Visualização Multi-dimensional:

Mineração Visual de Dados multidimensionais e aplicações Parte II – Variações e Aplicações

Rosane Minghim

2008-2012

Basic Concepts

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

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)*)

3

Result is a Vector Model

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

4

Vector Representation – term weighting

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

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

5

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

6

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

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

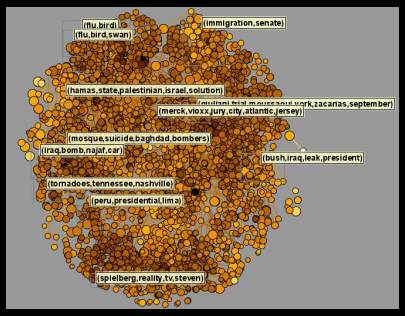
COSINE

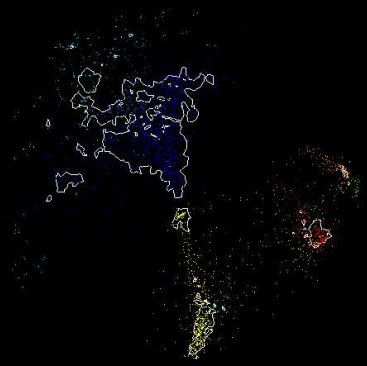
$$sim_{i,j} = \frac{(w_{i,1} \times w_{j,1}) + \ldots + (w_{i,k} \times w_{j,k})}{(w_{i,1}^2 + \ldots + w_{i,k}^2) \times (w_{j,1}^2 + \ldots + w_{j,k}^2)}$$

Alternative to Vector Representation

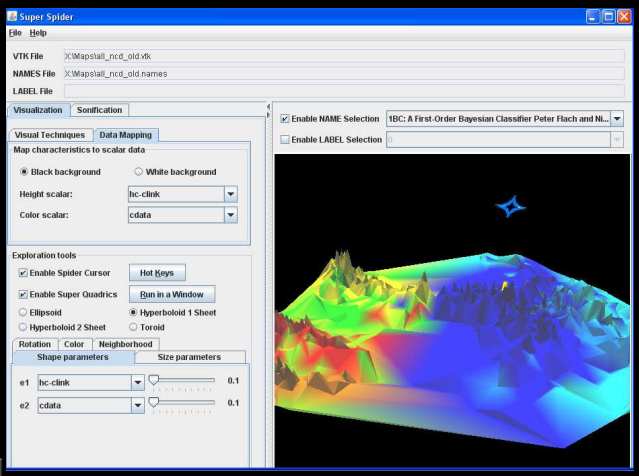
- Similarity Calculation text against text
  - Ex: NCD Normalized Compression Distance
    - Approximation of Kolmogorov Complexity

Visual representations: graphs, surfaces, volumes, triangulations





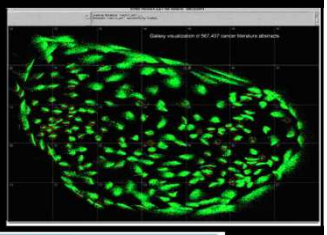
9

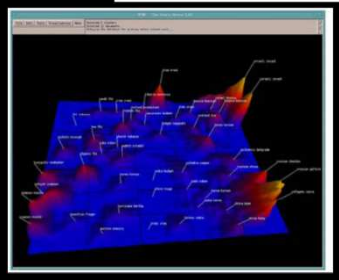


10

IN-SPIRE

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

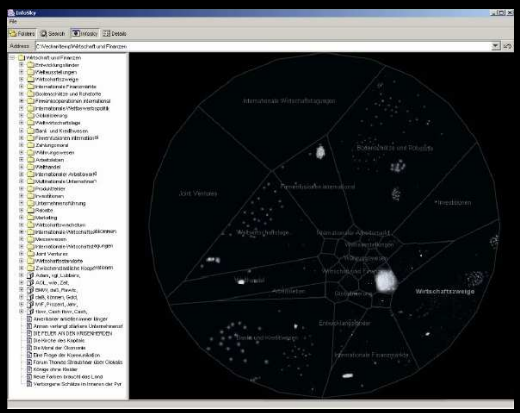




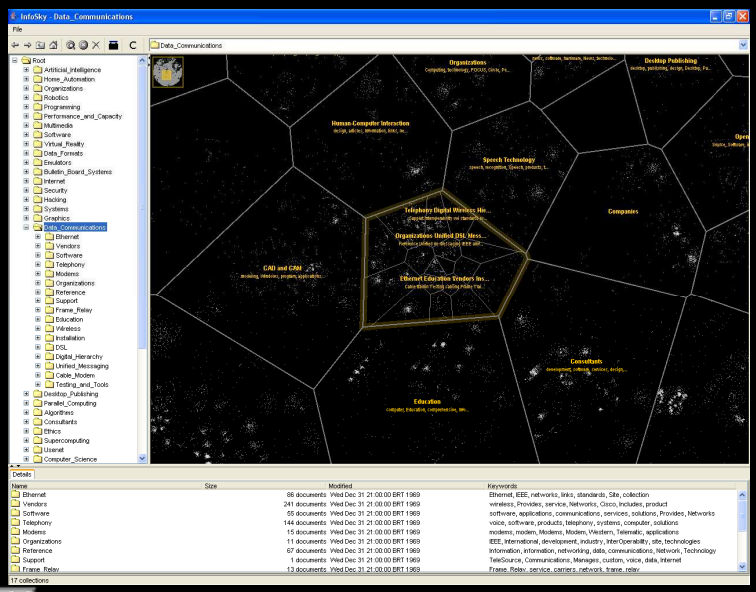
11

InfoSky

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



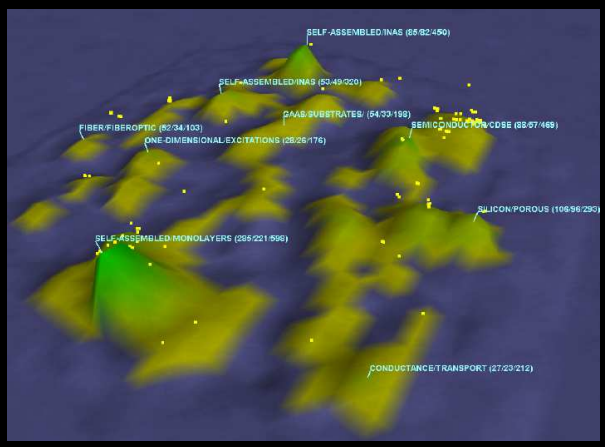
12



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

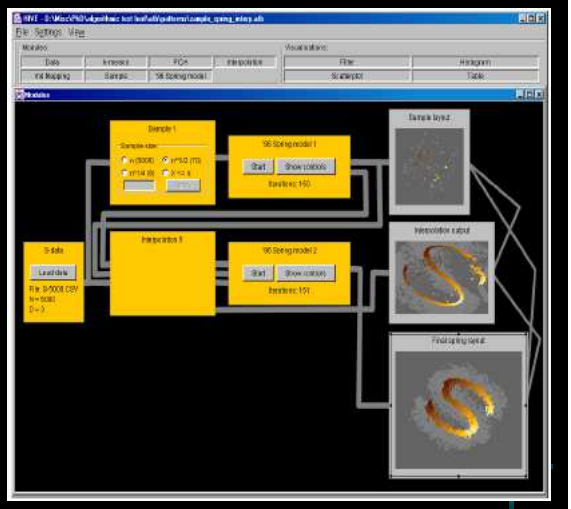
VxInsight

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



HIVE (Ross and Chalmers 2003)

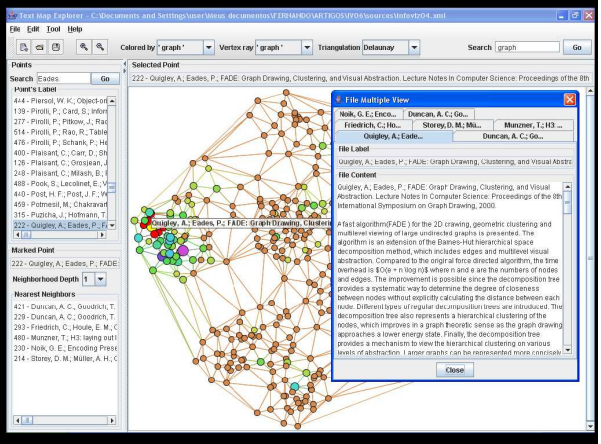
- Interconnected components:
  - Import
  - Transform
  - Render multi-dim data



15

Projection Explorer (PEx)

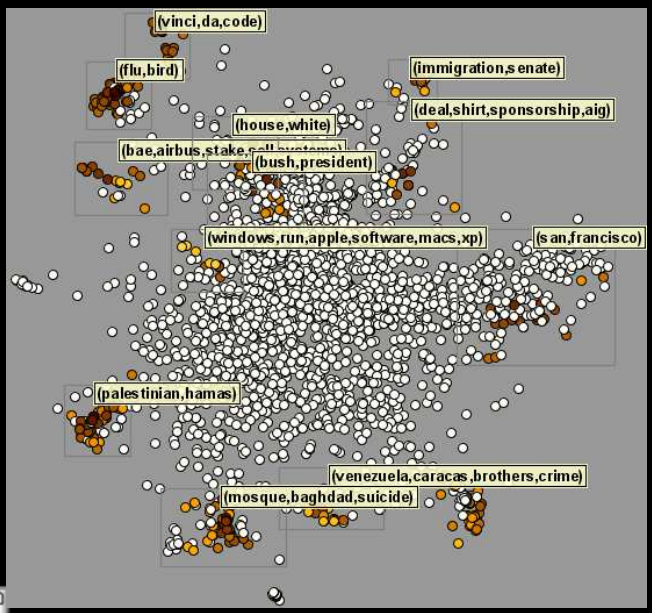
- Projection and Point placement
- Precision
- Graphs and surfaces (Super Spider)



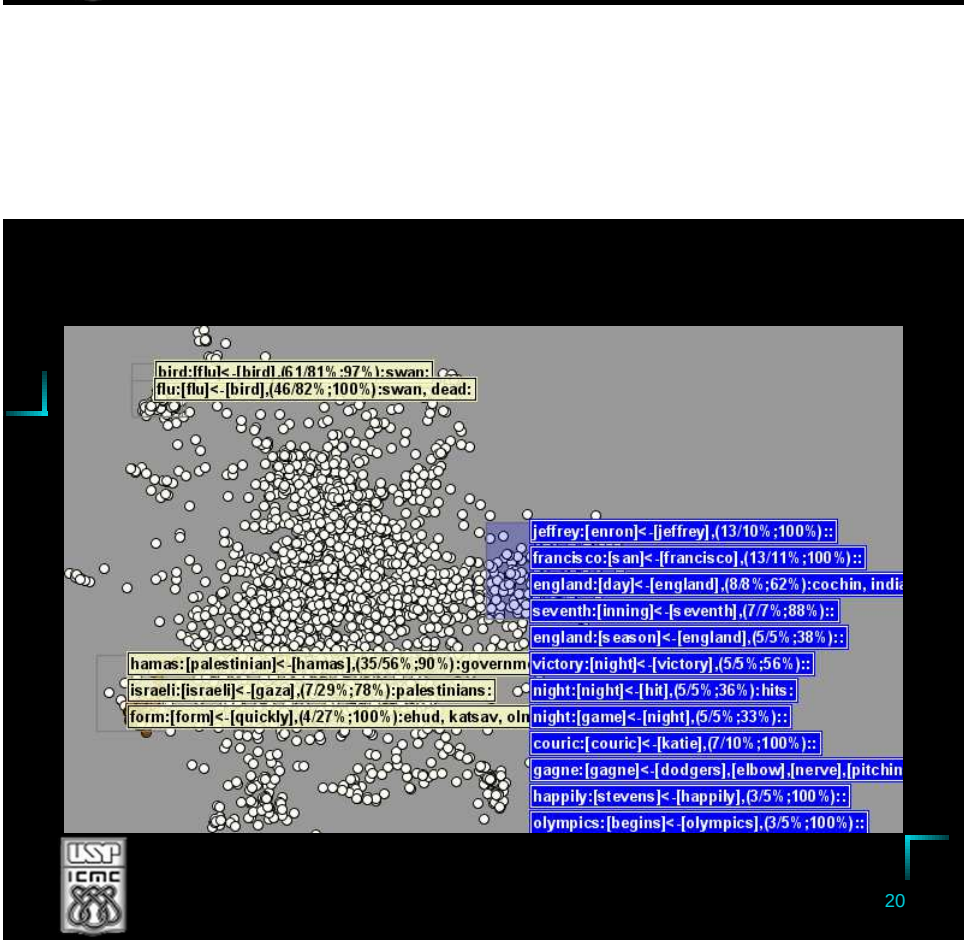
16

Mapping Text Collections via Projections and Point Placement

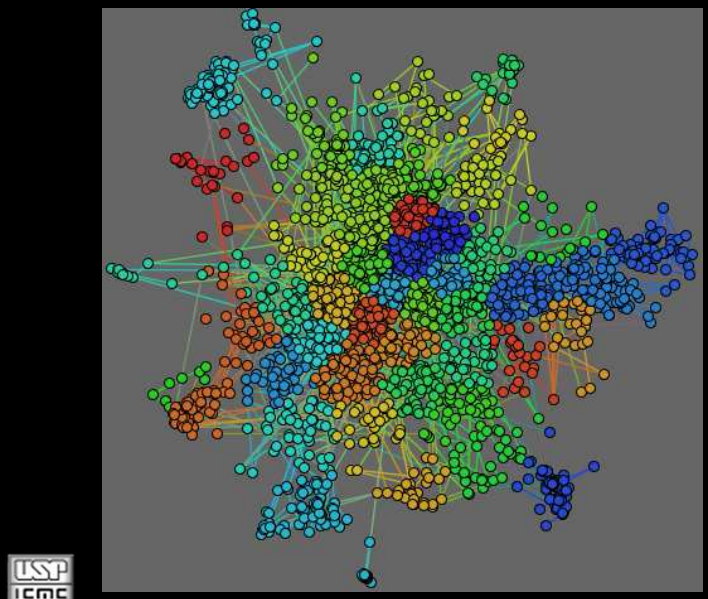
- Positioning and labeling



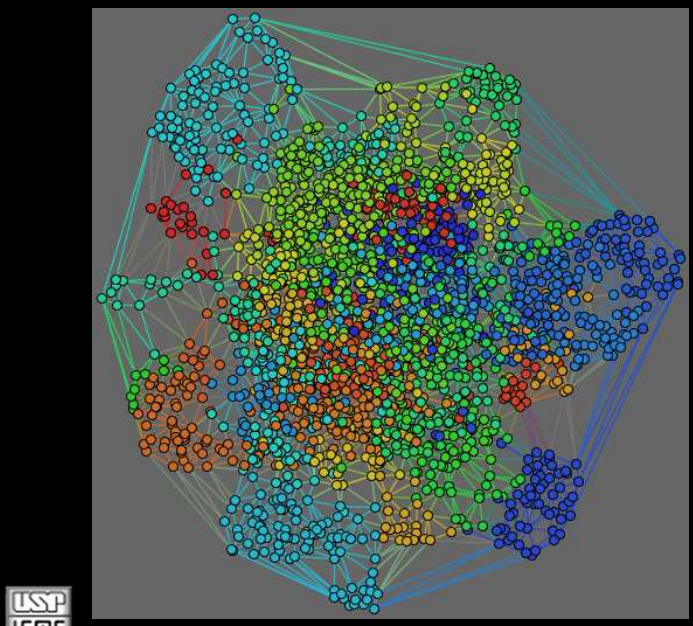
- Detailing topics



- Finding Relationships

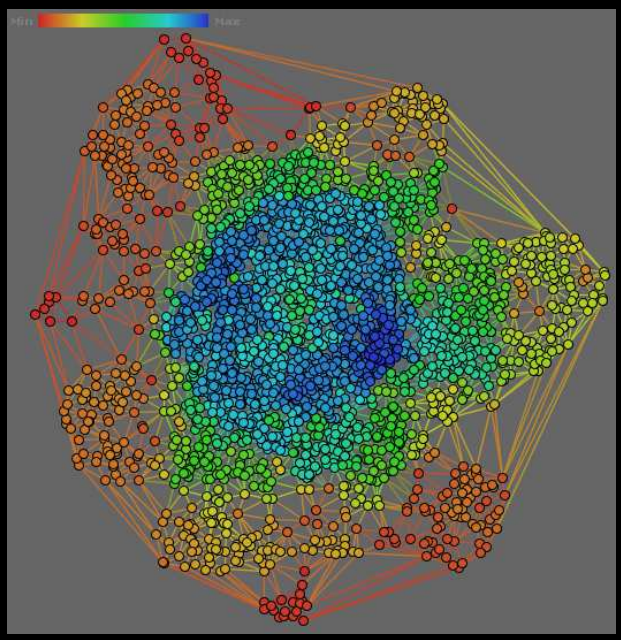


- Building a mesh



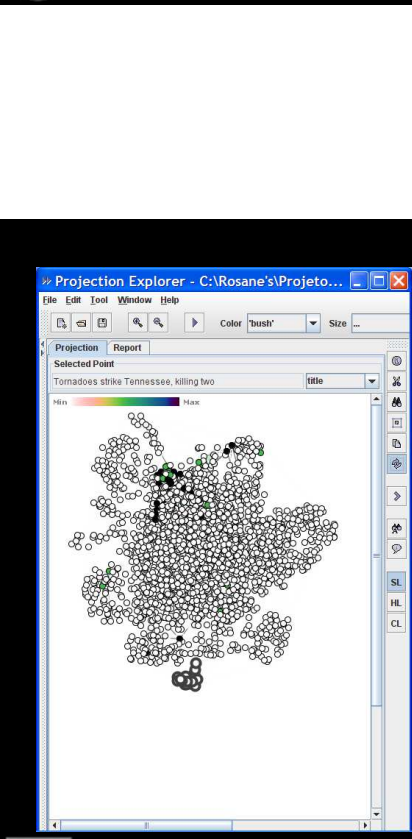
- Coloring by degree of proximity

25



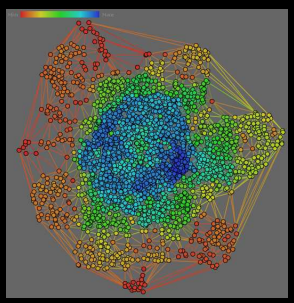
26

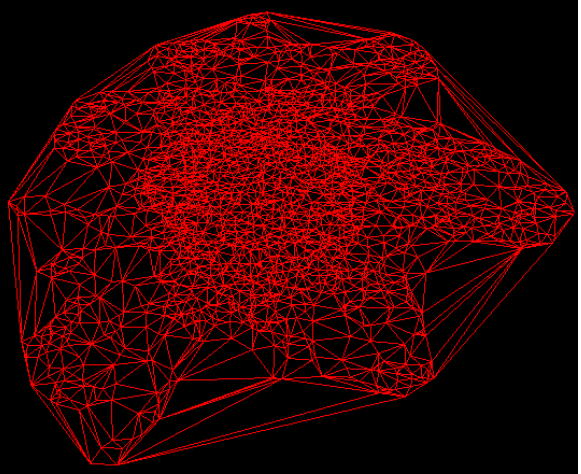
- Coordinating



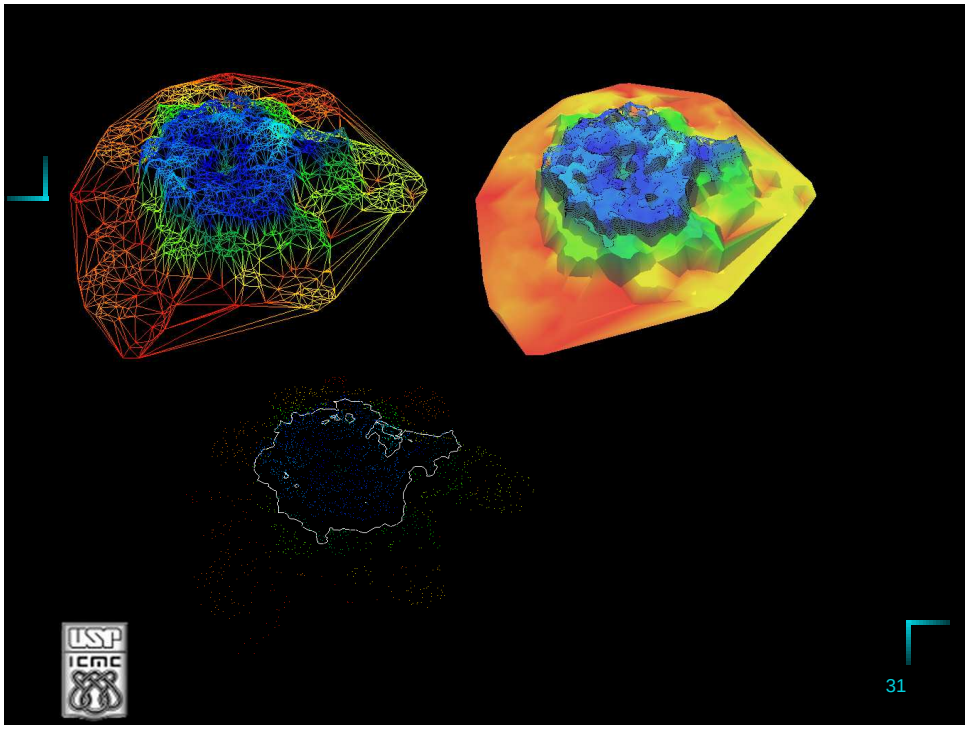
- Building a Surface

29



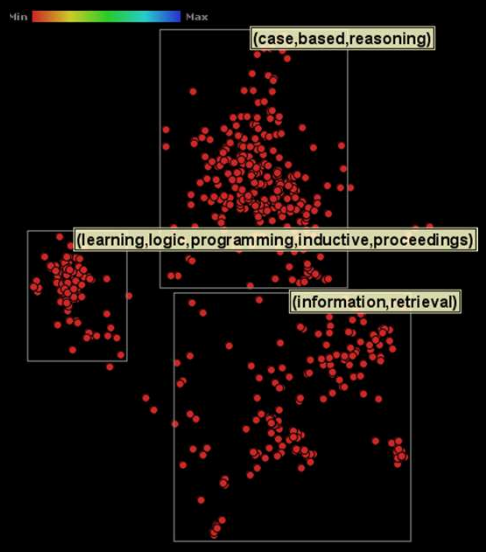


30



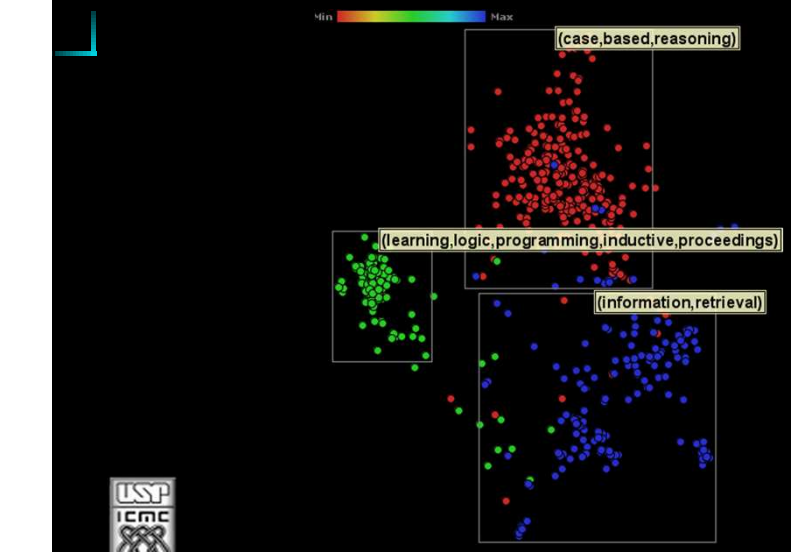
31

Exemplos de Mapas



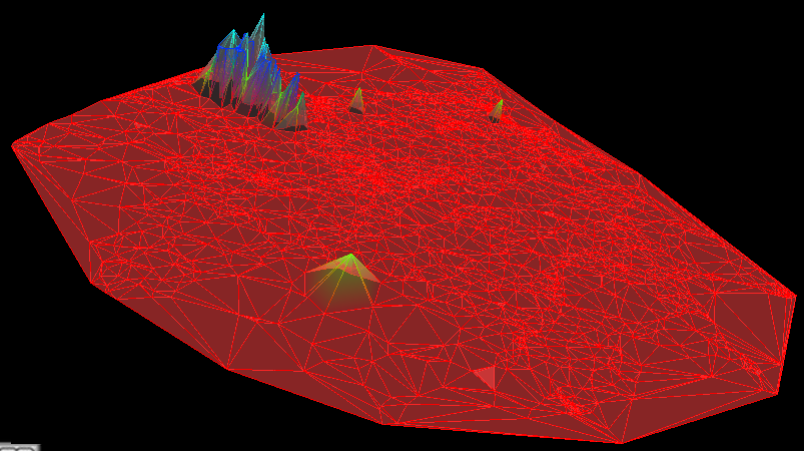
32

Exemplos de Mapas



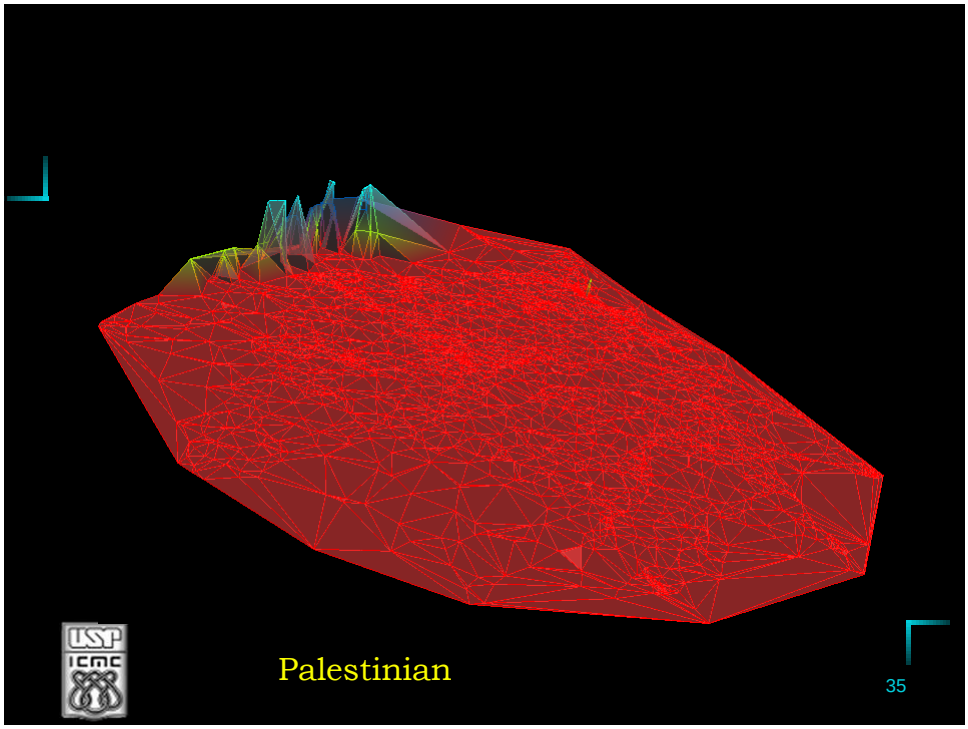
33

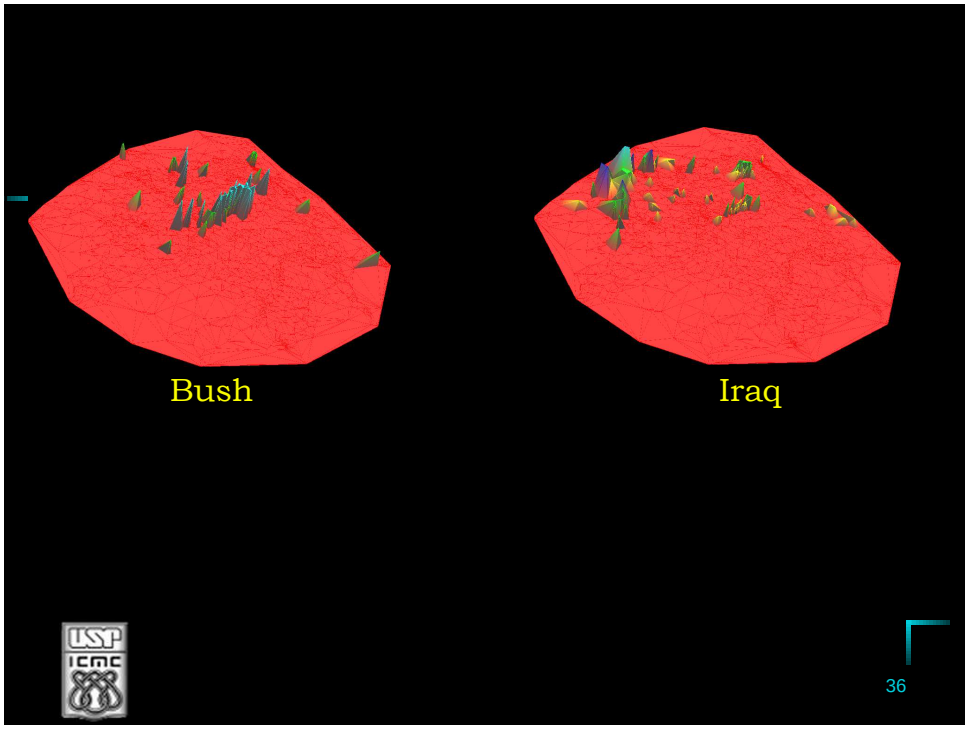
RSS News Flash

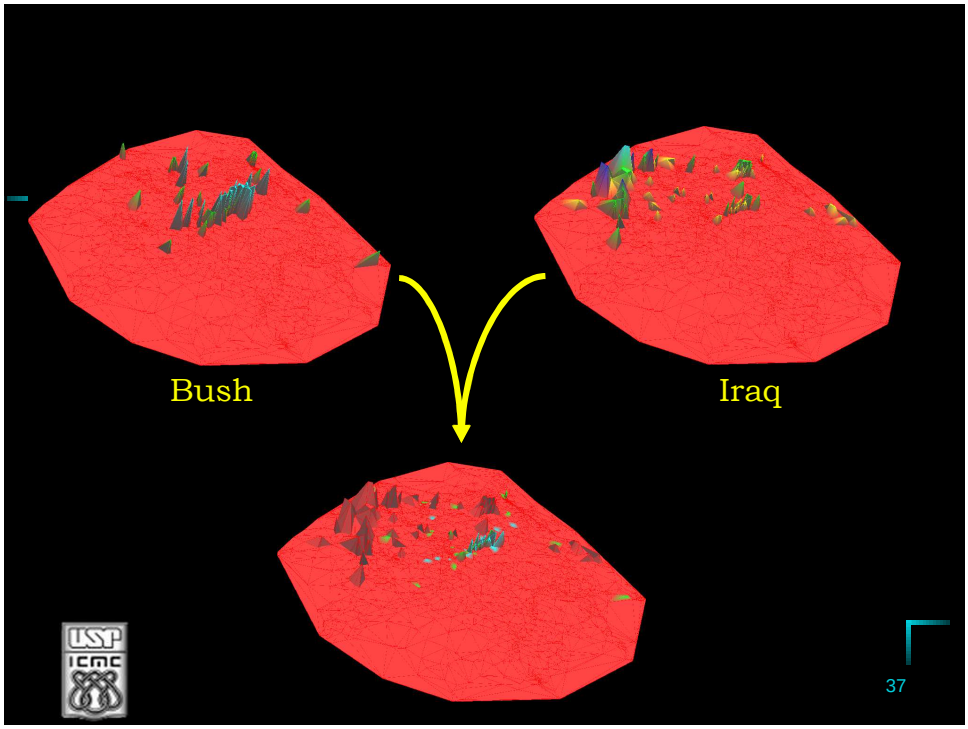


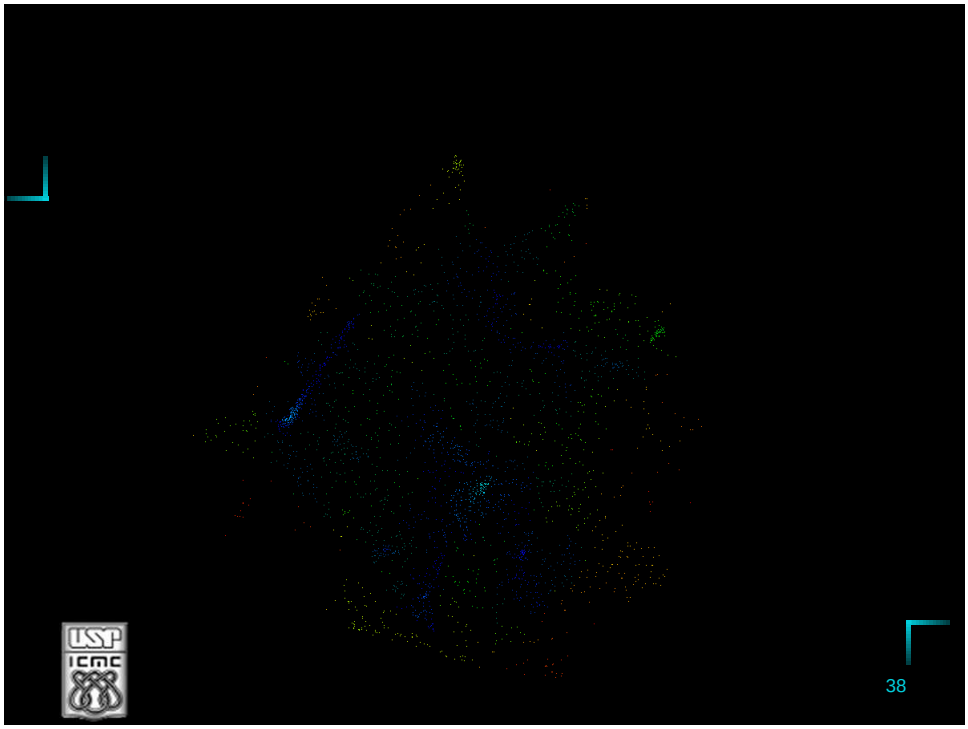
Bird and Flu

34

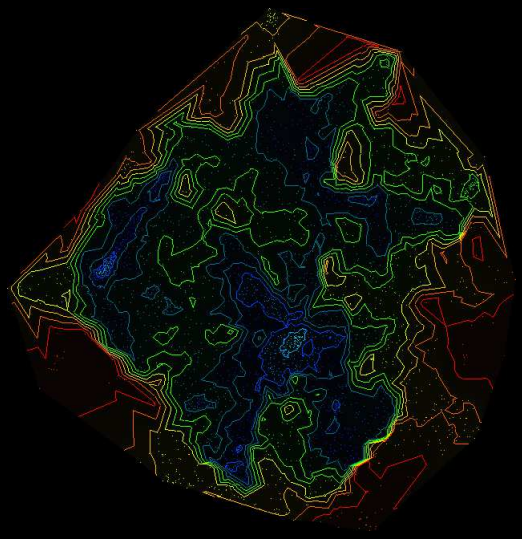

Palestinian
35


Bush
Iraq
36


Bush
Iraq
37


38

Curvas de Nível



Séries Temporais – Vazão em Hidrelétricas

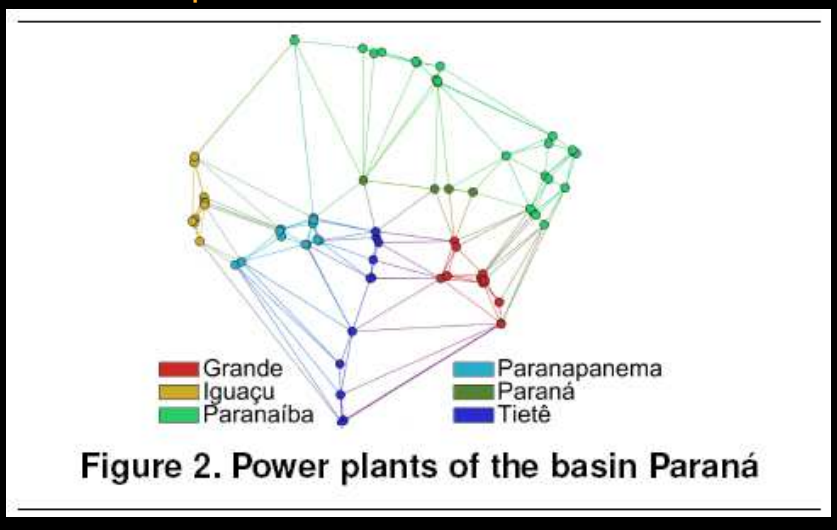
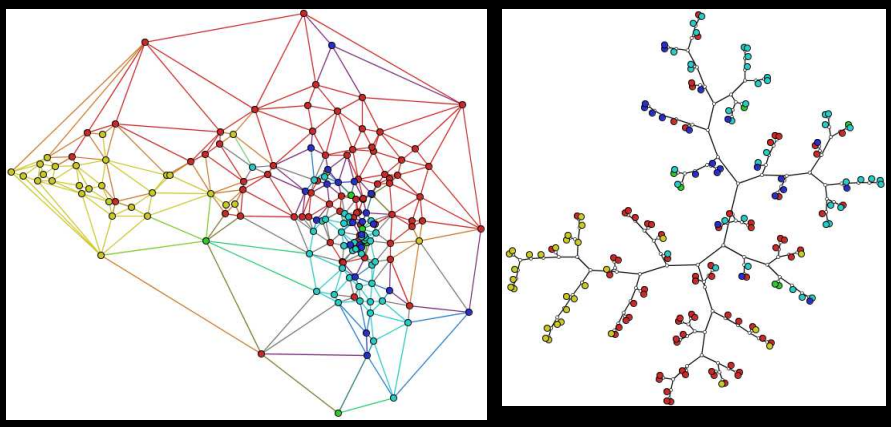


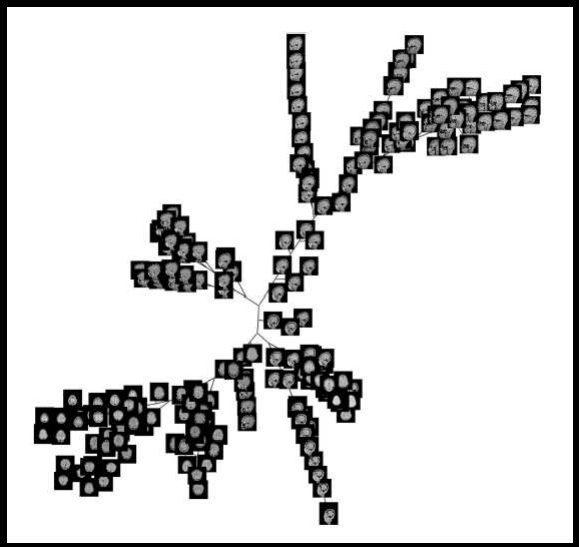
Figure 2. Power plants of the basin Paraná

Further Example - patents



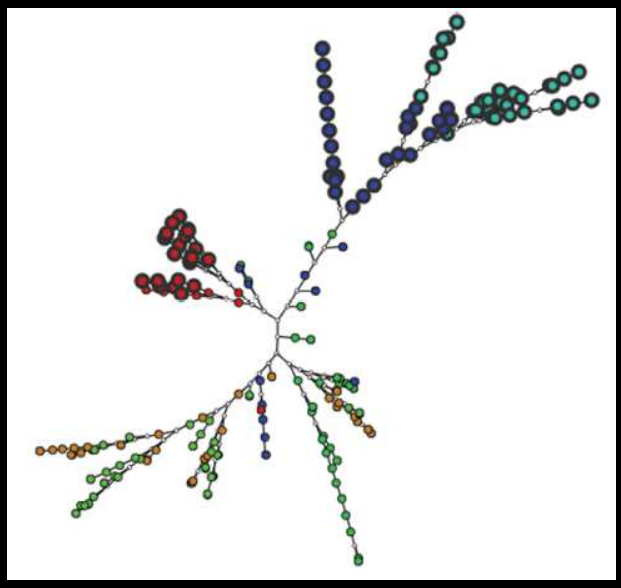
41

Images



42

Images



Exercícios

- Download Pex
- Usar os conjuntos de dados presentes no site e explorá-los.
- Registrar as conclusões

- Download Pex-Image
- Fazer o mesmo

  - Infoserver.lcad.icmc.usp.br/infovis2

Referências

- 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.

Referências

- 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.