UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

UNIVERSITY COLLEGE CORK

School of Computer Science and Information Technology

ICMC USP SÃO CARLOS
Instituto de Ciências Matemáticas e de Computação
University of São Paulo

DATA VISUALIZATION BASICS

Multidimensional Projections

Text / other applications

SCC0652 – Ter 20/10

Rosane Minghim

2020

Multidimensional Visualization Projections/Multidimensional Projections Document Collections Image Collections

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

Input

15														
6	2													
15	1	2												
15	1	2	1											
15	2	1	2	2										
15	3	10	8	15	15									
15	10	7	10	10	7	15								
15	15	10	15	15	10	15	16							
15	15	10	10	10	10	15	16	16						
15	15	15	15	15	15	15	15	3	15					
15	10	10	10	10	10	15	15	15	15	15				
15	10	10	10	10	10	15	15	15	1	15	15			
15	10	15	12	15	15	10	15	15	15	15	15	15		
15	10	10	10	10	10	15	15	15	15	15	2	8	15	
15	10	10	10	10	10	15	1	15	15	15	15	15	15	15

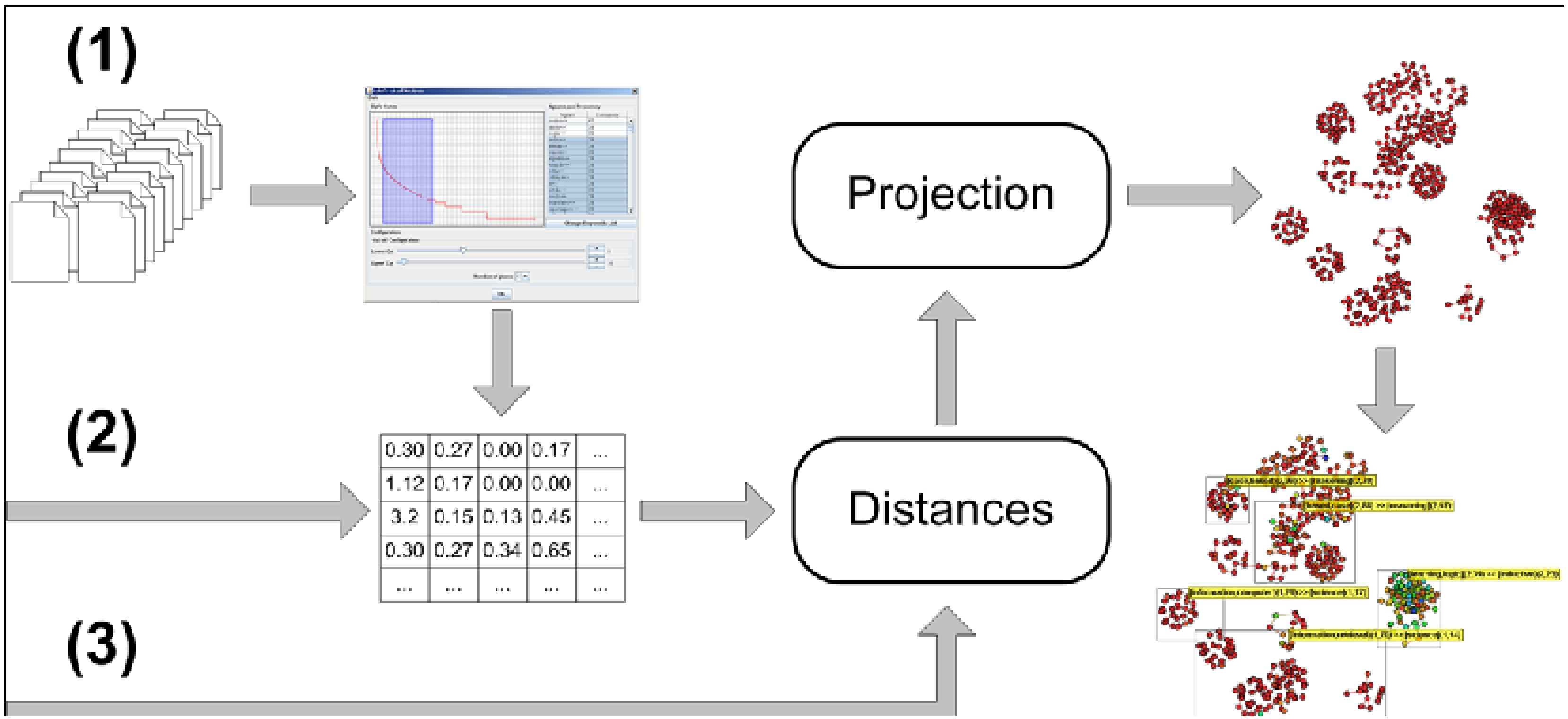
pairwise distances
or similarities

and/or

5	12	15	2	7	5	0	12	9	0	8
12	5	0	12	12	12	12	12	18	12	12
0	1	05	10	15	12	8	12	9	11	5
0	12	01	12	9	0	12	10	5	5	12
12	8	05	12	12	12	8	12	9	12	12
10	12	0	11	10	2	7	12	2	16	7
5	6	8	12	12	15	12	6	9	17	0
7	12	05	0	12	12	10	17	9	12	12
2	10	05	15	12	1	12	10	9	8	2
12	12	7	12	0	12	0	12	10	12	12
6	12	05	17	12	10	12	12	9	12	8
12	10	2	12	1	12	12	11	6	0	12
1	12	05	12	12	16	2	12	9	12	0
10	0	12	12	9	12	0	10	12	12	8
0	12	1	12	12	5	1	7	11	12	12
8	2	11	10	7	12	5	12	15	10	0

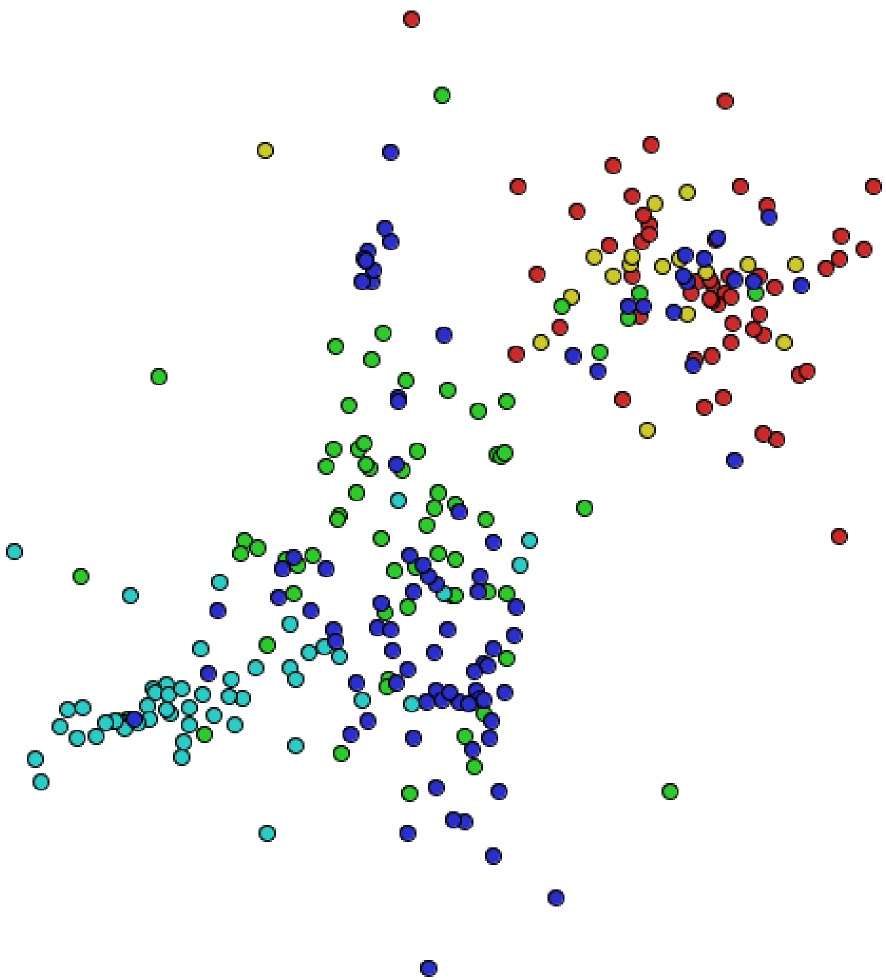
dimensional embedding
(feature space)

Content – based by Projections

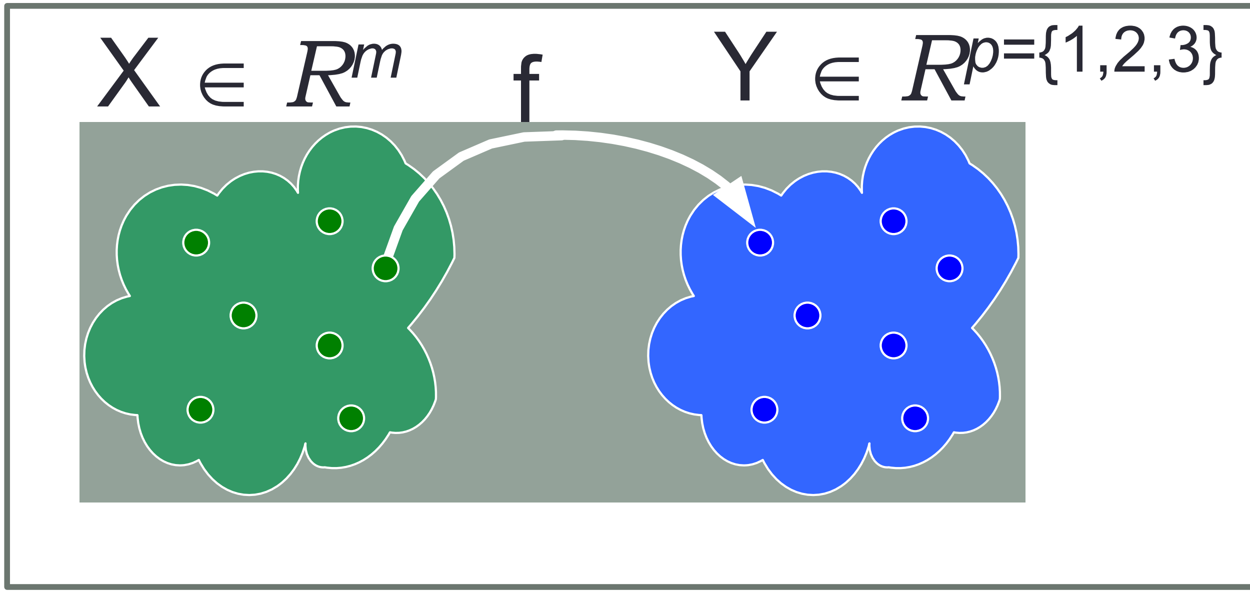


Mapping to a plane to allow exploration

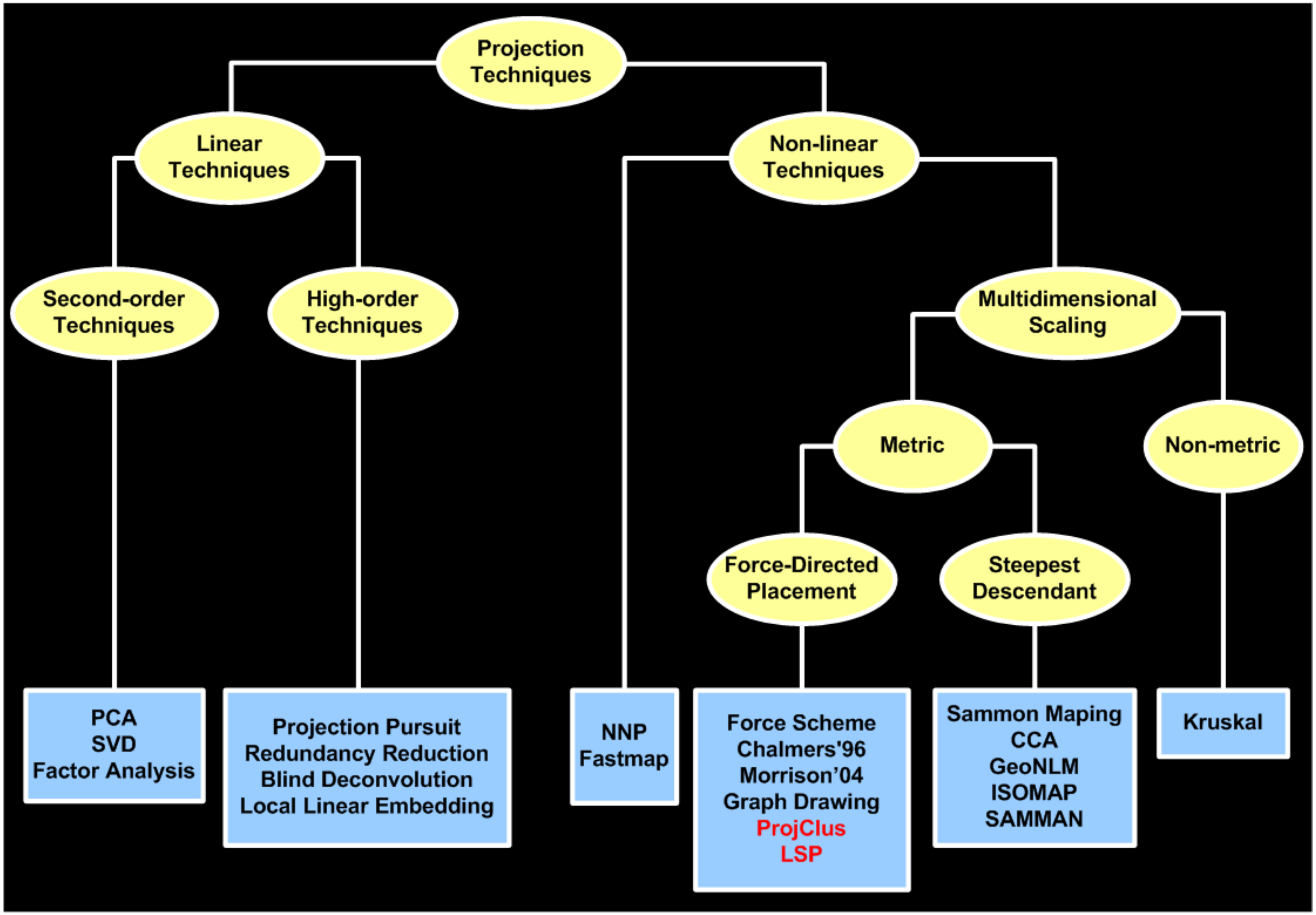
Ex: Patents surgery, drugs, molecular bio



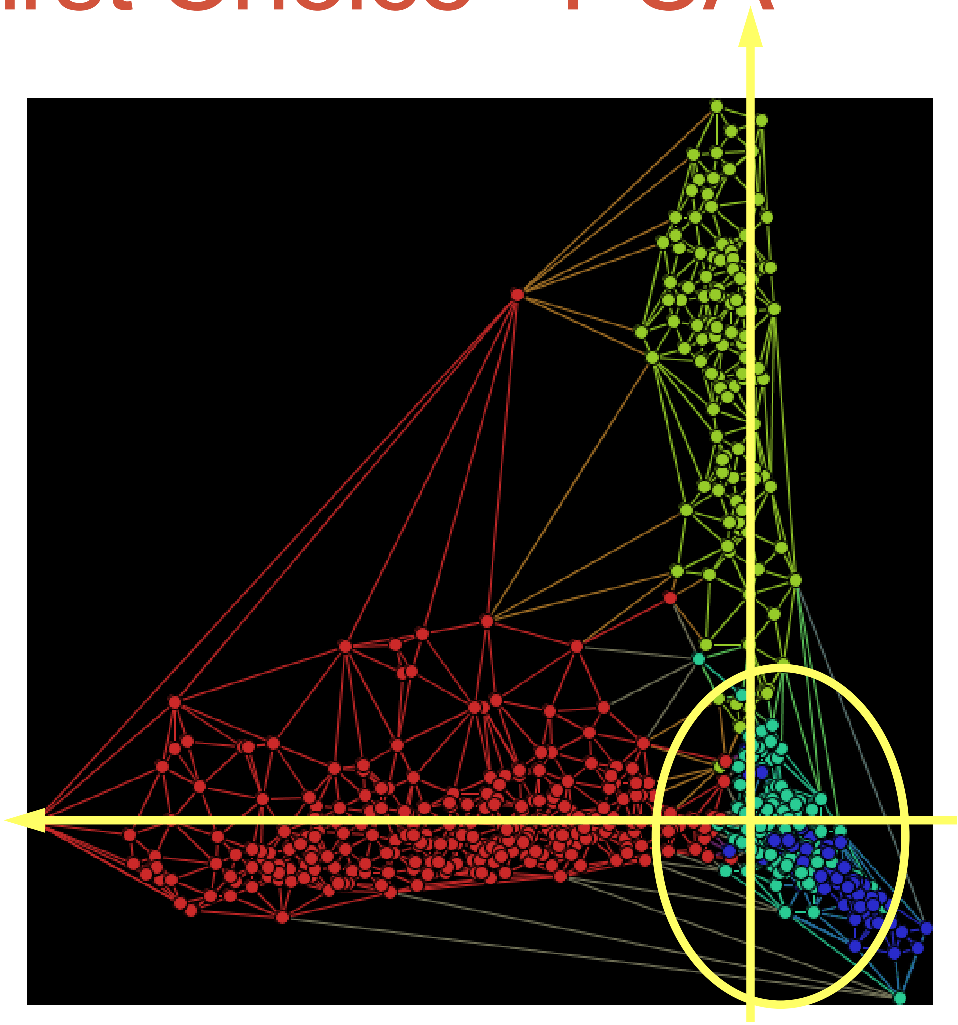
Projection Techniques



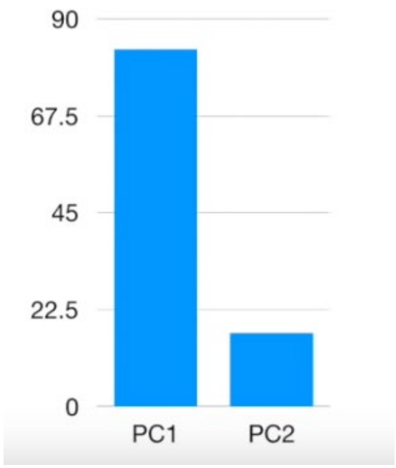
- $\delta: x_i, x_j \to R, x_i, x_j \in X$
- $d: y_i, y_j \to R, y_i, y_j \in Y$
- $f: X \to Y, |\delta(x_i, x_j) - d(f(x_i), f(x_j))| \approx 0, \forall x_i, x_j \in X$

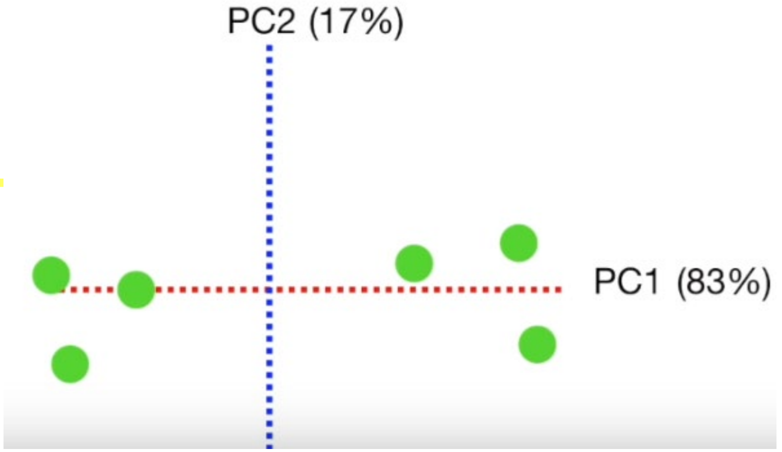

Projection Techniques
Linear Techniques
Non-linear Techniques
Second-order Techniques
High-order Techniques
Multidimensional Scaling
Metric
Non-metric
Force-Directed Placement
Steepest Descendant
PCA
SVD
Factor Analysis
Projection Pursuit
Redundancy Reduction
Blind Deconvolution
Local Linear Embedding
NNP
Fastmap
Force Scheme
Chalmers'96
Morrison'04
Graph Drawing
ProjClus
LSP
Sammon Maping
CCA
GeoNLM
ISOMAP
SAMMAN
Kruskal

First Choice - PCA



Source:own

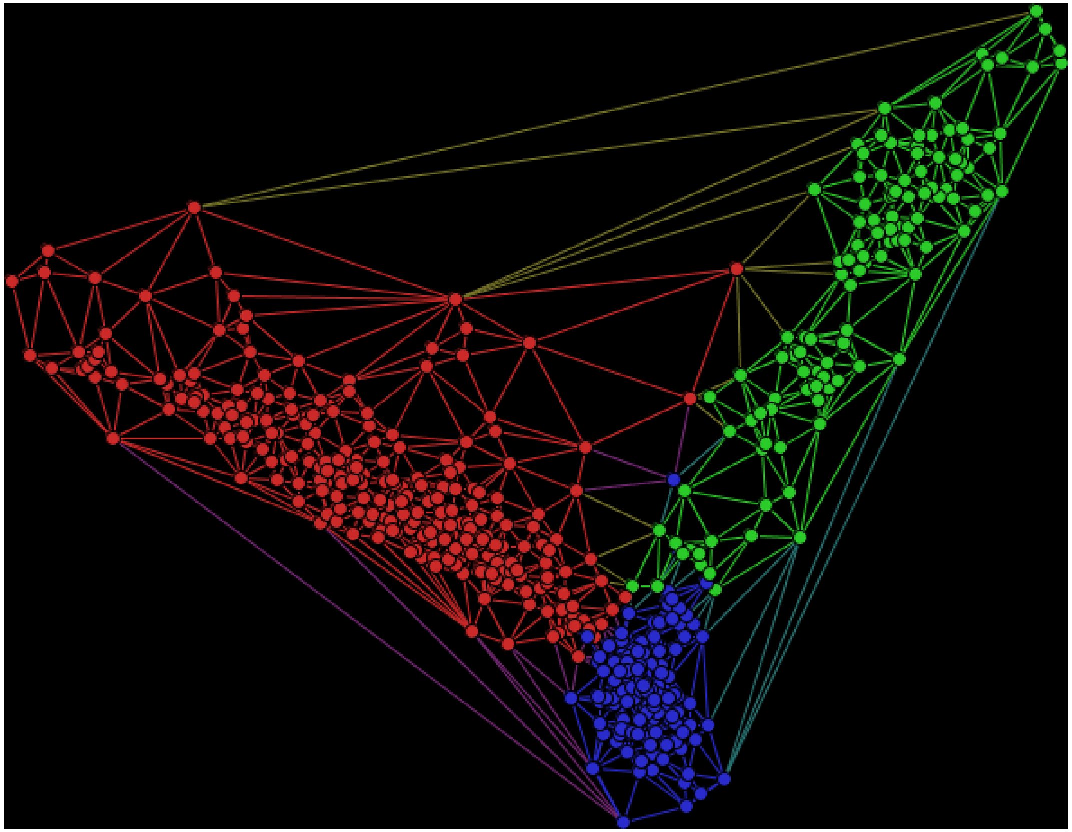




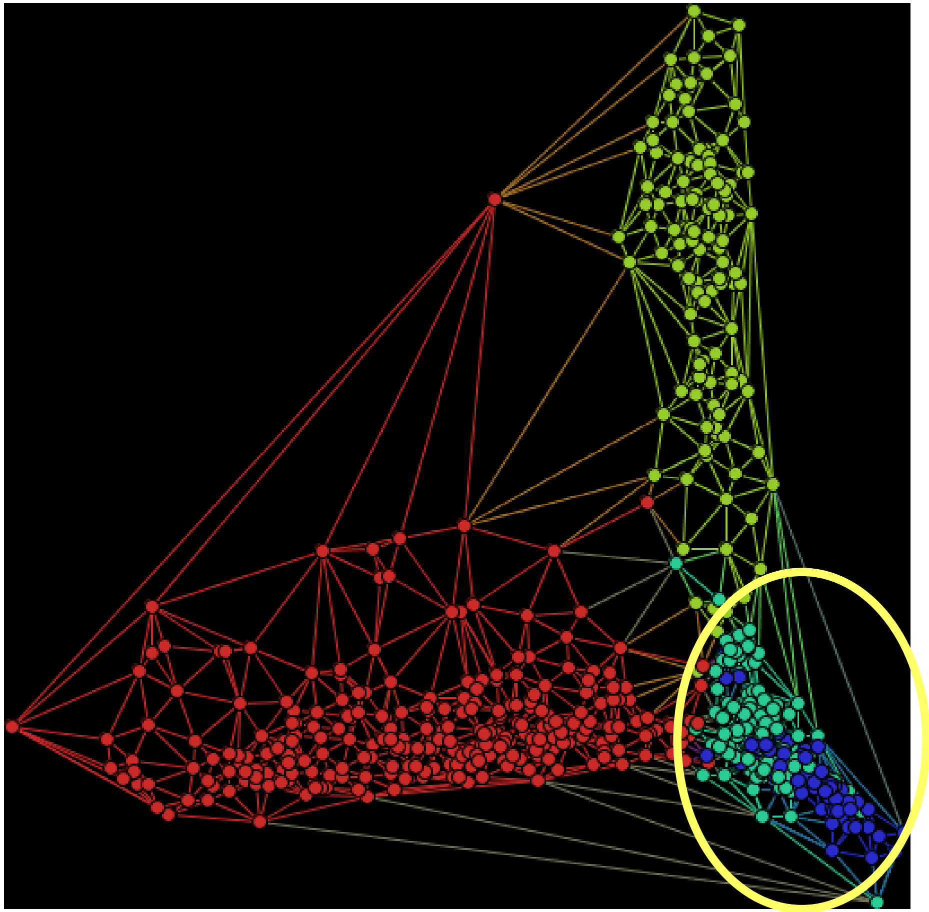
source: StatQuest

Problems PCA

390 dimensions



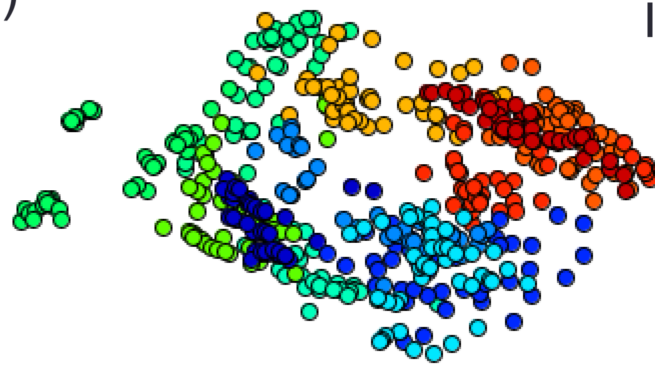
Problems PCA

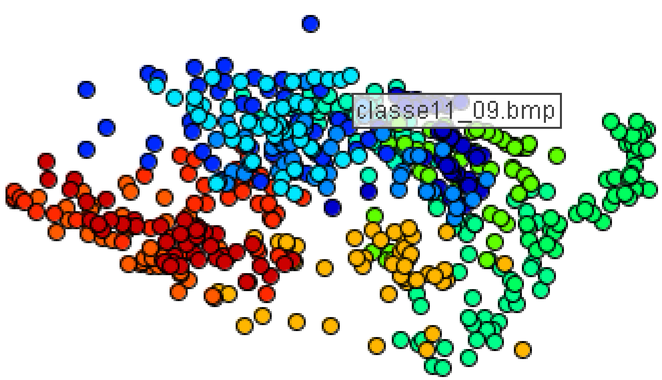


Projection as dimension reduction

- Classical
  - PCA (Principal Component Analysis)
  - Multidimensional Scaling (MDS)
  - LLE (Local Linear Embedding)
  - ISOMAP
  - Sammon´s mapping

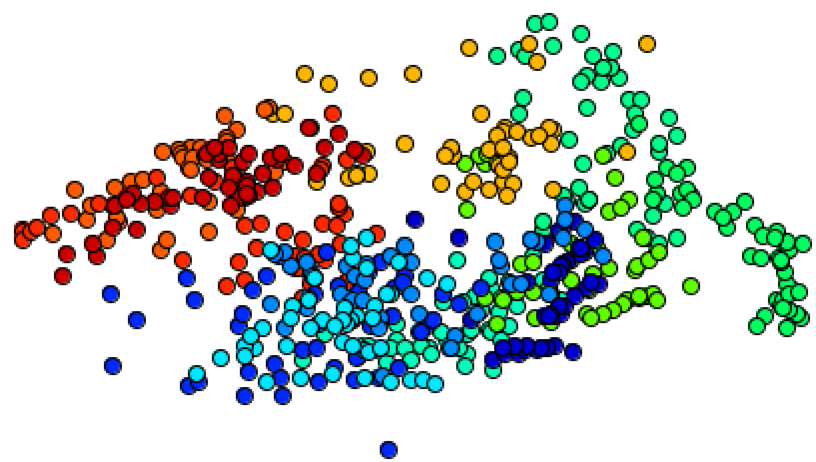
ISOMAP



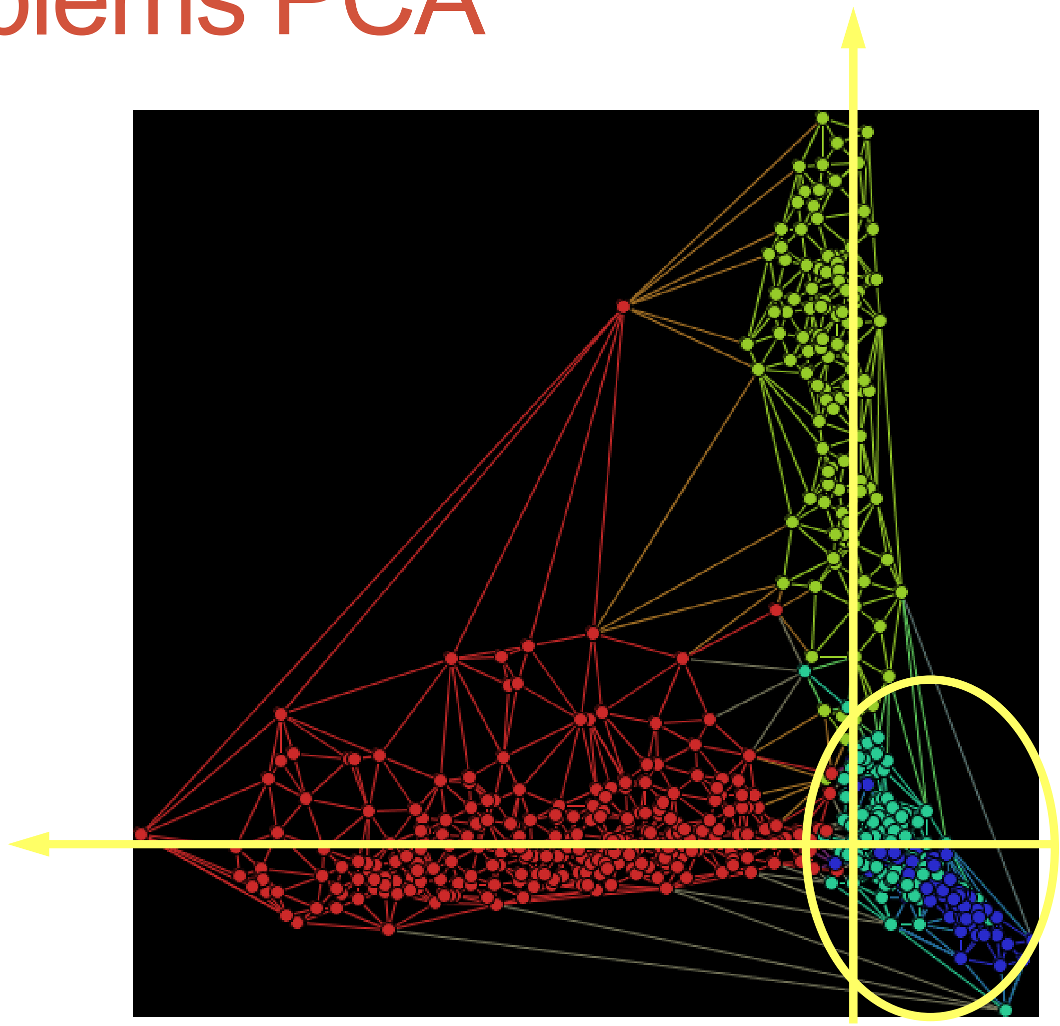


Classical Scaling

PCA

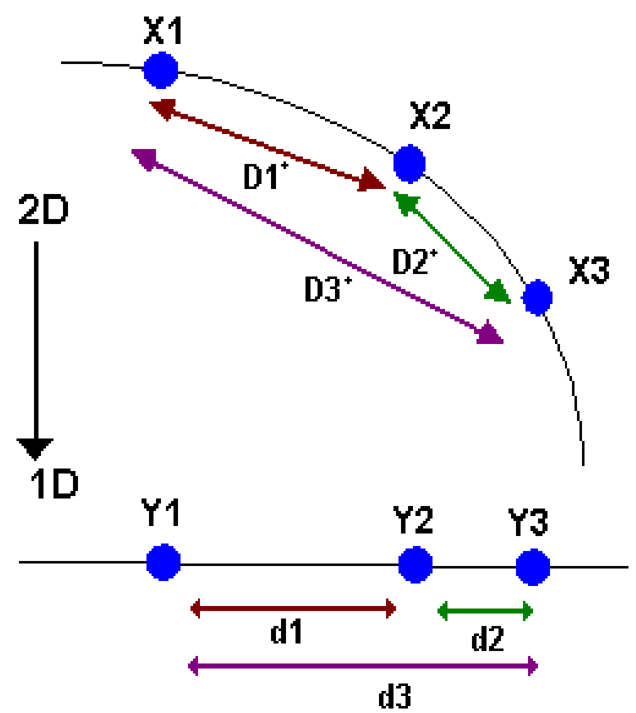


Problems PCA

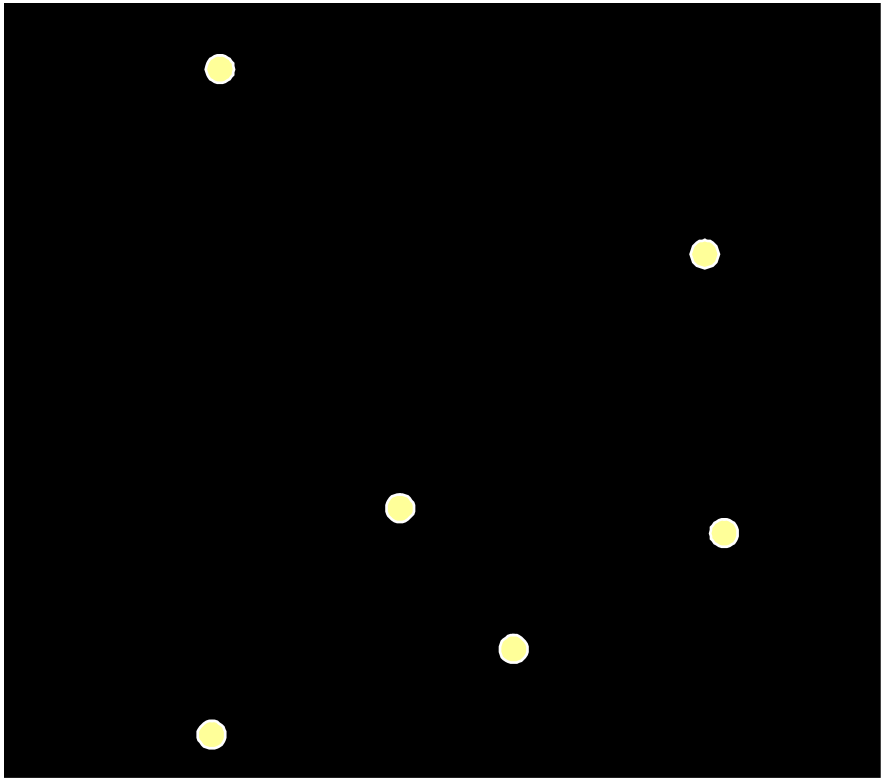


Ex: Sammon Mapping

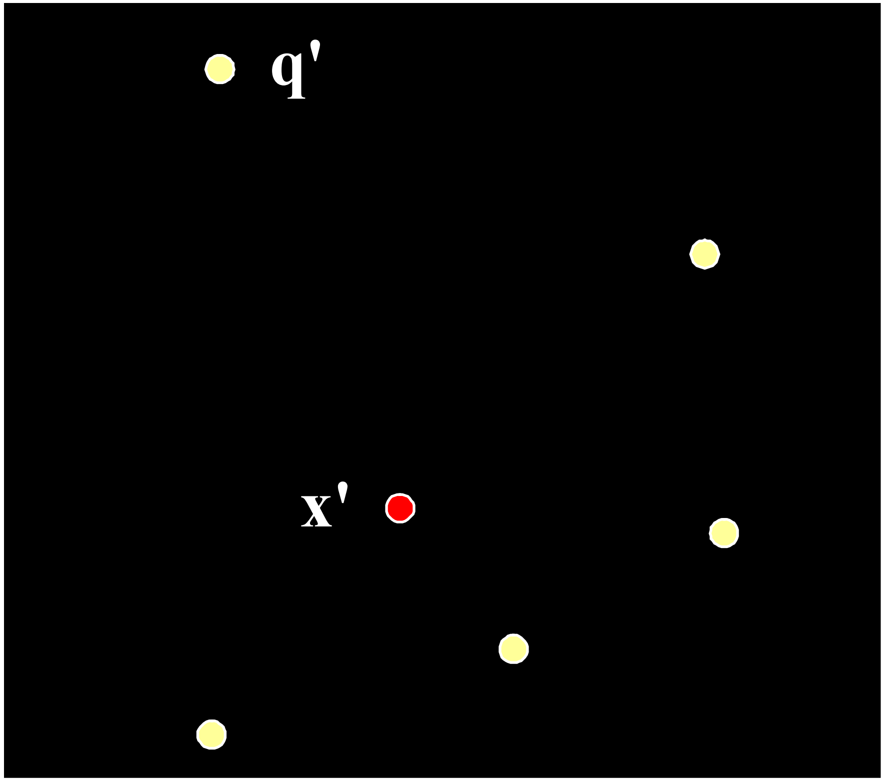
- Let **X** be the points in the original space R^n, we apply a distance measure d_{ij}* between Xi an Xj., and find **Y, the projected point**, ex. R^2 and d_{ij} the Euclidean distance between them.
- Sammon's method applies an error function to measure the target.



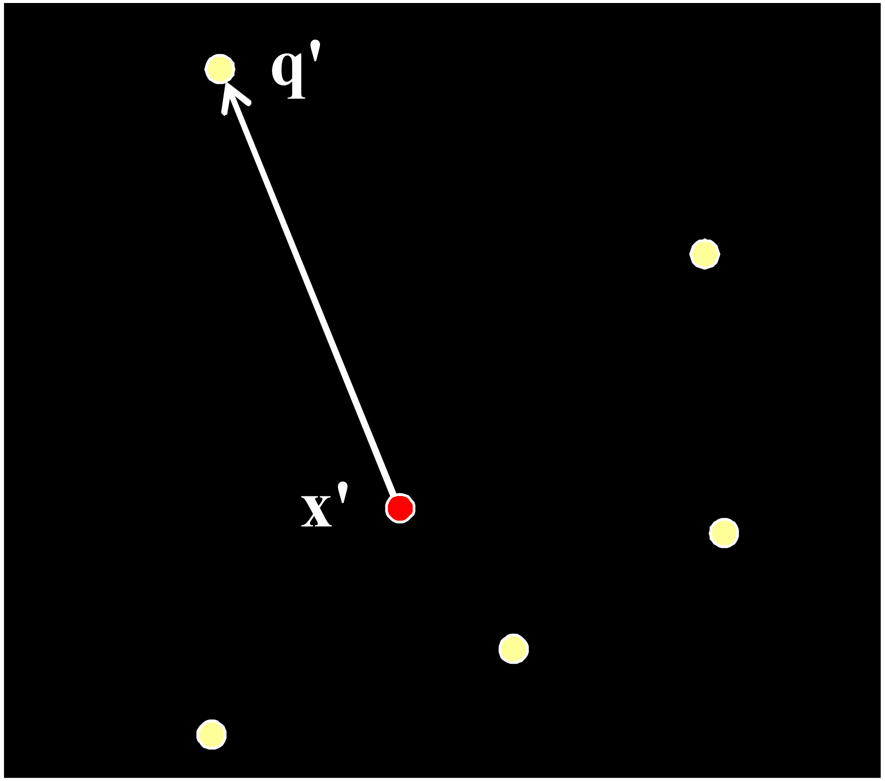
Force Based Point Placement



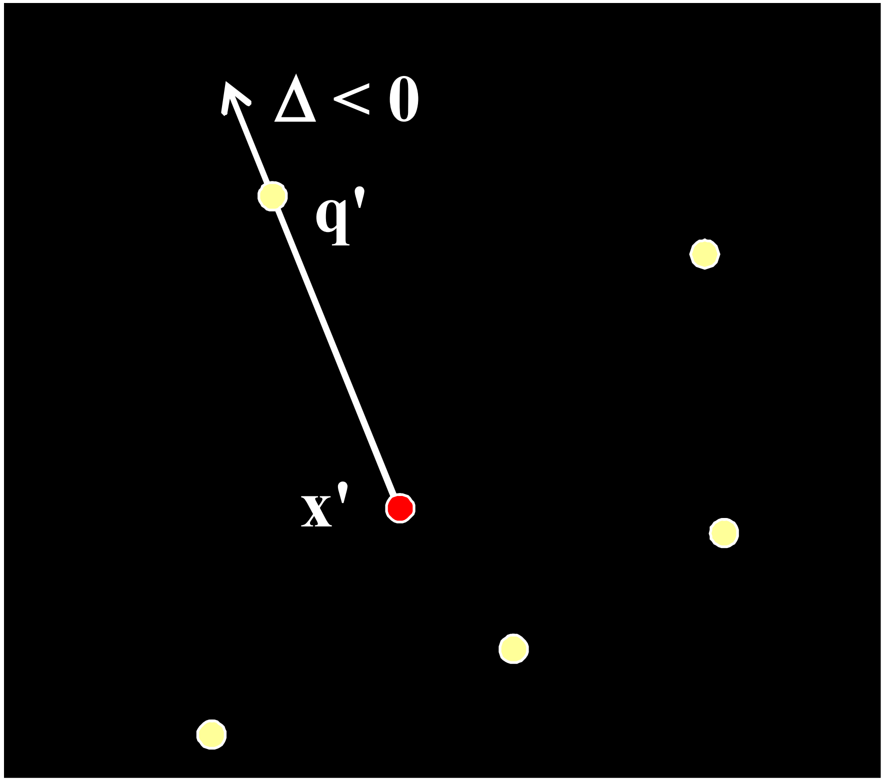
Force Scheme [Tejada et al., 2003]



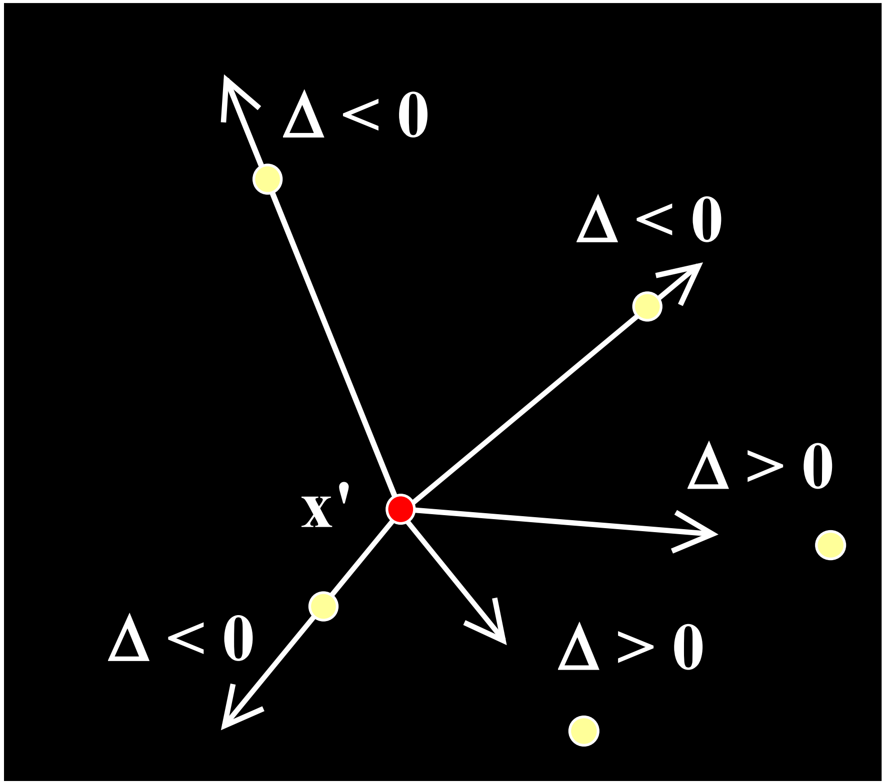
Force Scheme [Tejada et al., 2003]



Force Scheme [Tejada et al., 2003]



Force Scheme [Tejada et al., 2003]



Force Scheme [Tejada et al., 2003]

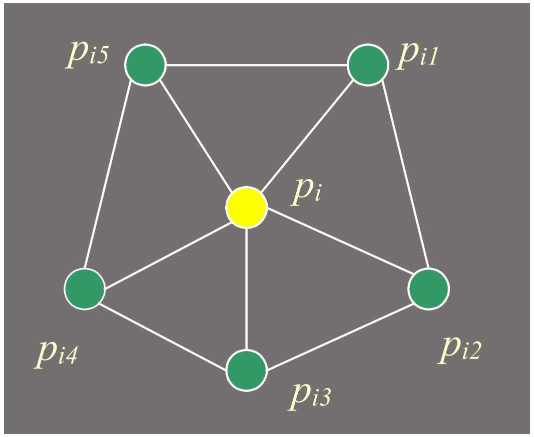
1. Map each point X to the plane (fastmap, nnp, etc.)
2. For each projected point x
    1. For each projected point q’ ≠ x’
        1. Compute the vector **v** of <x’ to q’>
        2. Move q’ in direction of **v**, one fraction of Δ

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

3. Normalize the coordinates between [0,1]

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.



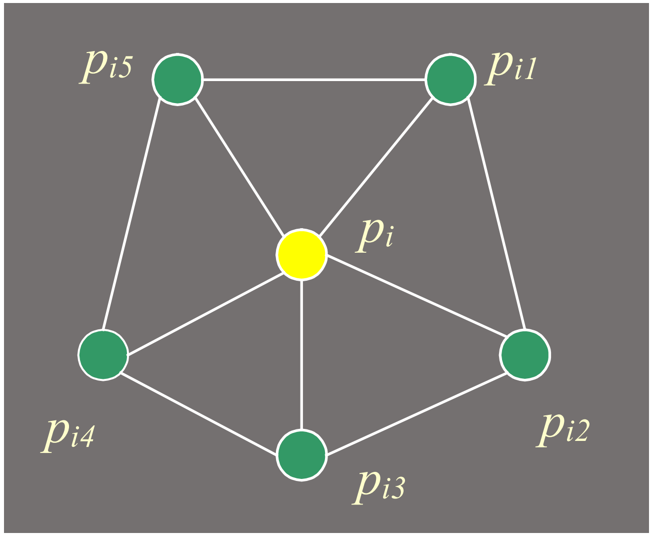
LSP [Paulovich et al., 2006/2008]

- 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 = {$p_{i1},\ldots,p_{iki}$} *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

LSP: Laplacian Matrix

$$L\mathbf{x_1}=0,\ L\mathbf{x_2}=0,\ \ldots,\ L\mathbf{x_p}=0$$

where x_1, $x_2, \ldots,$ x_p are vectors containing the Cartesian coordinates of the points

and L is the matrix defined by:

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

$$\begin{pmatrix} \boxed{\quad \mathrm{L} \quad} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 0 \end{pmatrix}$$

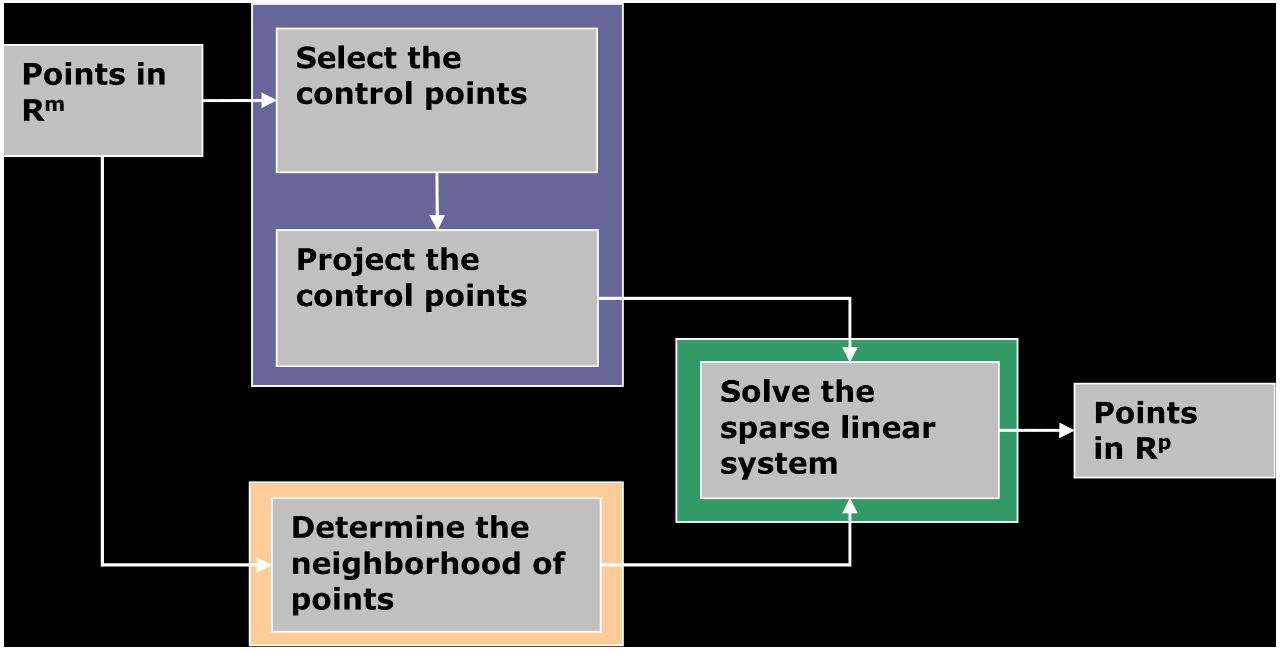
LSP: Adding control points

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

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

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

LSP: Overview



LSP: System example

$$v_1 = \{p_3 p_4 p_6\}$$

$$v_2 = \{p_5 p_4 p_6\}$$

$$v_3 = \{p_1 p_5 p_6\}$$

$$v_4 = \{p_1 p_6\}$$

$$v_5 = \{p_3 p_2 p_6\}$$

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

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

LSP: System example

$$v_1 = \{p_3 p_4 p_6\}$$

$$v_2 = \{p_5 p_4 p_6\}$$

$$v_3 = \{p_1 p_5 p_6\}$$

$$v_4 = \{p_1 p_6\}$$

$$v_5 = \{p_3 p_2 p_6\}$$

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

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

LSP: System example

$$v_1 = \{p_3 p_4 p_6\}$$

$$v_2 = \{p_5 p_4 p_6\}$$

$$v_3 = \{p_1 p_5 p_6\}$$

$$v_4 = \{p_1 p_6\}$$

$$v_5 = \{p_3 p_2 p_6\}$$

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

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

LSP: System example

$$v_1 = \{p_3 p_4 p_6\}$$
$$v_2 = \{p_5 p_4 p_6\}$$
$$v_3 = \{p_1 p_5 p_6\}$$
$$v_4 = \{p_1 p_6\}$$
$$v_5 = \{p_3 p_2 p_6\}$$
$$v_6 = \{p_1 p_2 p_4 p_5\}$$

$$pc = \{p_3 p_6\}$$

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

L

LSP: System example

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

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

$pc = \{p_3 p_6\}$

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

LSP: System example

$$v_1 = \{p_3 p_4 p_6\}$$

$$v_2 = \{p_5 p_4 p_6\}$$

$$v_3 = \{p_1 p_5 p_6\}$$

$$v_4 = \{p_1 p_6\}$$

$$v_5 = \{p_3 p_2 p_6\}$$

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

$$pc = \{p_3 p_6\}$$

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

LSP: System example

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

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

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

$pc = \{p_3 p_6\}$

LSP: System example

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

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

$pc = \{p_3 p_6\}$

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

LSP: System example

$v_1 = \{p_3 p_4 p_6\}$

$v_2 = \{p_5 p_4 p_6\}$

$v_3 = \{p_1 p_5 p_6\}$

$v_4 = \{p_1 p_6\}$

$v_5 = \{p_3 p_2 p_6\}$

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

$pc = \{p_3 p_6\}$

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

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

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

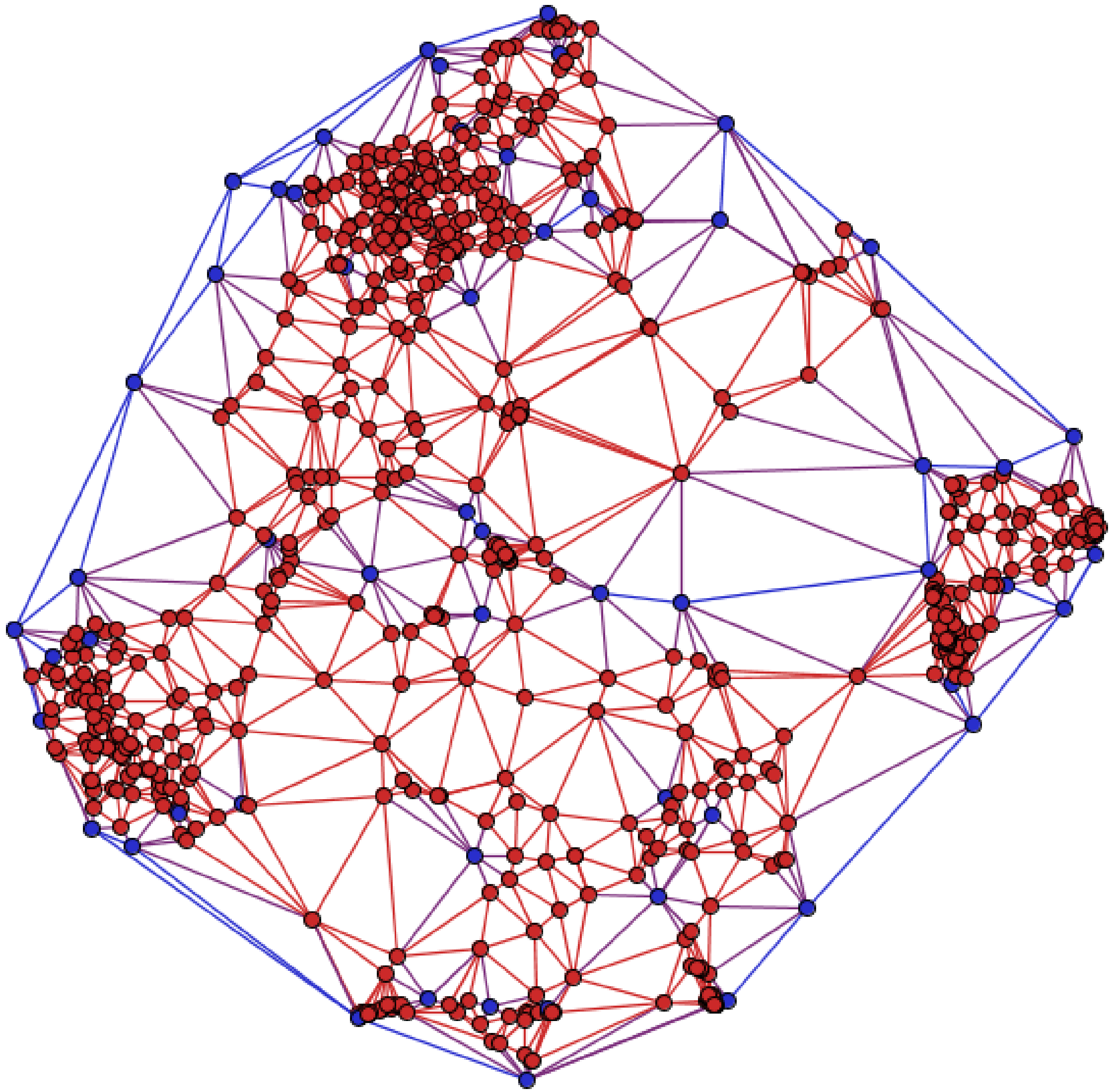
- $A^T A$ is symmetric and sparse and can be solved using the factorization of Cholesky

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

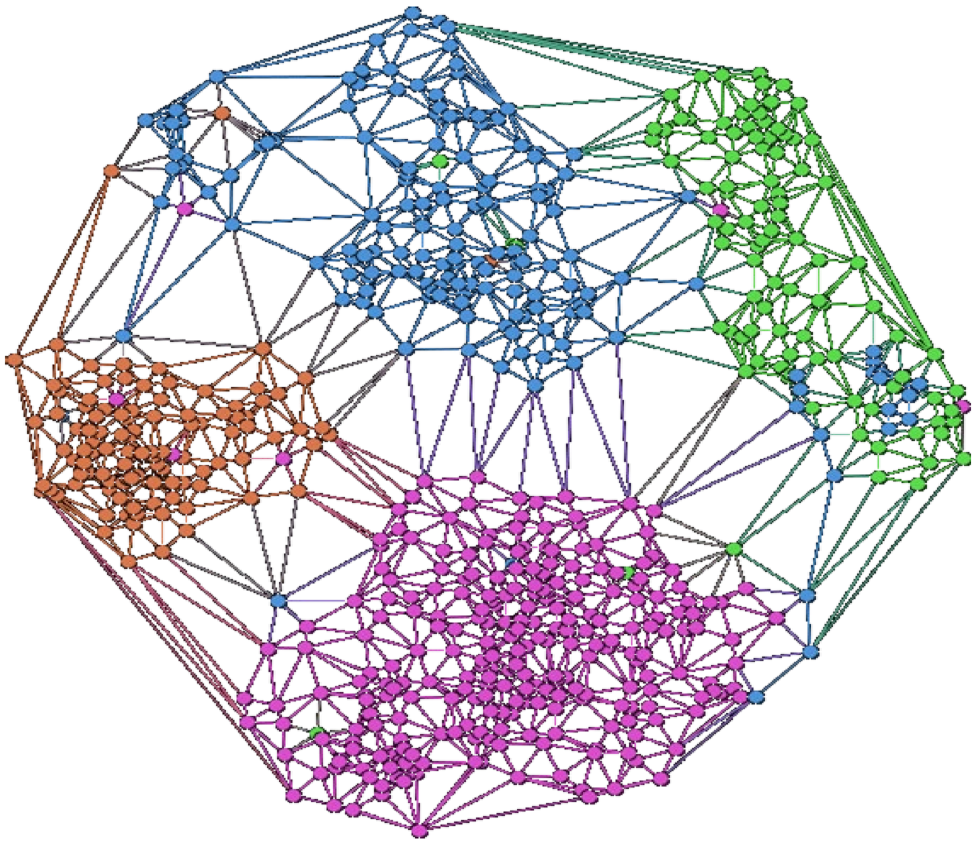
Choosing the Control Points

- 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

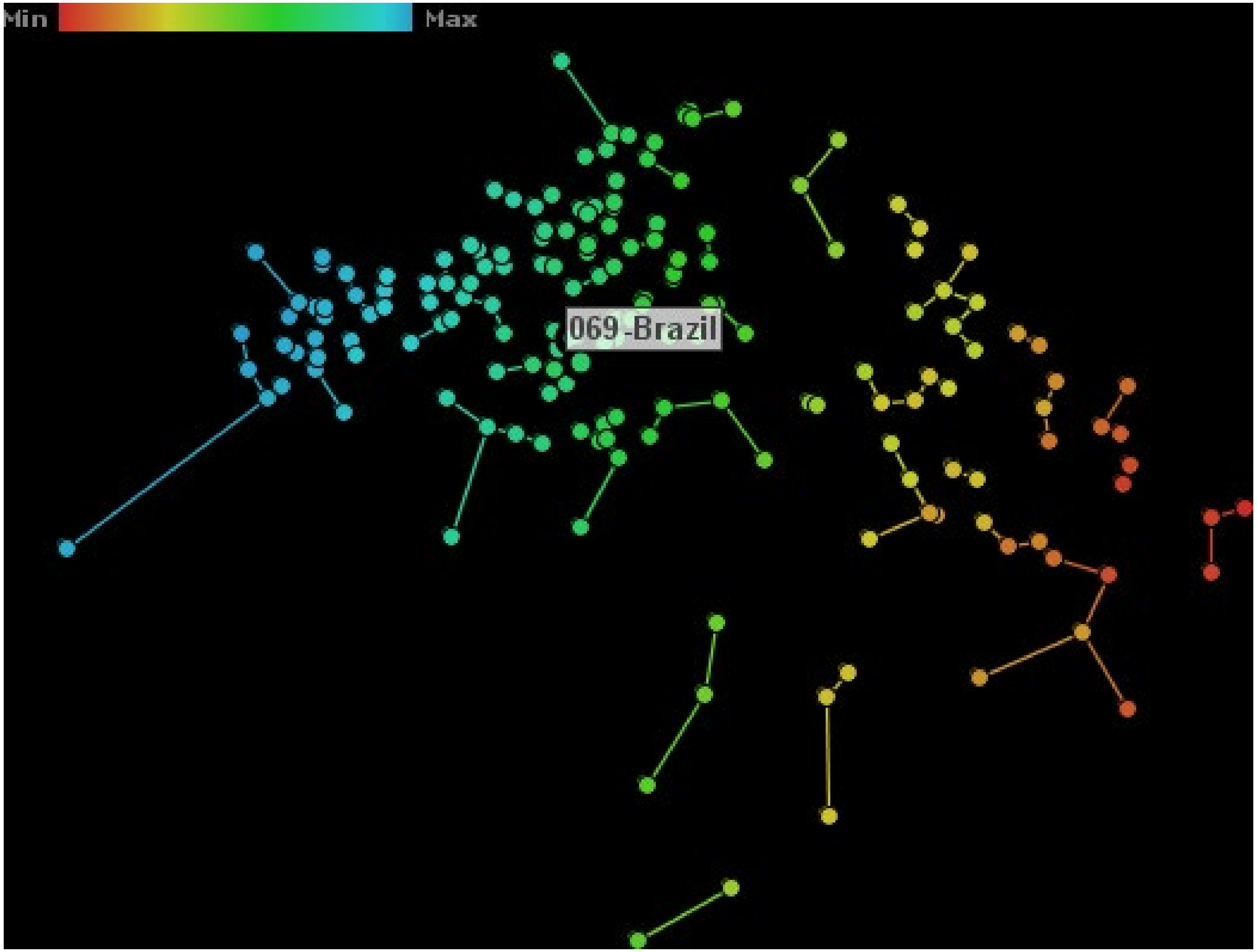


Control points in blue

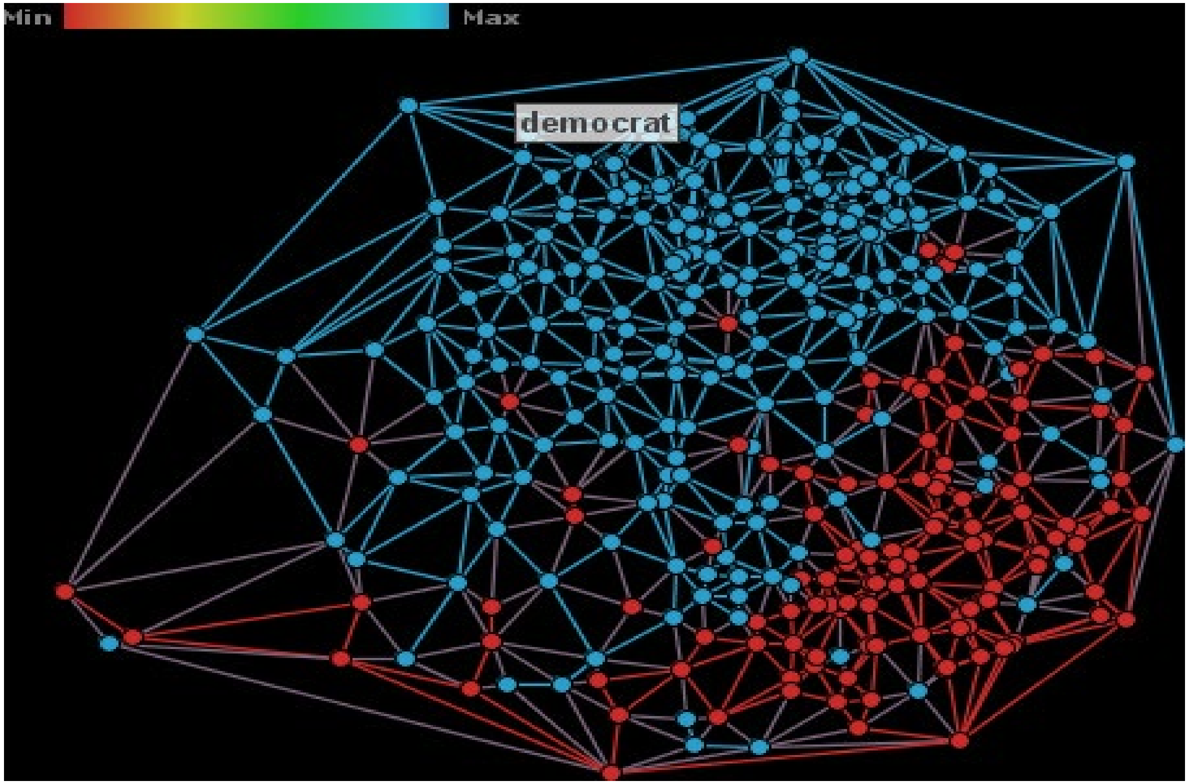
Projection



Projection: HDI



Projection: Voting



Stochastic Neighborhood Embedding sne and t-sne

- Distance in original space

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)},$$

- Distance in projected space

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}$$

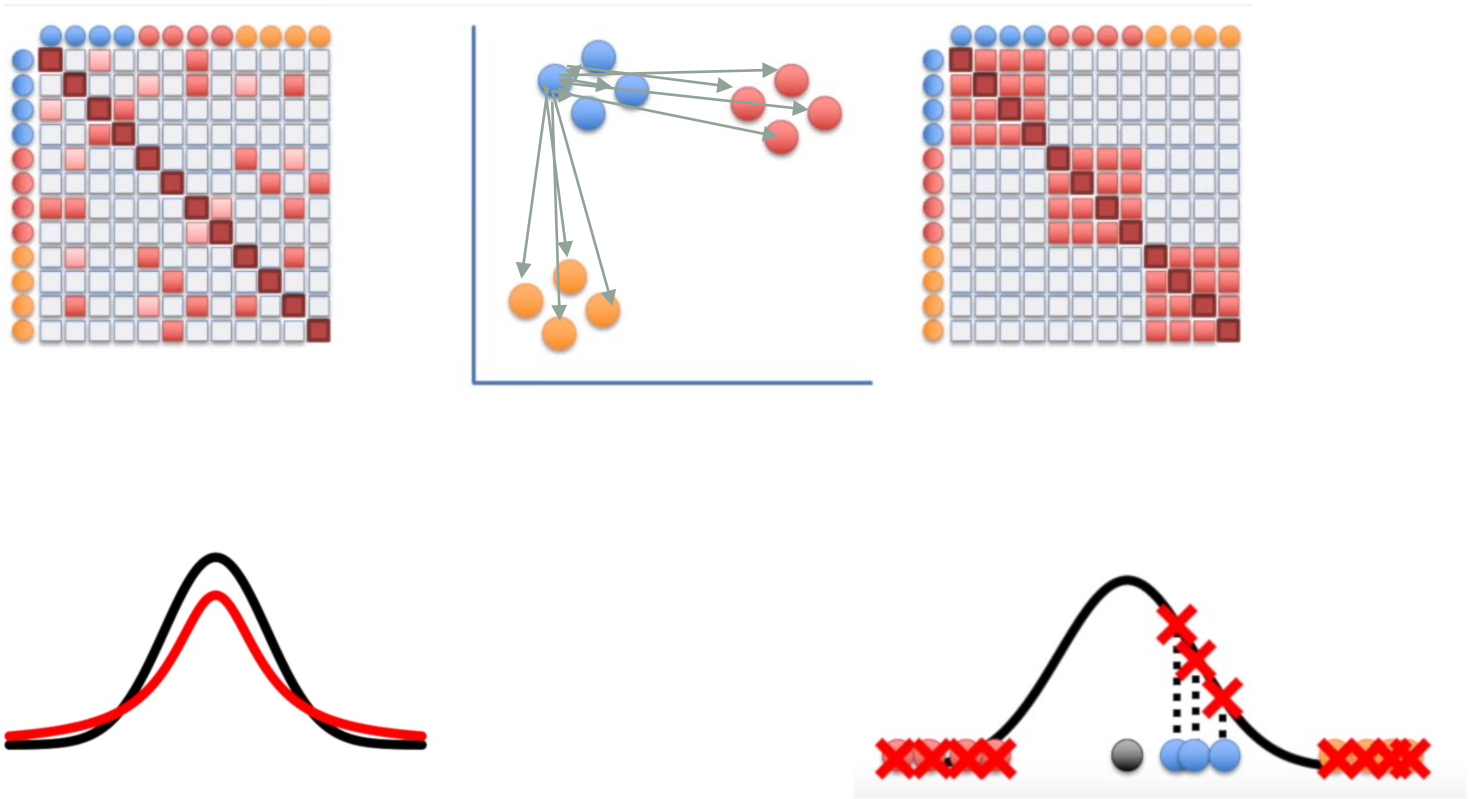
- Cost function

$$C = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

- Non-gaussian neighborhoods: t-sne

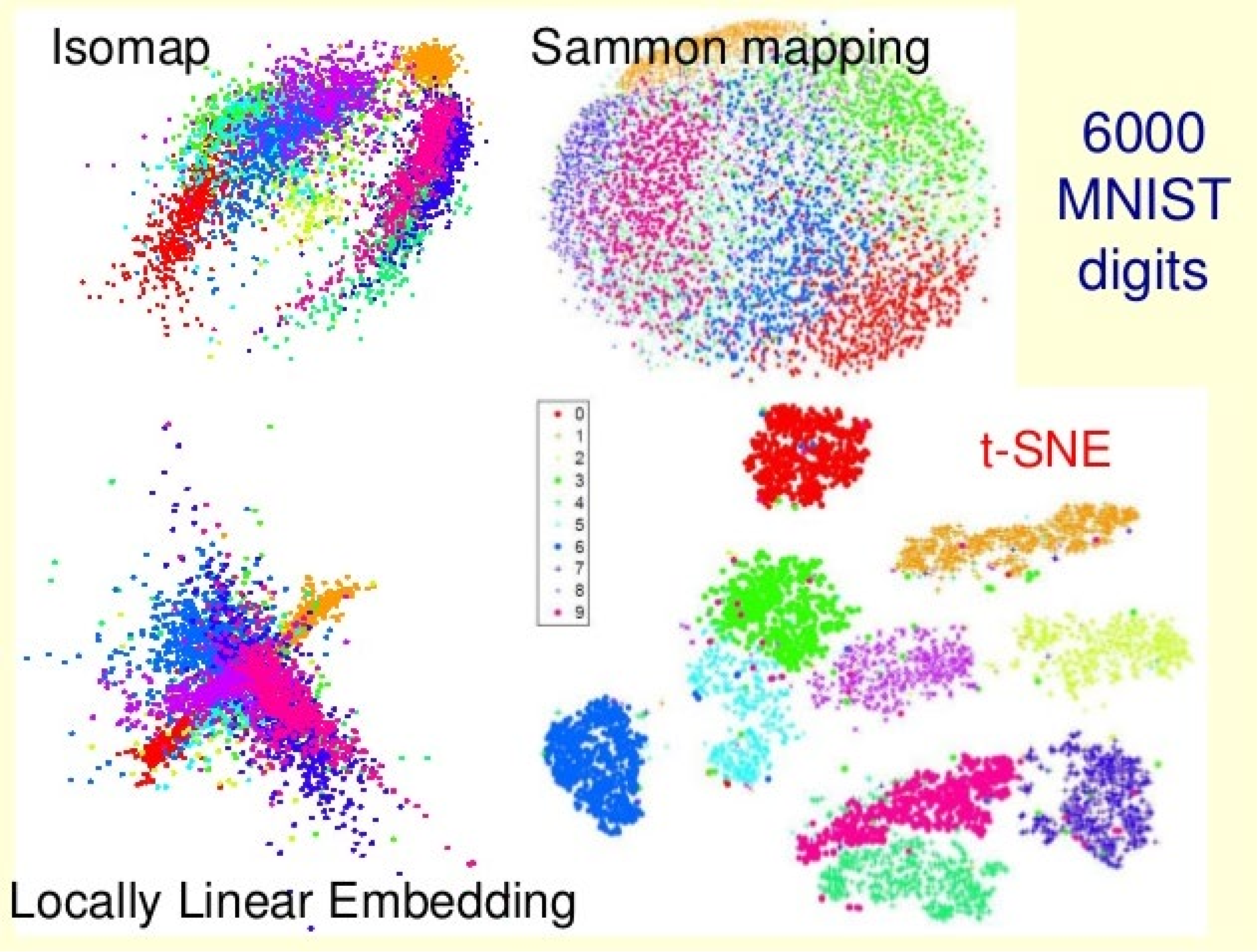
L.J.P. van der Maaten and G.E. Hinton. **Visualizing High-Dimensional Data Using t-SNE**. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008

Stochastic Neighborhood Embedding sne and t-sne



Source: StatQuest (adapted)

T-sne Examples



Source: https://www.slideshare.net/xuyangela/an-introduction-to-tsne

T-sne etc.

- Demo:

  https://projector.tensorflow.org/

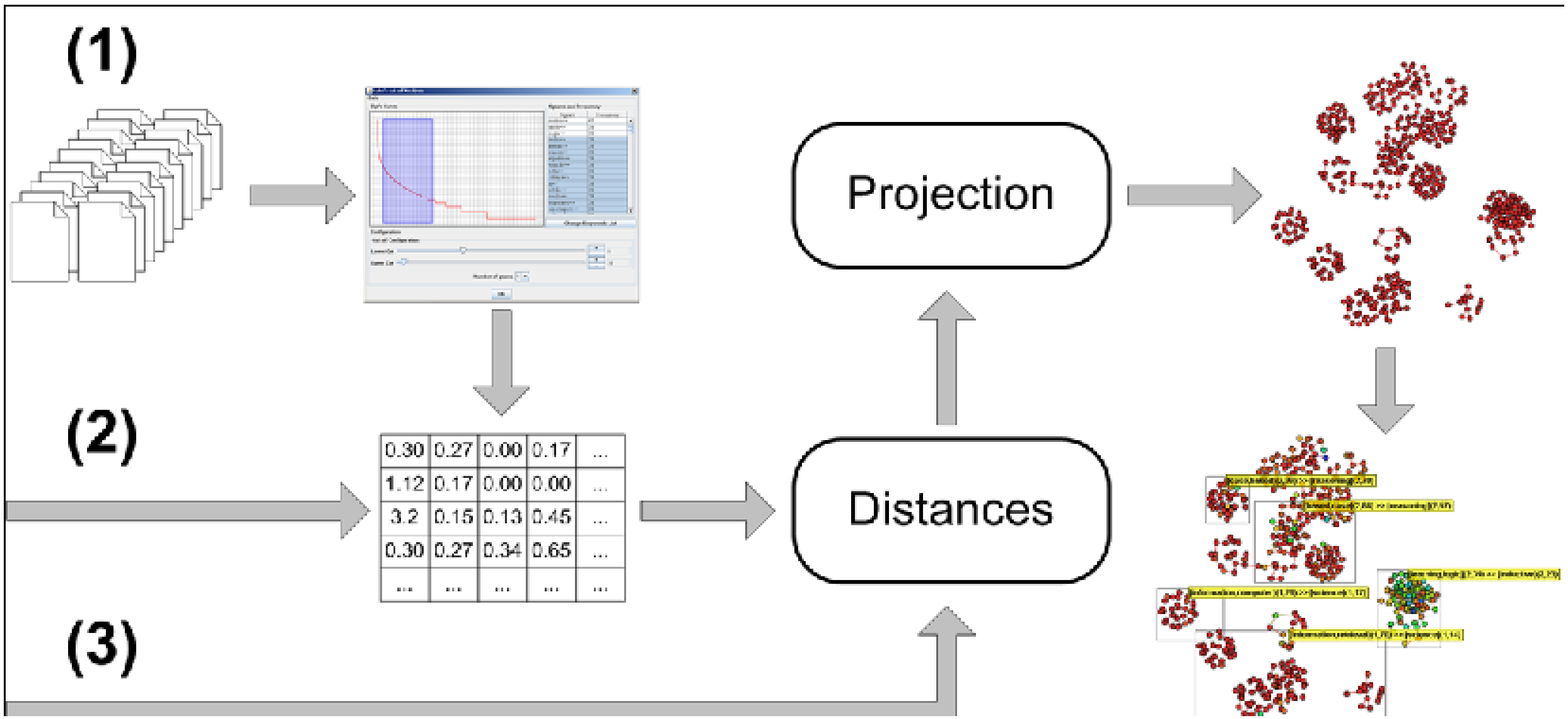
- Explanation (StatQuest from University of North Caroline – Genetics Department:

  https://www.youtube.com/watch?v=NEaUSP4YerM

- See also UMAP projection

  https://umap-learn.readthedocs.io/en/latest/

Visualization by Projections

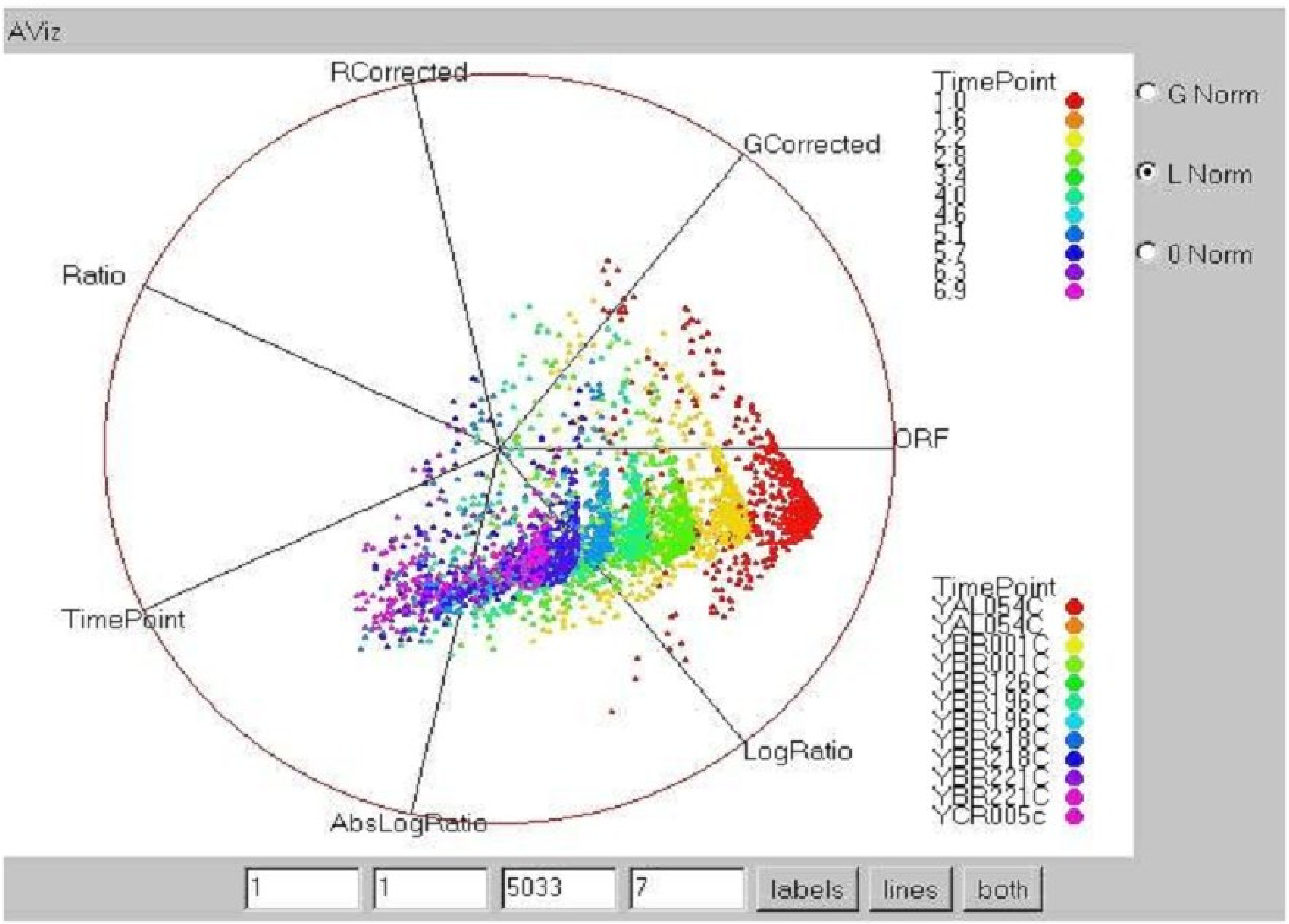


The case of document collections

- Applications
  - Teaching/Research
  - Search
  - Investigation

  - Patents
  - Medical reports
  - News

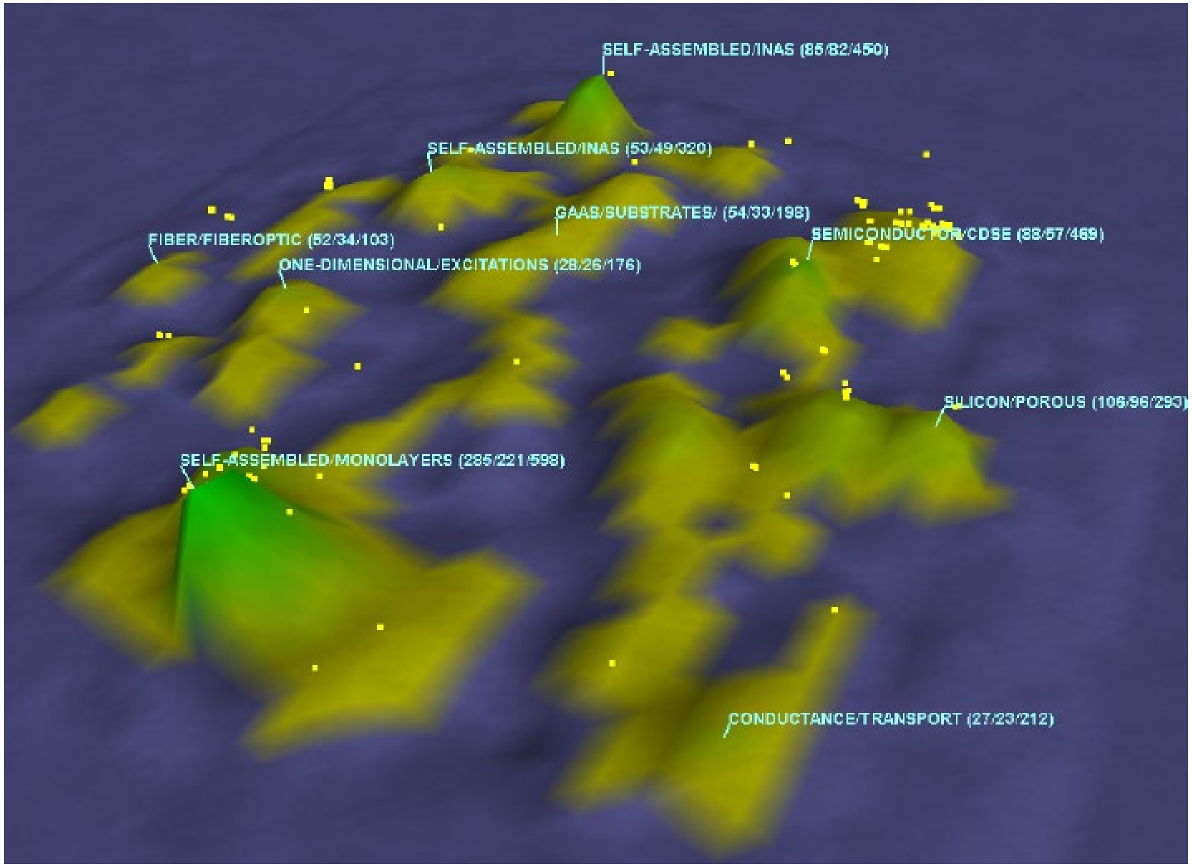
RadViz



P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. DNA visual and analytic data mining. In Proceedings of the 8th conference on Visualization '97, VIS '97, pages 437–441, Los Alamitos, CA, USA, 1997. IEEE Computer Society Press.

VxInsight

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



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

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

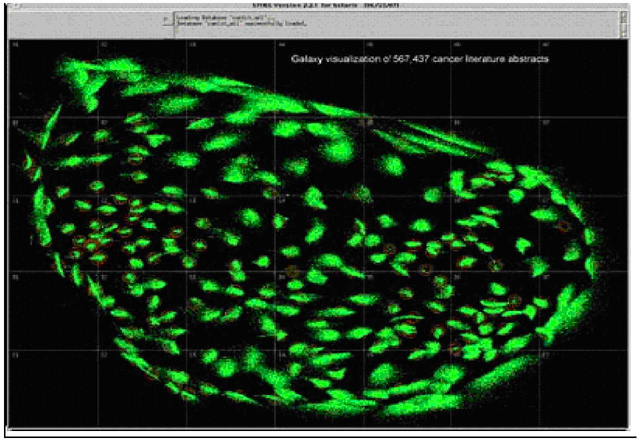
COSINE

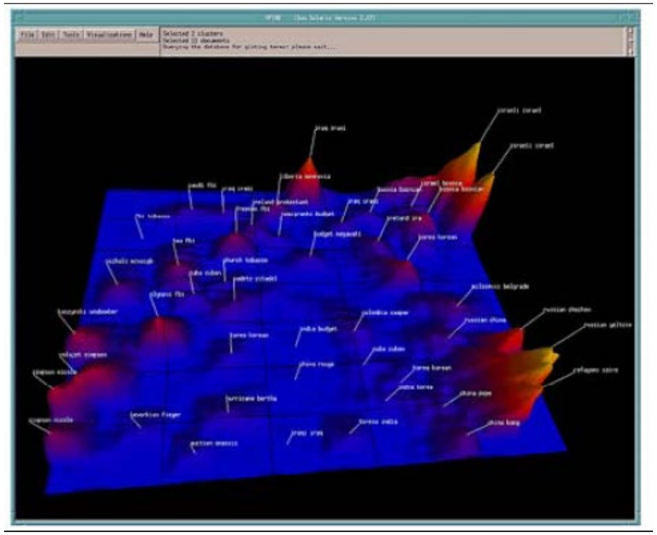
$$sim_{i,j} = \frac{(w_{i,1} \times w_{j,1}) + \ldots + (w_{i,k} \times w_{j,k})}{(w_{i,1}^2 + \ldots + w_{i,k}^2) \times (w_{j,1}^2 + \ldots + w_{j,k}^2)}$$

IN-SPIRE

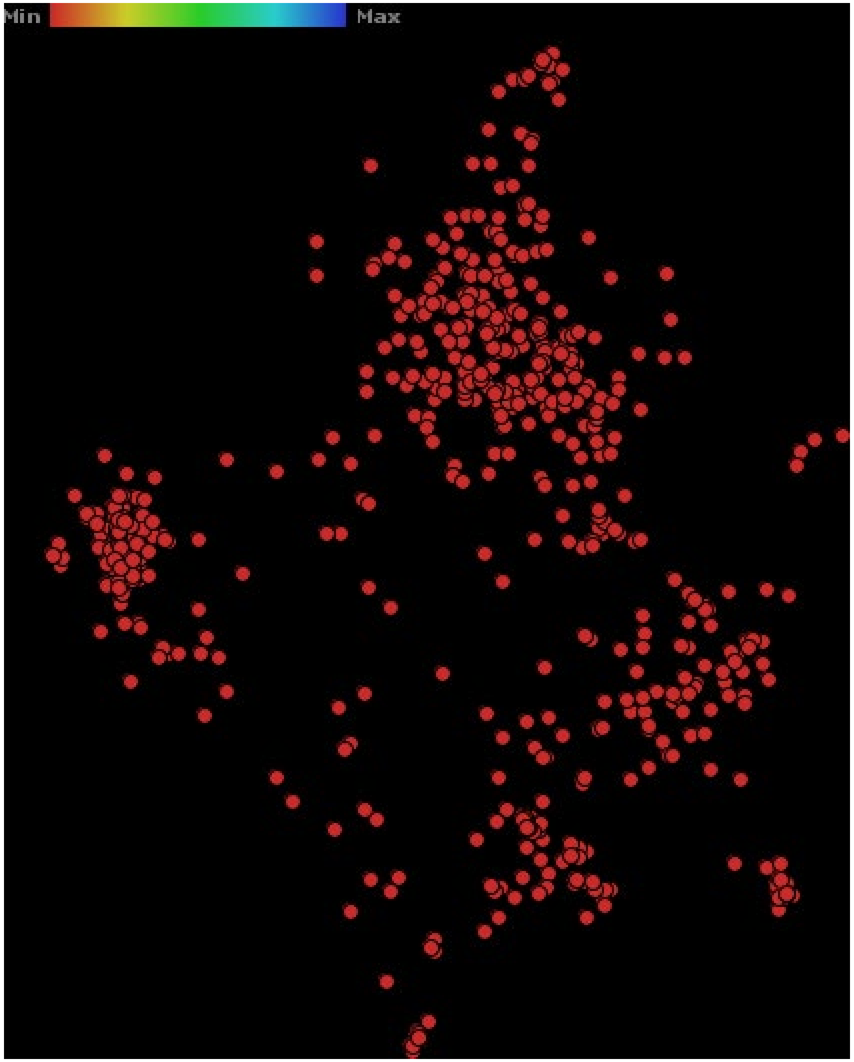
- Spatial Paradigm for Information Retrieval - Pacific Nortwest National Laboratories

- Two Visualization Metaphors:
  - Galaxies – dimensional reduction
  - Themescape



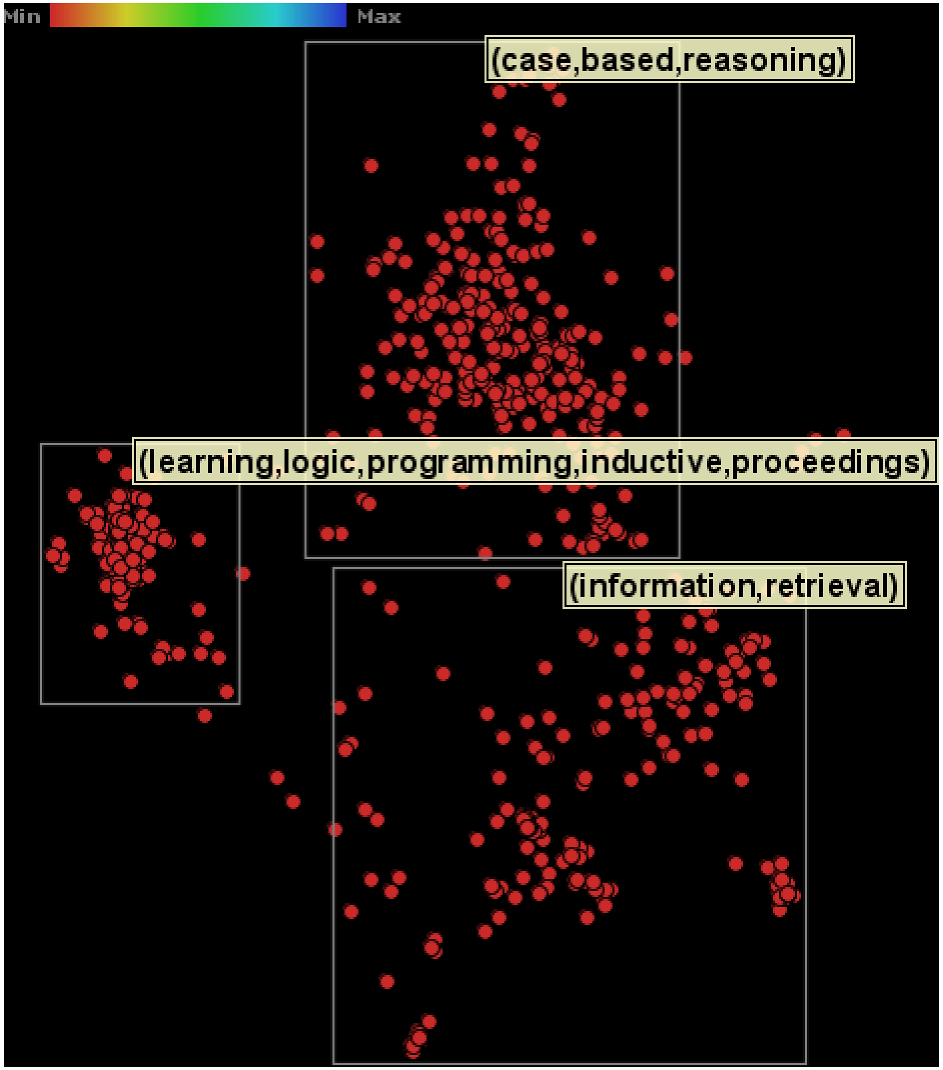


Exploring

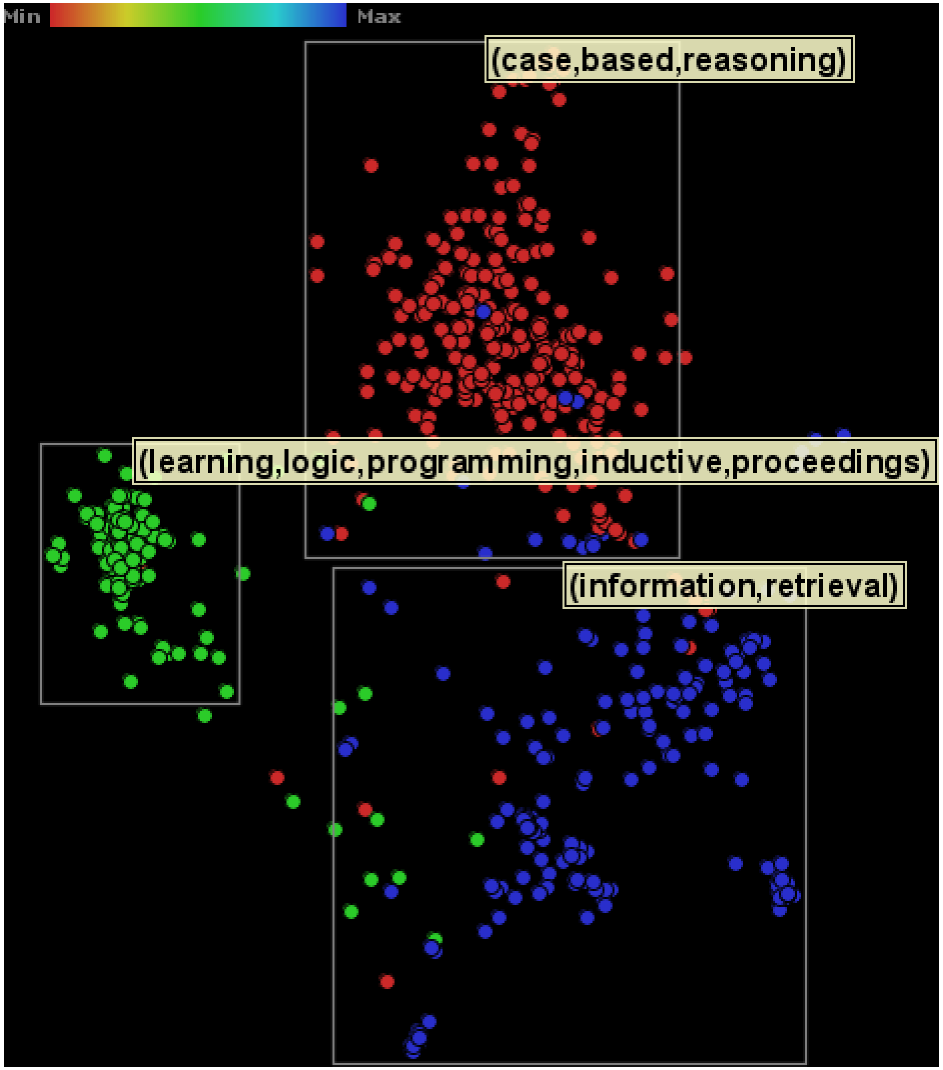


- Case-Base Reasoning
- Information Retrieval
- Inductive Logic Programming

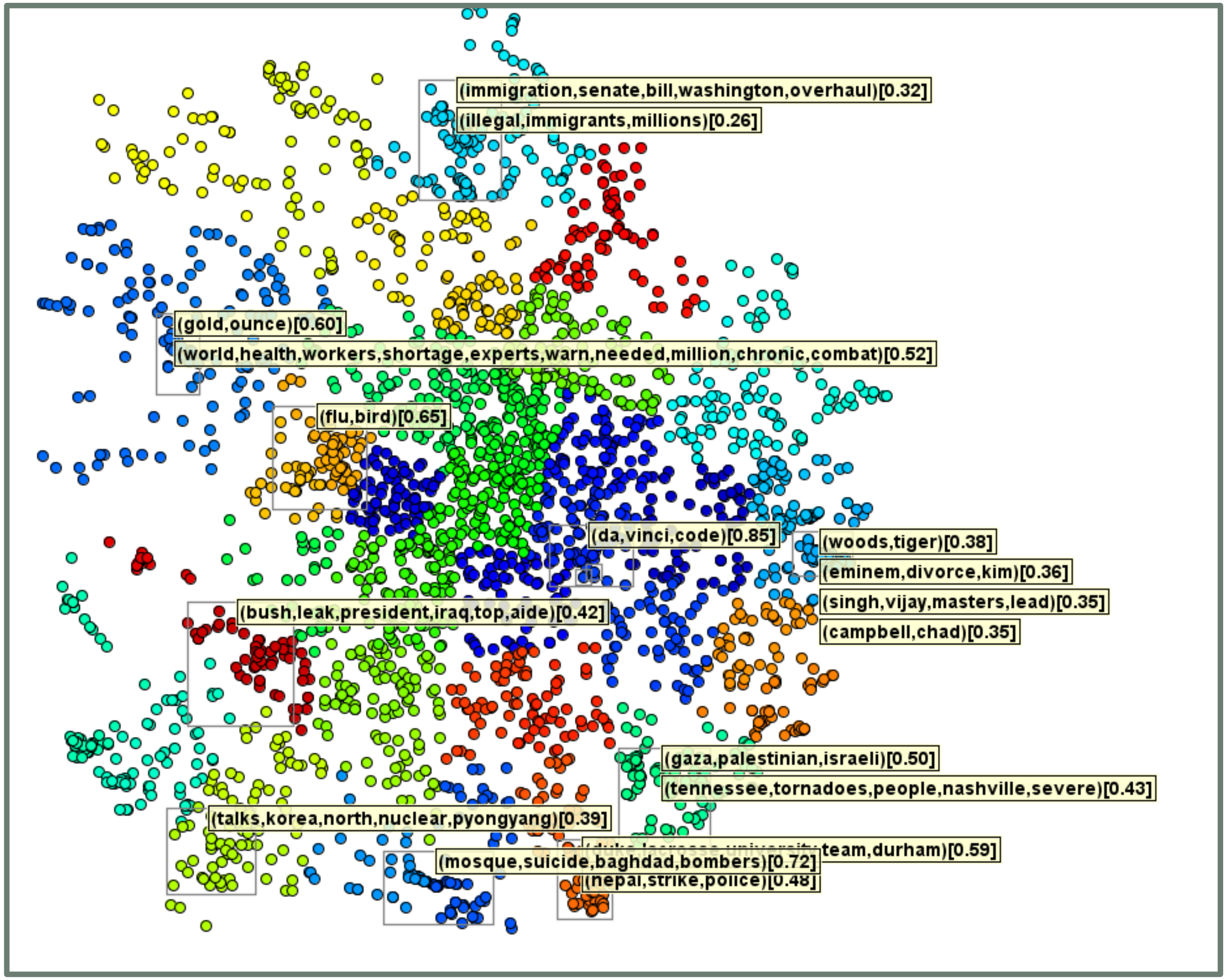
Examples


Min
Max
(case,based,reasoning)
(learning,logic,programming,inductive,proceedings)
(information,retrieval)

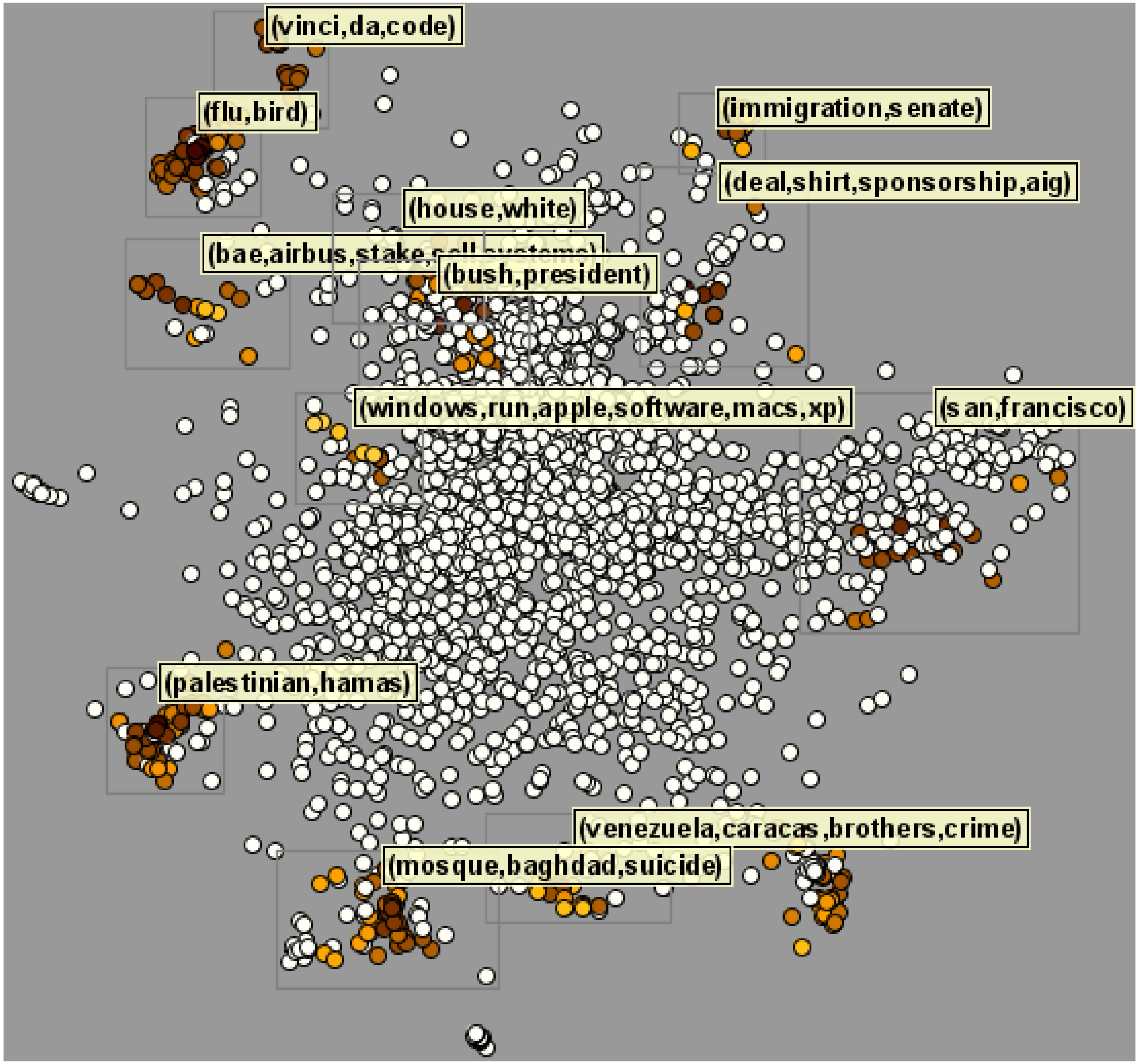
Examples


Min
Max
(case,based,reasoning)
(learning,logic,programming,inductive,proceedings)
(information,retrieval)

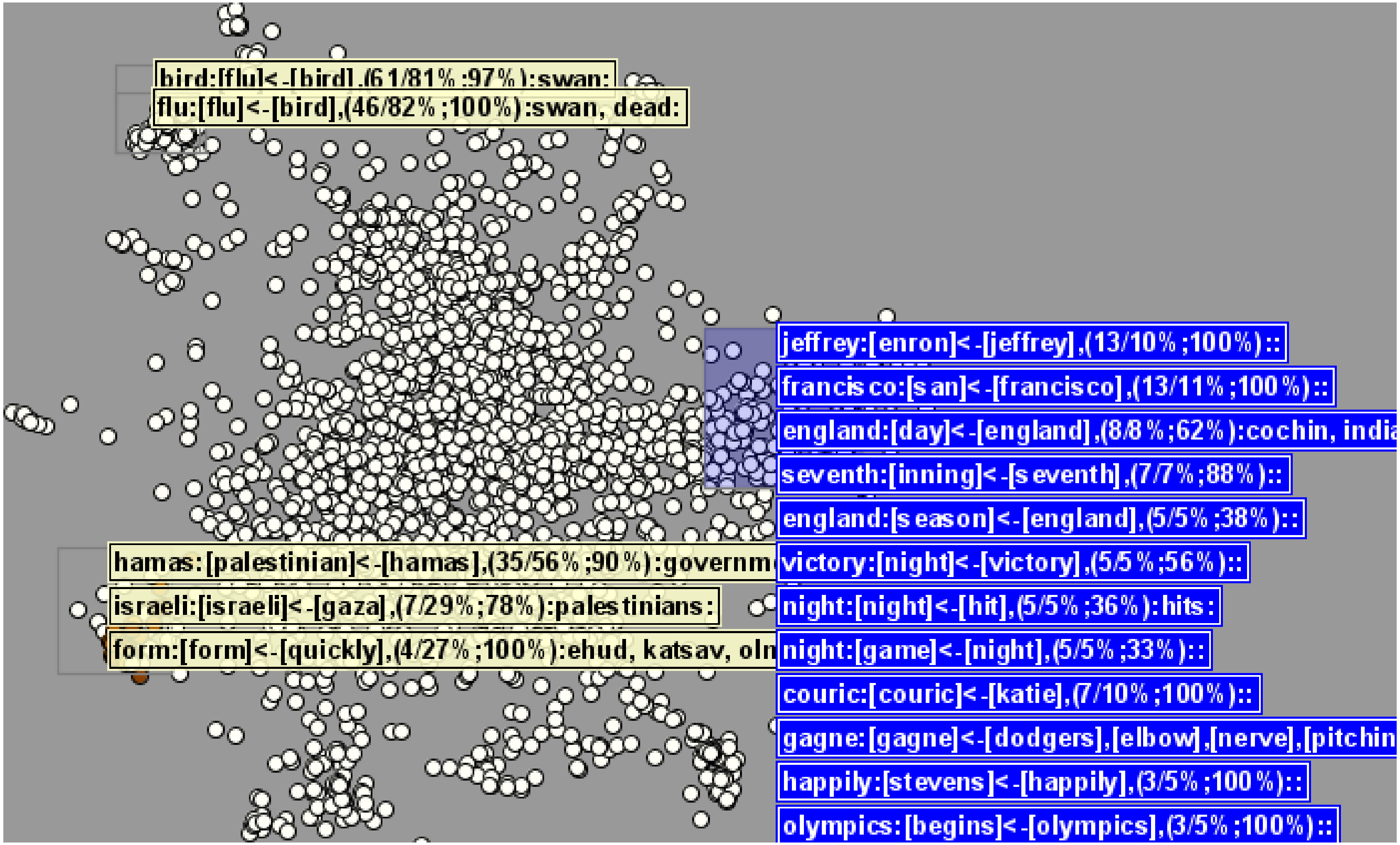
Example – LSP with text



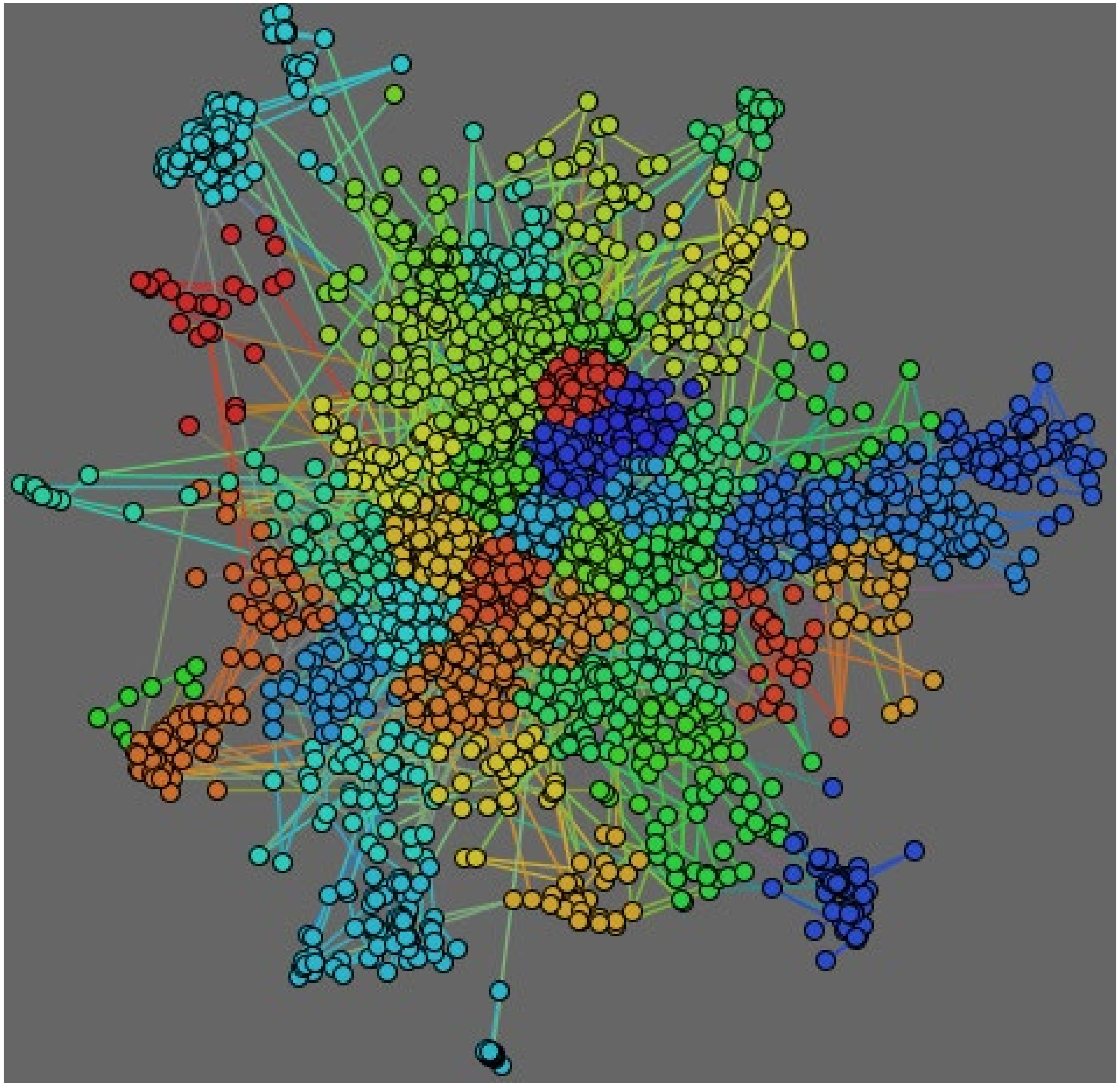
April 2006
News from
4 outlets


(vinci,da,code)
(flu,bird)
(immigration,senate)
(deal,shirt,sponsorship,aig)
(house,white)
(bush,president)
(windows,run,apple,software,macs,xp)
(san,francisco)
(palestinian,hamas)
(venezuela,caracas,brothers,crime)
(mosque,baghdad,suicide)

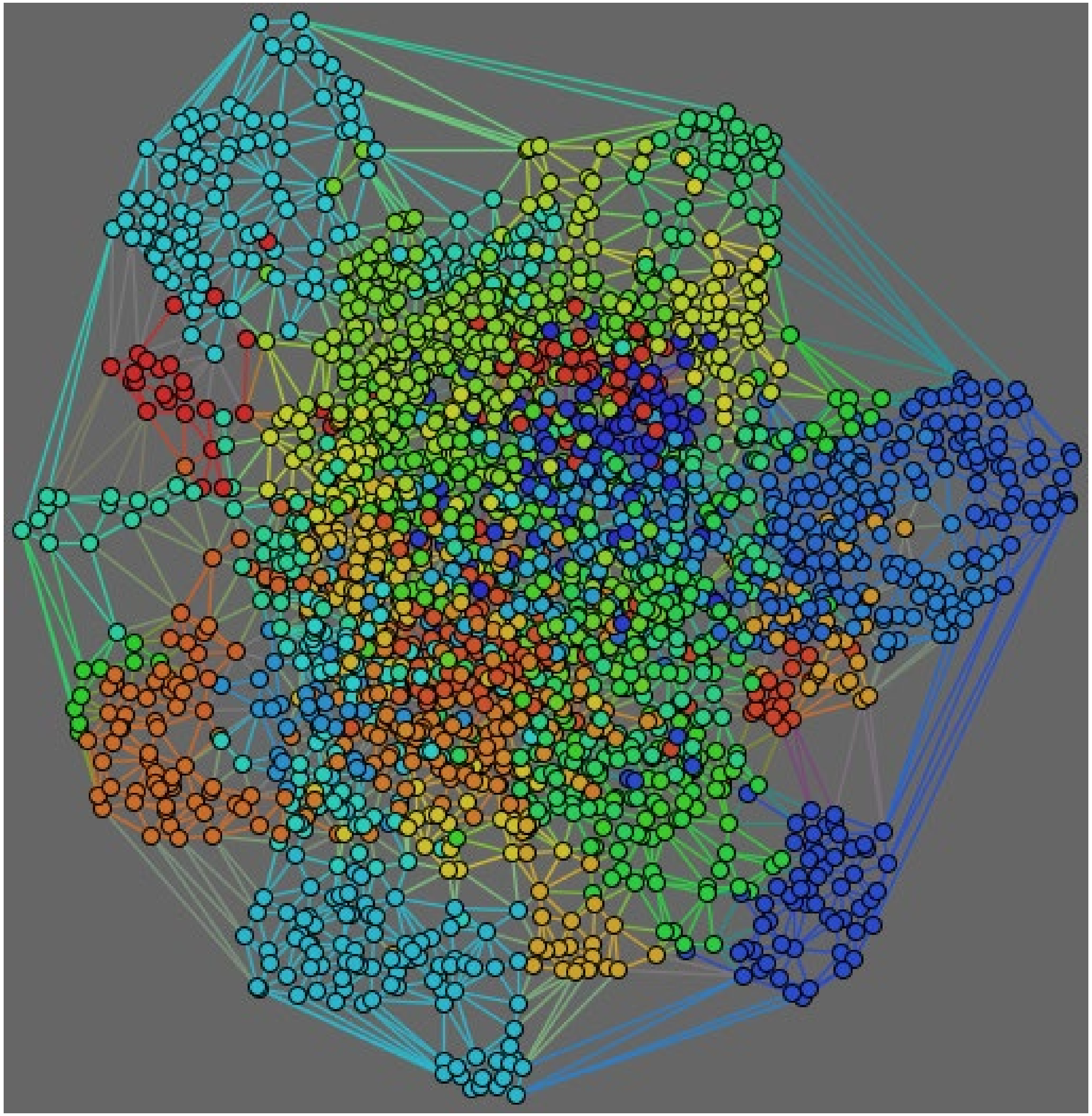
- Detailing topics


bird:[flu]<-[bird],(61/81%;97%):swan:
flu:[flu]<-[bird],(46/82%;100%):swan, dead:
hamas:[palestinian]<-[hamas],(35/56%;90%):governm
israeli:[israeli]<-[gaza],(7/29%;78%):palestinians:
form:[form]<-[quickly],(4/27%;100%):ehud, katsav, olm
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%)::
victory:[night]<-[victory],(5/5%;56%)::
night:[night]<-[hit],(5/5%;36%):hits:
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%)::

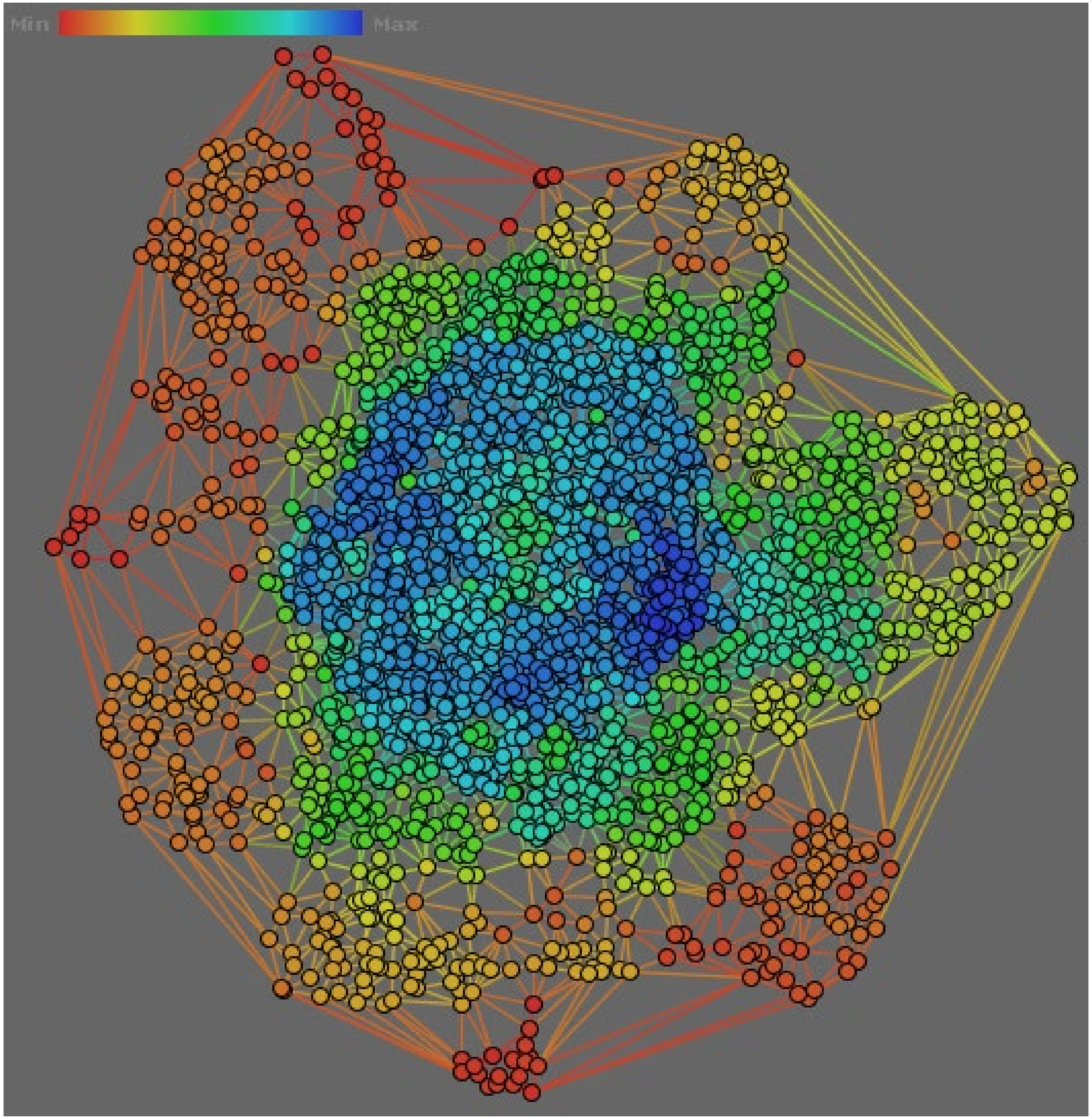
- Finding Relationships



- Building a mesh

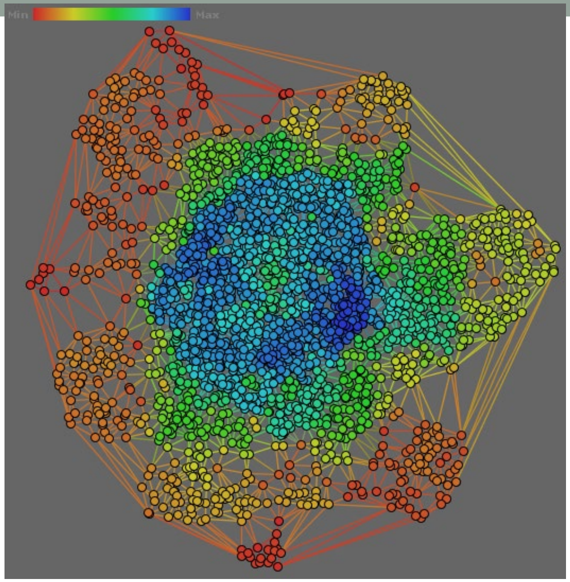


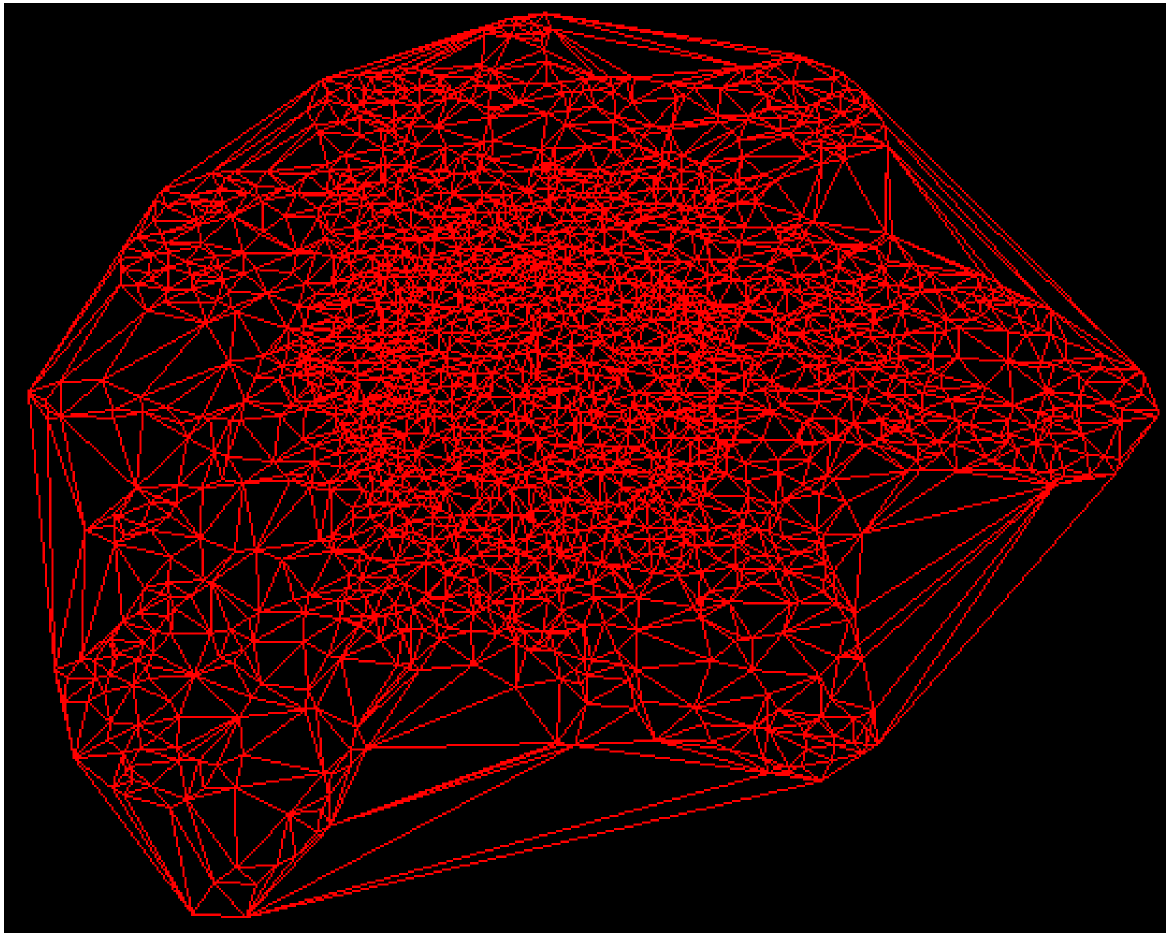
- Coloring by degree of proximity

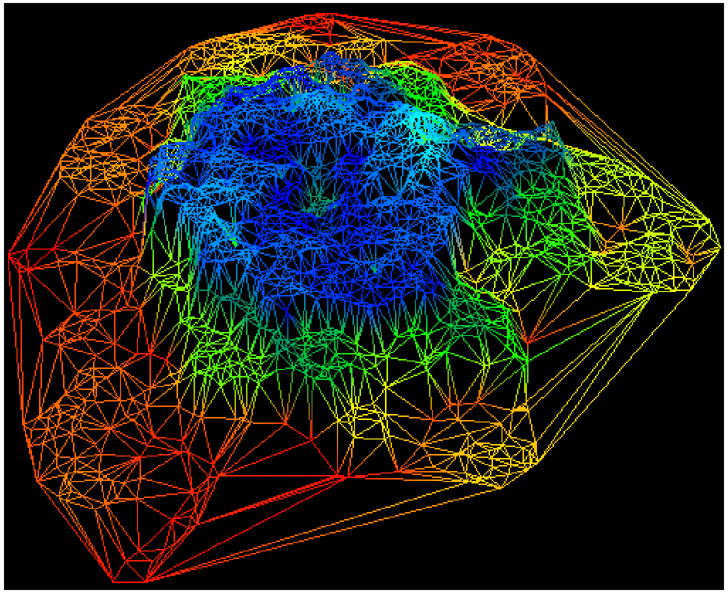

Min
Max

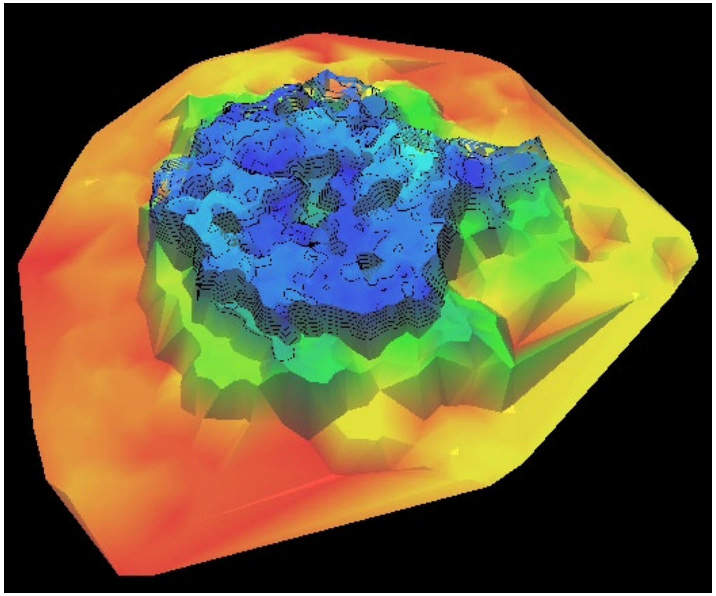
- Coordinating

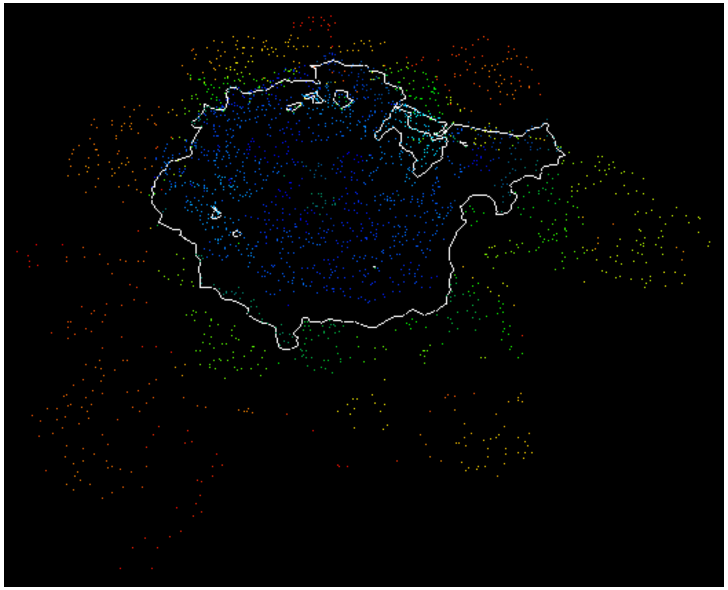
- Building a Surface


Min
Max

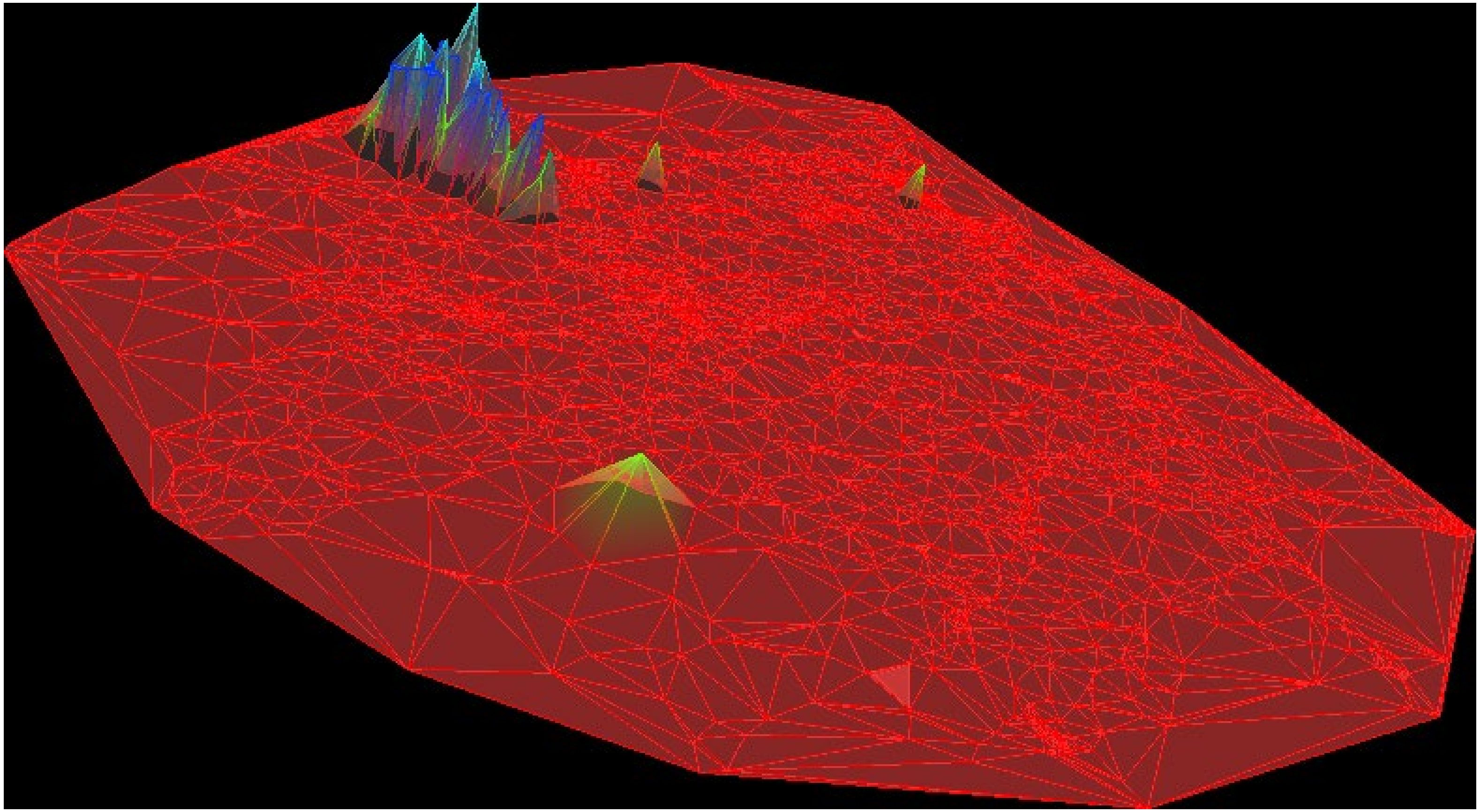




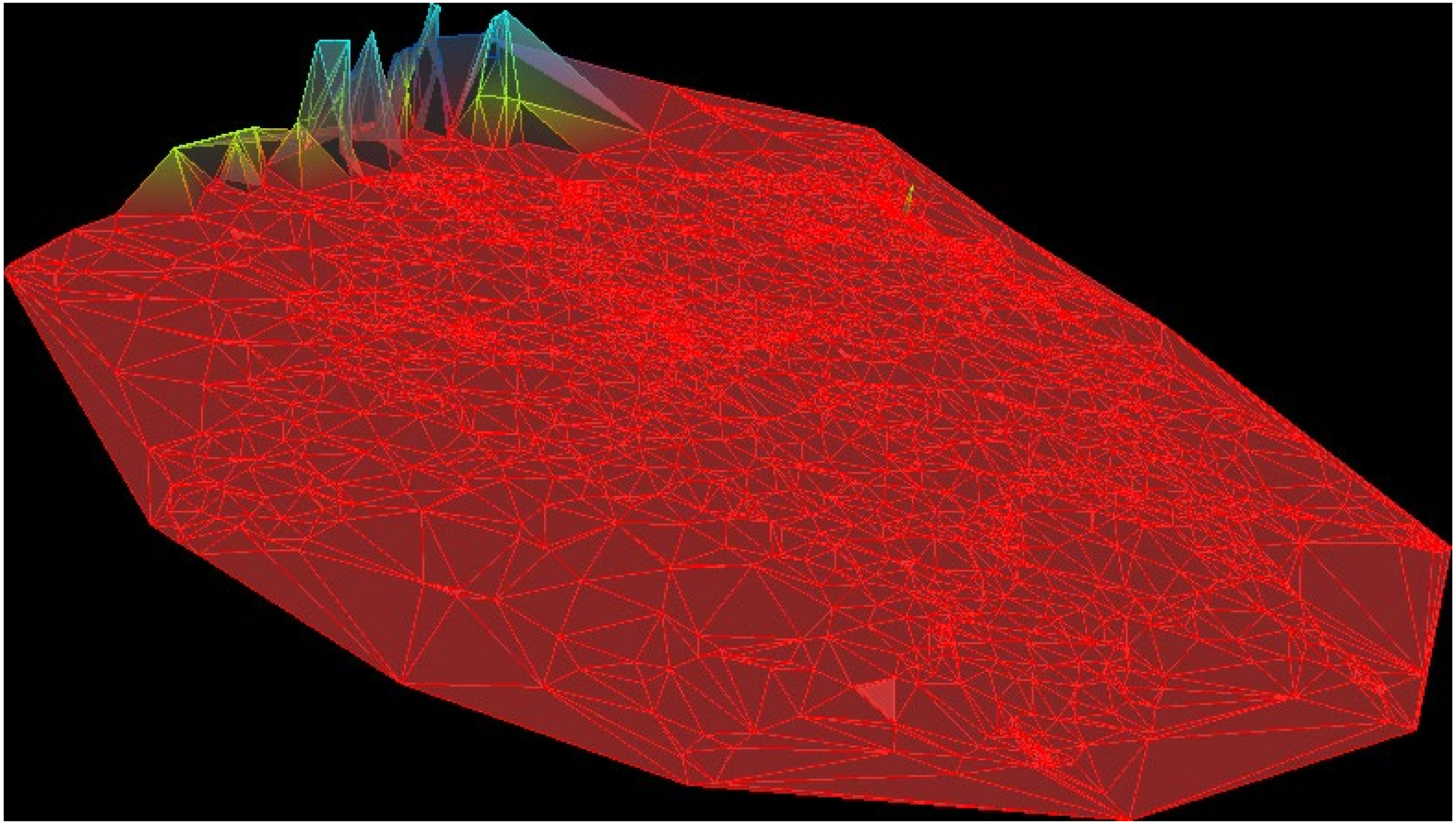




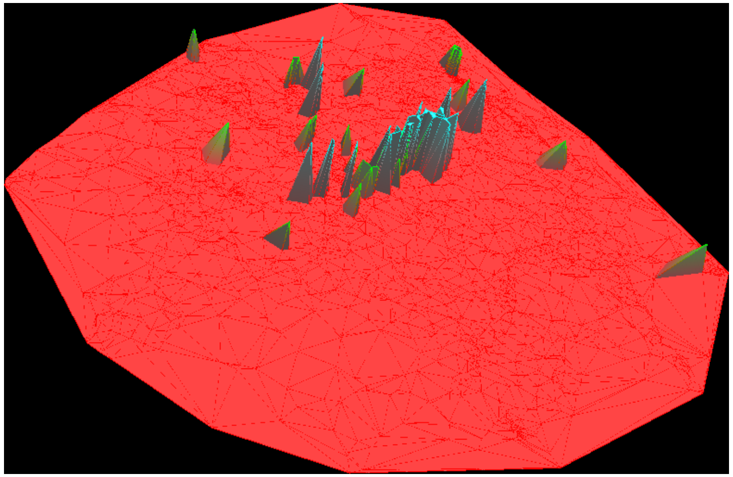
RSS News Flash



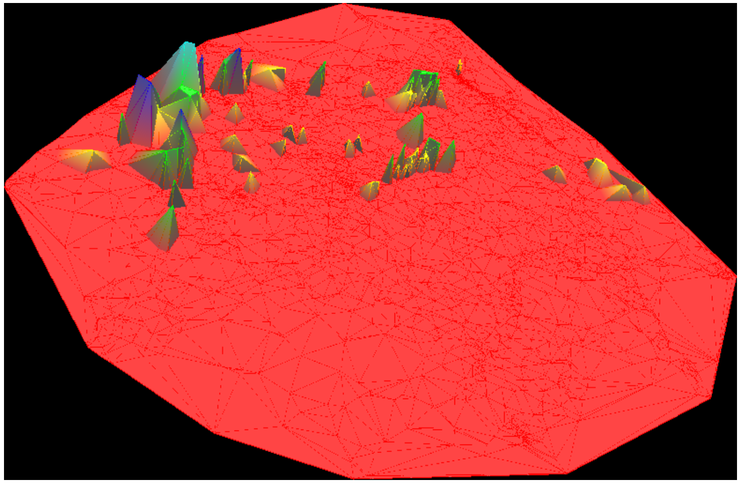
Bird and Flu



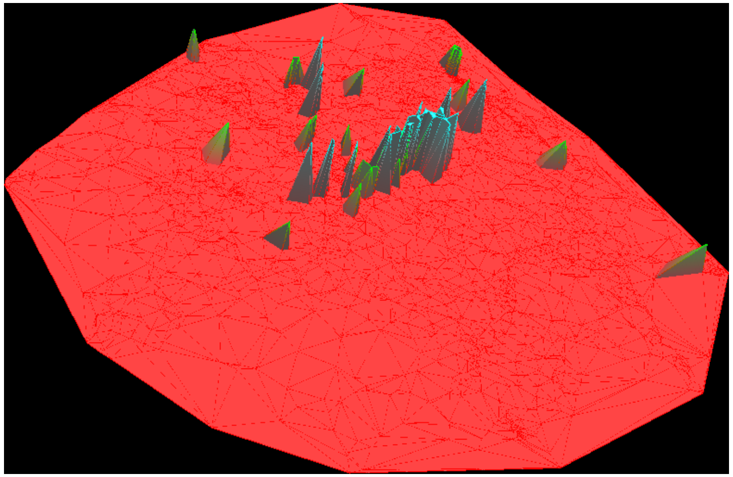
Palestinian



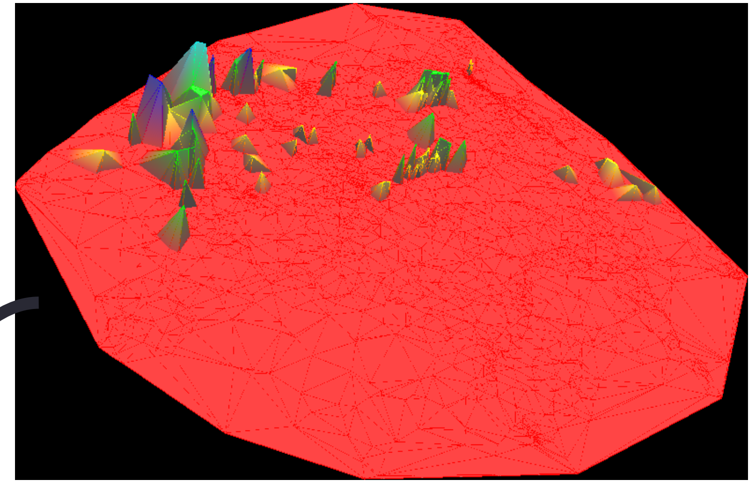
Bush



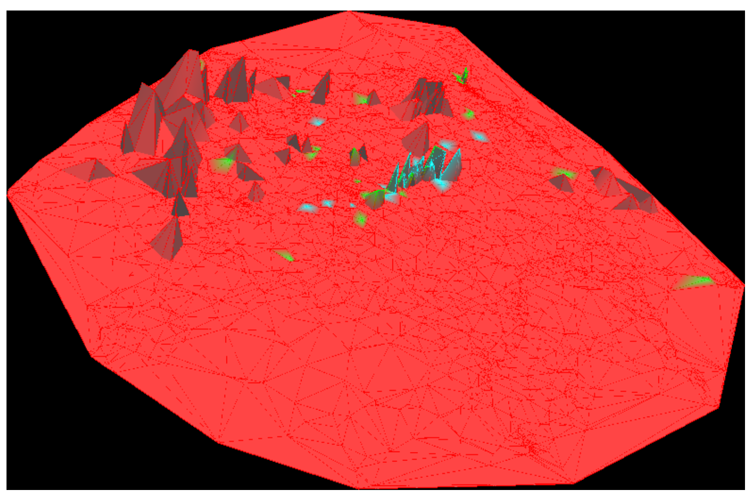
Iraq



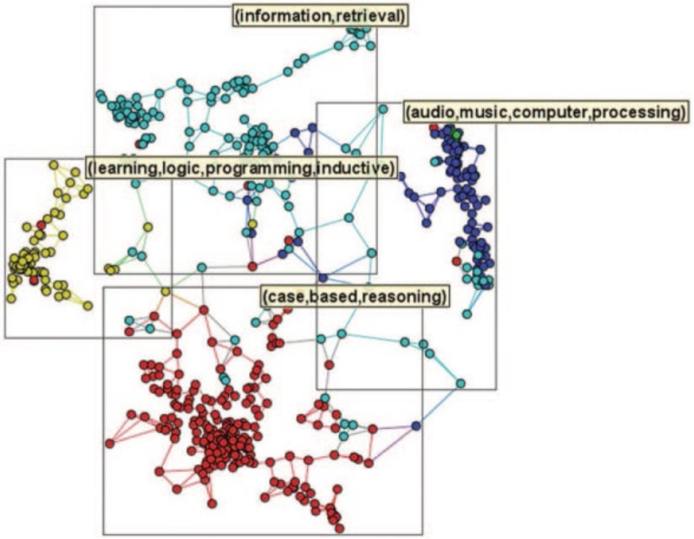
Bush



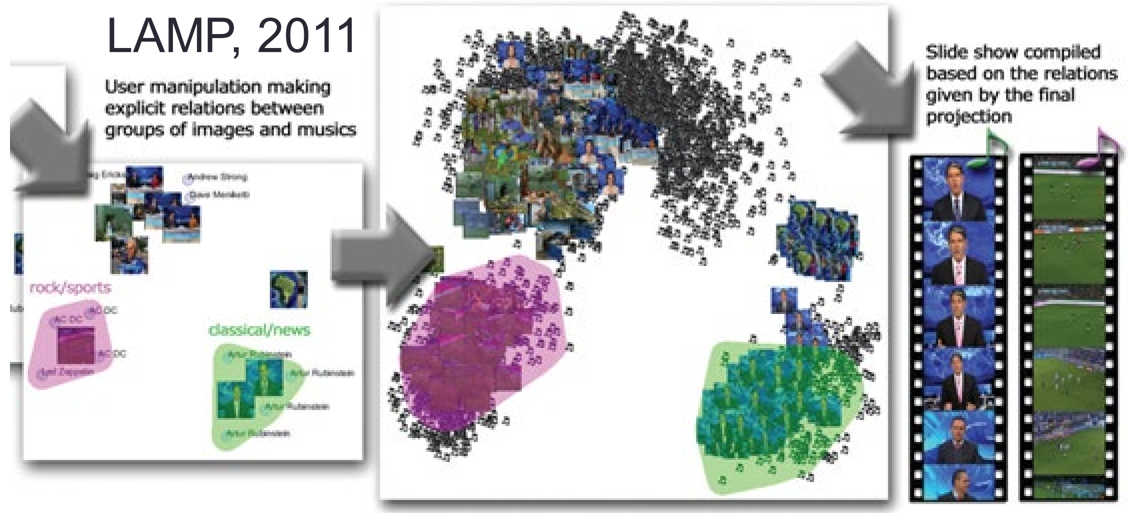
Iraq



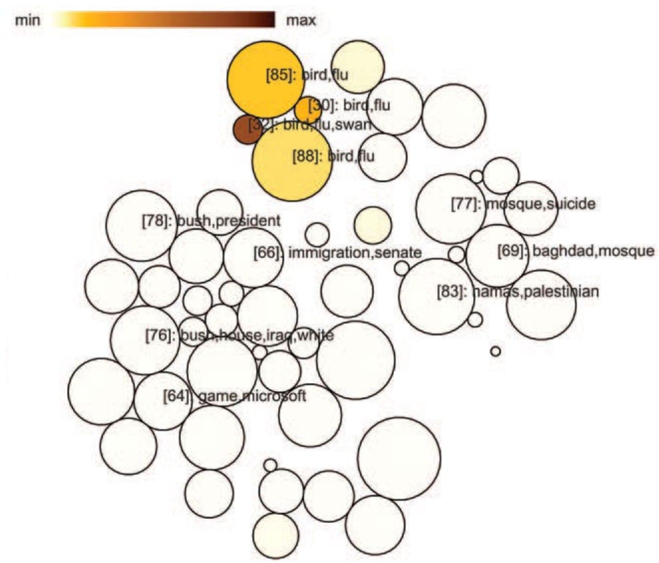
LSP, 2008



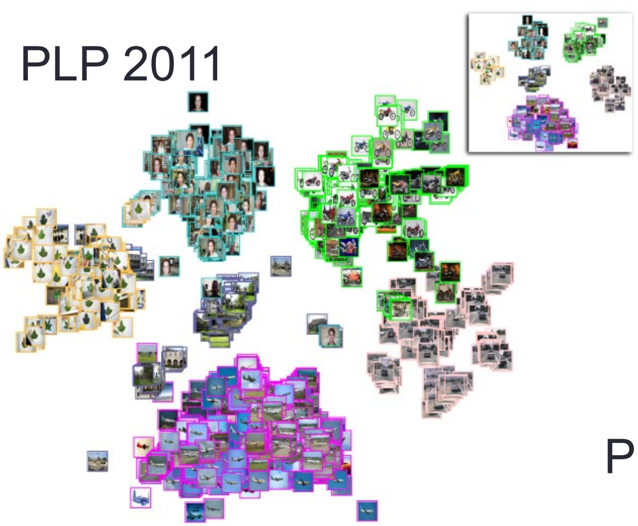
LAMP, 2011

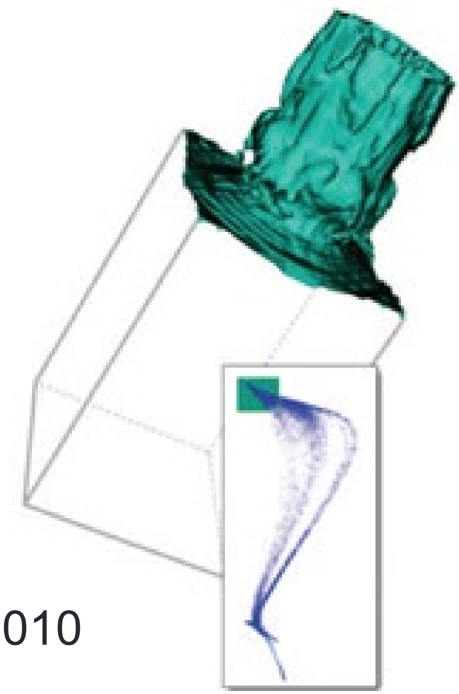


HiPP, 2008,2018



PLP 2011





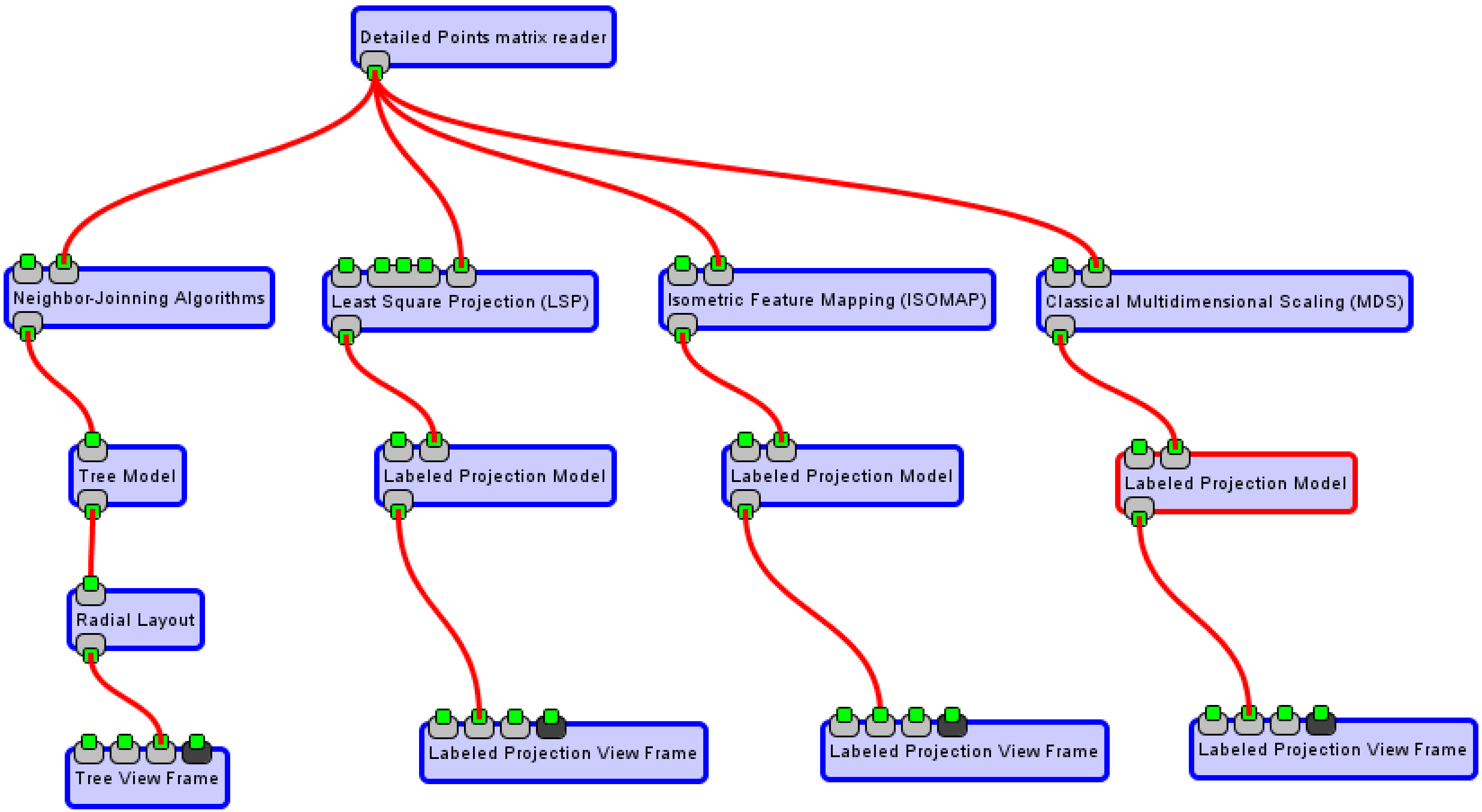
PLMP, 2010

Further Example

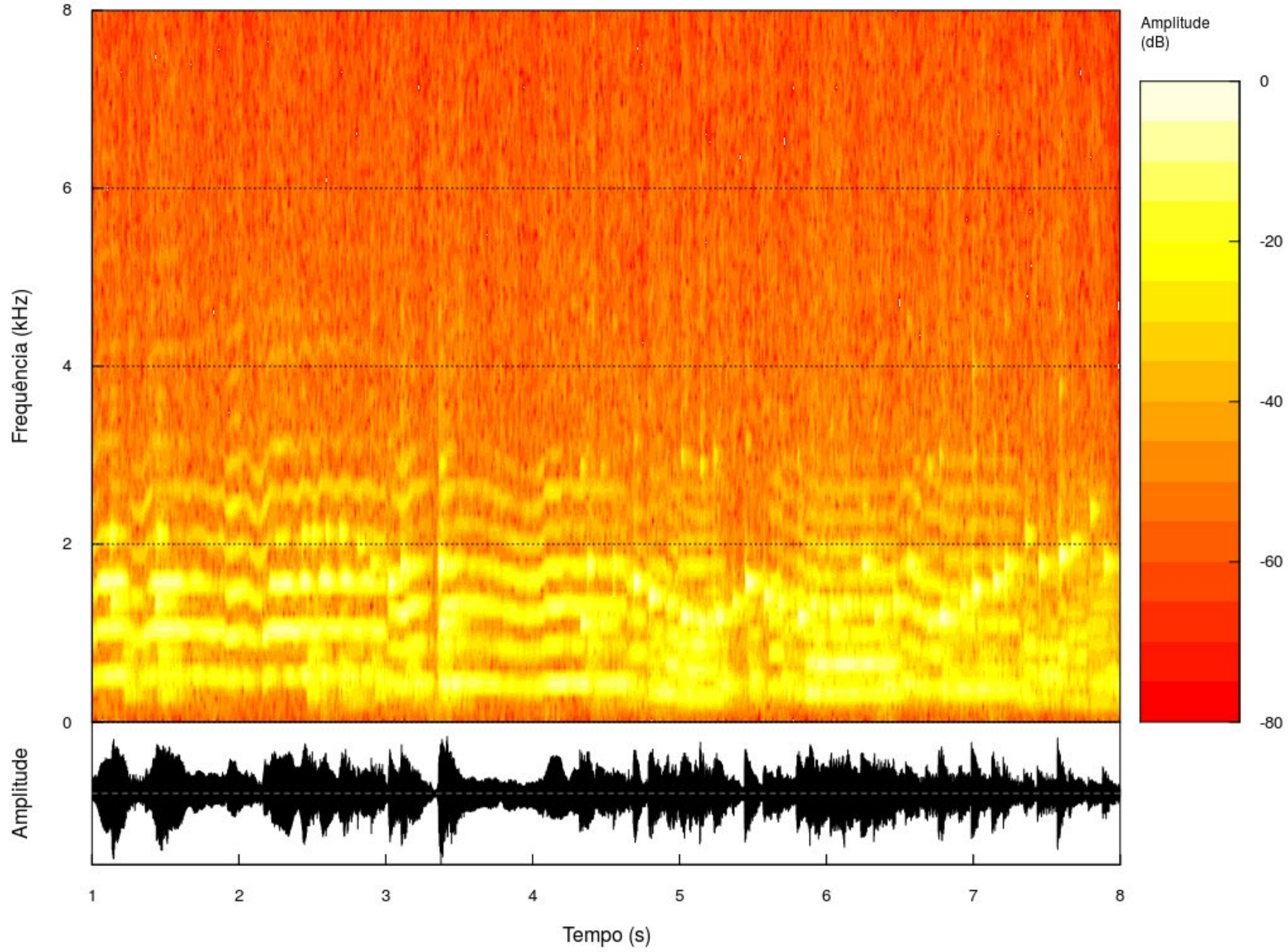
- Cattle performance data
  - Translated to text from categorical information, e.g.,
    - Ranges of weight to words such as:

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

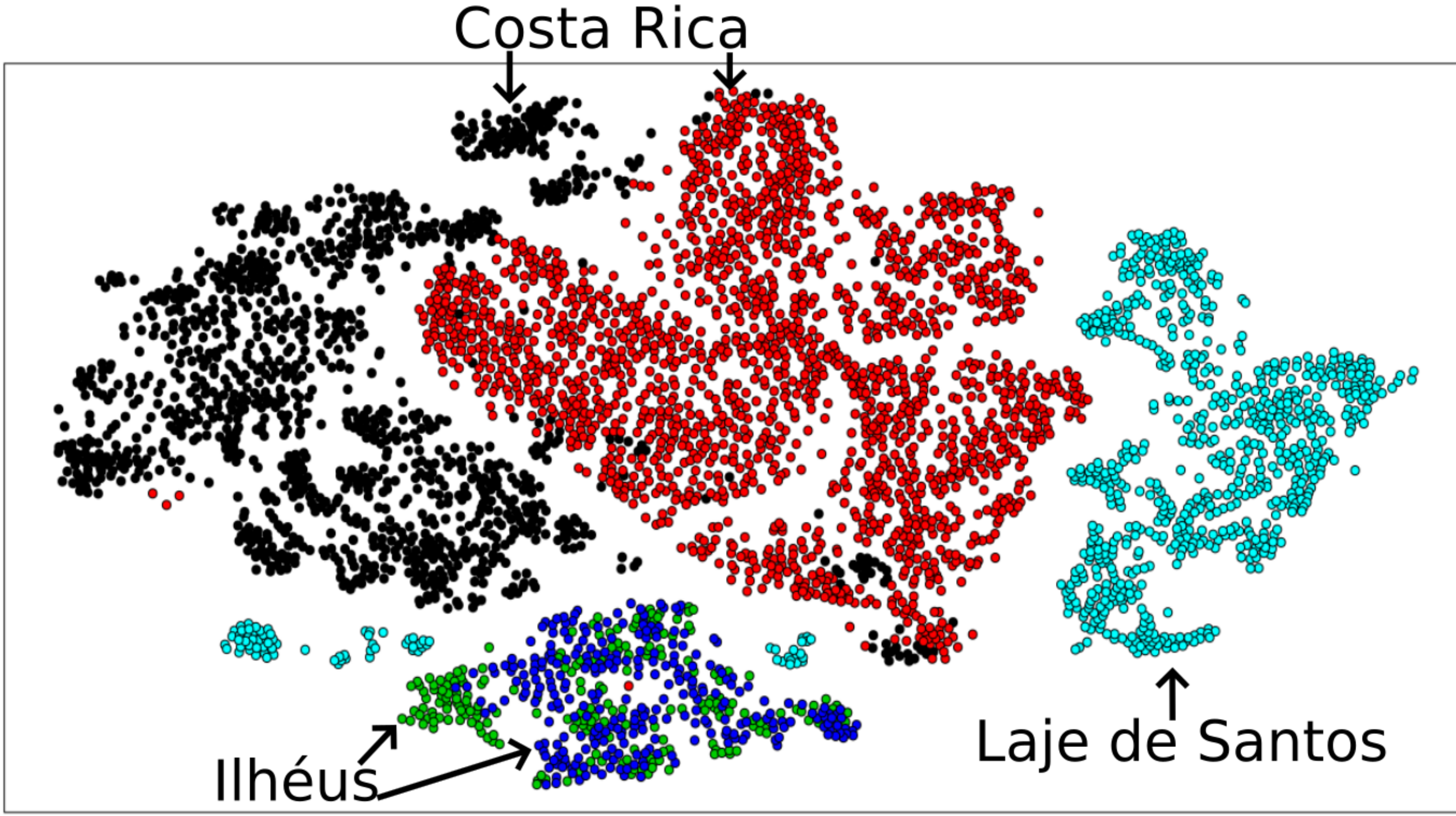
Before we continue... Vispipeline and other tools



Acoustic Landscape for Environmental Monitoring



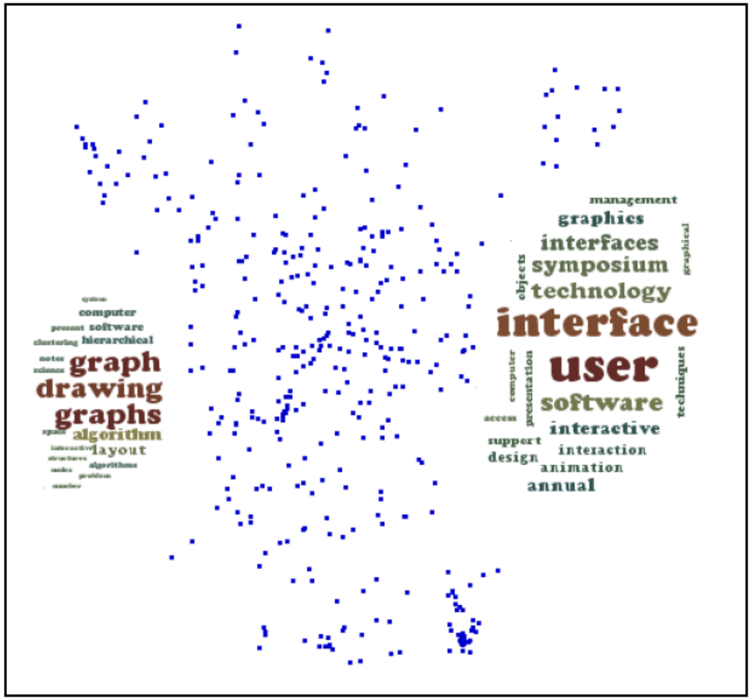
Distinction between landscapes



Dias, Pedrini, Minghim 2020

More Applications – Word clouds

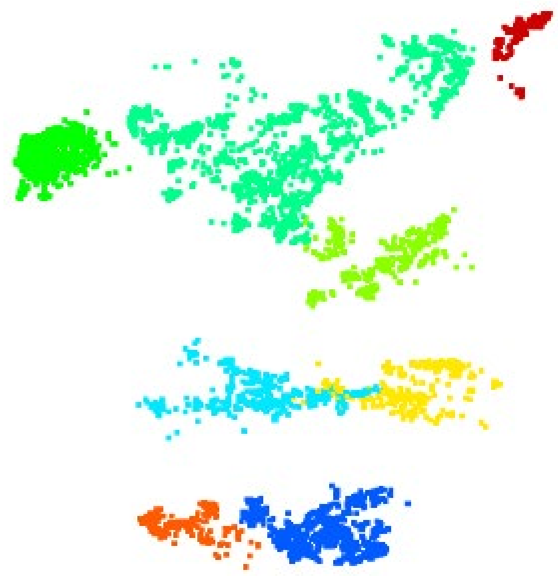
- Semantical ordering of keywords from projected points (new)
- Semantical filling of polygons over projections (new)



Paulovich, Telles, Toledo, Minghim, Nonato – Semantic Wordification of Document Collections, **Computer Graphics Forum, Eurovis 2012**

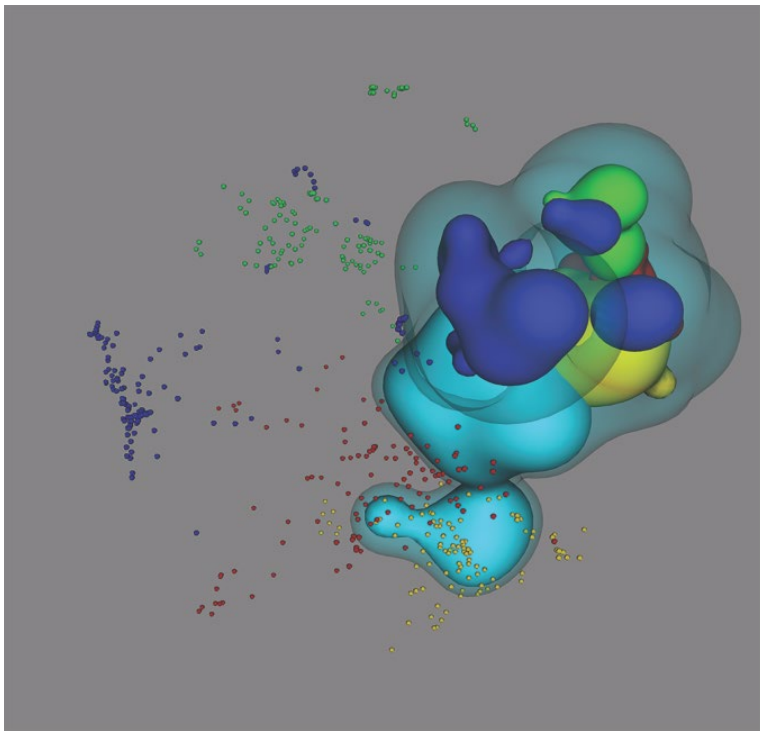
More Applications – Fiber Tracking

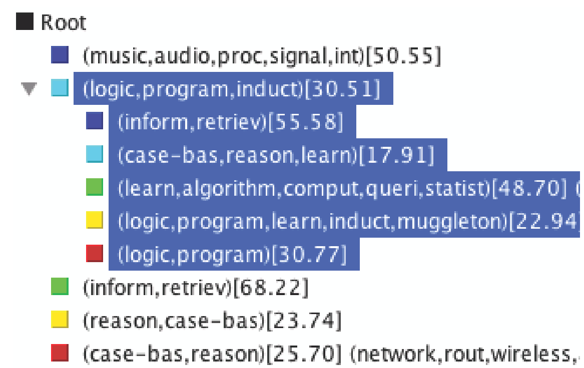
- Projection from fiber features
- Interaction through fast and reconfigurable projections (LAMP)
- Lines, Tubes and Surface Views



Poco, Eler, Paulovich, Minghim - Employing 2D projections for fast visual exploration of large fiber tracking data, **Computer Graphics Forum**, **Eurovis** 2012.

Visualization challenge: clutter



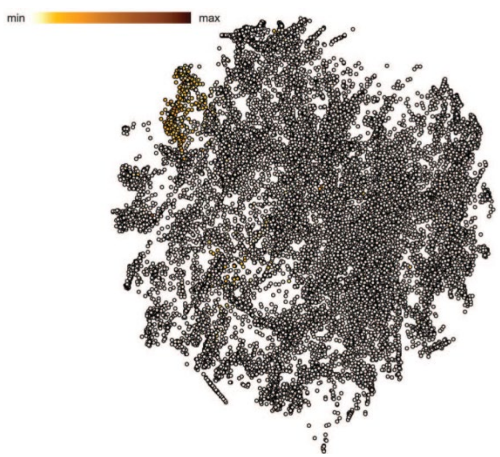


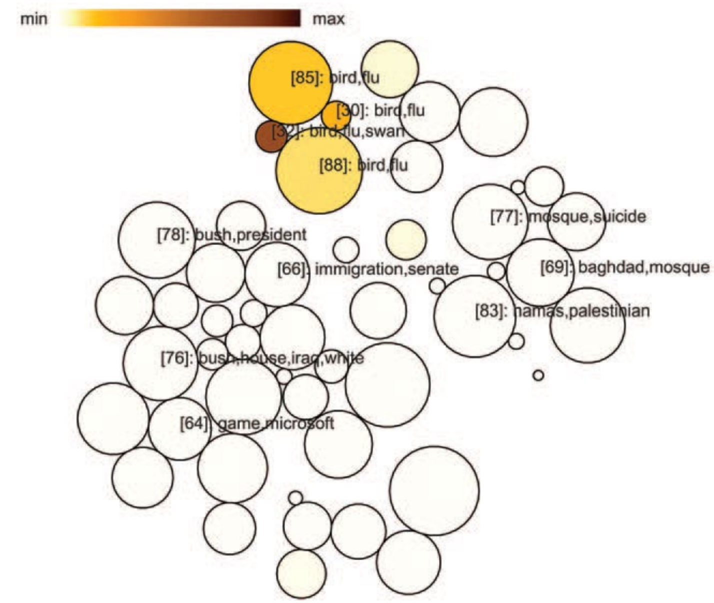
Poco; Etedmapour, Paulovich, Long, Rosenthal, Oliveira, Linsen, Minghim. A framework for exploring multidimensional data with 3D projections, ***Computer Graphics Forum***, Eurovis 2011.

112

Handling scalability? Xhipp

Partitioning and Projection





F. V. Paulovich and R. Minghim, "HiPP: A Novel Hierarchical Point Placement Strategy and its Application to the Exploration of Document Collections," in ***IEEE Transactions on Visualization and Computer Graphics***, vol. 14, no. 6, pp. 1229-1236, Nov.-Dec. 2008.

F. Dias and R. Minghim, "xHiPP: eXtended Hierarchical Point Placement Strategy",***31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI),*** Parana, Brazil, pp. 361-368, 2018, IEEE CS Press.

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

Homework

- Explore the data sets left available using:
  - Vispipeline
  - Any tool available (go fetch!!)
- 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 (see pex-manual for that)
  - Explore using both projections and trees.
  - Write and illustrate your findings in two pages.

References

- 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.

- Dias, F. F.; Pedrini, H.; Minghim, R; Soundscape segregation based on visual analysis and discriminating features, Ecological Informatics (accepted), 2020.

- Mihael Ankerst, Martin Ester, Hans-Peter Kriegel. Towards an effective cooperation of the user and the computer for classification. KDD 2000: 179-188.

- L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008.

- J.B. Tenenbaum, V. de Silva, and J.C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, vol. 290, no. 5500, pp. 2319-2323, Dec. 2000.

- L. G. Nonato and M. Aupetit, Multidimensional Projection for Visual Analytics: Linking Techniques with Distortions, Tasks, and Layout Enrichment, *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 8, pp. 2650-2673, 1 Aug. 2019, doi: 10.1109/TVCG.2018.2846735.