JIM JOURNÉES D'INFORMATIQUE MUSICALE 2017

STRUCTURING MUSIC BY MEANS OF AUDIO CLUSTERING AND GRAPH SEARCH ALGORITHMS

Frédéric Le Bel - http://repmus.ircam.fr/lebel - 2017.

ircam Centre Pompidou
UNIVERSITÉ PARIS 8
Fonds de recherche Société et culture Québec

PREMISE [SCHOENBERG 1911]

- **If** it is possible to create **patterns** out of tone colors that are differentiated **according to pitch** (melodies),

- **then** it must also be possible to create **patterns** out of tone colors that are differentiated **according to tone colors** (sound-color melodies)

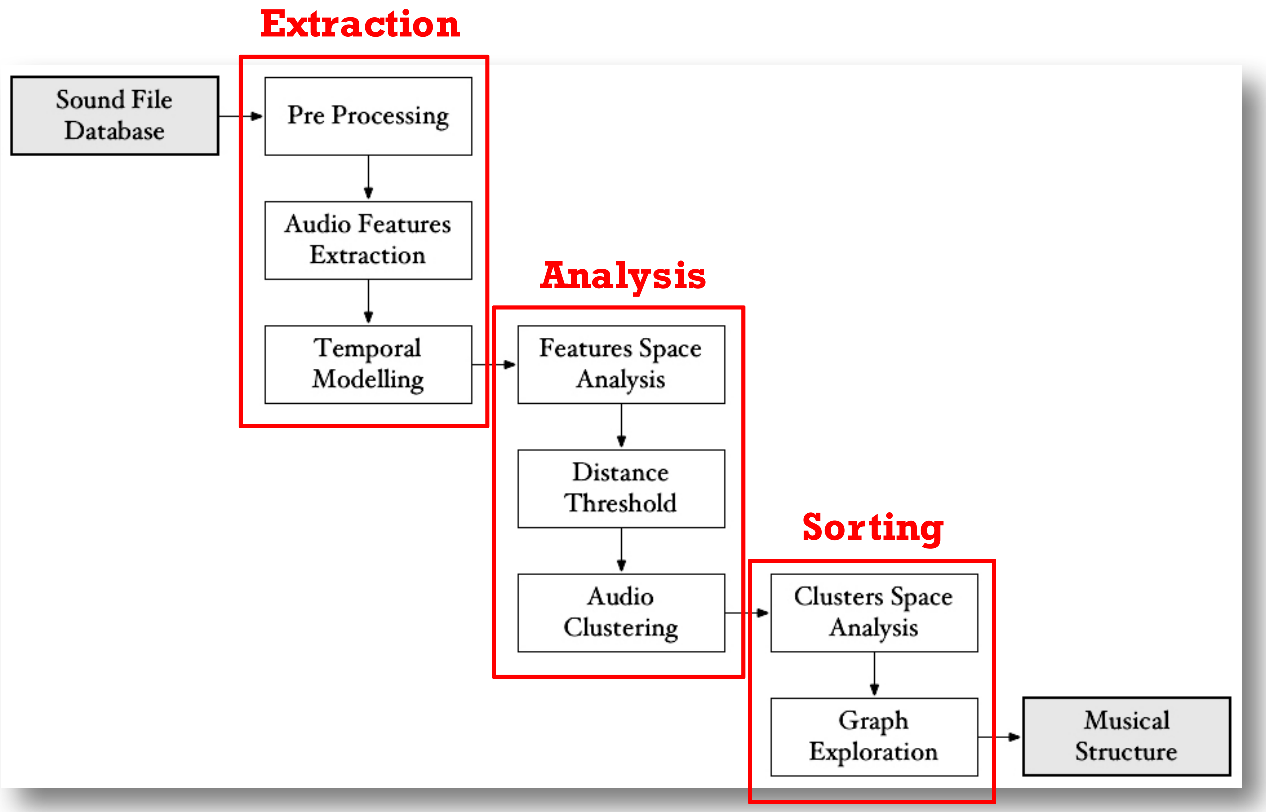
- whose relations with one another work with a kind of **logic entirely equivalent to** that logic which satisfies us in **the melody of pitches**.

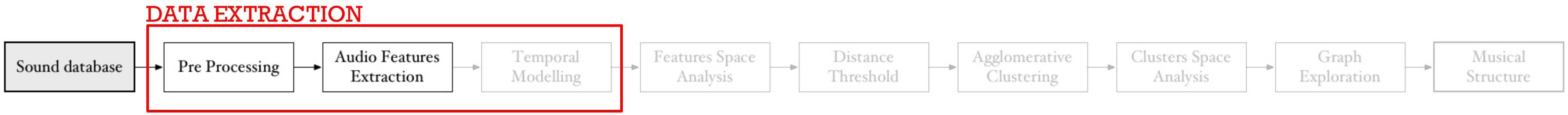
THE IDEA…

- Develop a **CAC framework,**
  - Not to generate but to **analyse**, to **explore** and to **understand** a given sound material,
  - An **analytical** rather than a generative **approach to composition**,
  - **Deduce** rather than induce **(musical) structure**,
  - Towards formalizing music [Assayag 1985].

- Based on the principle that **sounds are form-bearing elements** [McAdams 1985]
  - The framework proposes to:
    - exploit various **dimensions of timbre** (audio-features),
    - study <u>some</u> **relationships between sounds** (features-space),
    - discover **latent (musical) structures** (clusters-space).

STRUCTURAL OVERVIEW (METHODOLOGY)





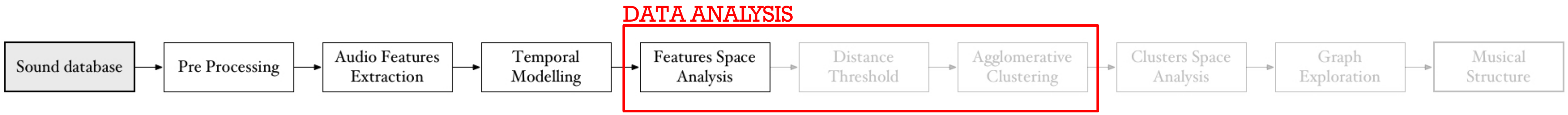
LOW-LEVEL DESCRIPTORS +

Physical model	Perceptual model
Instantaneous TEMPORAL descriptors	
• *Log attack time* • *Temporal increase* • *Temporal decrease* • *Amplitude modulation* • *Frequency modulation* • *Effective duration* • *Temporal centroid* • **Energy envelope** • **Auto-correlation** • **Zero crossing rate**	
Instantaneous ENERGY descriptors	
	• **Loudness** • **Loudness spread** • **Relative Specific Loudness**

out of scope…

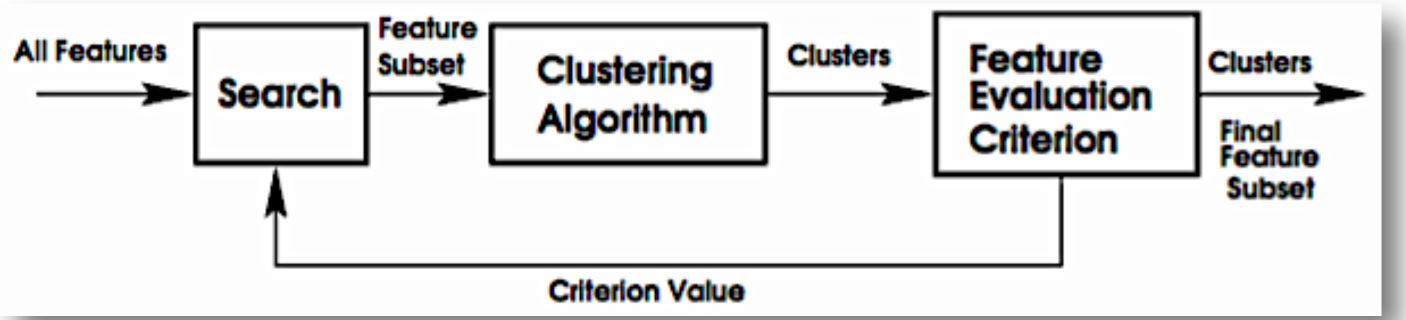
Physical model	Perceptual model
Instantaneous SPECTRAL descriptors	
	• **MFCC** • **Spectral centroid** • **Spectral spread** • **Spectral skewness** • **Spectral kurtosis** • **Spectral decrease** • **Spectral roll-off** • **Spectral variation** • **Spectral deviation** • **Spectral flatness** • **Spectral crest**
Instantaneous HARMONIC descriptors	
• **F0** • **Inharmonicity** • **Noisiness** • **Chroma** • CS-analysis • Partial tracking • Peak analysis • Masking effects	• **Spectral deviation** • **Odd to even ratio** • **Tristimulus**

higher-level…

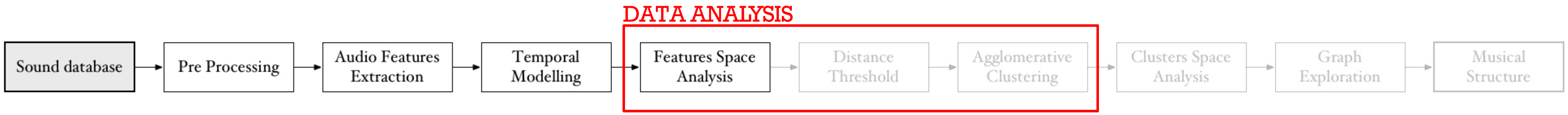


AUTOMATED FEATURES SELECTION

- **Why not use all the features we have?**
  - Some features may be **redundant**, **irrelevant** or even **misguiding**.
  - Reducing the number of features **increases interpretability**.
- **Feature evaluation criterion**
  - scatter separability = **minimize intra-cluster distances** * **maximize inter-cluster distances**

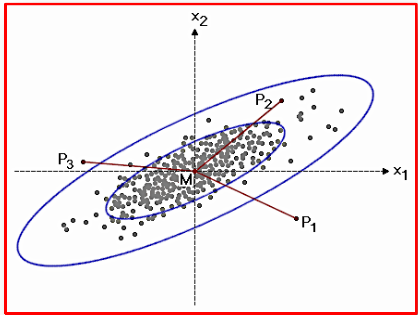


[Dy and Brodley 2004]

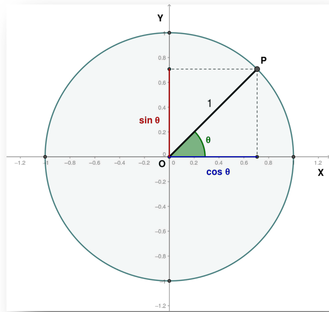


LEVELS OF SIMILARITY

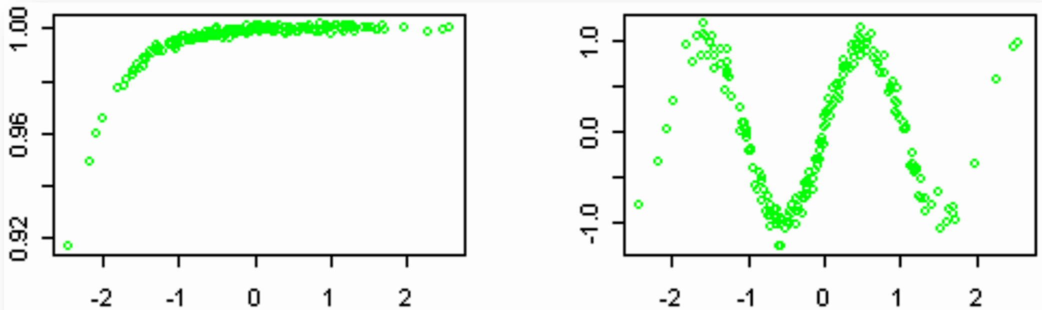
- **Distances** (magnitude)
  - Mahalanobis distance
  - *Euclidean distance*
- **Similarity** (orientation)
  - |Cosine similarity|
  - *Jaccard index*
- **Correlation** (dependency)
  - |Spearman coefficient|
  - *|Pearson coefficient|*



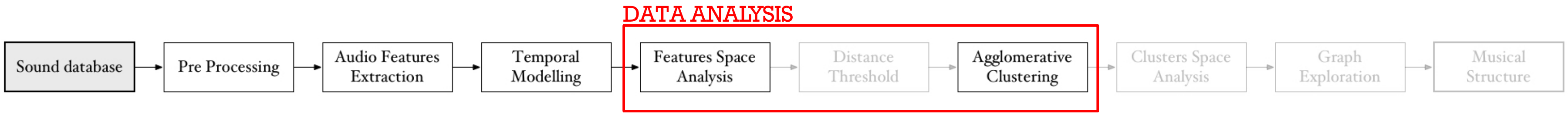
Mahalanobis [Mahalanobis 1936]



Cosine [Singhal 2001]

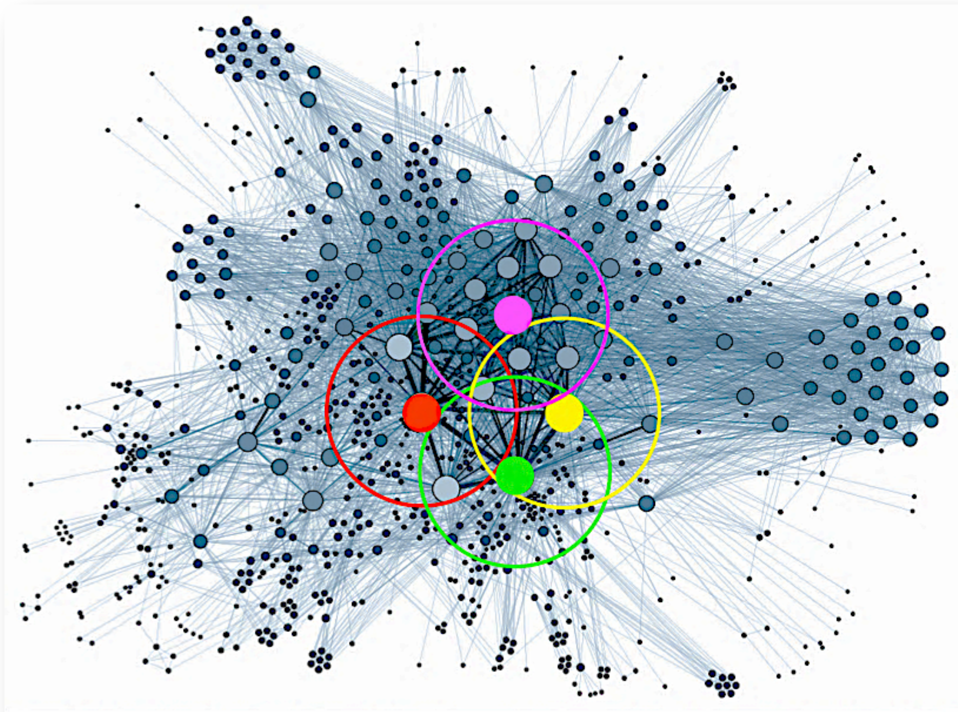


Spearman [Rakotomalala 2015]



FUZZY CLUSTERING (CENTROID-BASED)

- **Threshold constraint** (C_t)
  - Threshold = max distance (diametrical)
  - Kind of perceptual witness…
- If $d(x, y) \leq C_t$ = *true, then agglomerate,*
- else, *create new cluster.*
- Complexity = O(n!) = brute force
  - Speed is traded for accuracy...
- **Particularities**
  1. isomorphic clusters,
  2. overlapping clusters.



DATA ANALYSIS

Sound database → Pre Processing → Audio Features Extraction → Temporal Modelling → Features Space Analysis → Distance Threshold → Agglomerative Clustering → Clusters Space Analysis → Graph Exploration → Musical Structure

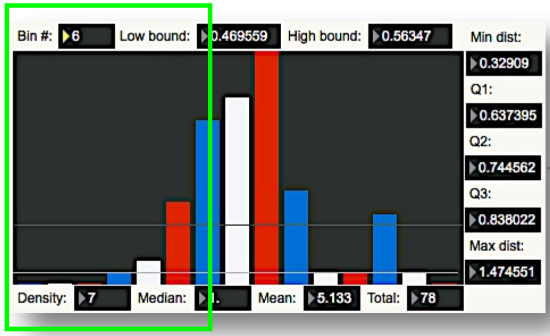
DISTANCE THRESHOLD

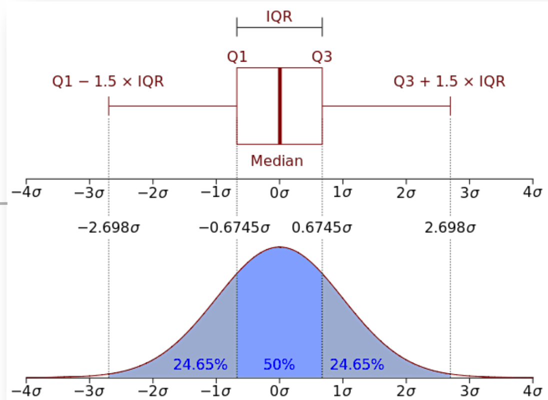
- **Histogram** (sparsity of x)
  - Freedman-Dicaonis rule [Freedman 1981]
  - **Adaptive bin width** (h)
  - For number of bins (k)

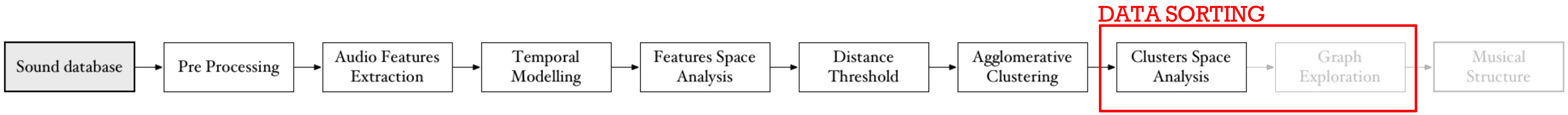
$$k = \left\lceil \frac{\max x - \min x}{h} \right\rceil$$

$$k = \left\lceil \frac{\max x - \min x}{2\frac{\mathrm{IQR}(x)}{\sqrt[3]{n}}} \right\rceil$$

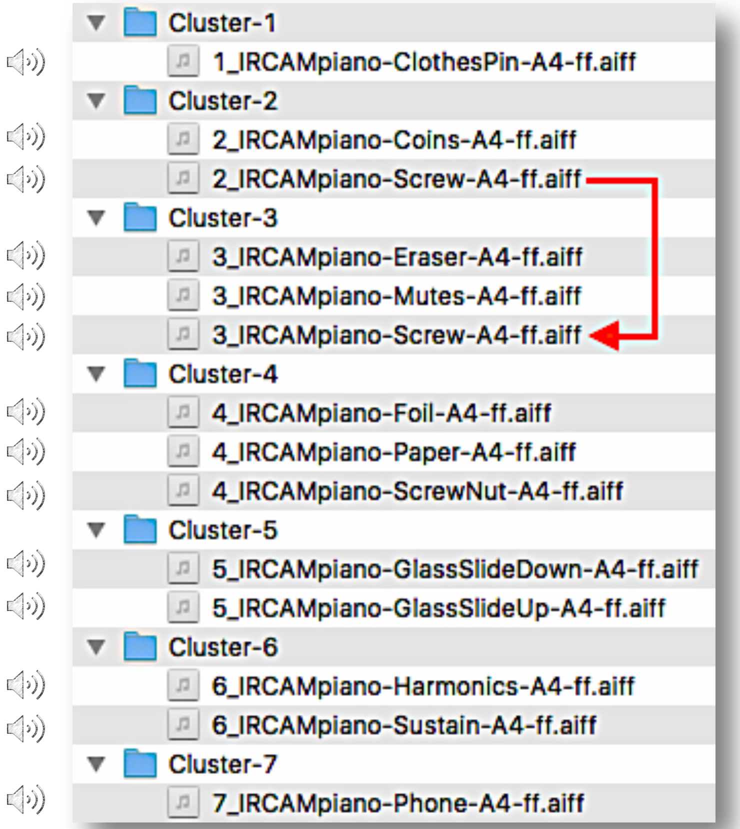
Q1



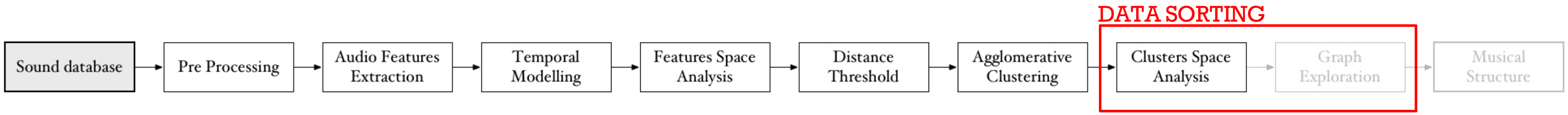




CLUSTERS SPACE ANALYSIS



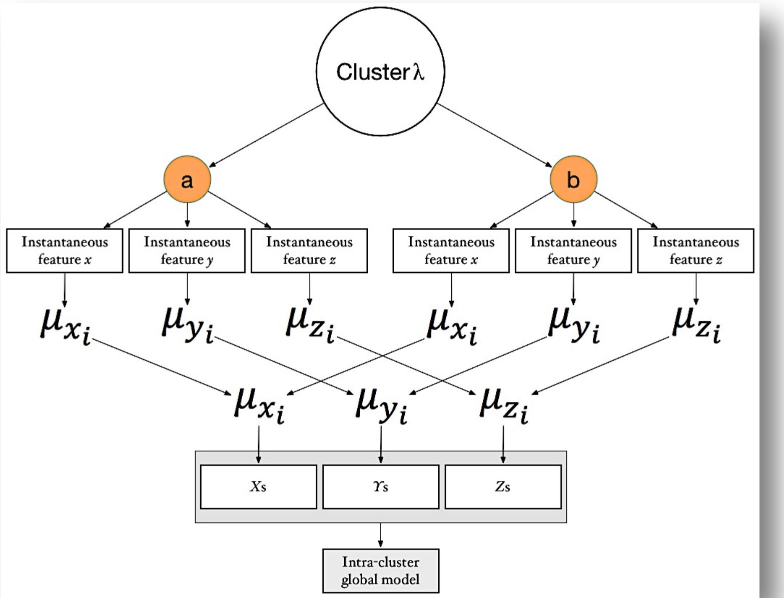
- What is the 'relationship' between clusters?
- Similar process as before (dist. oriented) but…
  1. Intra-cluster modelling
  2. Inter-cluster analysis
- To gain **more insight** on the resulting space
  - visualize the clusters space…

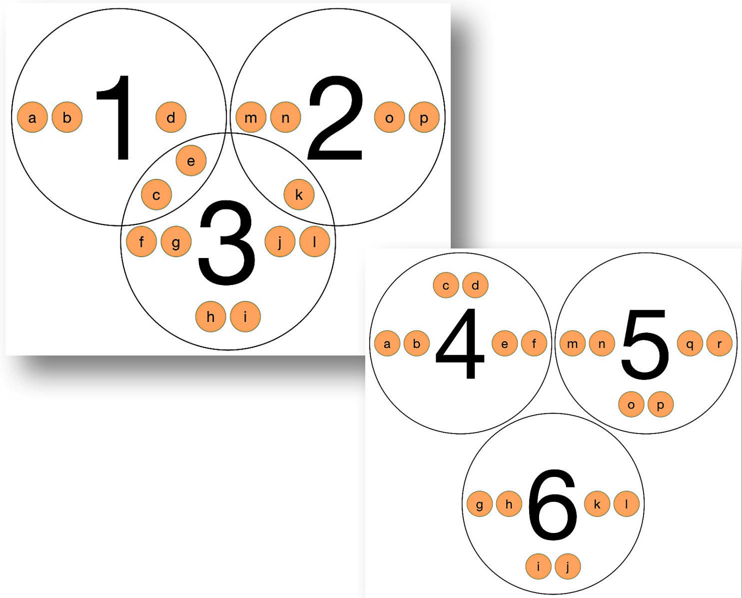


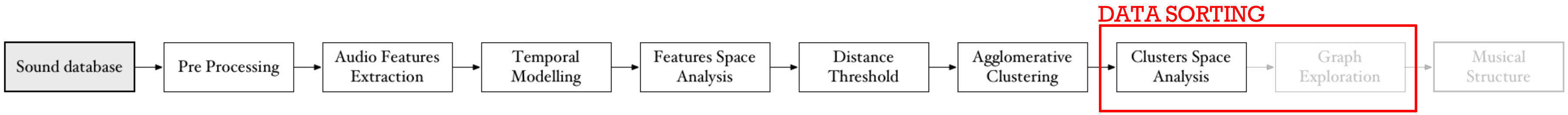
INTRA/INTER-CLUSTER ANALYSIS

- Find the theoretical centre of mass (barycentre)
- Barycentre = ‘multidimensional’ vector
  - Each data point = mean of means of each audio features

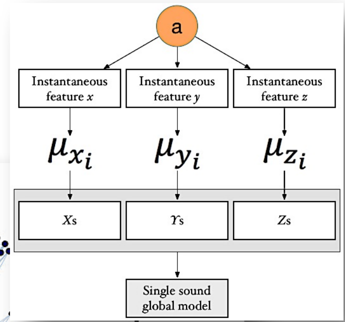
- To measure the distance between each **clusters barycentre**
  - **Mahalanobis** distance (magnitude) * ***Jaccard*** *index (similarity)*



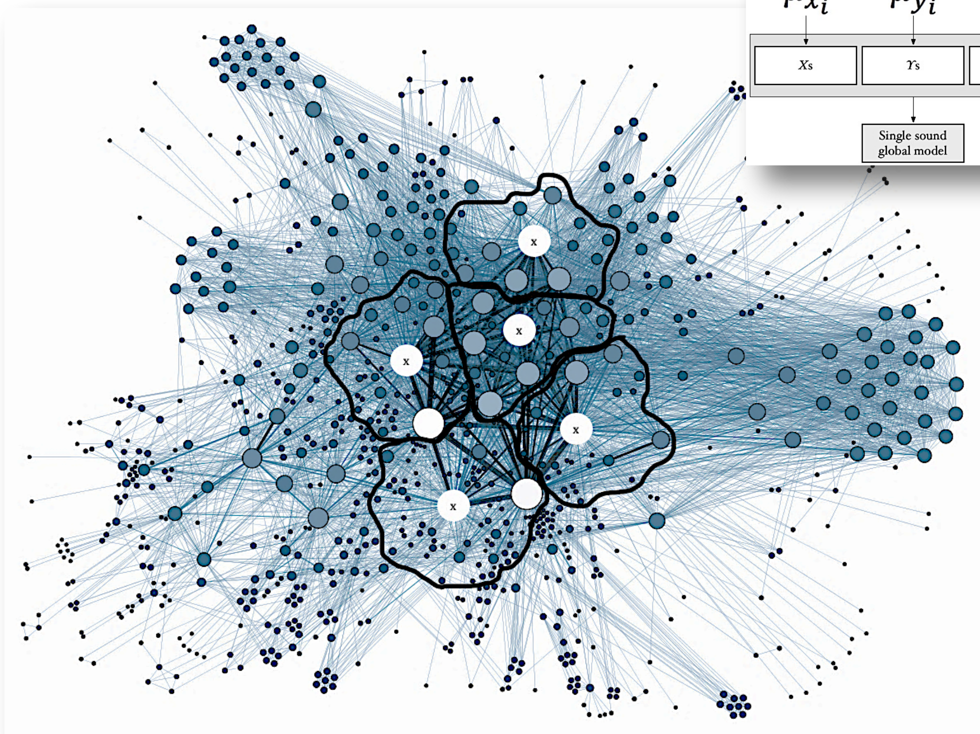


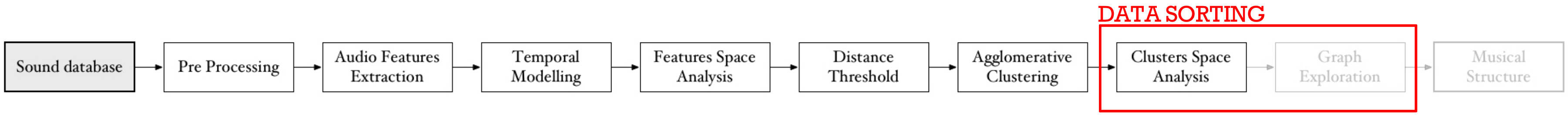


NEAT CLUSTERING (CENTROID-BASED)



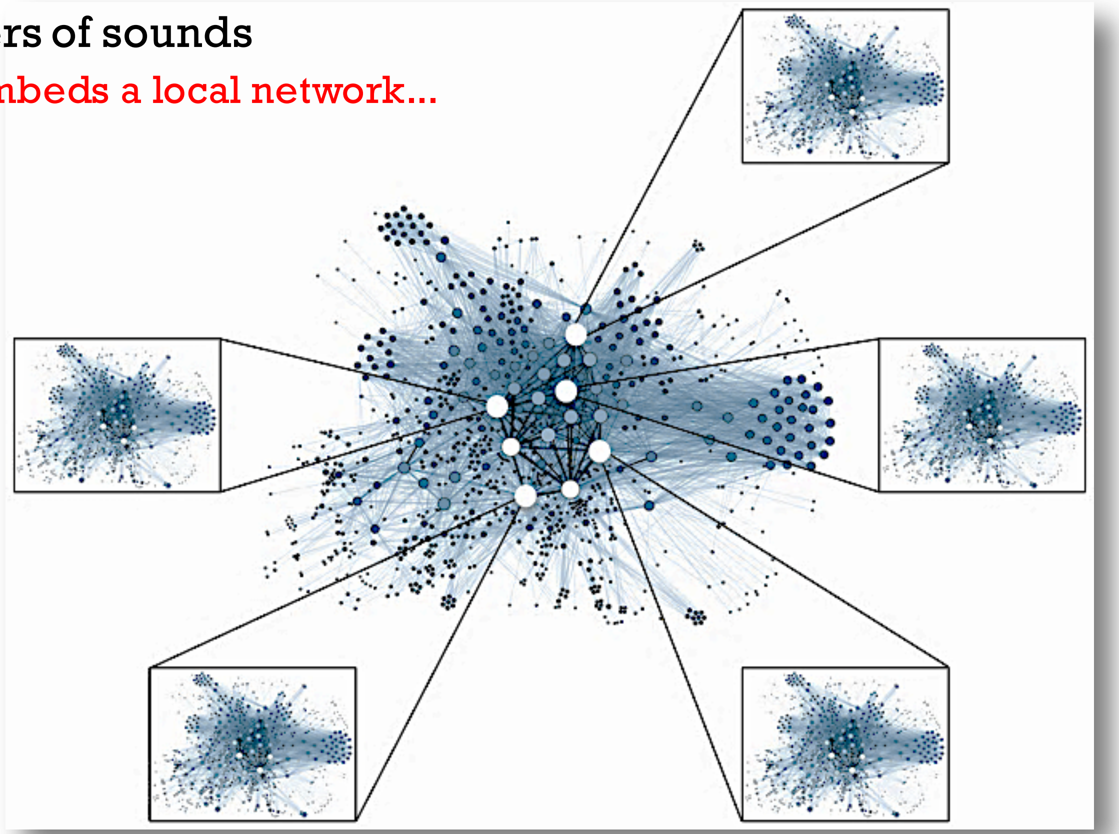
- Remove the fuzzy components
- Based on **maximum likelihood…**
- Each component belongs to a single cluster

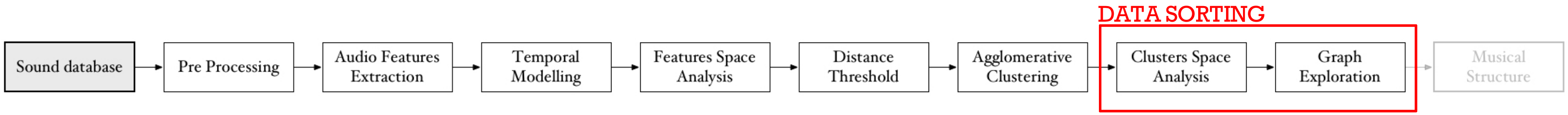




CLUSTERS SPACE NETWORK

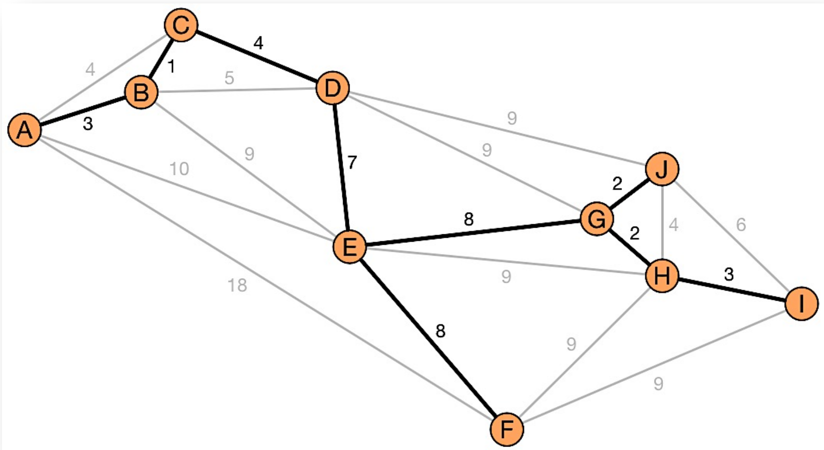
- nodes = clusters of sounds
  - Each node embeds a local network...



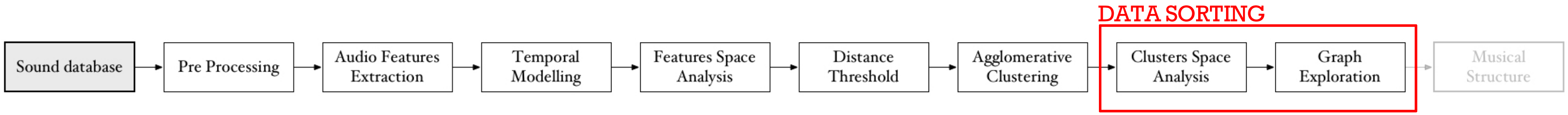


GRAPH EXPLORATION (PROBLEM SPECIFIC…)

- **Minimum spanning tree (MST)**
  - Similar to the TSP, a MST is a **subset of edges that connects all nodes together** with the **minimum possible total weight** without any cycles. [Graham 1985]
- From a compositional angle,
  - Some sort of an optimized road map for which the **total distance is minimal**…
  - Two connected clusters **may or may not** be the closest… (global optimum!)

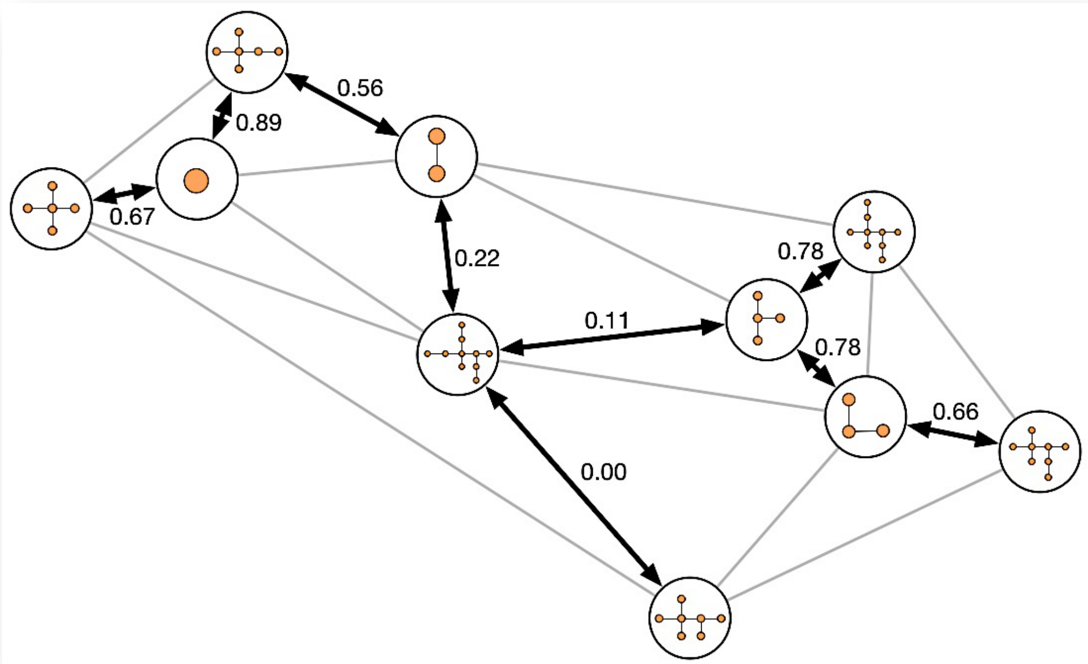


[Kruskal 1956]

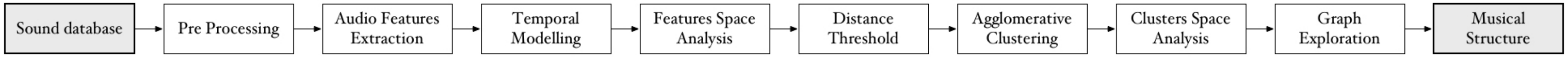


HIDDEN MARKOV MODEL (HMM)

- A Markov model is a stochastic model for which it is assumed that **future states depend only on the current state**, not on past events… (probabilistic forecasting)

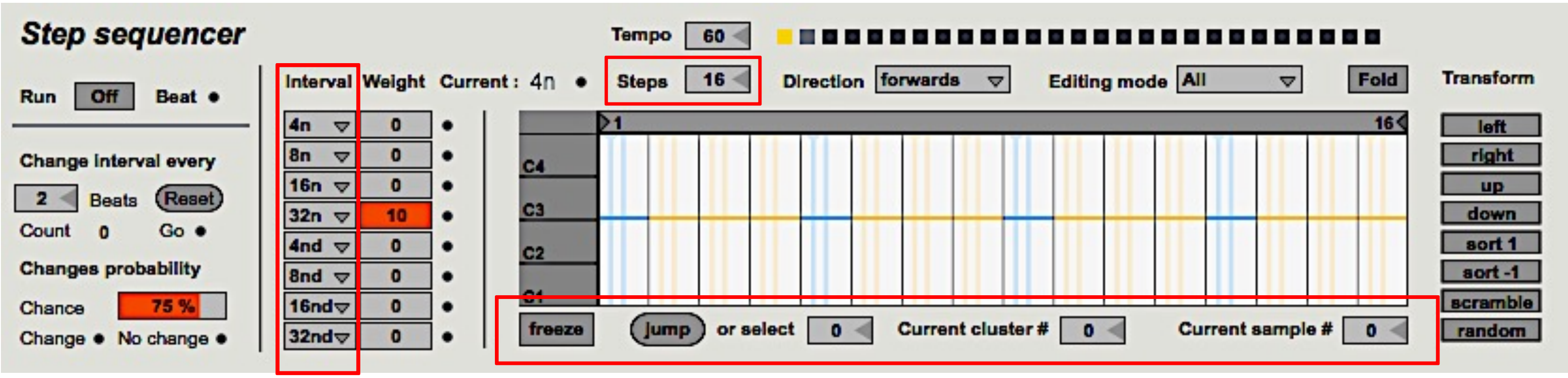


- Transition layer = **inter-clusters** MST
- Emission layer = **intra-clusters** MST



A SMART(ER) SEQUENCER

- Transition period (clusters) = sequence length of *n* steps (*n* beats)
- Emission period (sounds) = step length of interval *x* (*n* beats subdivisions)



- Automatic exploration OR manual exploration (freeze, next, select)

DEMO TIME

- **Sound corpus,**
  - 47 alto saxophone multiphonics.

- **Parameters,**
  - Clustering type: Neat
  - Distance type(s): Mahalanobis
  - Audio features:
    - k=5: Noisiness + Kurtosis + <u>Spectral Variation</u> + CS Analysis + Masking FX
    - k=4: <u>Noisiness + Kurtosis</u> + CS Analysis + Masking FX

    - k=3: Spectral Centroid + <u>CS Analysis</u> + Masking FX
    - k=2: <u>Spectral Centroid + Masking FX</u>

    - k=1: Spectral Variation

CONCLUSIONS…

- **Latent problematics:**
  - Temporal alignment (duration vs. perception)
  - **Automated feature selection (evaluation criterion vs. perception)**
  - Levels of similarity (measurements vs. perception)
  - Implicit space (L^p space vs. perception)
  - **Graph exploration (optimization vs. perception)**

- **Notable limitations:**
  - Audio signal as input (quality dependency)
  - **Audio features as variables (low level interpretation...)**
  - **Unsupervised clustering (no qualitative information...)**
  - Complexity of use (knowledge required…)

…CONCLUSIONS

- **Future works:**
  - Automated feature selection **evaluation criterion**
  - **Graph search strategies** and related algorithms
  - Add and develop **different audio features**
  - **Semi-supervised learning** approach using concept target models
    - **qualitative scales** = higher level understanding…

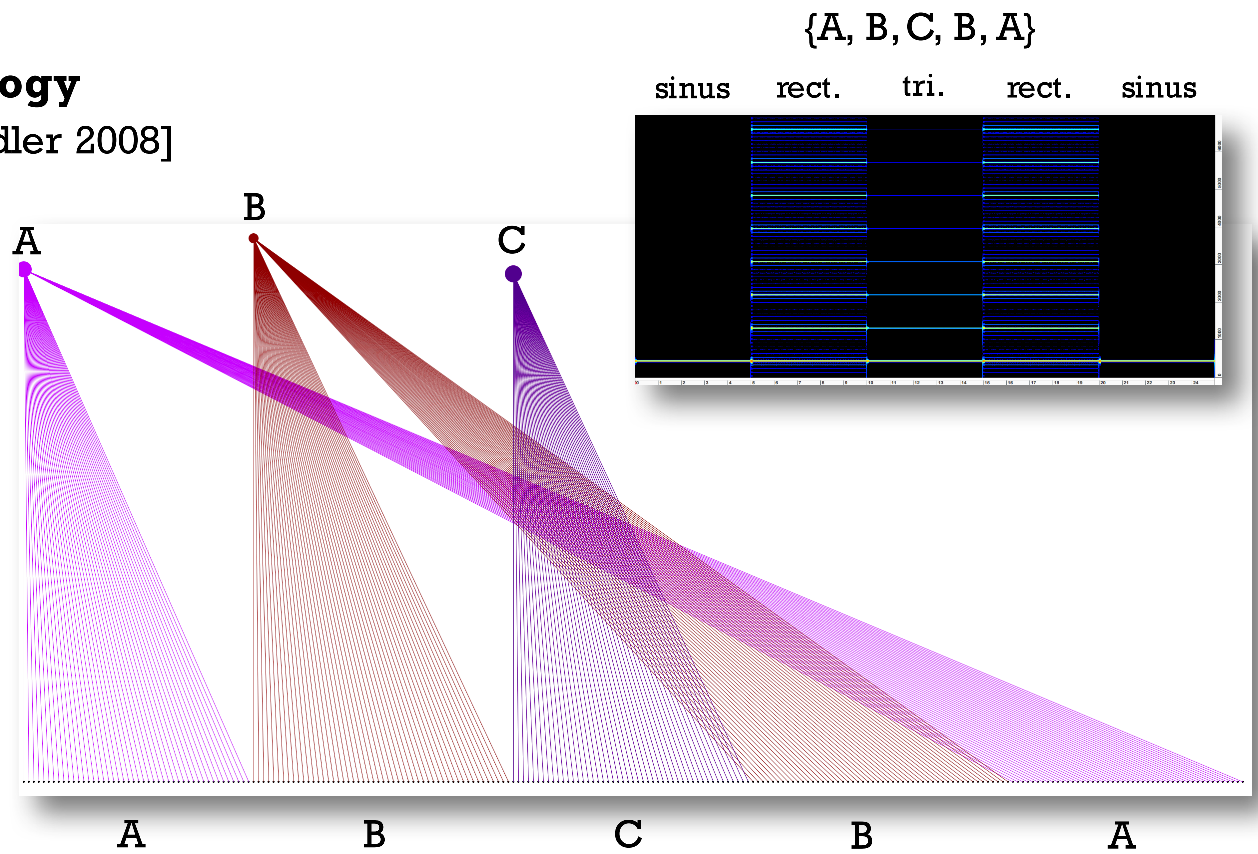
  - Unsupervised **musical structure analysis/extraction** from audio signal

FOR MORE INFORMATION

- http://repmus.ircam.fr/lebel
  - slideshow, paper, demo software, audio examples

STRUCTURE ANALYSIS (PROOF OF CONCEPT...)

- **Computational musicology**
  - Temporal clustering [Sandler 2008]



Assumption:
two consecutive frames from the same section should be more similar than two consecutive frames from different sections…