



Machine Learning for Data Science

Mohamed Nadif

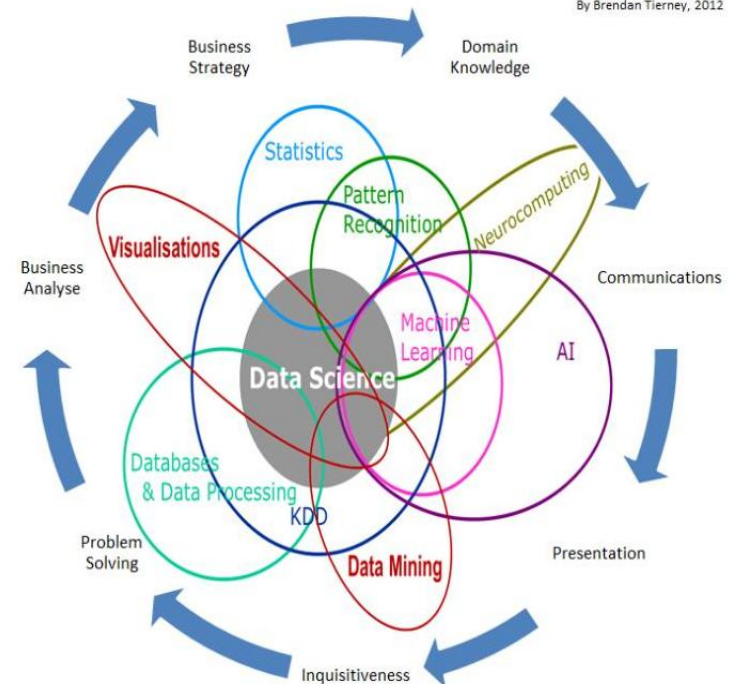
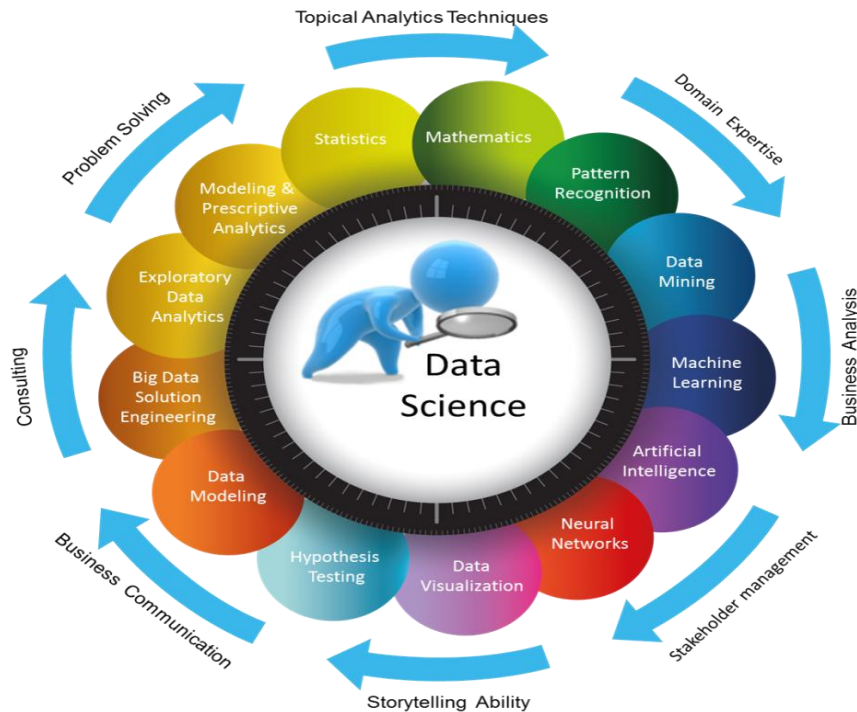
Journée Big Data@USPC, 30 novembre 2015

- **LIPADE**
Université Paris Descartes
45 rue des Saints Pères
75270 Paris Cedex 06
- Equipe **M**achine **L**earning for **D**ata **S**cience (MLDS)
<http://www.math-info.univ-paris5.fr/~nadifmoh/Accueil.html>
 - Effectif : 11
 - 5 enseignants-Chercheurs (1 PR et 4 MCF)
FX Jollois, L. Labiod, F. Role, N. Rogovschi et M. Nadif
 - 6 doctorants
A. Ailem, K. Allab, N. Médoc, Ch. Laclau, M.C Dani, A. Salah
- Responsable de l'équipe : Mohamed Nadif

Data Science

Data Science Is Multidisciplinary

By Brendan Tierney, 2012



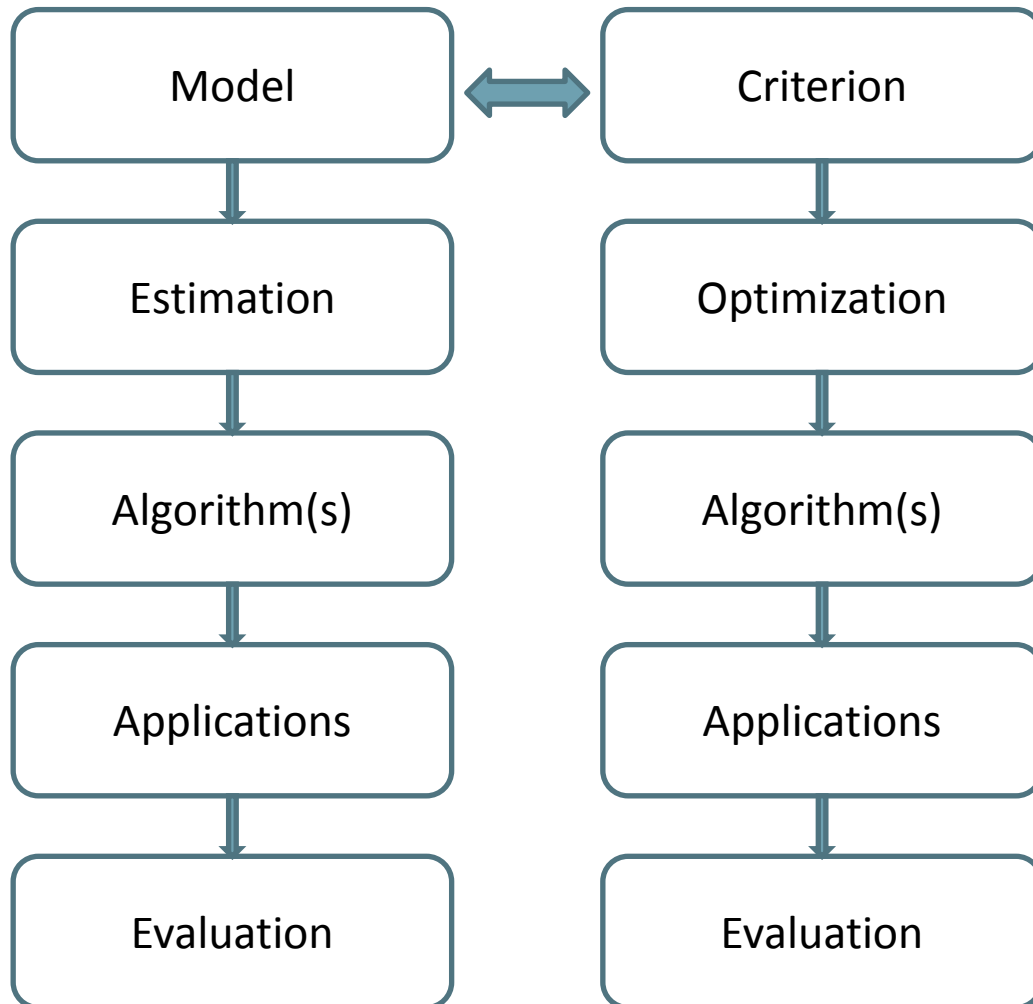
**Compétences de
Equipe MLDS**

Thème de recherche

→ Keywords :

- Unsupervised learning
- Supervised learning
- Semi-supervised learning
- Visual analytics
- Data Analysis
- Data mining
- Missing data
- Recommender systems
- Text mining

Démarche d'ordre théorique et appliquée

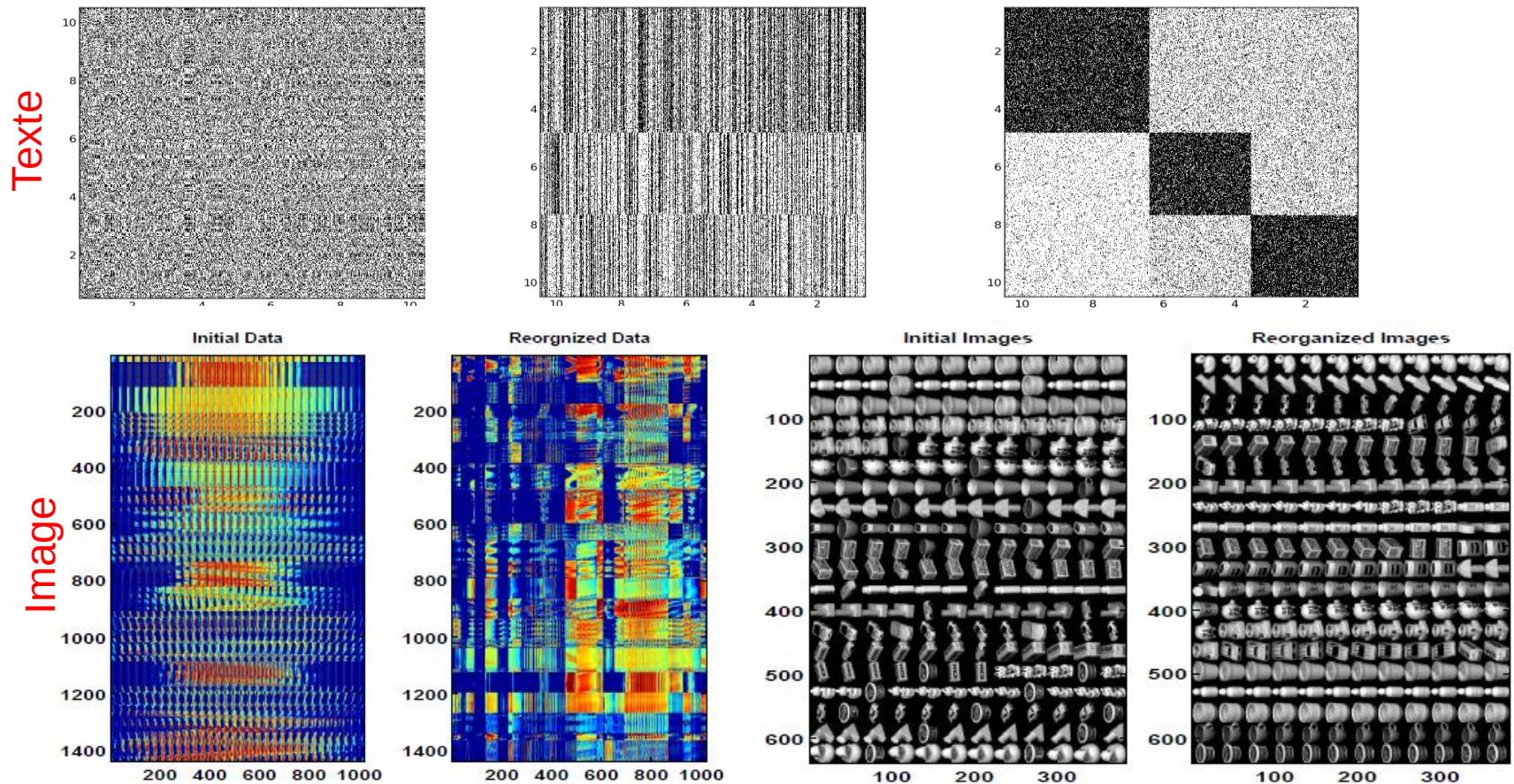


Récentes collaborations :

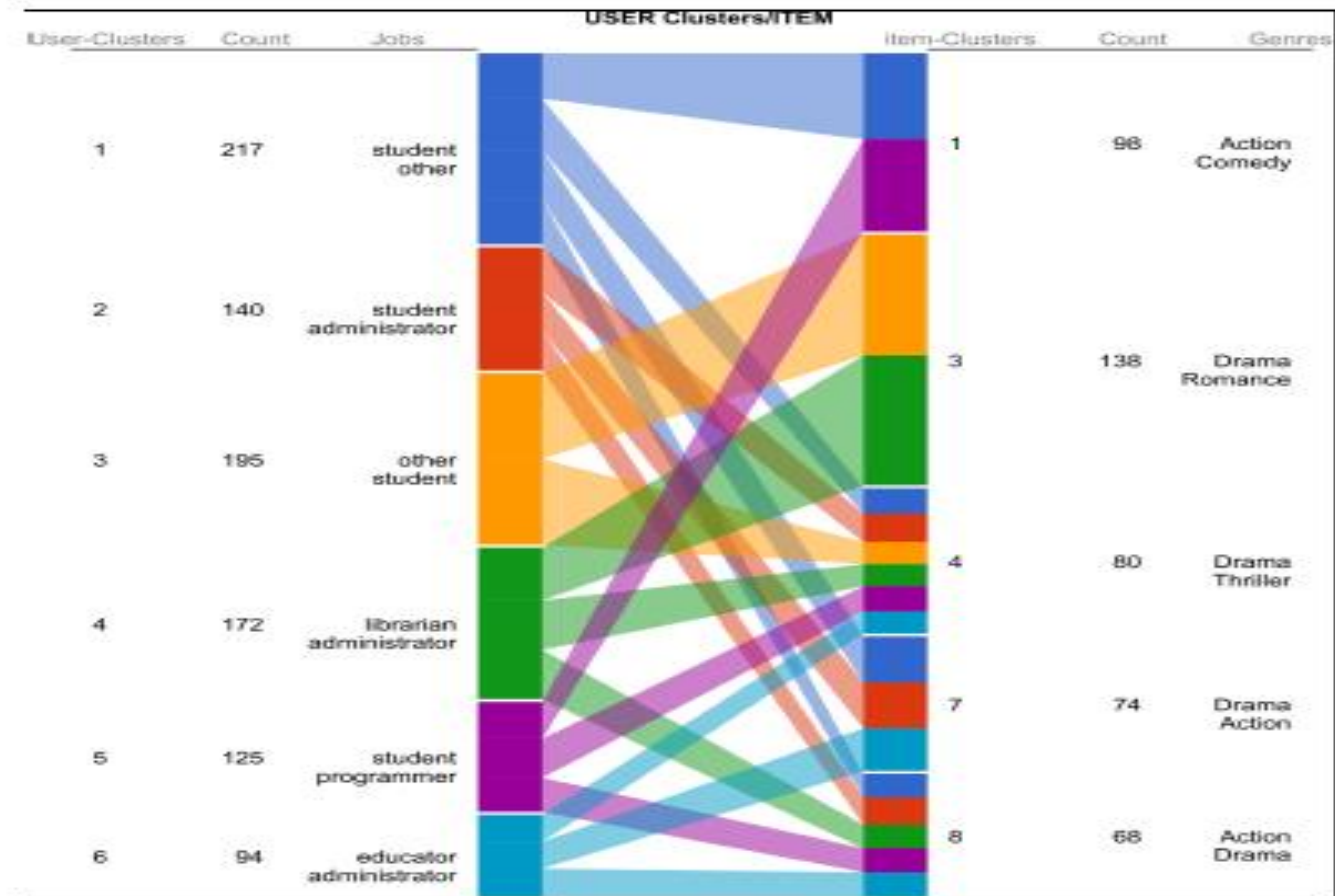
- Text-mining : PubMed database (25 millions abstracts).
INSERM **U946** P7
- Gestion des déchets : Trinov
- Détection d'anomalies : Airbus
- Analyse de flux textuel
RTL (Luxembourg)
- Systèmes de recommandation

Différents types de données : illustrations

- Clustering and Co-clustering



Systèmes de recommandation



Visualisation interactive

Proportions

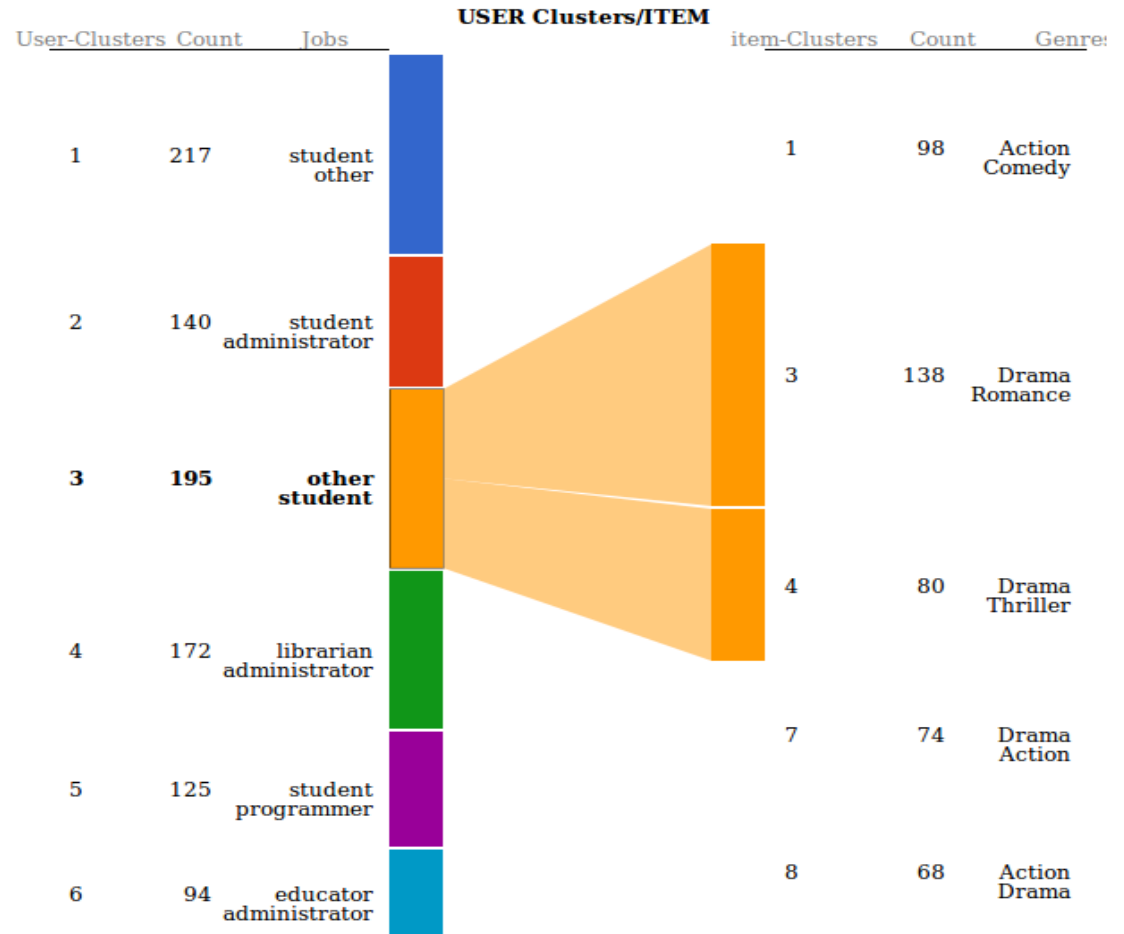
- other: 0.35
- student: 0.15
- writer: 0.15
- artist: 0.05
- educator: 0.05
- executive: 0.05
- healthcare: 0.05
- librarian: 0.05
- programmer: 0.05
- retired: 0.05

Gender proportions

- Male: 0.85
- Female: 0.15

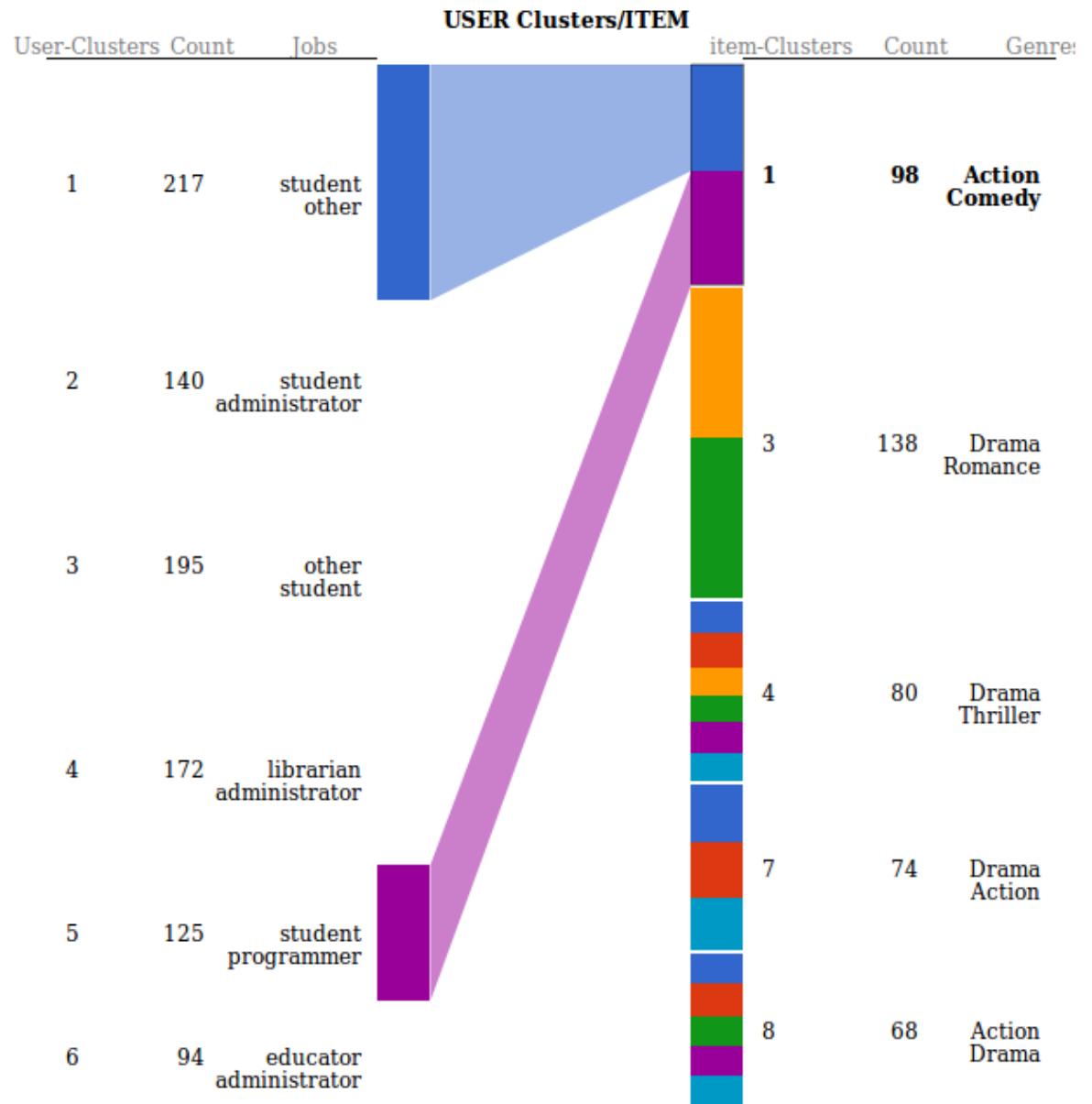
Age proportions

- [7-->17]: 0.1
- [18-->25]: 0.15
- [26-->34]: 0.4
- [35-->44]: 0.05
- [45-->54]: 0
- [55-->75]: 0.3



Proportions

- Action: 0.17
- Comedy: 0.15
- Romance: 0.13
- Adventure: 0.09
- Children: 0.08
- Drama: 0.08
- Musical: 0.08
- Thriller: 0.08
- Animation: 0.06
- Crime: 0.04
- Sci-Fi: 0.04
- Western: 0.02



Approches tenant compte du volume des données

→ Probabiliste

- Modèles de mélanges
- Modèles des blocs latents

→ Métrique

→ Factorisation

- NMF
- NMTF
- Semi-NMF et Semi-NMTF

→ Approche combinant réduction de la dimension et clustering

Master 2 IA, parcours « MLDS »

Master M2 IA, parcours « Machine Learning » depuis 2010

- **Tronc commun**
 - Apprentissage orienté agent
 - Théorie de la décision
 - Anglais
 - Méthodologies de la recherche et innovation
- **Parcours MLDS**
 - Apprentissage non supervisé
 - Apprentissage supervisé
 - Visual Analytics / Exploration visuelle des données
 - Modèles de mélanges
 - Classification croisée sous différentes approches
 - Science des données
- **UE libre**

Master 2 IA, parcours « MLDS »

- **Master « Machine Learning for Data Science »**

- 14 étudiants/an
- Métier : Data Scientist
- Poursuite en doctorat : 25-30%
 - Sous contrats CIFRE
 - Laboratoires de recherche universitaires
 - INSERM

Actuellement 4/6 étudiants en thèse au sein de MLDS

- **Points forts du master**

- Une maîtrise de certaines méthodes de ML tant au niveau théorique qu'appliqué
- Tous les membres de l'équipe interviennent

Synthèse, network et collaborations au sein de l'USPC

