraction using Association Rules (example)

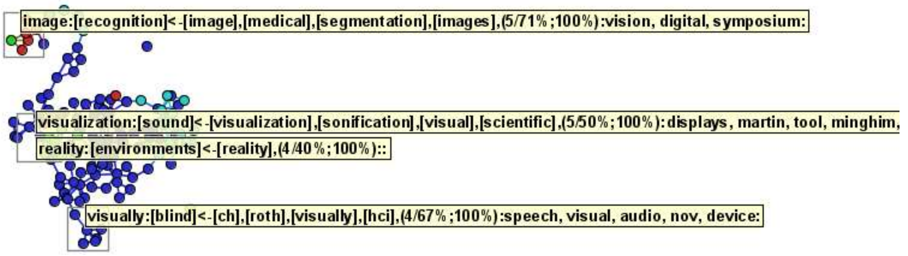


image:[recognition]<-[image],[medical],[segmentation],[images],(5/71%;100%):vision, digital, symposium:
visualization:[sound]<-[visualization],[sonification],[visual],[scientific],(5/50%;100%):displays, martin, tool, minghim,
reality:[environments]<-[reality],(4/40%;100%)::
visually:[blind]<-[ch],[roth],[visually],[hci],(4/67%;100%):speech, visual, audio, nov, device:

Topic Extraction using Association Rules

- Topics using AR
  - Term co-occurance in documents <=> subject
  - Transaction => Document
  - Item => term

14
1

Topic Extraction using Association Rules

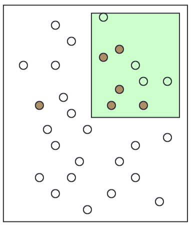
- Issues
  - Discovered rules amount
  - Term relevance (items)
  - Rule relevance measure (filtering)

  - High Sup. & conf. => few interesting rules
  - Low Sup. & conf. => huge amount of rules

14
2

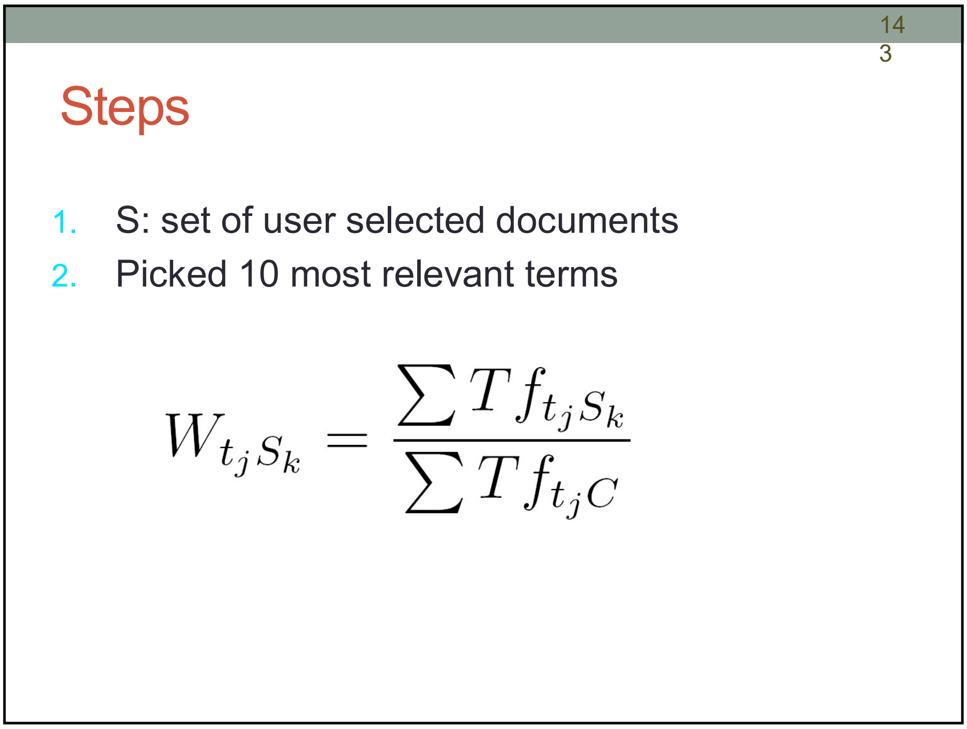
Locally weighted and seeded AR

- Weighting Terms and Rules



$$w_{i,j} = \frac{\sum_{j=1}^{k} Tf_{i,j}}{\sum_{j=1}^{k} Tf_i}$$

$wi,_s = 5/6 = \mathbf{0.83333}$

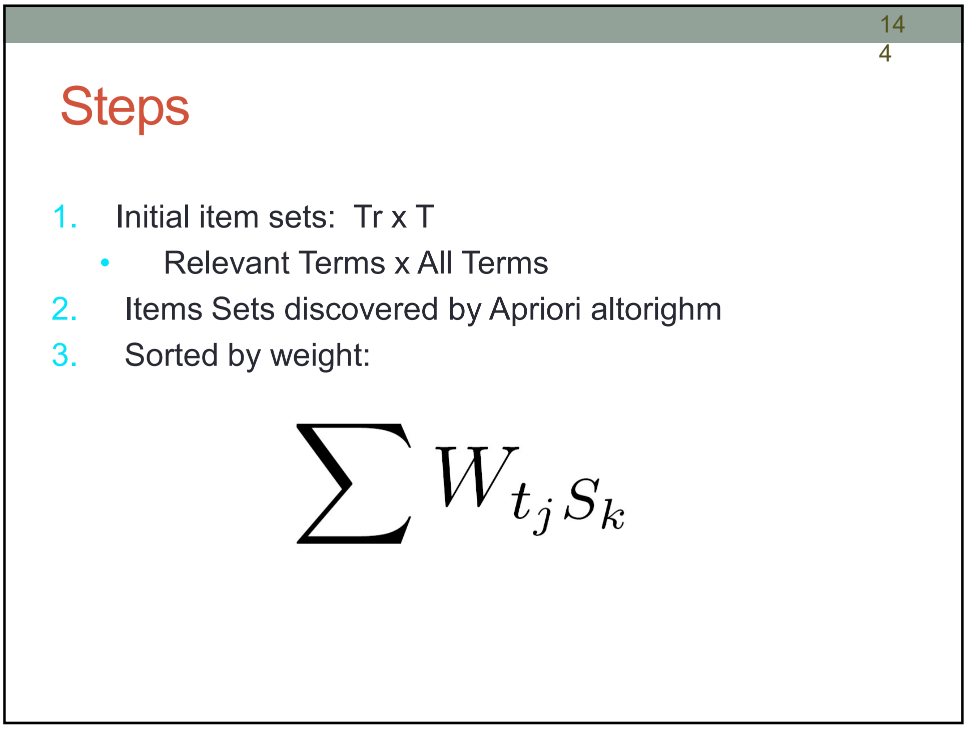


143

Steps

1. S: set of user selected documents
2. Picked 10 most relevant terms

$$W_{t_j S_k} = \frac{\sum Tf_{t_j S_k}}{\sum Tf_{t_j C}}$$

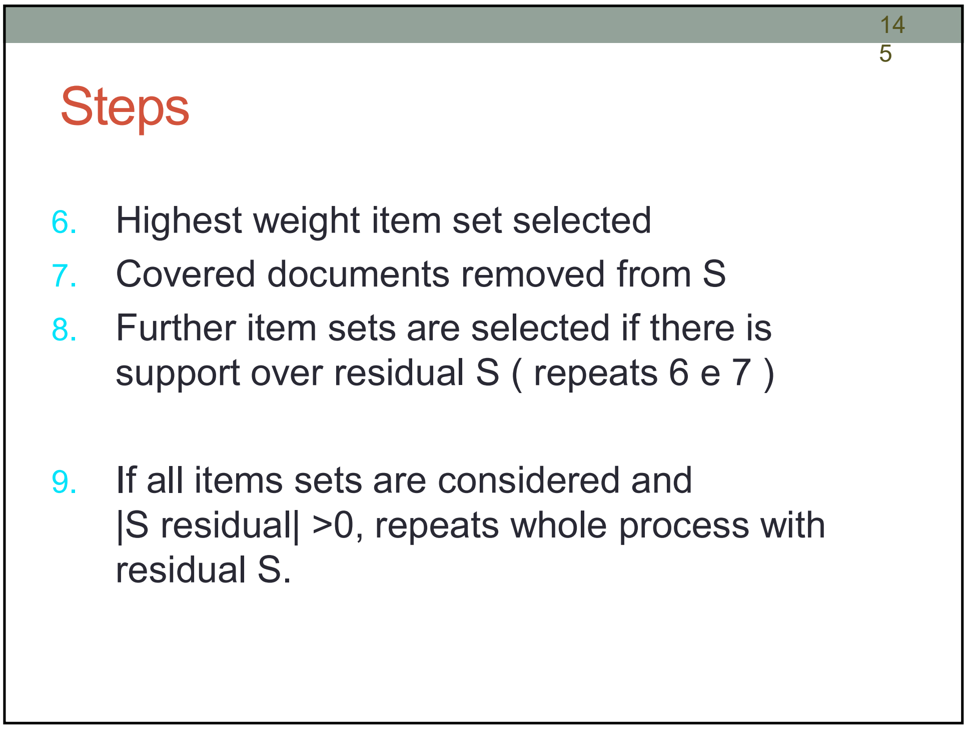


144

Steps

1. Initial item sets: Tr x T
   - Relevant Terms x All Terms
2. Items Sets discovered by Apriori altorighm
3. Sorted by weight:

$$\sum W_{t_j S_k}$$

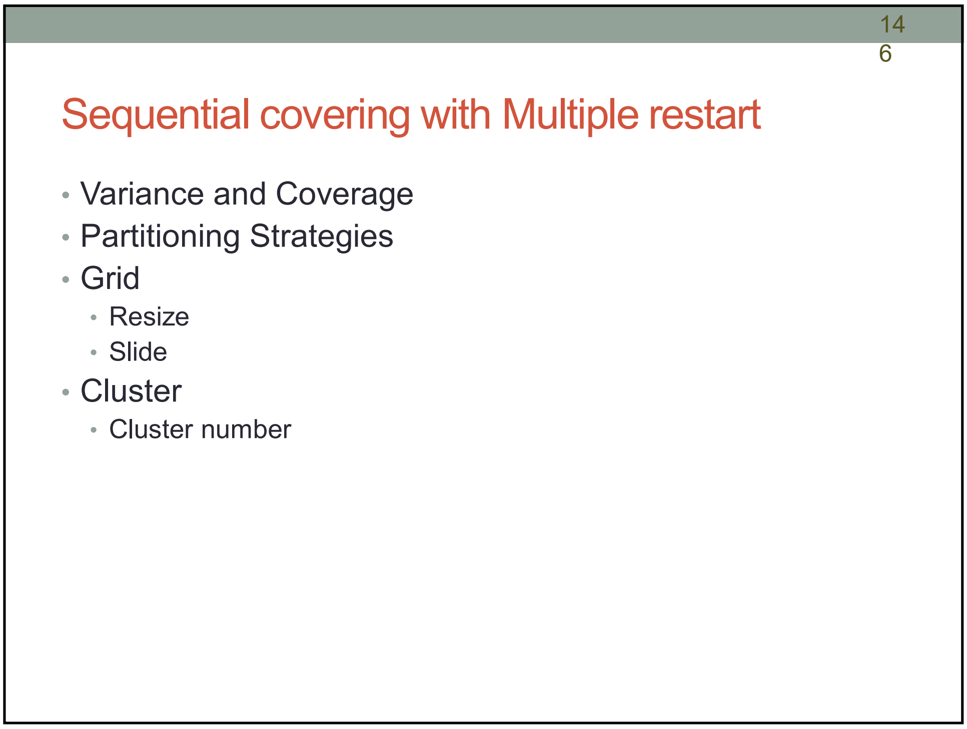


145

Steps

6. Highest weight item set selected
7. Covered documents removed from S
8. Further item sets are selected if there is support over residual S (repeats 6 e 7)

9. If all items sets are considered and |S residual| >0, repeats whole process with residual S.



146

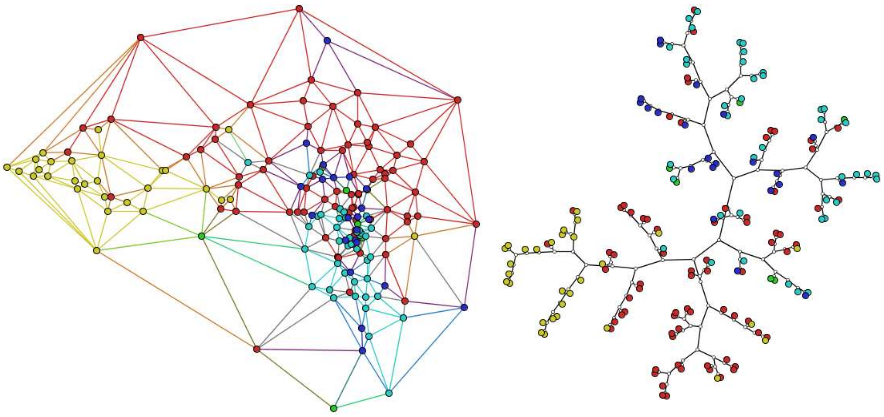
Sequential covering with Multiple restart

- Variance and Coverage
- Partitioning Strategies
- Grid
  - Resize
  - Slide
- Cluster
  - Cluster number

5.4 Further Examples

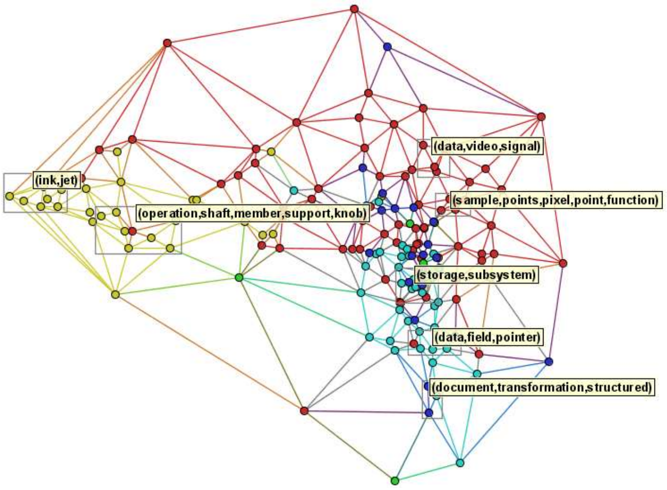
- RSS Patent Data, recovered from the Web
  http://www.freepatentsonline.com/
- Case 1:
  - 170 files
  - Graphics processing, printer, database, document, ai

Further Examples



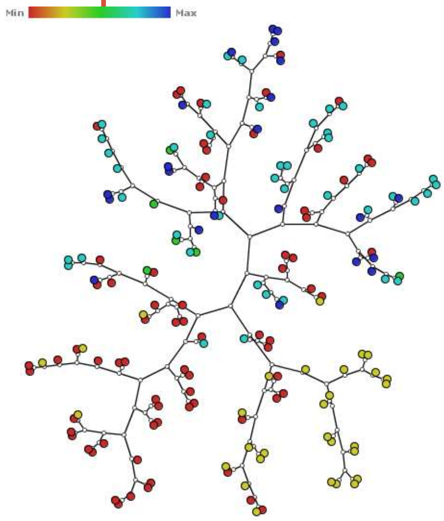
149

Further Examples

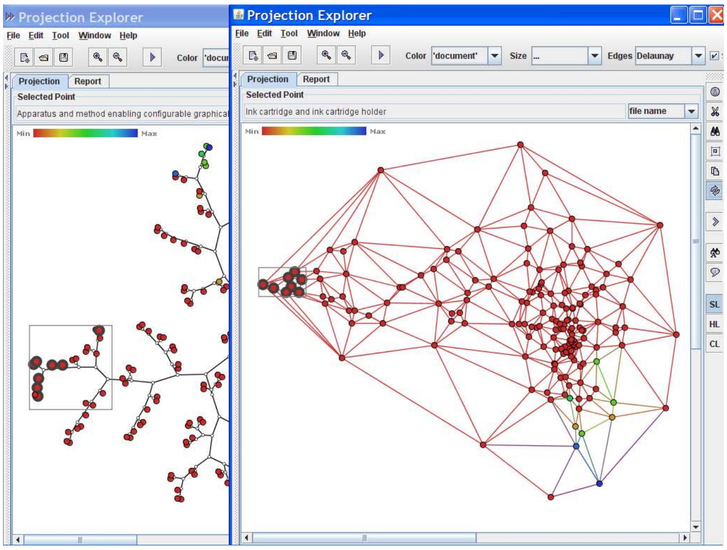

(data,video,signal)
(ink,jet)
(operation,shaft,member,support,knob)
(sample,points,pixel,point,function)
(storage,subsystem)
(data,field,pointer)
(document,transformation,structured)

150

Further Examples



15
1



Selected Point

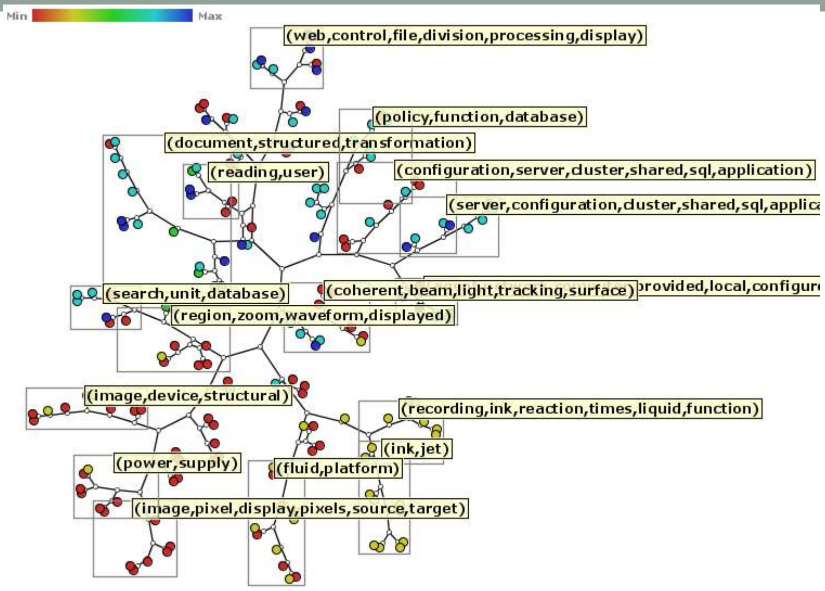
Apparatus and method enabling configurable graphical

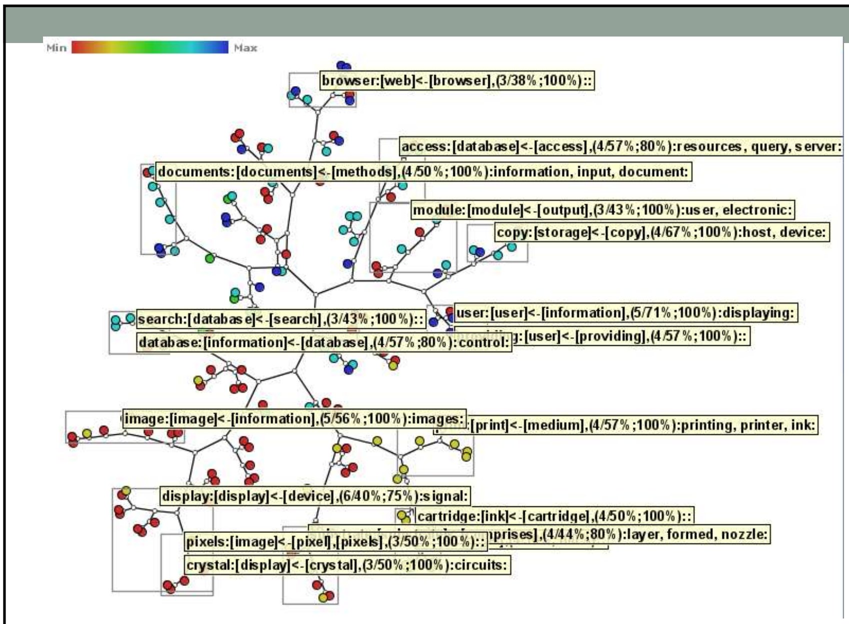
Selected Point

Ink cartridge and ink cartridge holder

(ink jet, document)

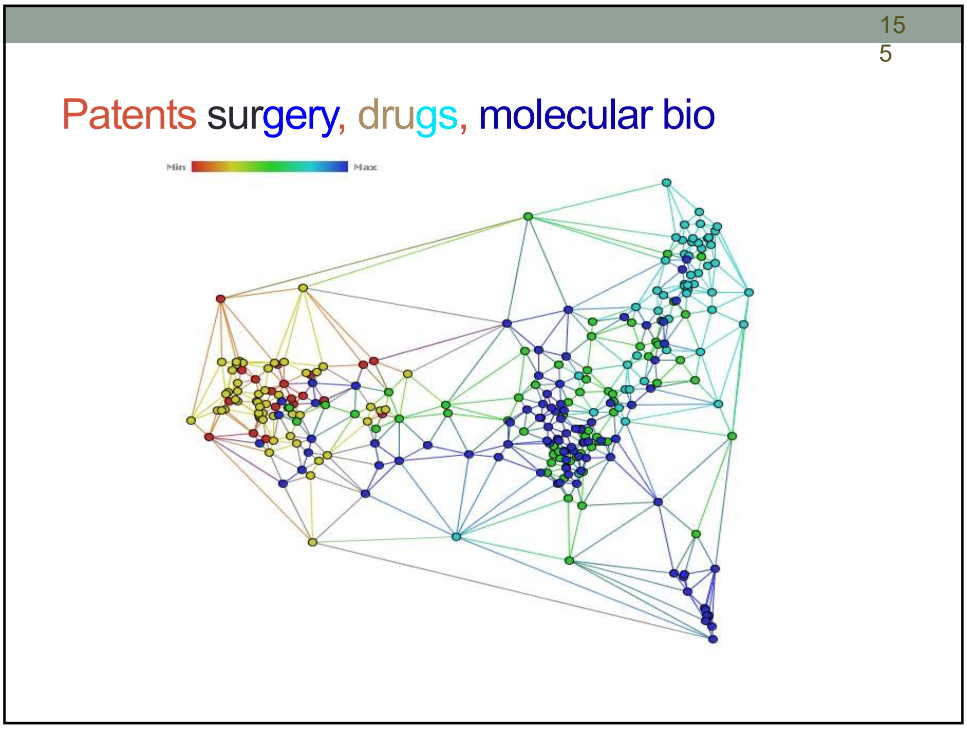
15
2


Min
Max
(web,control,file,division,processing,display)
(policy,function,database)
(document,structured,transformation)
(reading,user)
(configuration,server,cluster,shared,sql,application)
(server,configuration,cluster,shared,sql,applica
(search,unit,database)
(coherent,beam,light,tracking,surface)
provided,local,configur
(region,zoom,waveform,displayed)
(image,device,structural)
(recording,ink,reaction,times,liquid,function)
(ink,jet)
(power,supply)
(fluid,platform)
(image,pixel,display,pixels,source,target)

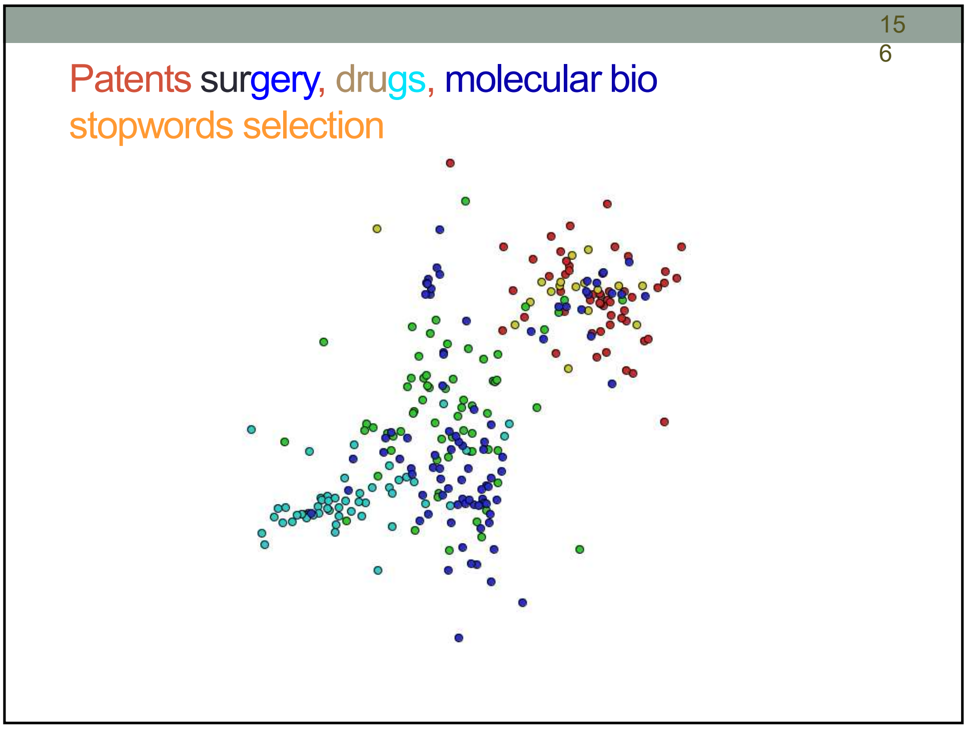

Min
Max
browser:[web]<-[browser],(3/38%;100%)::
access:[database]<-[access],(4/57%;80%):resources, query, server:
documents:[documents]<-[methods],(4/50%;100%):information, input, document:
module:[module]<-[output],(3/43%;100%):user, electronic:
copy:[storage]<-[copy],(4/67%;100%):host, device:
search:[database]<-[search],(3/43%;100%)::
user:[user]<-[information],(5/71%;100%):displaying:
database:[information]<-[database],(4/57%;80%):control:
g:[user]<-[providing],(4/57%;100%)::
image:[image]<-[information],(5/56%;100%):images:
:[print]<-[medium],(4/57%;100%):printing, printer, ink:
display:[display]<-[device],(6/40%;75%):signal:
cartridge:[ink]<-[cartridge],(4/50%;100%)::
prises],(4/44%;80%):layer, formed, nozzle:
pixels:[image]<-[pixel],[pixels],(3/50%;100%)::
crystal:[display]<-[crystal],(3/50%;100%):circuits:

Patents – case 2

- http://www.freepatentsonline.com/
- 172 files
- surgery (2), drugs(2), molecular biology



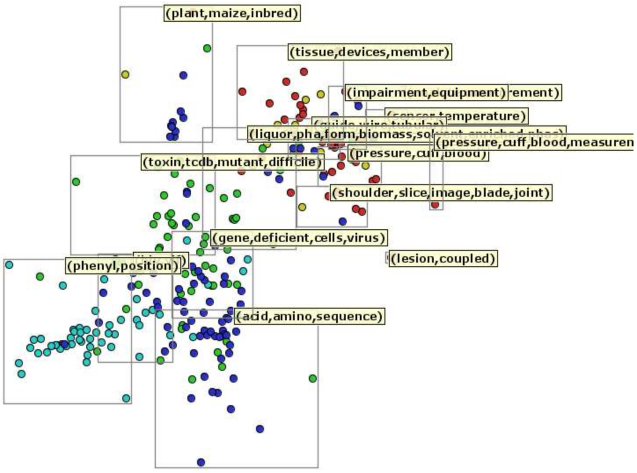
Patents surgery, drugs, molecular bio



Patents surgery, drugs, molecular bio stopwords selection

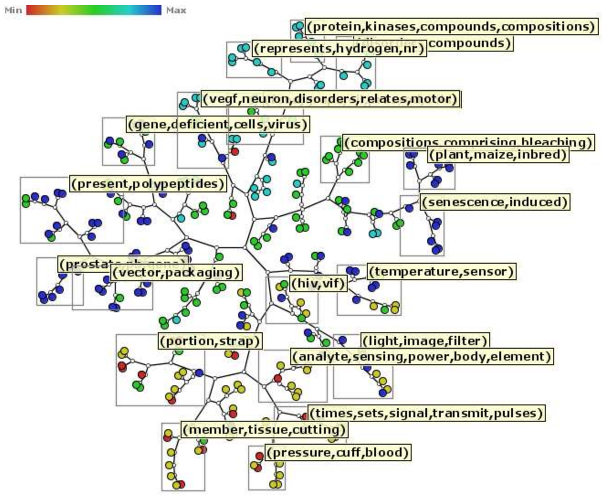
157

Patents surgery, drugs, molecular bio topics

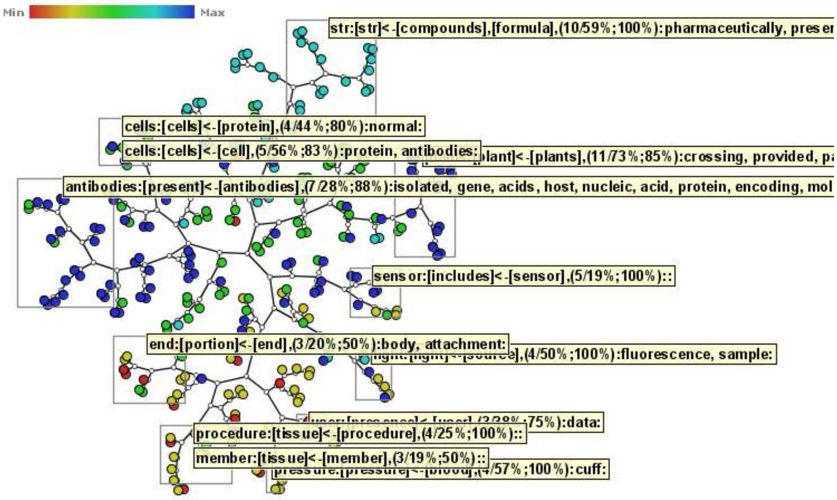


158

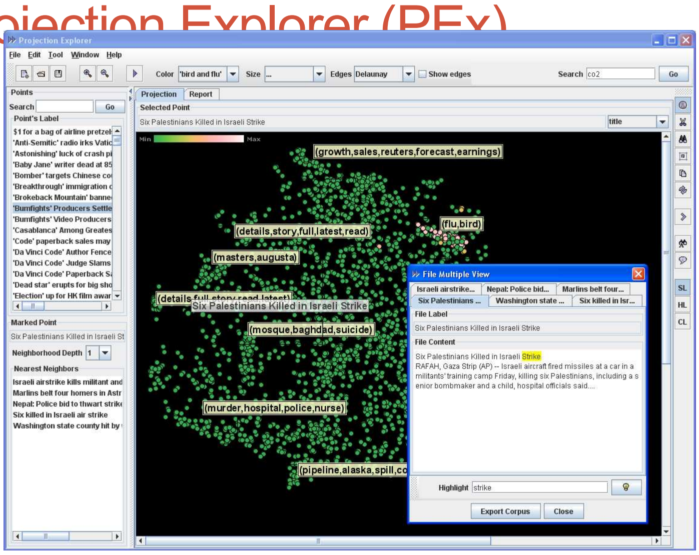
Patents surgery, drugs, molecular bio



Patents surgery, drugs, molecular bio


Min
Max
str:[str]<-[compounds],[formula],(10/59%;100%):pharmaceutically, preser
cells:[cells]<-[protein],(4/44%;80%):normal:
cells:[cells]<-[cell],(5/56%;83%):protein, antibodies:
lant]<-[plants],(11/73%;85%):crossing, provided, pa
antibodies:[present]<-[antibodies],(7/28%;88%):isolated, gene, acids, host, nucleic, acid, protein, encoding, mol
sensor:[includes]<-[sensor],(5/19%;100%)::
end:[portion]<-[end],(3/20%;50%):body, attachment:
],(4/50%;100%):fluorescence, sample:
;75%):data:
procedure:[tissue]<-[procedure],(4/25%;100%)::
member:[tissue]<-[member],(3/19%;50%)::
(4/57%;100%):cuff:

Projection Explorer (PEx)



Color 'bird and flu' Size ... Edges Delaunay □ Show edges Search co2 Go

Points

Search Go

Point's Label

$1 for a bag of airline pretzel
'Anti-Semitic' radio irks Vatic
'Astonishing' luck of crash pi
'Baby Jane' writer dead at 85
'Bomber' targets Chinese co
'Breakthrough' immigration c
'Brokeback Mountain' banne
'Bumfights' Producers Settle
'Bumfights' Video Producers
'Casablanca' Among Greates
'Code' paperback sales may
'Da Vinci Code' Author Fence
'Da Vinci Code' Judge Slams
'Da Vinci Code' Paperback Sa
'Dead star' erupts for big sho
'Election' up for HK film awar

Marked Point

Six Palestinians Killed in Israeli St

Neighborhood Depth 1

Nearest Neighbors

Israeli airstrike kills militant and
Marlins belt four homers in Astr
Nepal: Police bid to thwart strike
Six killed in Israeli air strike
Washington state county hit by

Selected Point

Six Palestinians Killed in Israeli Strike title

File Label

Six Palestinians Killed in Israeli Strike

File Content

Six Palestinians Killed in Israeli Strike
RAFAH, Gaza Strip (AP) -- Israeli aircraft fired missiles at a car in a militants' training camp Friday, killing six Palestinians, including a s enior bombmaker and a child, hospital officials said....

Highlight strike

http://infoserver.lcad.icmc.usp.br/