

# Khiops CoViz

a tool for visual exploratory analysis  
of k-coclustering results

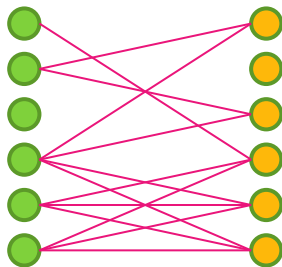


Bruno Guerraz  
Marc Boullé  
Dominique Gay  
Fabrice Clérot

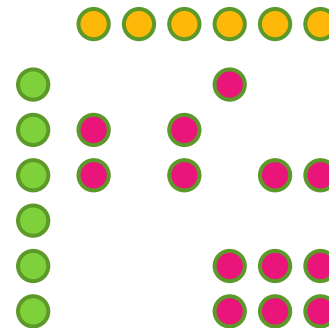
# Les données

---

- ▶ Deux entités reliées par des observations
  - ▶ Des documents, des mots
    - ▶ Ou des cookies, des sites
    - ▶ Ou ...
  - ▶ Une observation : un document donné utilise un mot donné
  - ▶ Graphe bi-parti dont les arêtes sont les observations



- ▶ Représentation matricielle équivalente



# Qu'est-ce qu'une co-partition ?

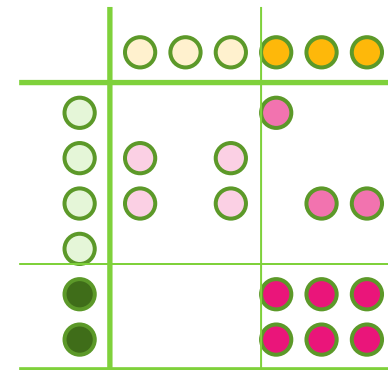
---

- ▶ C'est la partition induite sur les observations par le produit cartésien de deux partitions
  - ▶ Une partition par entité

# Qu'est-ce qu'une co-partition ?

- ▶ C'est la « partition » induite sur les observations par le produit cartésien de deux partitions
  - ▶ Une partition par entité

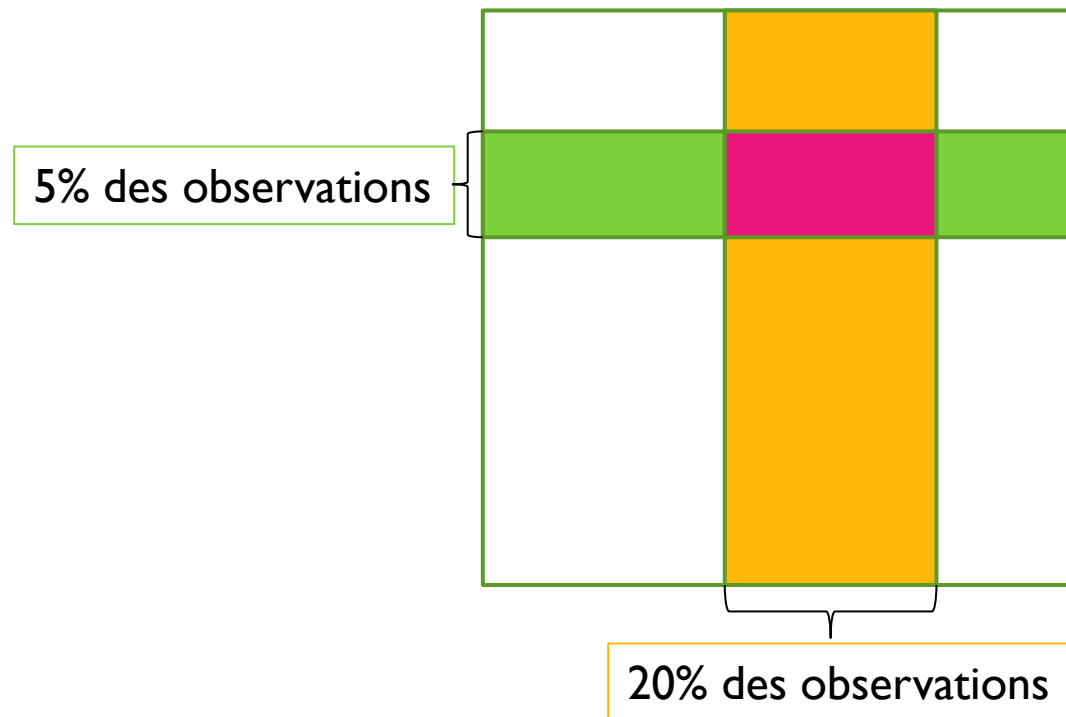
- ▶ Par exemple :
  - ▶ Une partition en 2 parties pour chaque entité
  - ▶ Une « partition » en 4 parties des observations
    - ▶ Dont une vide ...



# Qu'est-ce qu'un co-clustering ?

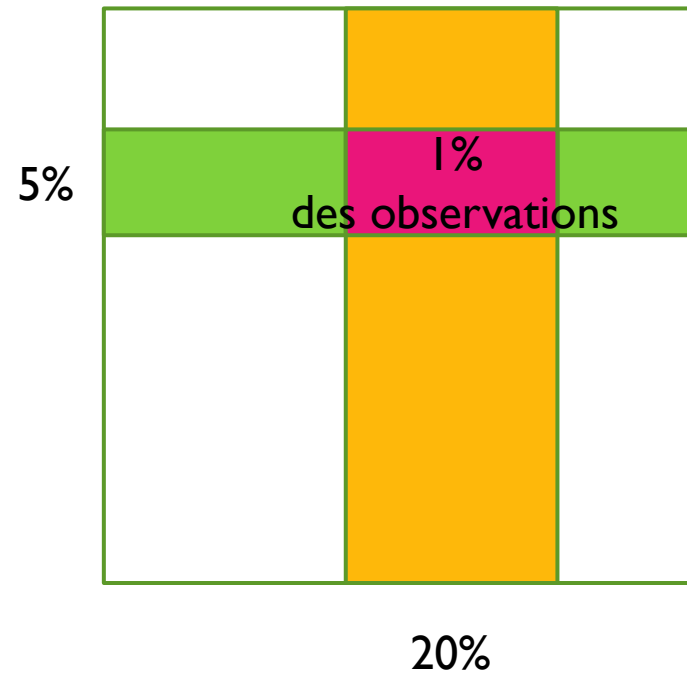
---

- Soit une co-partition :



# Qu'est-ce qu'un co-clustering ?

- Quelle information porte cette co-partie ?



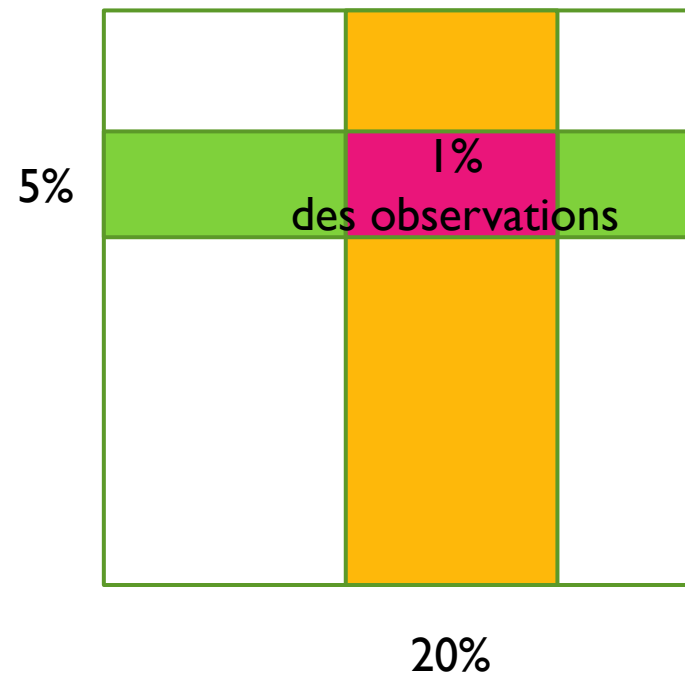
# Qu'est-ce qu'un co-clustering ?

---

## ► Quelle information porte cette co-partie ?

- Aucune !
- C'est ce qu'on attendrait sous l'hypothèse d'indépendance des variables (ligne, colonne)
  - Distribution des observations « au hasard » respectant les comptes des parties ligne et colonne (« marginales »)

□  $5/100 \times 20/100 = 1/100$  !



# Qu'est-ce qu'un co-clustering ?

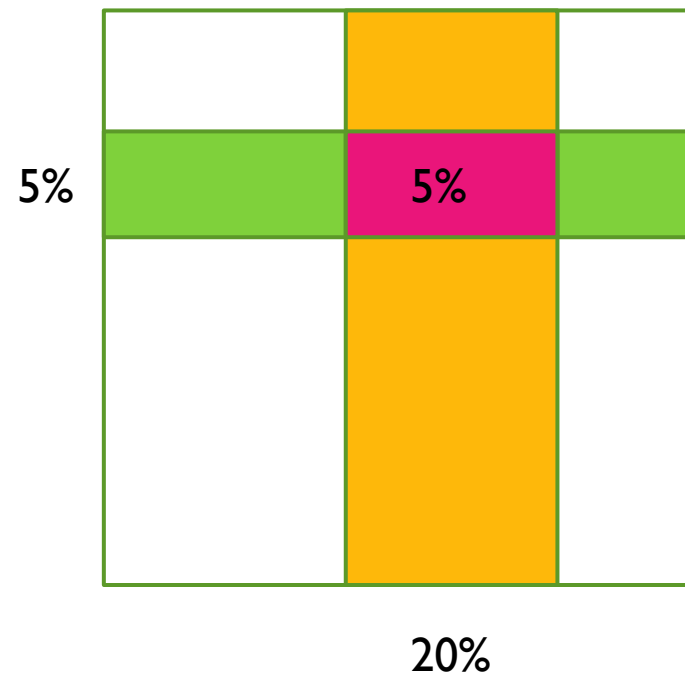
---

- ▶ Un co-clustering est une co-partition qui maximise le contraste entre la distribution des co-parties et la distribution espérée sous l'hypothèse d'indépendance (connaissant les marginales)

# Qu'est-ce qu'un co-clustering ?

---

- ▶ Un co-clustering est une co-partition qui maximise le contraste entre la distribution des co-parties et la distribution espérée sous l'hypothèse d'indépendance (connaissant les marginales)
- ▶ Par exemple :

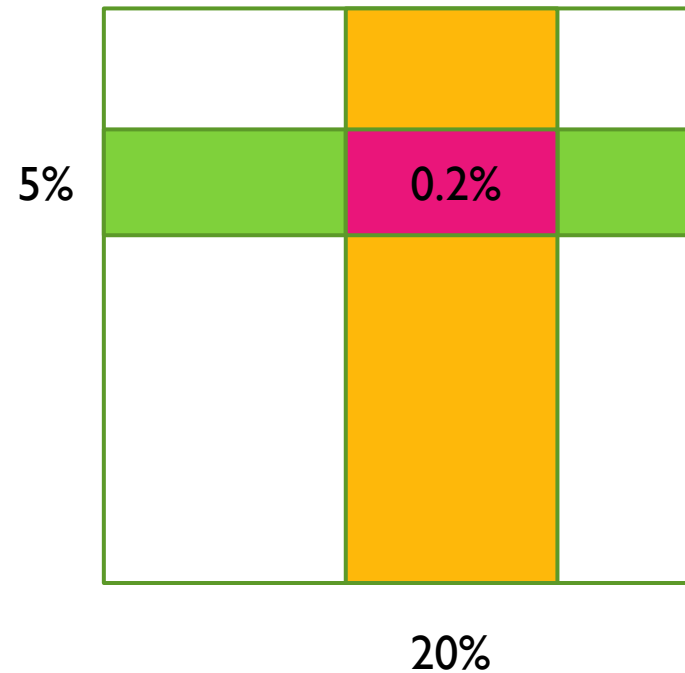


# Qu'est-ce qu'un co-clustering ?

---

- Un co-clustering est une co-partition qui maximise le contraste entre la distribution des co-parties et la distribution espérée sous l'hypothèse d'indépendance (connaissant les marginales)

- Ou encore :



# Deux produits

► Formalisation  
mathématique du  
problème

+

► Algorithmique

+

► Visualisation interactive

► Khiops co-clustering

► Khiops co-visualisation



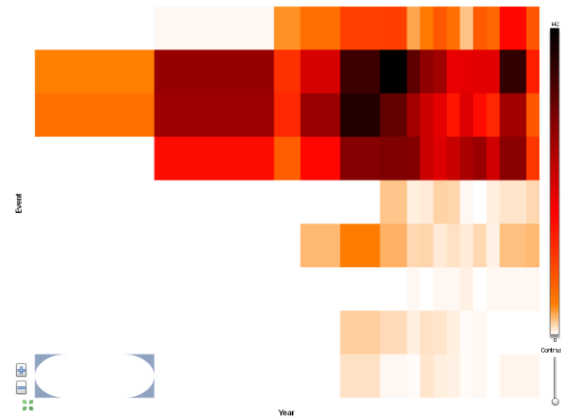
# Exploration & Visualisation of joint distribution of input variables using data grid models

k-D Data

| X        | Y        |
|----------|----------|
| $V_{1X}$ | $V_{3Y}$ |
| $V_{4X}$ | $V_{5Y}$ |
| ...      | ...      |

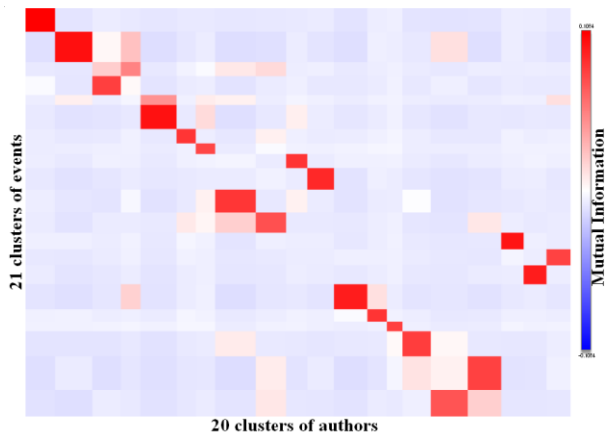


Khiops  
Coclustering &  
Visualisation



Several visualisation  
criteria for various  
insights in the data

Exploration of results at  
multiple granularities



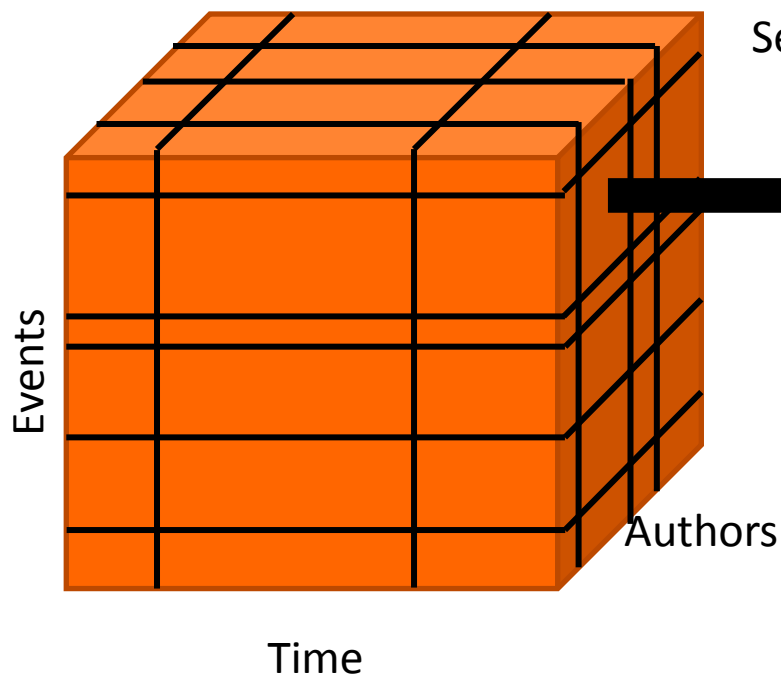
## Types of data

- Text Data (T , W)
- Functional Data (F , X , Y)
- (Temporal) Graph Data (O , D , T)
- Temporal Sequences (S , T , E)
- any other k-adic/co-occurrence data
- ...

## Applications

- Summarization/(co)-clustering
  - Texts
  - Temporal sequences/series
  - Graphs

# Hints for 3D-visualization



Selection of a  
cluster of  
authors

2D-Visualization of a cluster of authors  
(Time x Event) visualization

