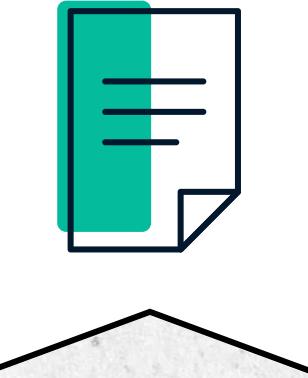




BORCELLE  
UNIVERSITY

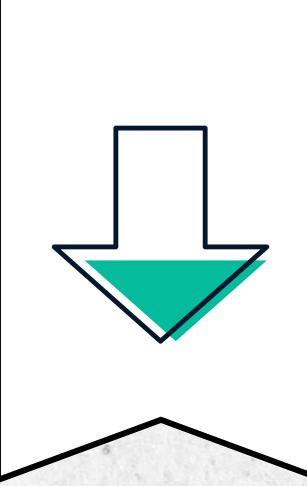


Cherkaoui Nabil/Ait Youb Abdelmoughit

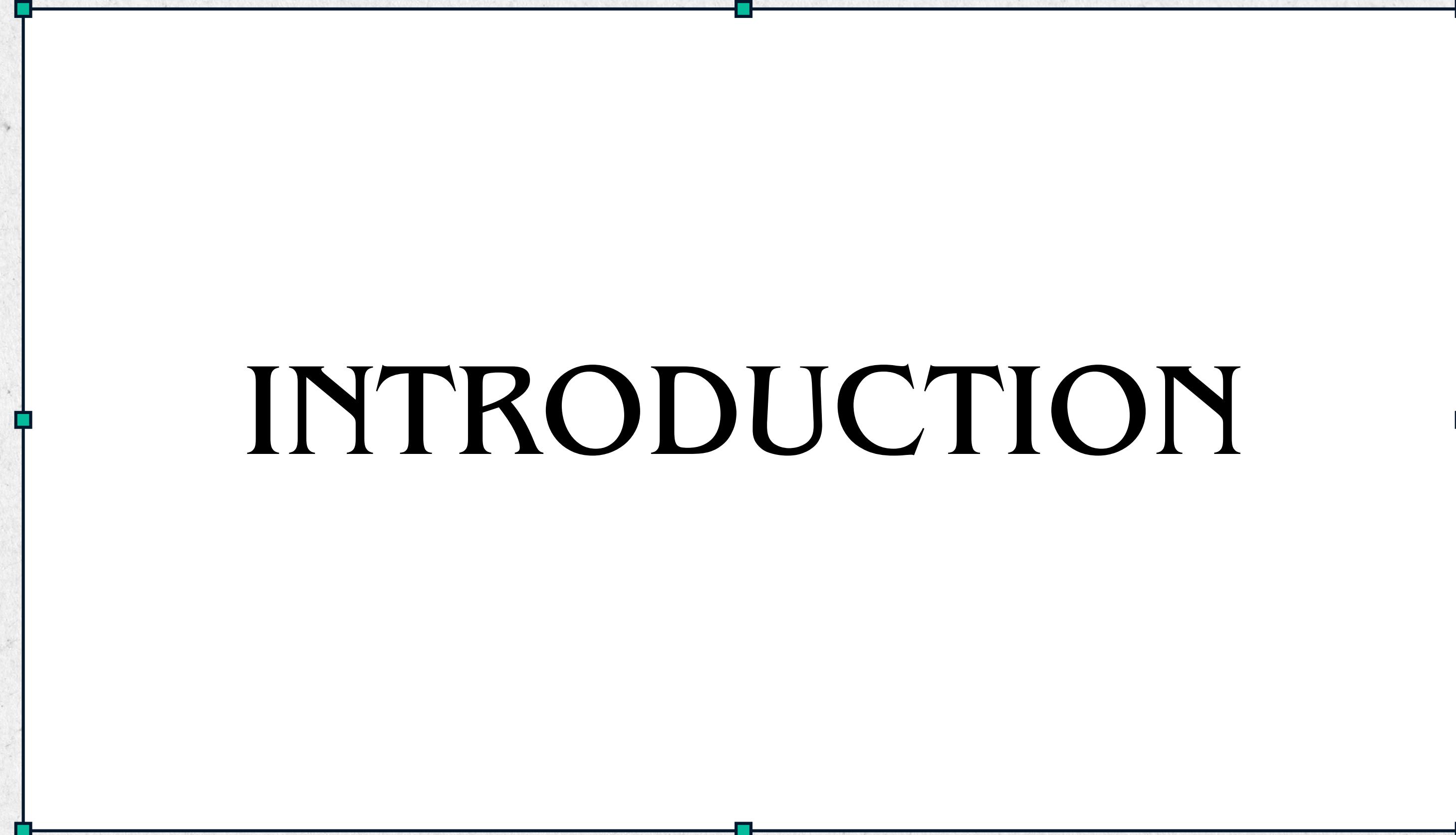


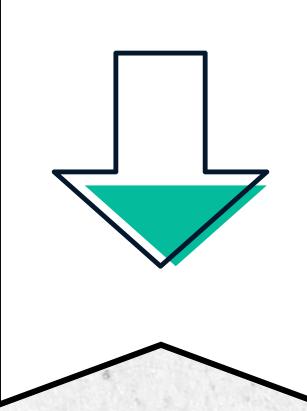
# SOMMAIRE

- Introduction
- Objectifs RD
- Graphe EuroRoads
- PCA
- LLE
- UMAP
- Conclusion



# INTRODUCTION



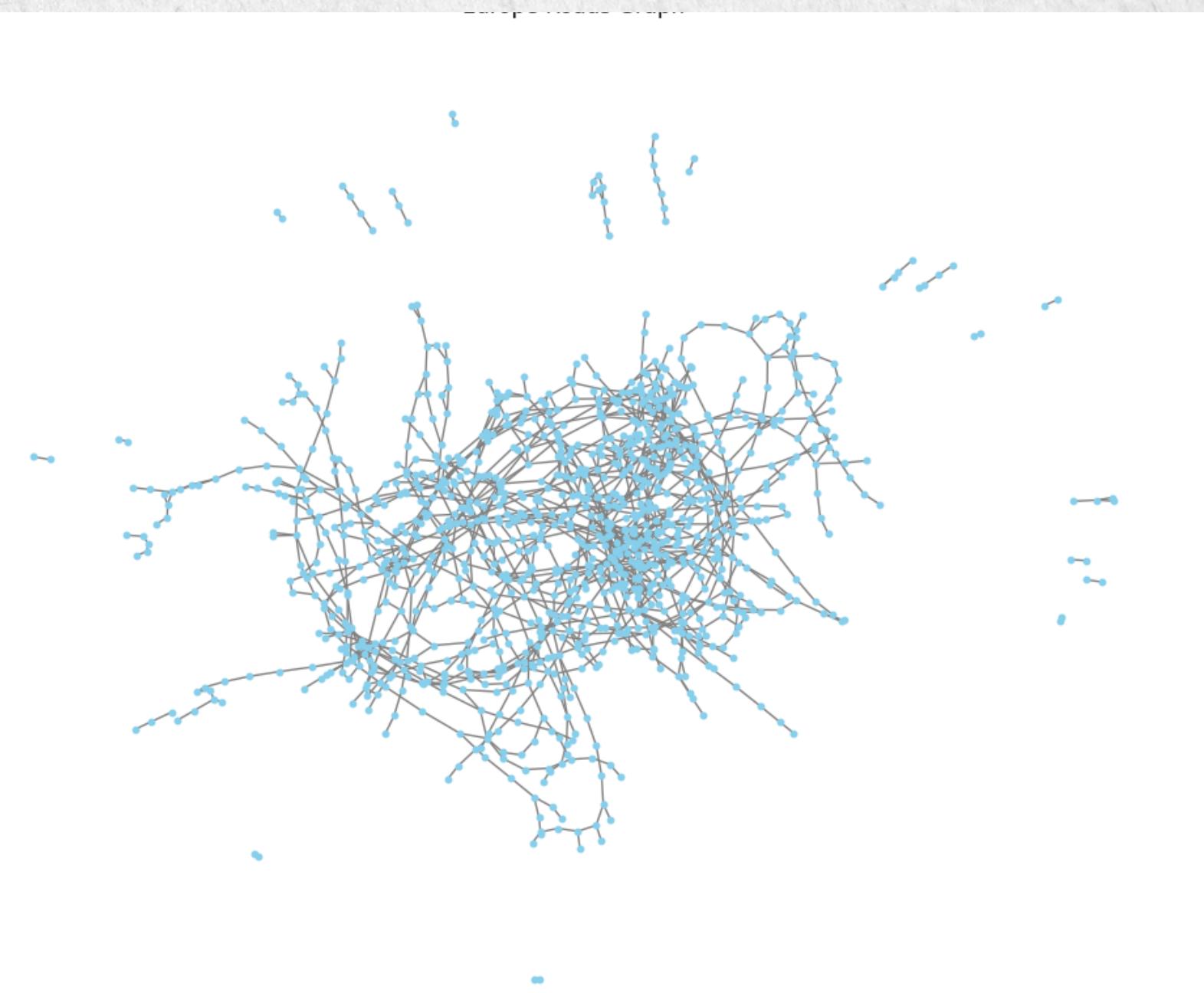
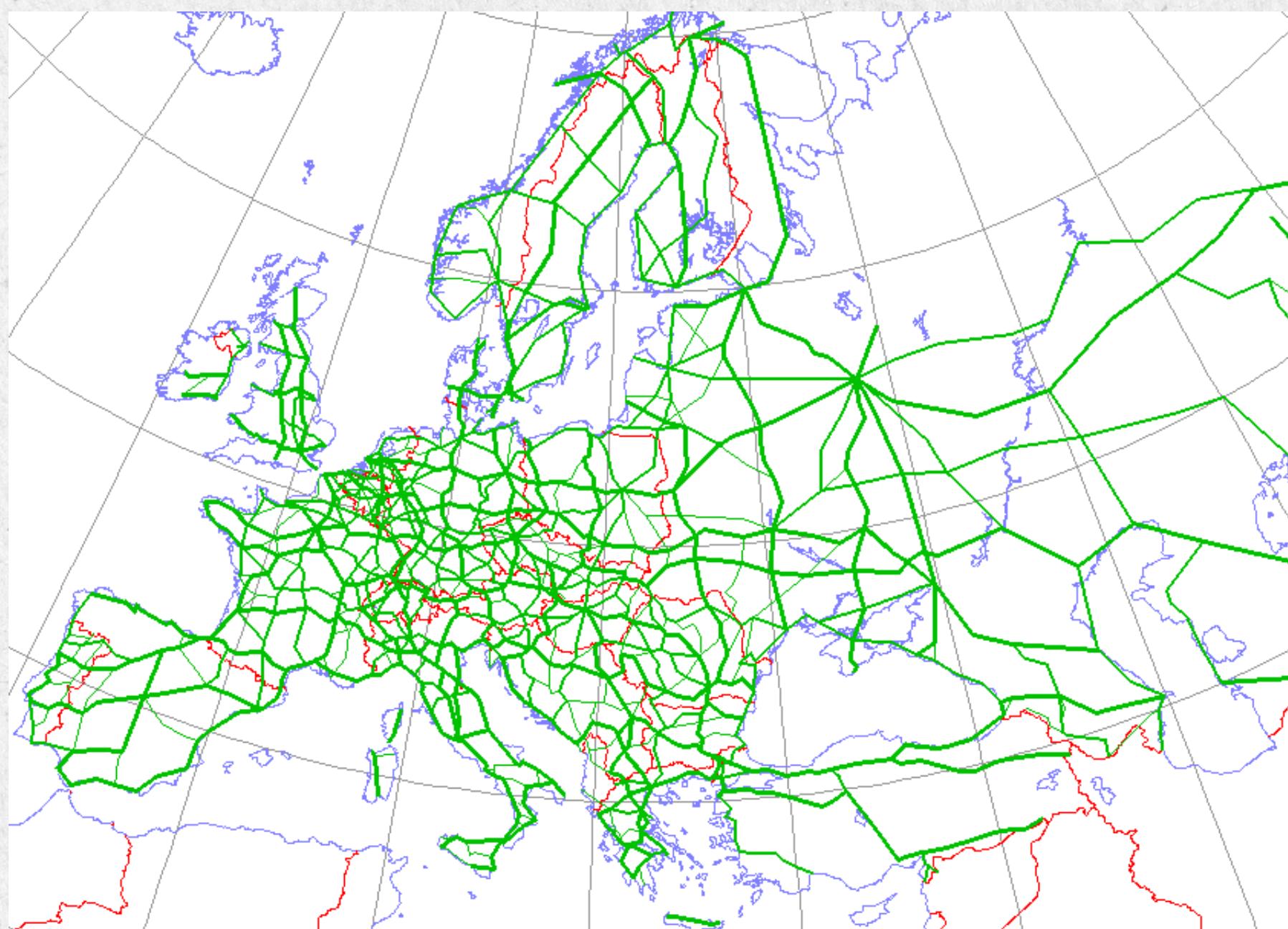


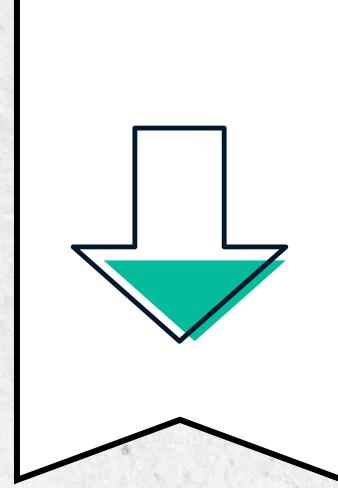
# OBJECTIFS DE LA RD

- Visualisation Intuitive
- Simplification des données et réduction de bruits
- Gains d'espace et de temps
- Détection de structures

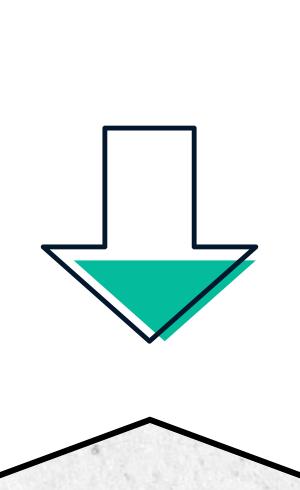


# GRAPHE EUROADS



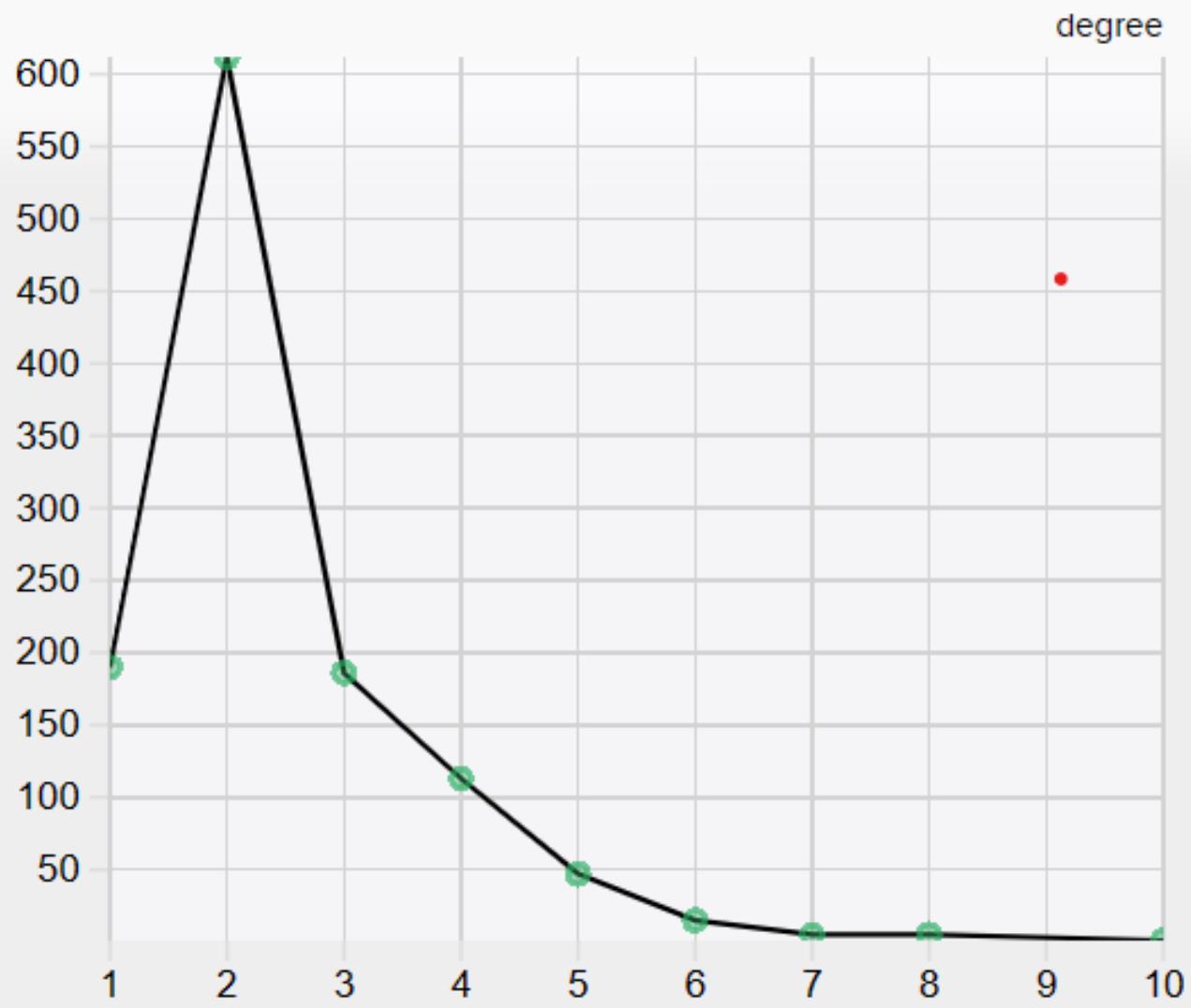


- Nombre de Noeuds : 1174
- Nombre d'Arêtes : 1417
- Densité : 0,00205794
- Degré Maximum : 10
- Degré Minimum : 1
- Degré Moyen : 2

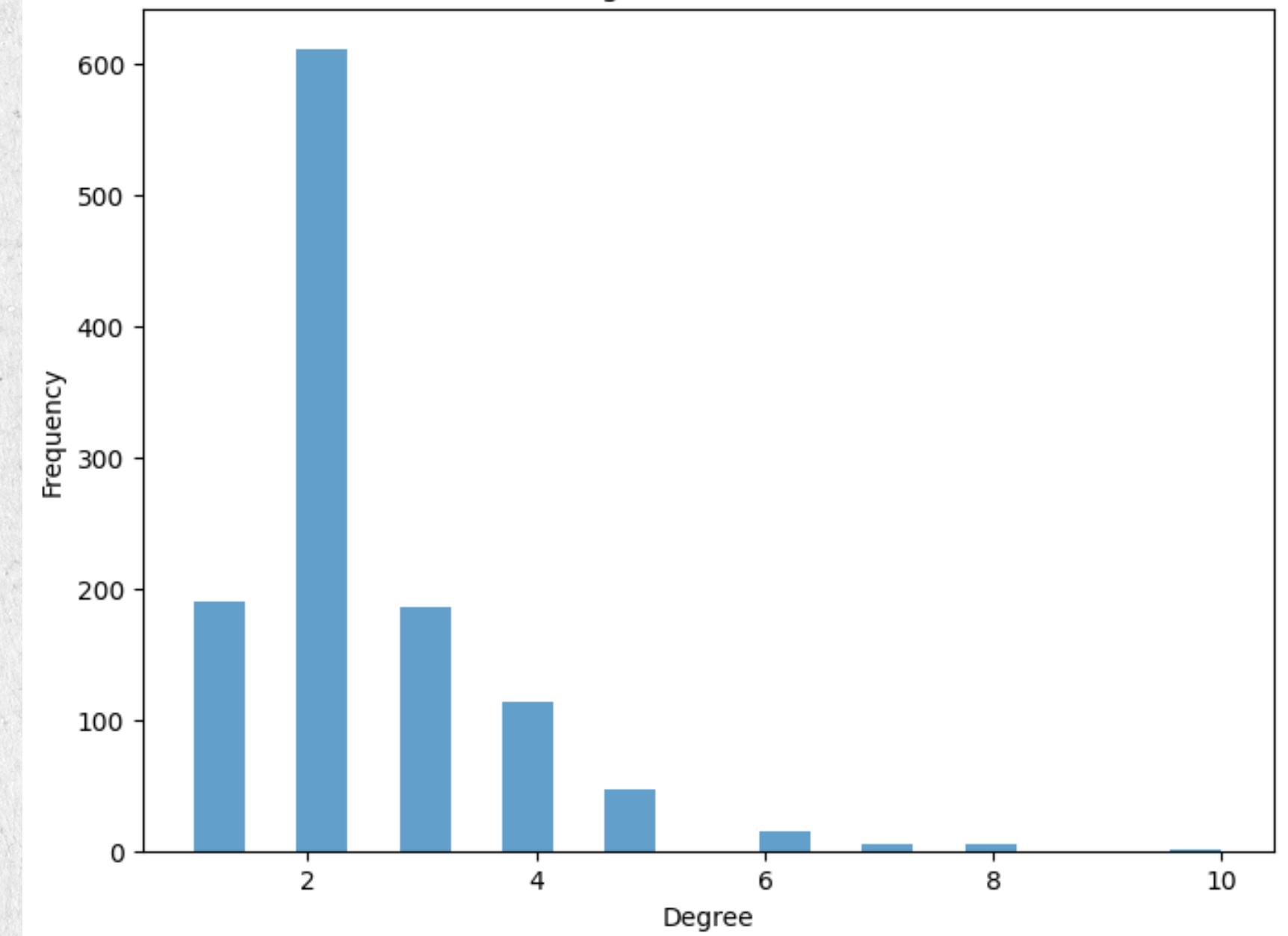


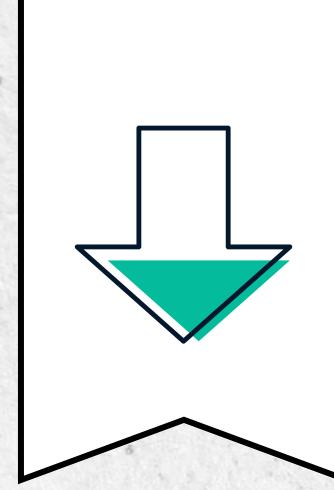
# DEGREE DISTRIBUTION

degree distribution



Degree Distribution





# PCA

PCA (Analyse en Composantes Principales): est une méthode de réduction de dimensionnalité qui vise à trouver une nouvelle base de vecteurs orthogonaux, appelés **composantes principales**, qui capturent la variance maximale des données.



# FONCTIONNEMENT

- Standardisation des données
- Calcul de la Matrice de Covariance
- Calcul des Vecteurs Propres et des Valeurs Propres
- Sélection des Composantes Principales

Formule des Composantes Principales :

$$PC = X \cdot V$$

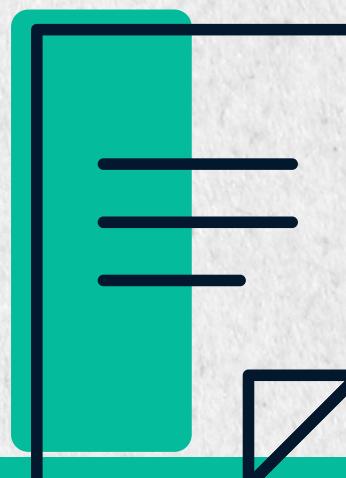
Formule de la Matrice de Covariance :

$$C = \frac{1}{m-1} \cdot (X - \bar{X})^T \cdot (X - \bar{X})$$

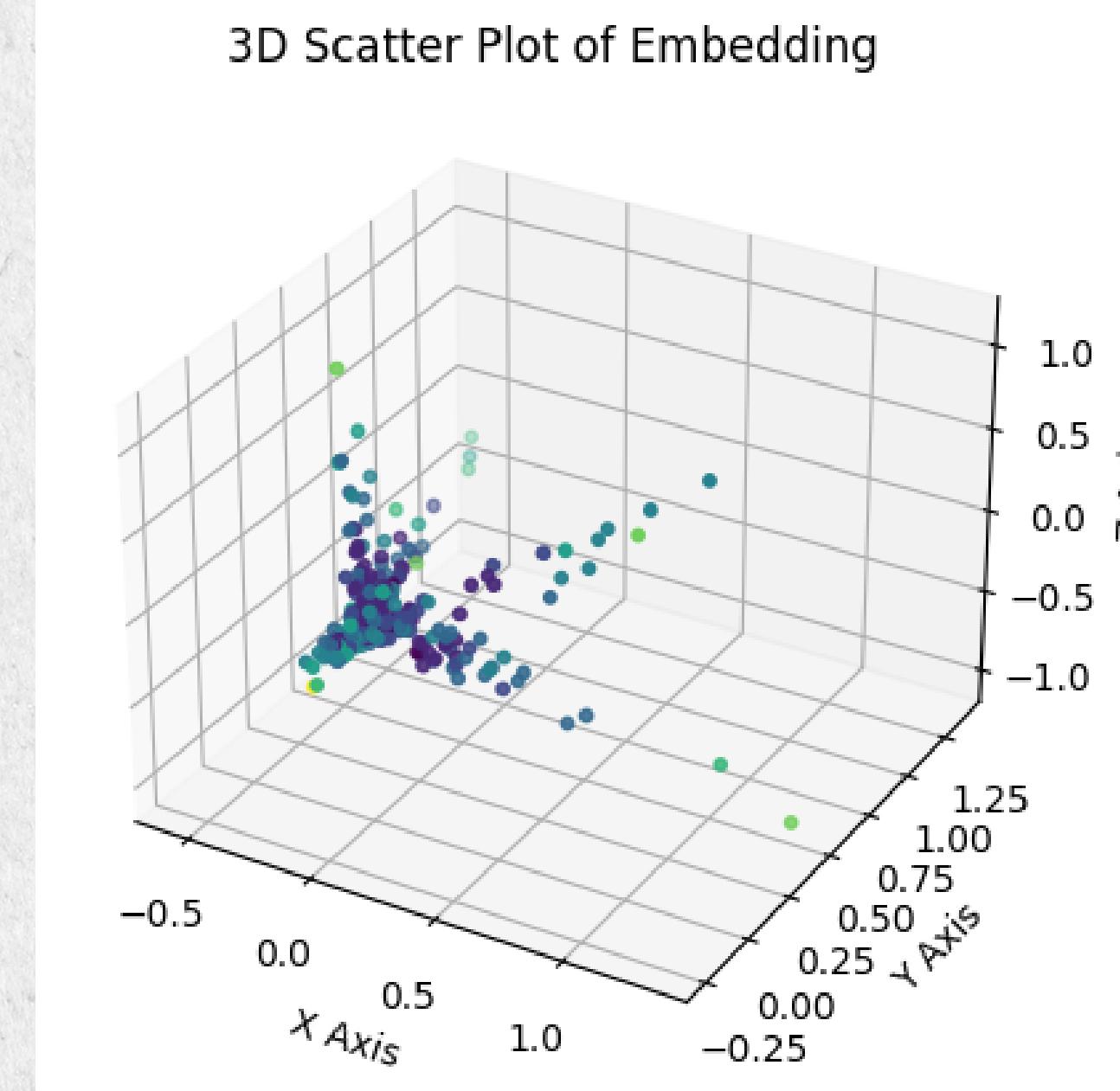
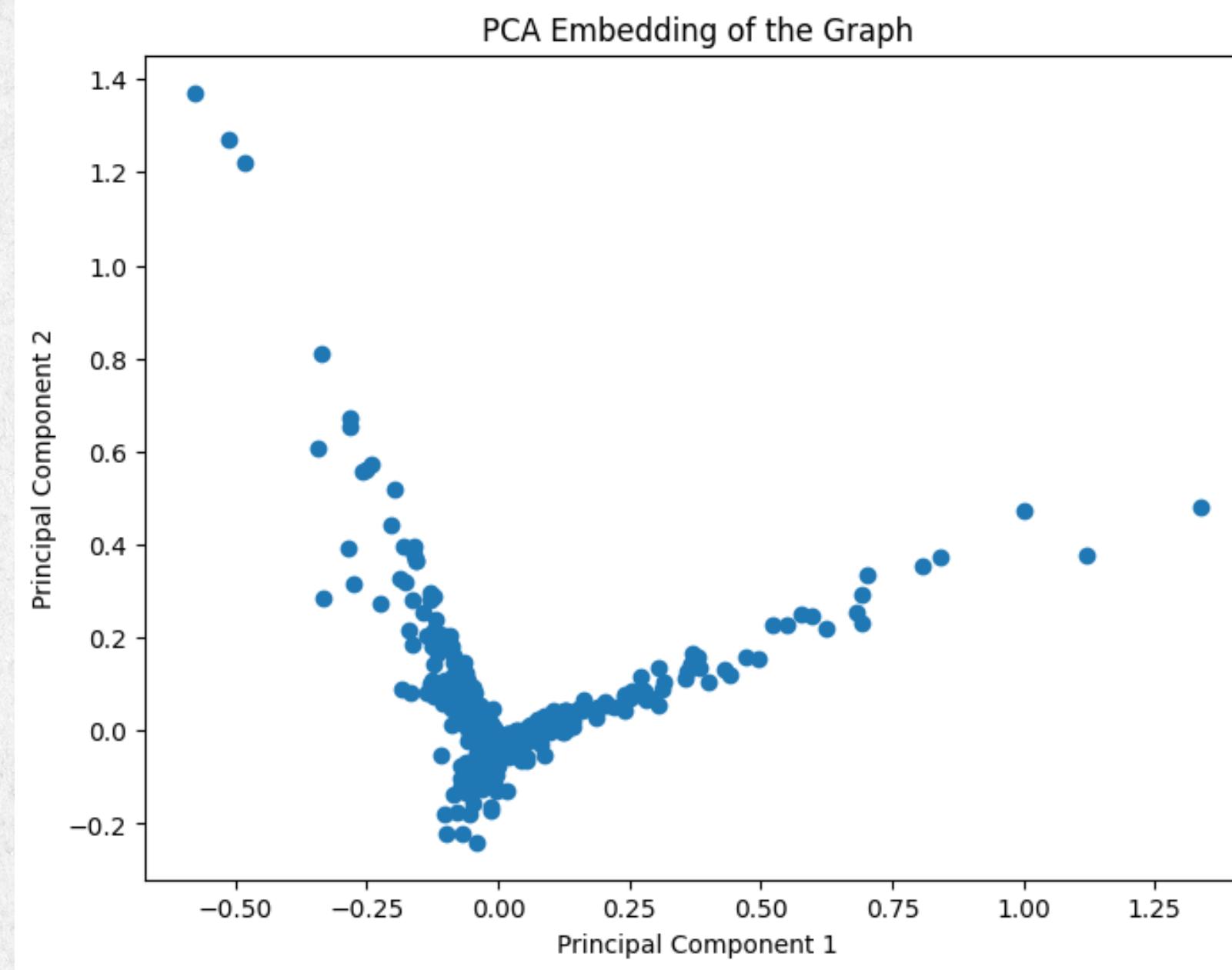
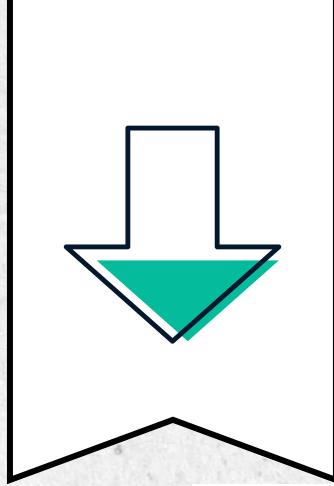
Formule des Valeurs Propres :

$$C \cdot v_i = \lambda_i \cdot v_i$$

Où  $V$  est la matrice des vecteurs propres (composantes principales) et  $\lambda_i$  est la i-ème valeur propre correspondante.



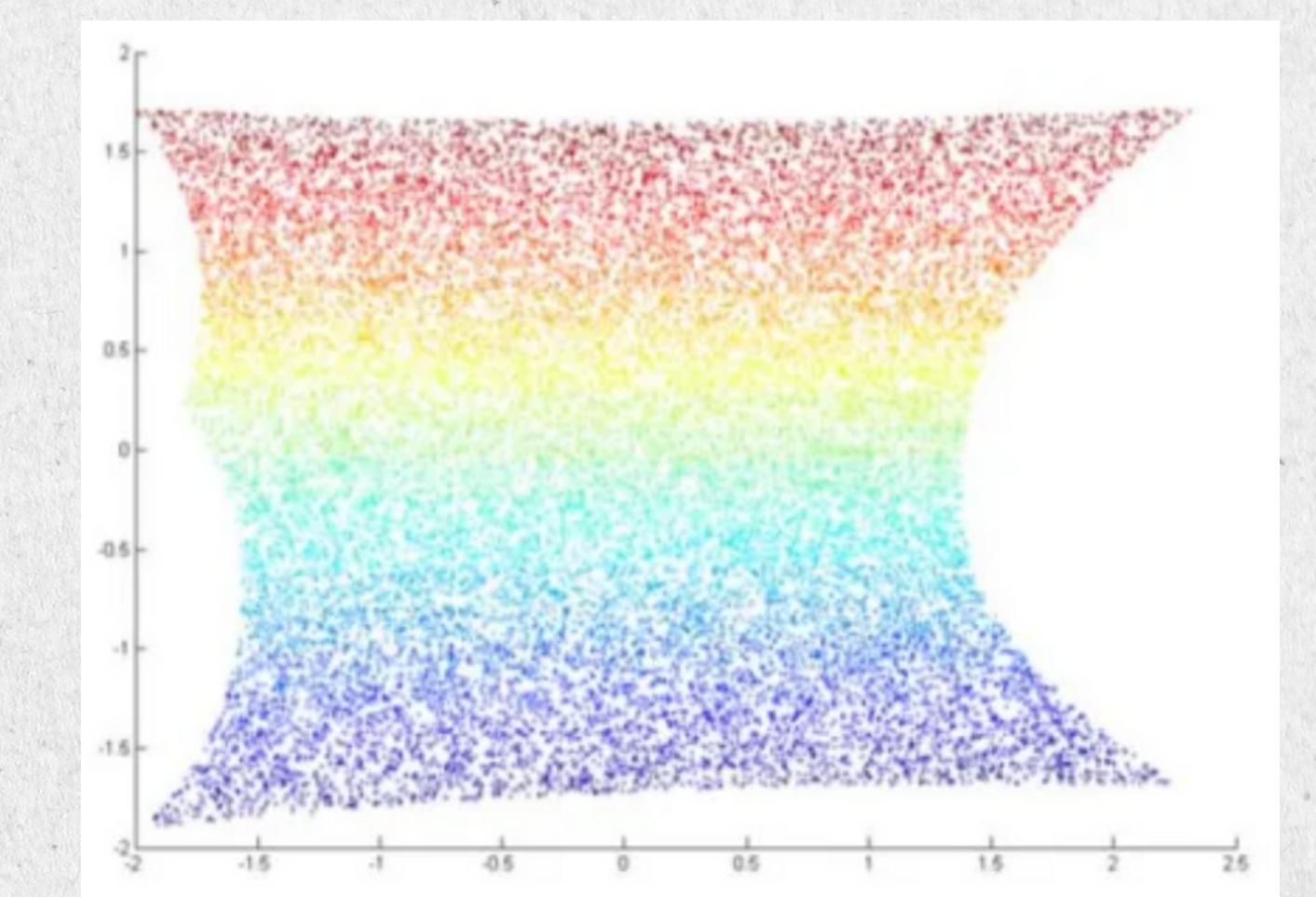
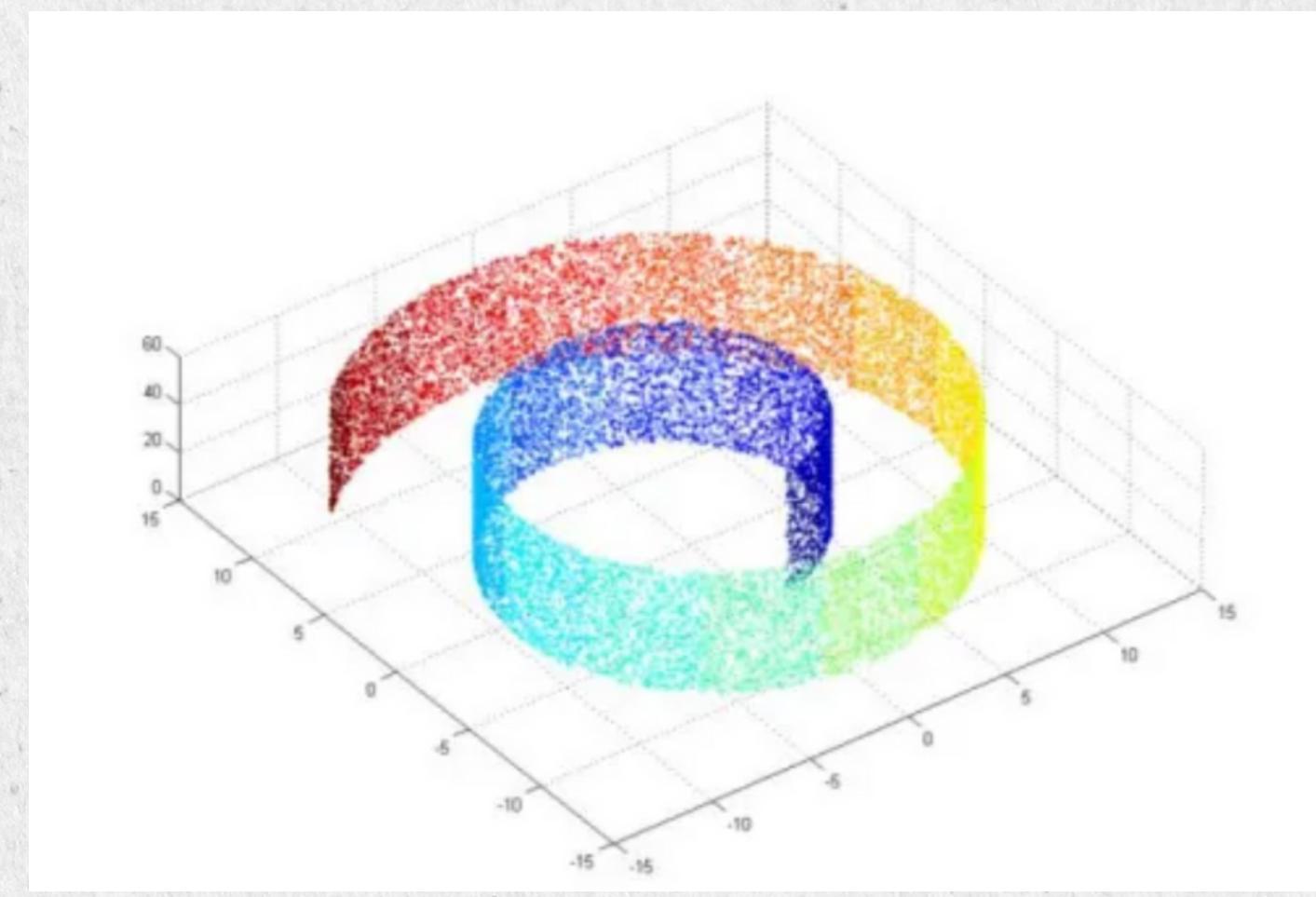
# PCA APPLICATION



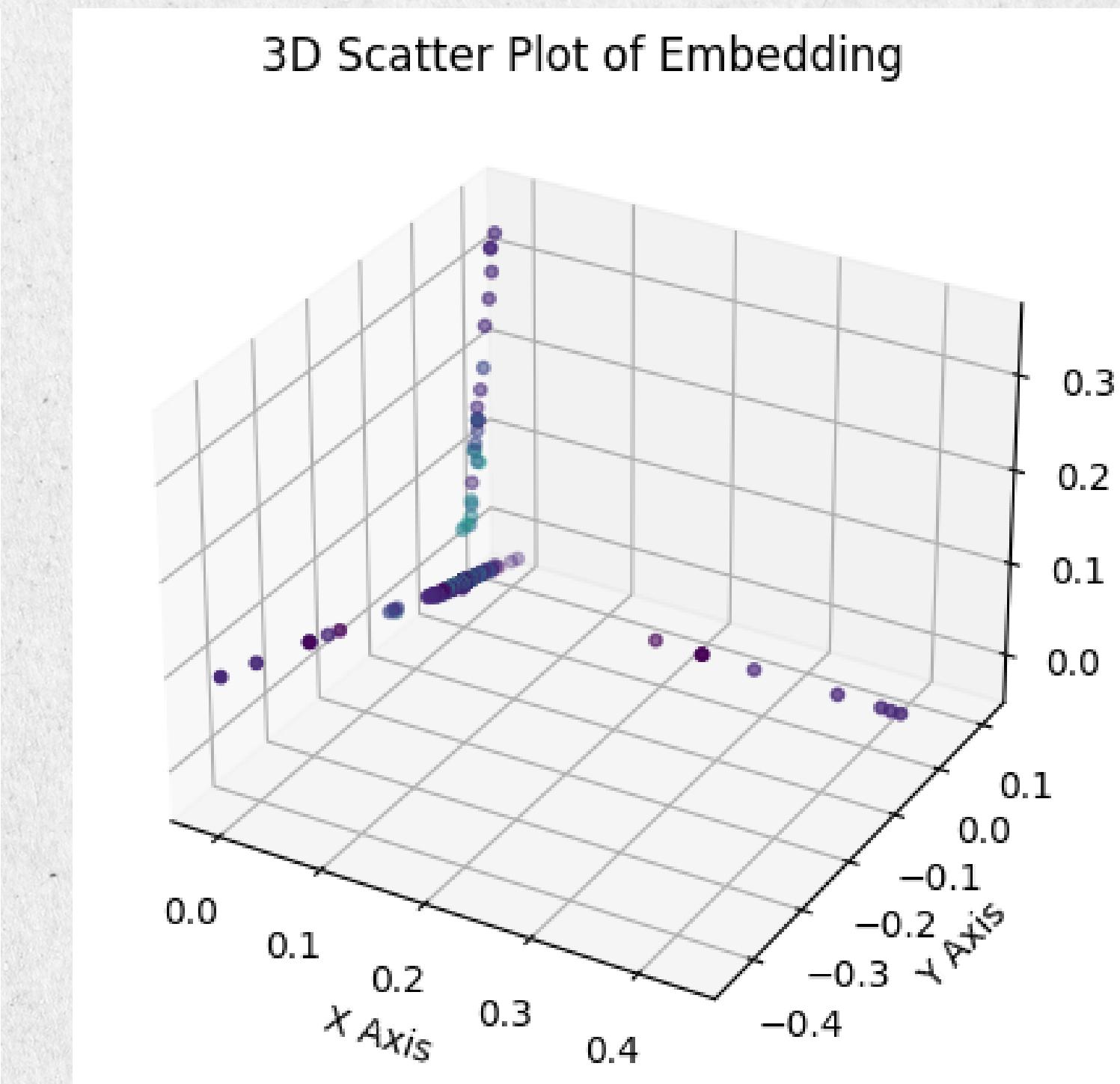
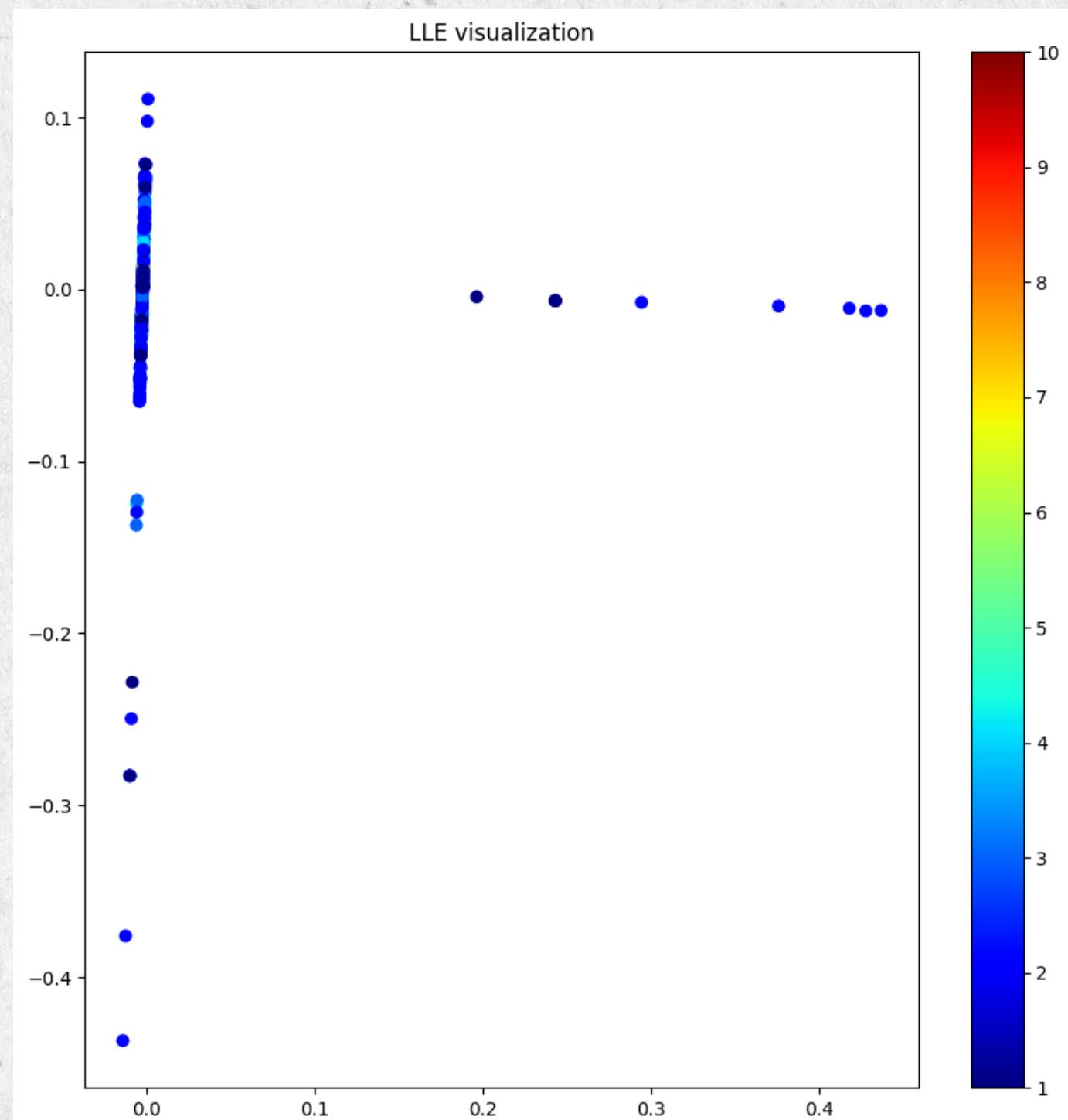
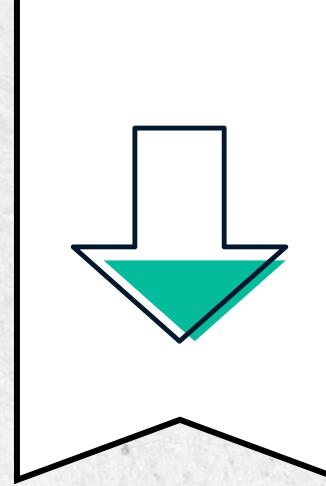


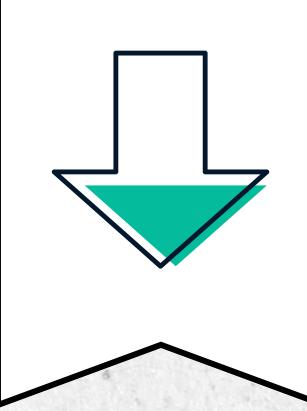
# LLE

LLE (Locally Linear Embedding) : est une méthode non linéaire qui préserve les relations locales entre les nœuds du graphe. Bien qu'il n'y ait pas de formule simple, LLE repose sur des calculs de graphe complexes pour créer une représentation de faible dimension.



# LLE APPLICATION





# UMAP

UMAP (Uniform Manifold Approximation and Projection) est un algorithme de réduction de dimensionnalité non linéaire qui s'est révélé être un outil puissant pour la visualisation et l'analyse de données de haute dimension.



# FONCTIONNEMENT

- Similarité dans l'ensemble de départ
- Similarité dans l'ensemble réduit d'arrivée
- Optimisation avec SGD



# SIMILARITÉ DANS L'ENSEMBLE DE DÉPART

- Concept similaire à LLE avec comme différence une notion de voisinage probabiliste.
- On calcule la probabilité qu'un point  $x_j$  soit voisin de  $x_i$  pour tout  $(i,j)$

$$p_{j|i} = \exp\left(-\frac{\|x_i - x_j\|_2 - \rho_i}{\sigma_i}\right)$$

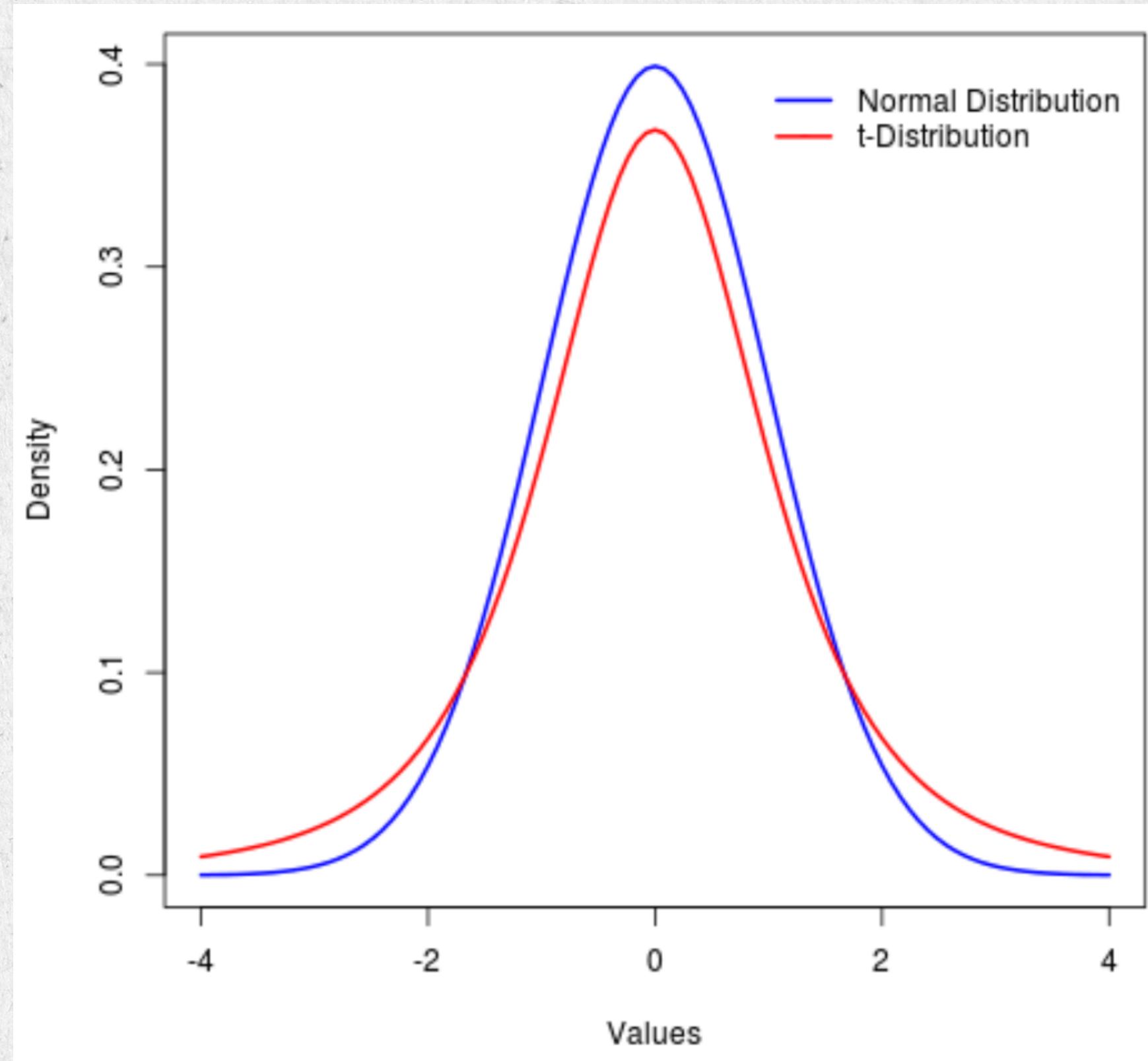


# SIMILARITÉ DANS L'ENSEMBLE D'ARRIVÉE

- La probabilité utilisée dans l'ensemble de départ pose des complications dans l'ensemble d'arrivée
- UMAP va utiliser une nouvelle probabilité inspirée de la loi de t-Student

$$q_{i,j} = \frac{1}{1 + a(y_i - y_j)^{2b}}$$

# GAUSS VS T-STUDENT



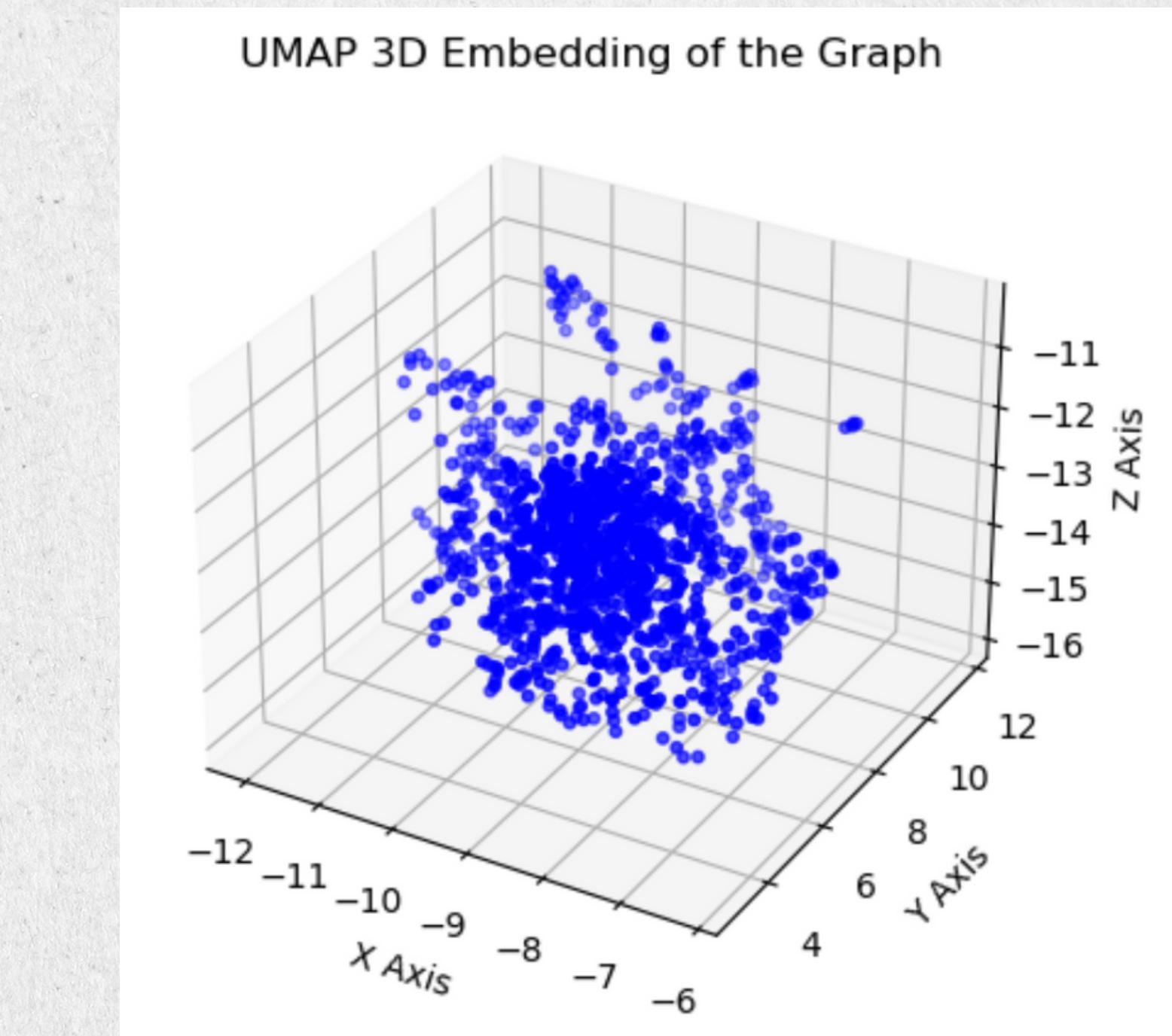
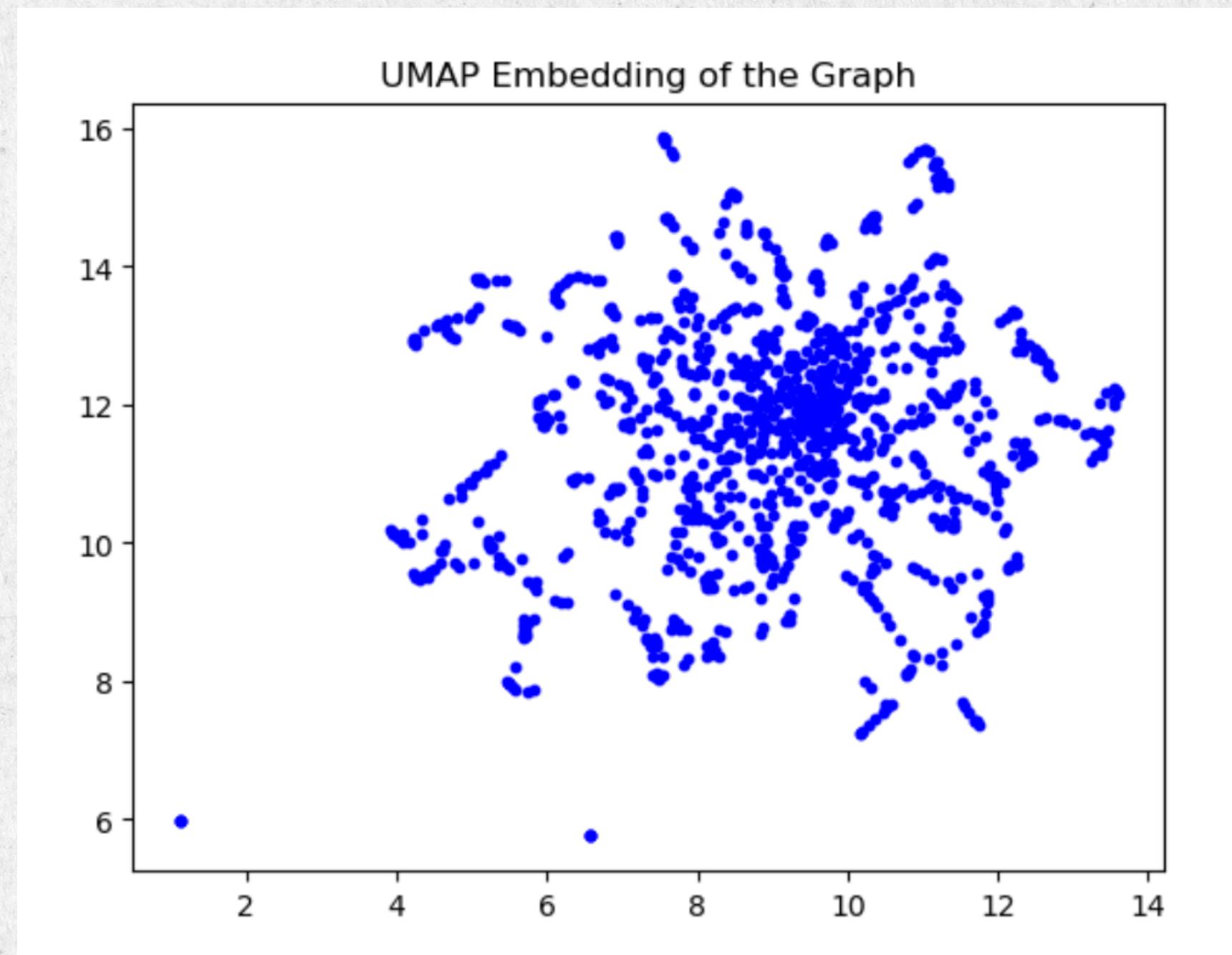
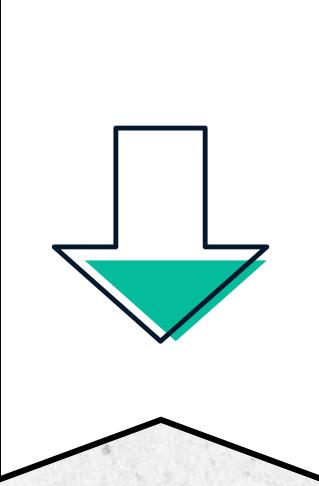


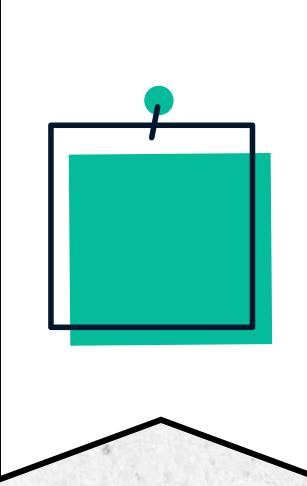
# OPTIMISATION

- L'objectif est de faire tendre la distribution de  $q$  vers celle de  $p$
- Cela revient à minimiser une fonction objectif d'entropie croisée
- SGD est utilisé comme algo d'optimisation

$$CE(X, Y) = \sum_{i=1}^N \sum_{j=1}^N \left[ p_{ij} \log \left( \frac{p_{ij}}{q_{ij}} \right) + (1 - p_{ij}) \log \left( \frac{1 - p_{ij}}{1 - q_{ij}} \right) \right]$$

# UMAP APPLICATION





**MERCI  
DE VOTRE  
ATTENTION**

