

# Inférer le sexe et l'âge d'individus à partir du petit mobilier funéraire mérovingien

Gaël Poux-Médard\*

Schneider É., Touzet A., Fischbach T., Santos F., Chenal F., Abert F., Médard F., Barrand-Emam H.

\*Université de Lyon, France  
Lyon 2, ERIC UR 3083

7 octobre 2022



# Introduction

- PCR Mérovingien Grand Est – Axe « Identité biologique »

COORDINATION	SPÉCIALITÉ	ORGANISME DE RATTACHEMENT
BARRAND EMAM Hélène	Archéo-anthropologue	Antea, UMR 7044
ABERT Franck	Archéologue, haut Moyen Âge	Archéologie Alsace
CHENAL Fanny	Archéo-anthropologue	Inrap, UMR 7044

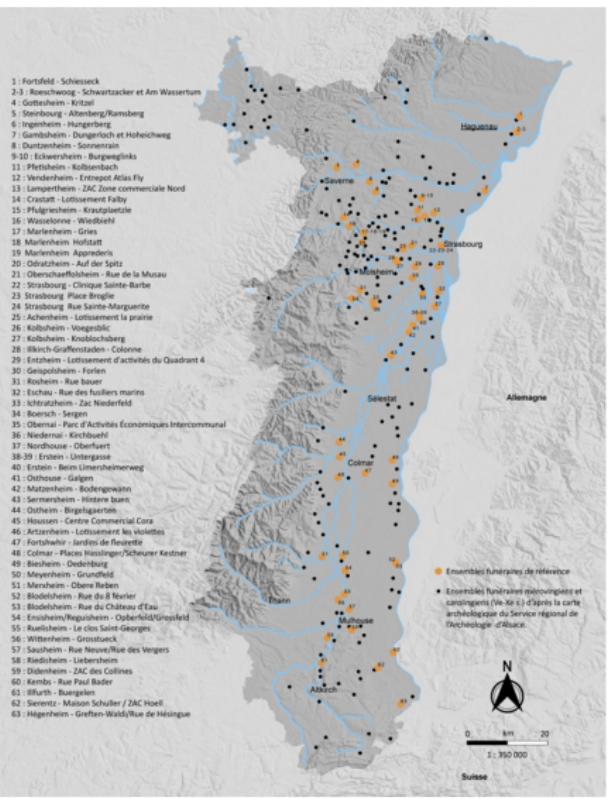
ÉQUIPE DE RECHERCHE 2019 - 2021		
ABERT Franck	Archéologue, haut Moyen Âge	Archéologie Alsace
BARRAND EMAM Hélène	Archéo-anthropologue	Antea, UMR 7044
BOURY Loïc	Archéozoologie	Antea
CHATELET Madeleine	Archéologue, haut Moyen Âge	Inrap, UMR 7044
CHENAL Fanny	Archéo-anthropologue	Inrap, UMR 7044
FISCHBACH Thomas	Archéologue, haut Moyen Âge	Antea, Doctorant Uni. Strasbourg
MEDARD Fabienne	Spécialiste textiles et matières organiques	ANATEX, UMR 7044
PEYTREMANNE Edith	Archéologue, haut Moyen Âge	Inrap, UMR 6273
PELISSIER Amélie	Archéo-anthropologue	Archéologie Alsace
POUX-MEDARD Gael	Mathématiques appliquées	Université de Lyon 2
PRUVOST Mélanie	Paleogénéticienne	Univ. Bordeaux ,CNRS, UMR 5199
PUTELAT Olivier	Archéozoologue	Archéologie Alsace, UMR 7041
SANTOS Frédéric	Statisticien	Université Bordeaux, UMR 5199
TOUZET Amandine	Archéologue, base de données, site internet	Antea

ÉTUDIANTS PARTICIPANTS AU PROJET		
NEELS Louise	Etudiante Master II (2019-2020)	Université de Strasbourg
SCHNEIDER Elisa	Etudiante Master II puis Doctorat	Université de Paris I
VAUX Océane	Etudiante Master I et II (2019-2021)	Université de Strasbourg
CHEIK MOKTAR Fofana	Etudiant Master I (2020-2021)	Université de Strasbourg
PAVIN Anais	Etudiante Master II (2020)	Université de Bordeaux
LA ROCCA Fanny	Etudiante Master II (2020)	Université de Bordeaux
CUNY Mathéo	Etudiant Master I (2020-2021)	Université de Strasbourg

Archéologie

Autres

## Introduction



Nom de l'opération		ODRATZHEIM Sandgrube, Lotissement du Wehland	
Responsable de l'opération	Agnieszka KOZIOL, PAIR		Numéro d'OA
		Numéro structure :	1062
Numéro individu		Numéro objet	2
Reste(s) archéologique(s)		Champs obligatoires	
4 rivets + 89 clous en alliage cuivreux + 1 applique		Statut	identifié
Objet interprété/restitué		Precision	de scramasaxe
Matériaux	<input checked="" type="checkbox"/> alliage cuivreux <input type="checkbox"/> ambre <input type="checkbox"/> améthyste <input type="checkbox"/> argent <input type="checkbox"/> bois <input type="checkbox"/> bois de cervidé <input type="checkbox"/> calcaire <input type="checkbox"/> coquille/coquillage		
Décors	<input type="checkbox"/> ajouré <input type="checkbox"/> céramique tournée <input type="checkbox"/> céramique non tournée <input type="checkbox"/> céramique en montage mixte <input type="checkbox"/> cloisonné <input type="checkbox"/> moulé		
Techniques et décors			
Typeologie			
Localisation	associé au corps		
Association			
Remarques			
Mobilier	<input type="checkbox"/> H < 4 sur 11 <input type="checkbox"/> H > 4 sur 11 <input type="checkbox"/> Autres filtres   Rechercher		
Structure	<input type="checkbox"/> H < 36 sur 110 <input type="checkbox"/> H > 36 sur 110 <input type="checkbox"/> Autres filtres   Rechercher		
Site	<input type="checkbox"/> 1 sur 1 <input type="checkbox"/> Tous <input type="checkbox"/> Filtrer   Rechercher		

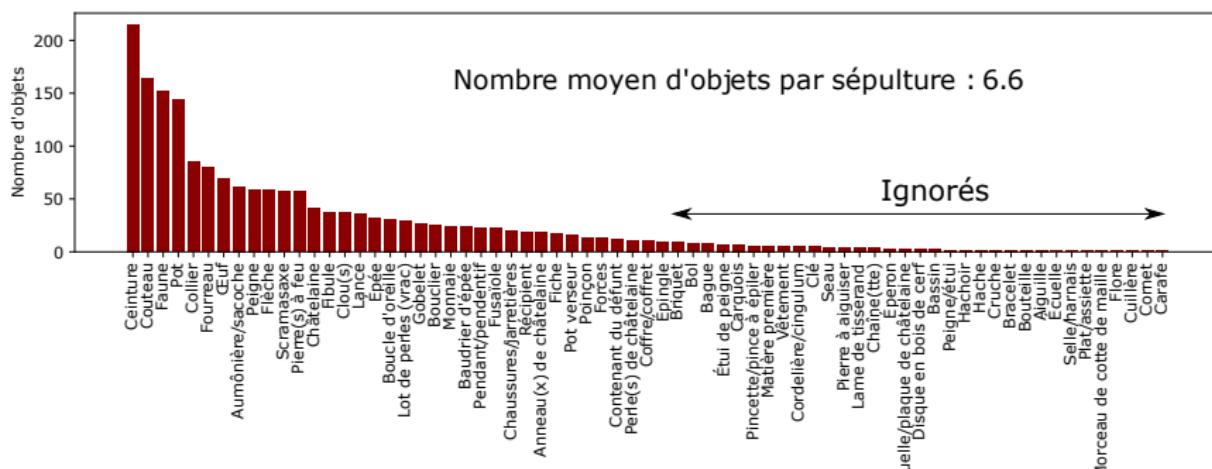
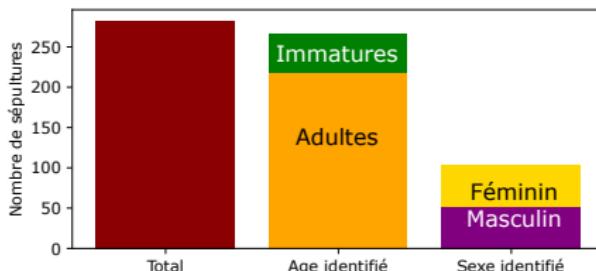
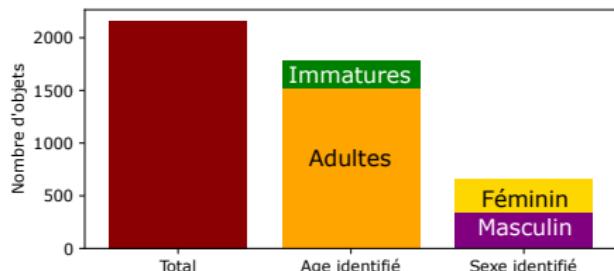
# Base de données

Nom	n° Structure	Statut	Objet	Localisation	Sexe_Bio	Sexe_Genetic	Sexe_Archeo	Classe	AgeEstime	ClasseAge
ARTZENHEIM	5	identifié	ceinture	associé au corps				Adulte	plus de 15 ans [15-19]/[>20]	
ARTZENHEIM	5	possible	contenant du	indéterminé				Adulte	plus de 15 ans [15-19]/[>20]	
ARTZENHEIM	14	identifié	peigne	indéterminé				Adulte	>20 ans	[>20]
ARTZENHEIM	14	non identifié	indéterminé	indéterminé				Adulte	>20 ans	[>20]
ARTZENHEIM	14	non identifié	indéterminé	indéterminé				Adulte	>20 ans	[>20]
ARTZENHEIM	15	identifié	baudrier d'épée	associé au corps				Adulte	>20 ans	[>20]
ARTZENHEIM	15	identifié	ceinture	associé au corps				Adulte	>20 ans	[>20]
ARTZENHEIM	15	identifié	flèche	associé au corps				Adulte	>20 ans	[>20]
ARTZENHEIM	16	non identifié	non identifié	indéterminé				Adulte	>20 ans	[>20]
ARTZENHEIM	20	non identifié	indéterminé	secondaire (perturbé, p)	Féminin	Féminin		Adulte	>20 ans	[>20]
ARTZENHEIM	24	identifié	ceinture	indéterminé	Féminin	Féminin		Adulte	> 60 ans	[>20]
ARTZENHEIM	25	possible	châtelaine	indéterminé				Adulte	>20 ans	[>20]
ARTZENHEIM	25	identifié	anneau(x) de	associé au corps				Adulte	>20 ans	[>20]
ARTZENHEIM	25	identifié	rouelle/plaqué	associé au corps				Adulte	>20 ans	[>20]

Figure 1 – Échantillon des données brutes

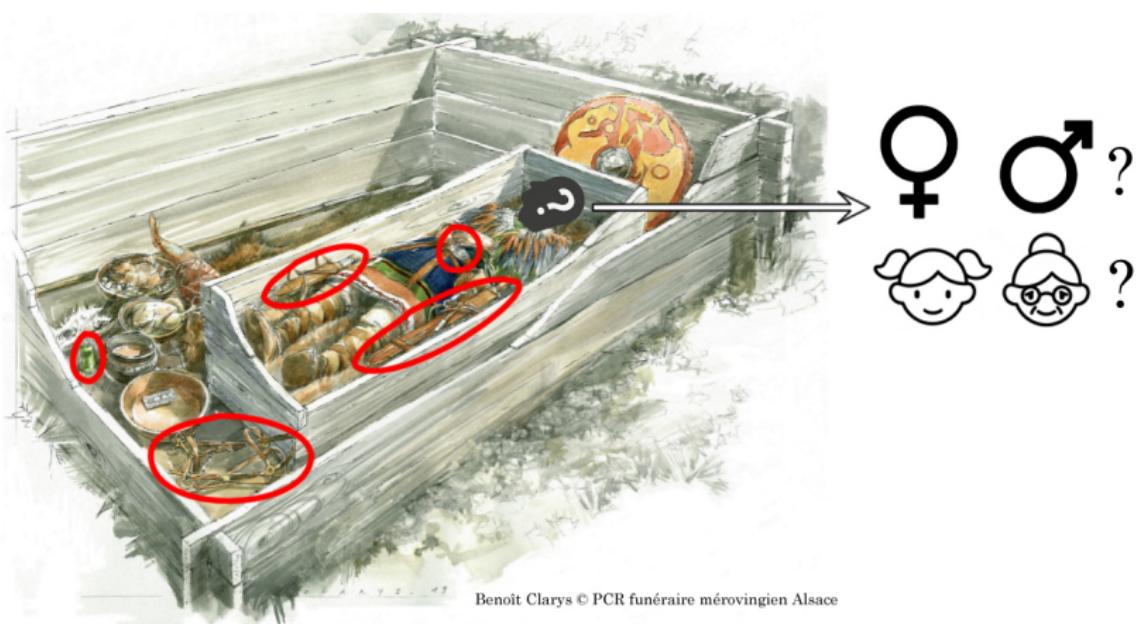
- Objectif : donner une liste d'objets à un programme qui fournit le sexe/âge de l'individu
- Pour le moment : établir l'existence de corrélations

# Base de données - Caractéristiques



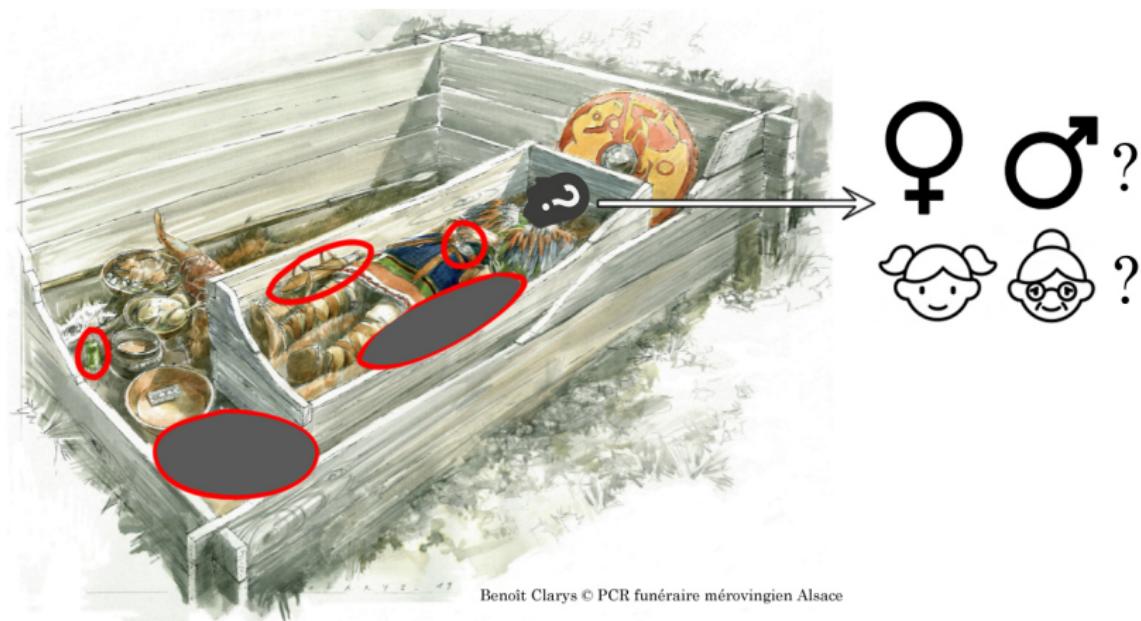
# Tâche

- À partir d'une liste d'objets :
  - Probabilité d'une classe d'âge
  - Probabilité d'un sexe



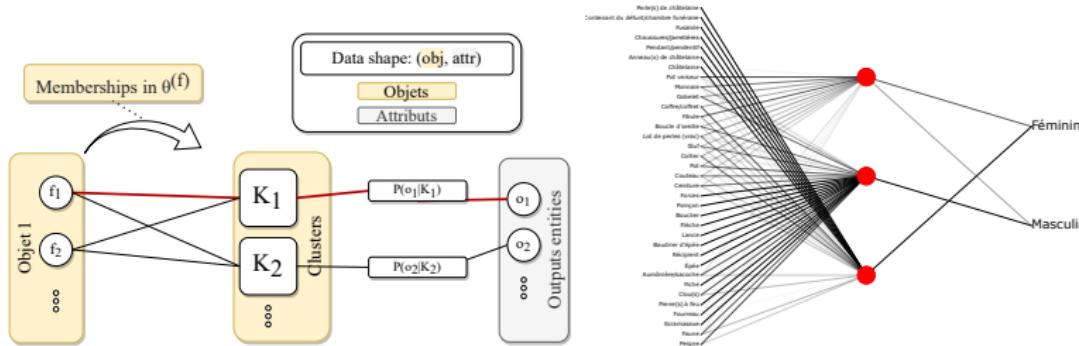
# Tâche

- Objets manquants : pas grave, le modèle recherche des régularités
- Données imparfaites : pareil → Intérêt des modèles d'apprentissage



## Approche

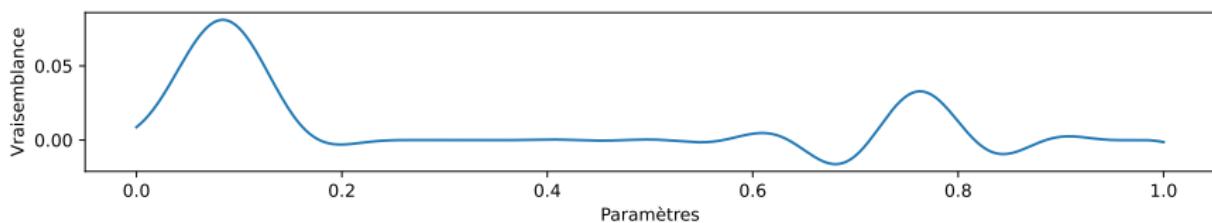
- Apprentissage automatique :
    - Pas d'étude archéologique
    - Recherche de **corrélations**
  - Liste d'objets + attributs associés :
    - Grouper les objets similaires par rapport aux attributs dans des clusters
    - Relier les clusters à la probabilité d'un attribut



**Figure 2 –** Objets, clusters et attributs (noeuds) fixés. Le modèle infère les liens entre ces noeuds.

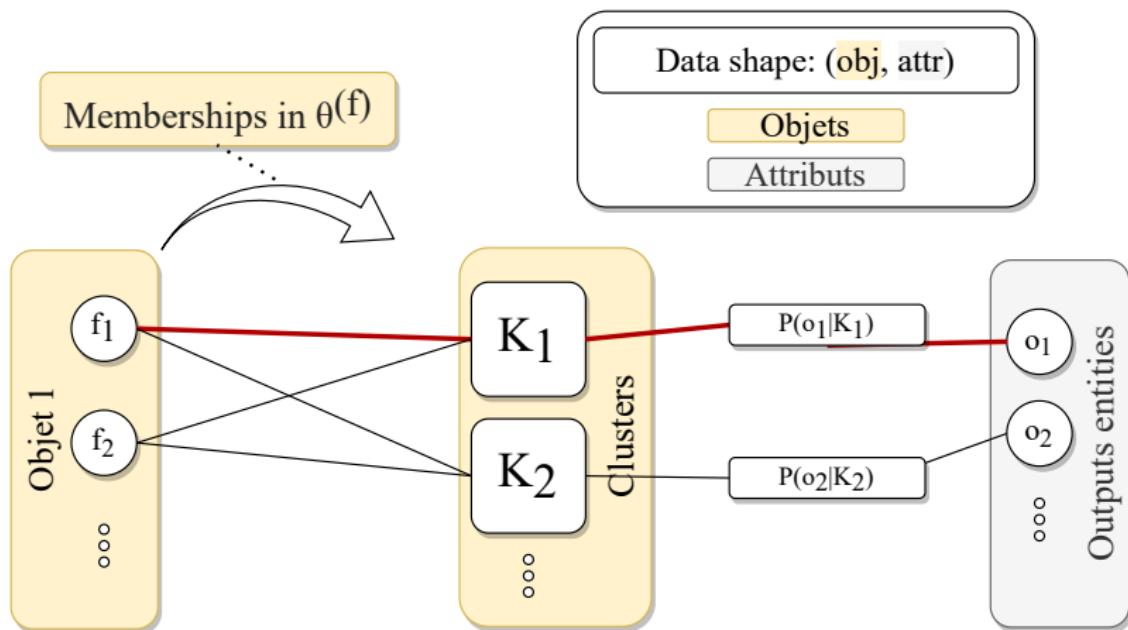
# Instant technique...

- $P(attr|obj) = \underbrace{\theta_{obj,clus}}_{P(obj \in cluster)} \times \underbrace{p_{clus,attr}}_{P(attr|cluster)}$
- Vraisemblance du modèle :  $\mathcal{L} = \prod_{(obj, attr) \in obs} P(attr|obj)$ 
  - Probabilité des observations étant données les paramètres
  - Maximiser  $\mathcal{L}$  en fonction des  $\theta$  et  $p$  ( $\sim 300$  paramètres)



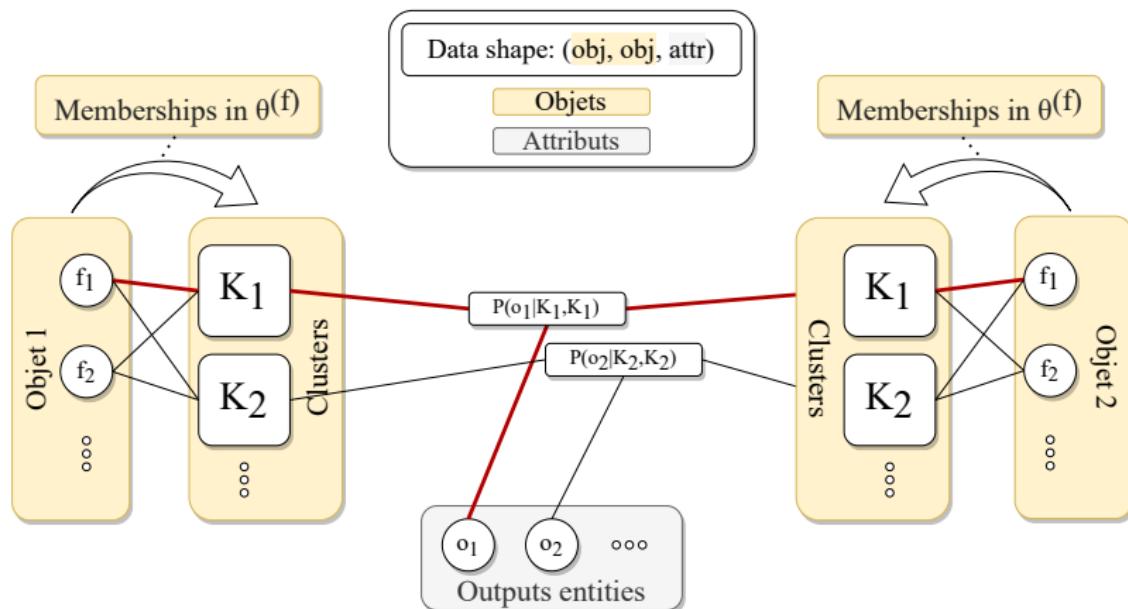
En vérité... 
$$\begin{cases} \mathcal{L} = \sum_{\vec{k} \in \vec{K}} p_{\vec{k}}(o) \prod_{n \in N} \theta_{f_n, k_n}^{(a(f_n))} & \omega_{\vec{f}, o}(\vec{k}) = \frac{p_{\vec{k}}(o) \prod_{n \in N} \theta_{f_n, k'_n}^{(a(f_n))}}{\sum_{\vec{k}' \in \vec{K}} p_{\vec{k}'}(o) \prod_{n \in N} \theta_{f_n, k'_n}^{(a(f))}} \\ \theta_{mn}^{(a(m))} = \frac{\sum_{(\vec{f}, o) \in \partial m} \sum_{\vec{k} \in \partial n} c_k(n) \omega_{\vec{f}, o}(\vec{k})}{c_m} & p_r(s) = \frac{\sum_{(\vec{f}, o) \in \partial s} \omega_{\vec{f}, o}(\vec{r})}{\sum_{(\vec{f}, o) \in R^o} \omega_{\vec{f}, o}(\vec{r})} \end{cases}$$

# SIMSBM(1)



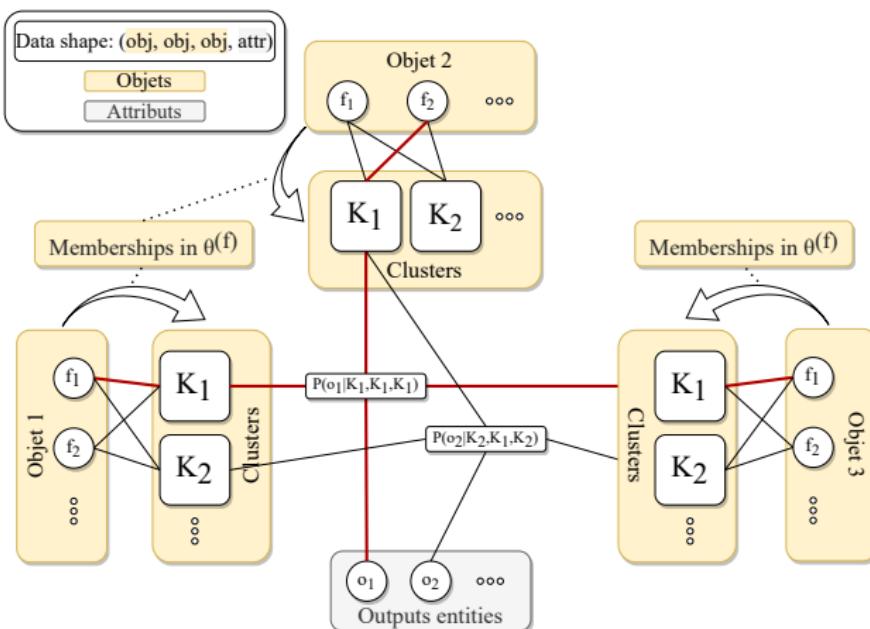
**Figure 3 – SIMSBM(1) – Deviner un attribut à partir d'un seul objet. Par exemple (fusaïole → féminin)**

# SIMSBM(2)



**Figure 4 – SIMSBM(2) – Deviner un attribut à partir d'une paire d'objets. Par exemple (fusaïole + flèche → masculin)**

# SIMSBM(3)



**Figure 5 – SIMSBM(3) – Deviner un attribut à partir de triplets d'objets. Par exemple (fibule + monnaie + oeuf → féminin)**

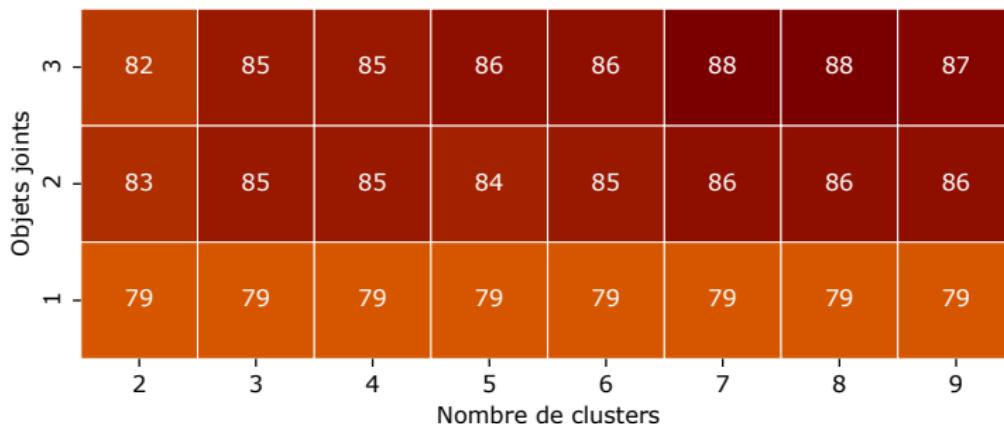
# Cross validation

- Peu de données d'entraînement et de test
- Vérifier la consistance des résultats : cross-fold validation
  - Élimine les biais statistiques



## Résultats - Sexe

### % de précision - Inférence du sexe



**Figure 6 – Résultats en prédiction du sexe pour différents nombres de clusters et d'interactions (en % de précision).**

- Corpus avec 50% d'individus masculin et 50% d'individus féminins
  - Fortes corrélations entre petit mobilier et sexe de l'individu

## Résultats - Âge

### % de précision - Inférence de l'âge

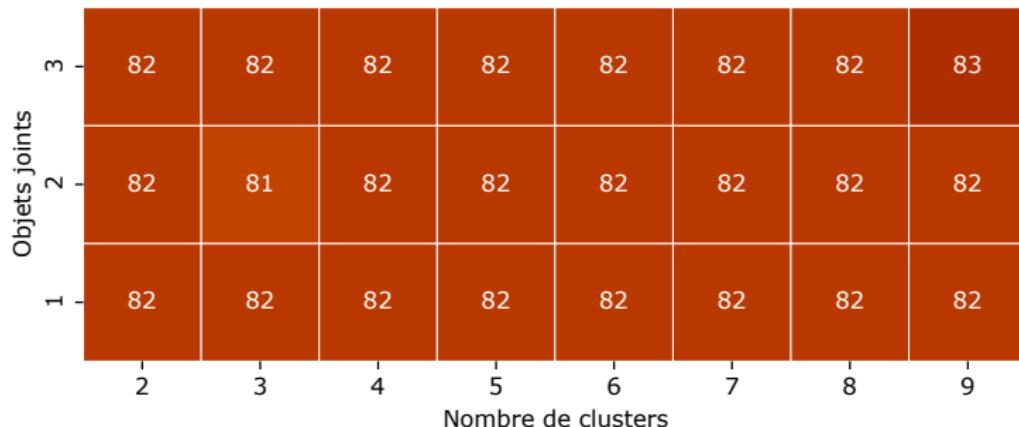


Figure 7 – Résultats en prédiction de la classe d'âge pour différents nombres de clusters et d'interactions (en % de précision).

- Corpus avec 81% d'adultes et 19% d'immatures
- Pas de corrélations détectées entre petit mobilier et classe d'âge

# Visualisation des données - SIMSBM(1)

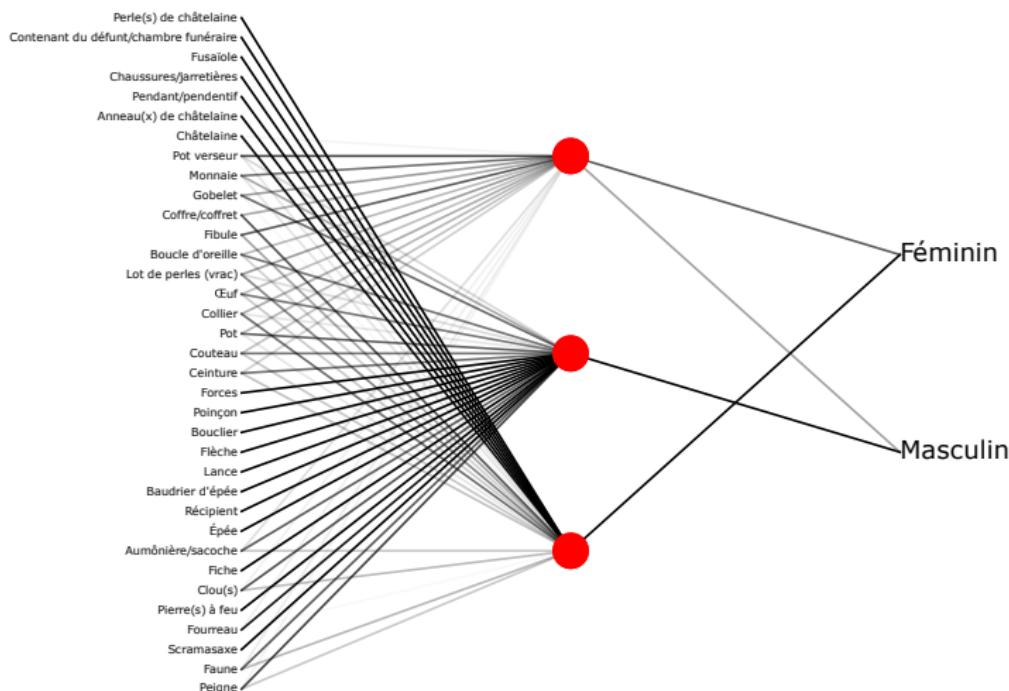


Figure 8 – Graphe inféré pour 3 clusters.

# Visualisation des données - SIMSBM(1)

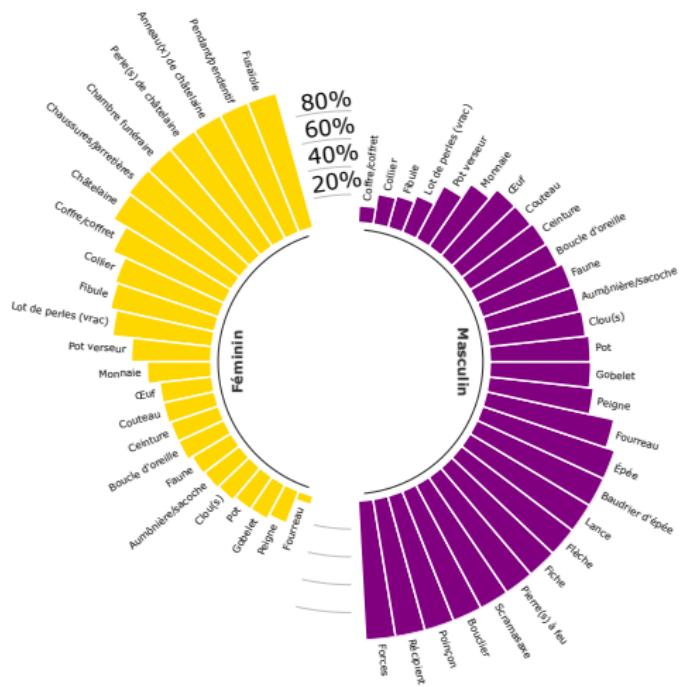


Figure 9 – Lien entre un objet individuel et le sexe d'un individu (% de prob.)

## Visualisation des données - SIMSBM(2)

## Probabilité d'un sexe

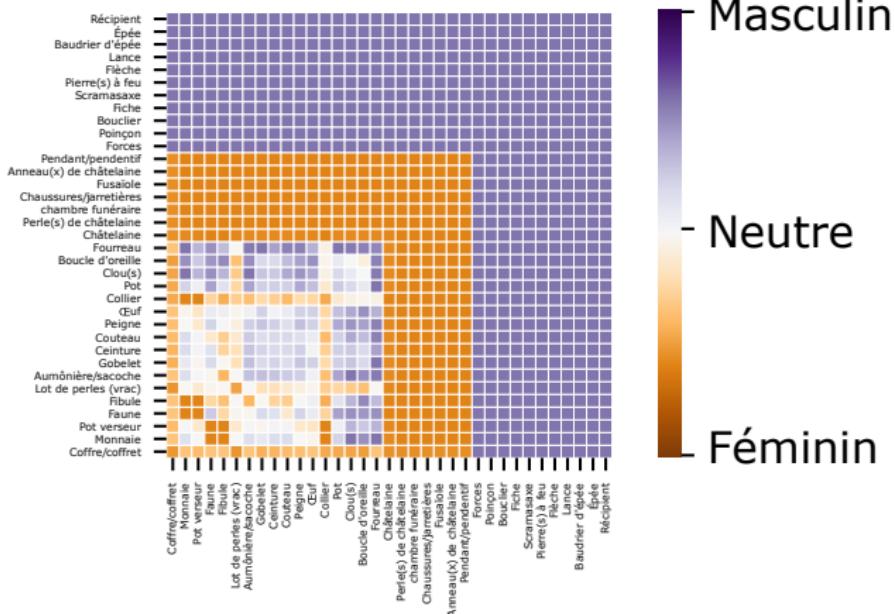
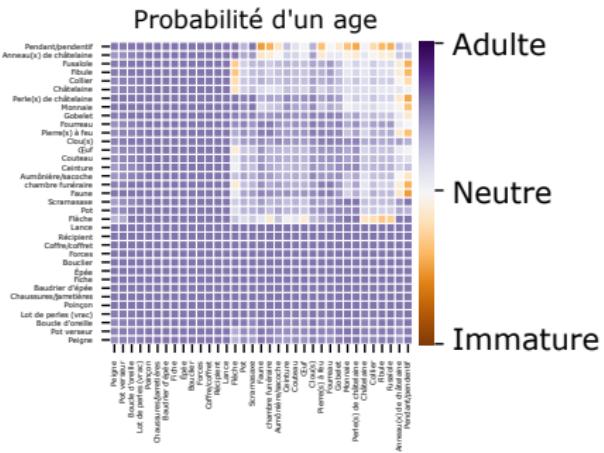
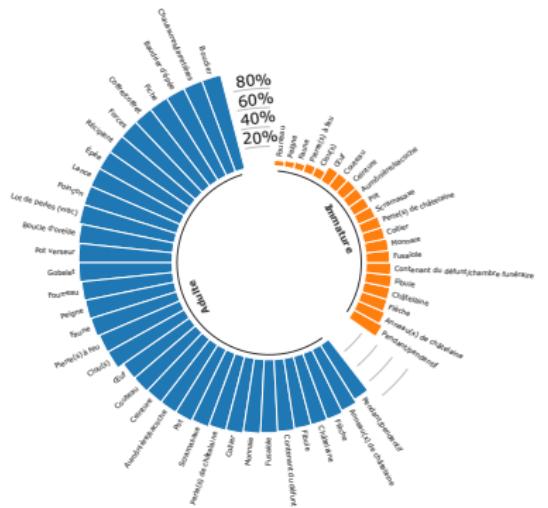


Figure 10 – Lien entre une paire d'objets et le sexe (% de prob). Bcp d'objets individuels 100% masculins, bcp de paires d'objets 100% féminines.

## Écueils

- Cas de l'inférence de la classe d'âge : plusieurs possibilités
    - Données non-équilibrées (trop peu de données)
    - Pas de corrélation avec le petit mobilier



**Figure 11 – SIMSBM(1) et SIMSBM(2) pour la classe d'âge.** Apparemment pas d'objets déterminants pour l'inférence de la classe d'âge.

# Conclusion

- Mise en valeur de :
  - Corrélations entre petit mobilier funéraire et le sexe d'un individu
  - Non-corrélations entre petit mobilier et classe d'âge
- Un outil d'apprentissage automatique
  - Identifier des relations potentiellement invisibles à l'oeil nu
  - Nouvelle méthode à disposition de l'archéologie
- Interprétation des résultats
  - Nécessité d'un regard critique (mathématique et archéologique) sur les résultats obtenus automatiquement

# Perspectives

- Promouvoir la transdisciplinarité :
  - Artefacts, Clauss-Slaby, PCR (IDF, CA, GE), ...
  - Apprentissage automatique comme outil
    - Logiciel de prédiction d'attributs à partir du petit mobilier ?
  - Formations universitaires en HN
- Augmenter la taille des corpus :
  - Apprentissage plus solide
  - Modèles plus détaillés
    - Dimension temporelle/spatiale
    - Données complexes (position de l'objet, matériaux, âge exact, etc.)
  - S'ouvrir à de nouvelles méthodologies (petit florilège "plug&play" ↓)

## 1.1. Linear Models

- 1.1. Ordinary Least Squares
- 1.1.1. Ridge regression and classification
- 1.1.2. Lasso
- 1.1.3. Multi-task Lasso
- 1.1.4. Multi-task Lasso
- 1.1.5. Elastic Net
- 1.1.6. Multi-task Elastic Net
- 1.1.7. Least Angle Regression
- 1.1.8. LARS Lasso
- 1.1.9. Orthogonal Matching Pursuit (OMP)
- 1.1.10. Bayesian regression
- 1.1.11. Logistic regression
- 1.1.12. Generalized Linear Regression
- 1.1.13. Stochastic Gradient Descent - SGD
- 1.1.14. Perceptron
- 1.1.15. Support Vector Machines
- 1.1.16. Robust regression: outliers and modeling
- 1.1.17. Quantile Regression

## 1.2. Linear and Quadratic Discriminant Analysis

- 1.2.1. Dimensionality reduction using Linear Discriminant Analysis
- 1.2.2. Mathematical formulation of the LDA and QDA classifiers
- 1.2.3. Mahalanobis formulation of LDA dimensionality reduction
- 1.2.4. Shrinkage and Covariance Estimator
- 1.2.5. Estimation algorithms

## 1.3. Kernel ridge regression

- 1.4.1. Classification
- 1.4.2. Regression
- 1.4.3. Density estimation, novelty detection
- 1.4.4. Complexity
- 1.4.5. Feature selection and Feature Use
- 1.4.6. Kernel functions
- 1.4.7. Mathematical formulation
- 1.4.8. Implementation details

## 1.5. Stochastic Gradient Descent

- 1.5.1. Classification
- 1.5.2. Regression
- 1.5.3. Multi-class One-Class SVM
- 1.5.4. Stochastic Gradient Descent for sparse data
- 1.5.5. Completeness
- 1.5.6. Stopping criterion
- 1.5.7. Tips on Practical Use
- 1.5.8. Mathematical formulation
- 1.5.9. Implementation details

## 1.6. Nearest Neighbors

- 1.6.1. Unsupervised Nearest Neighbors
- 1.6.2. Nearest Neighbors Classification
- 1.6.3. Nearest Neighbors Regression
- 1.6.4. Nearest Neighbor Algorithm
- 1.6.5. Nearest Centroid Classifier
- 1.6.6. Nearest Neighbors Transformer
- 1.6.7. Neighborhood Components Analysis

## 1.7. Gaussian Processes

- 1.7.1. Gaussian Process Regression (GPR)
- 1.7.2. GPR examples
- 1.7.3. Gaussian Process Classification (GPC)
- 1.7.4. GPC example
- 1.7.5. Kernels for Gaussian Processes

## 1.8. Cross decomposition

- 1.8.1. PLS/PCR
- 1.8.2. PLS/ID
- 1.8.3. PLS/Regression
- 1.8.4. Corinical Correlation Analysis

## 1.9. Naïve Bayes

- 1.9.1. Gaussian Naïve Bayes
- 1.9.2. Multinomial Naïve Bayes
- 1.9.3. Complement Naïve Bayes
- 1.9.4. Bernoulli Naïve Bayes
- 1.9.5. Categorical Naïve Bayes

## 1.10. Decision Trees

- 1.10.1. Classification
- 1.10.2. Regression
- 1.10.3. Multi-output problems
- 1.10.4. Complexity
- 1.10.5. Tips on practical use
- 1.10.6. Tree algorithms: ID3, C4.5, CS and CART

## 1.10.7. Minimum Cost-Complexity Pruning

## 1.11. Ensemble methods

- 1.11.1. Bagging meta-estimator
- 1.11.2. Forests of randomized trees
- 1.11.3. AdaBoost
- 1.11.4. Gradient Boosting
- 1.11.5. Histogram-Based Gradient Boosting
- 1.11.6. Voting Classifier
- 1.11.7. Voting Regressor
- 1.11.8. Stacked generalization

## 1.12. Multiclass and multioutput algorithms

- 1.12.1. Multiclass classification
- 1.12.2. Multilabel classification
- 1.12.3. Multiclass/multioutput classification
- 1.12.4. Multiclass regression

## 1.13. Feature selection

- 1.13.1. Recursive feature elimination
- 1.13.2. Feature selection via cross-validation
- 1.13.3. Sequential Feature Selection
- 1.13.4. Feature selection as part of a pipeline

## 1.14. Semi-supervised learning

- 1.14.1. Self training
- 1.14.2. Label Propagation

# Merci de votre attention !

- Pour creuser ce type de modèles :
  - Godoy-Lorite, Guimera, Moore, Sales-Pardo, *Accurate and scalable social recommendation using mixed-membership stochastic block models*, PNAS, 2016
  - Poux-Médard, Velcin, Loudcher, *Information Interactions in Outcome Prediction : Quantification and Interpretation using Stochastic Block Models*, RecSys 2021
  - Poux-Médard, Velcin, Loudcher, *Serialized Interacting Mixed Membership Stochastic Block Model*, ICDM 2022
- Plus de résultats sur mon GitHub : [https://github.com/GaelPouxMedard/SIMSBM\\_Merovingien](https://github.com/GaelPouxMedard/SIMSBM_Merovingien)
- Mail : gael.poux@yahoo.fr