

# Algorithmes de clustering

...

Projet Sciences des Réseaux - ENSEEIHT  
Années 2024-2025 - Semestre 8

# Plan

- Le problème - Mathis
- Algorithmes de clustering - Ambre
- Démonstration + Visualisation des résultats - Ambre/Mathis
- Remise en question - Capucine
- Autres pistes - Capucine/Ambre

# Le problème

# Le problème initial

- Ce projet vise à construire des clusters d'utilisateurs qui serviront à pointer les faisceaux d'une constellation de satellites à orbites basses.
- Données fournies : la carte des demandes (CSV file)

## Hypothèses et contraintes

- Clusters de 90km de diamètre
- Débits max (1 Gbps, 2 Gbps et 4 Gbps)

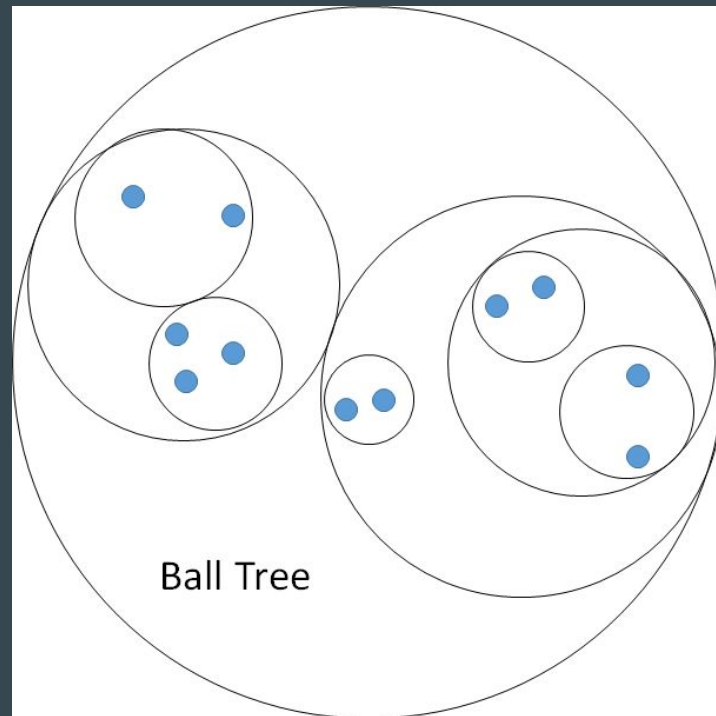
# Algorithmes de clustering

# Algorithme Ball-Tree

- Utilisation d'une structure de données arborescente pour séparer les points
- Clustering en fonction des séparations

## Propriétés de l'algorithme

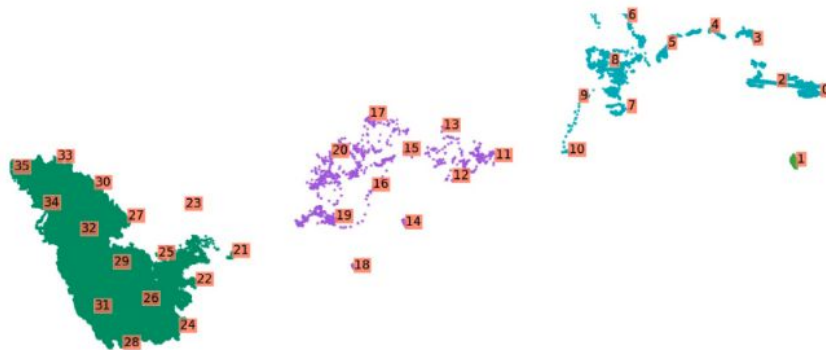
- Algorithme récursif (séparation récursive des points pour former des sous-arbres)
- Priorité aux points les plus densément entourés
- Si un point n'est pas encore clusterisé, on essaie de former un cluster à partir de ses voisins



# Algorithme Ball-Tree - Etat de l'art

- Ivan Šimeček and Ivan Šimeček -> A parallel algorithm for approximating the silhouette using a ball tree [lien](#)
- Xinye Chen and Stefan Güttel -> Fast and explainable clustering based on sorting
  - 2028780 data points to cluster with ball tree [lien](#)

Output:



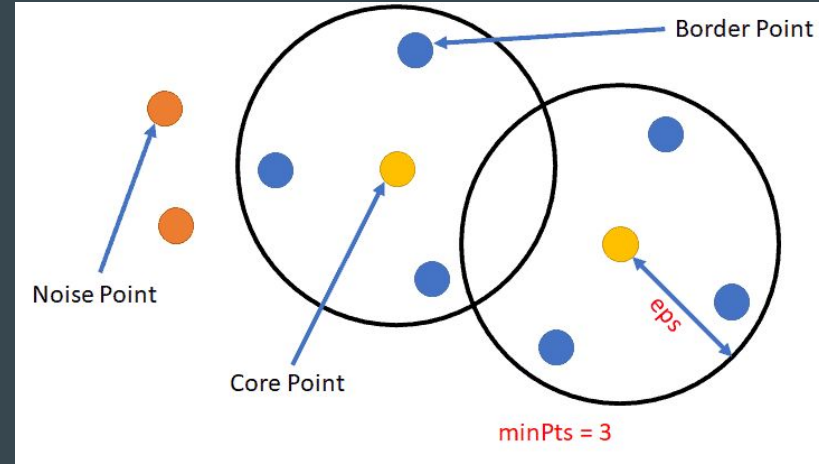
A clustering of 2028780 data points with 2 features has been performed.

# Algorithme DBSCAN

- Density-Based Spatial Clustering of Applications with Noise
- Regroupe les points en cluster s'ils sont suffisamment proches et nombreux
- Tri des points selon 3 catégories (core, border, noise), puis création des clusters / agrégation

## Propriétés de l'algorithme

- Traitement des points dans l'ordre du fichier fourni
- Agrégation par propagation (exploration des proches voisins)





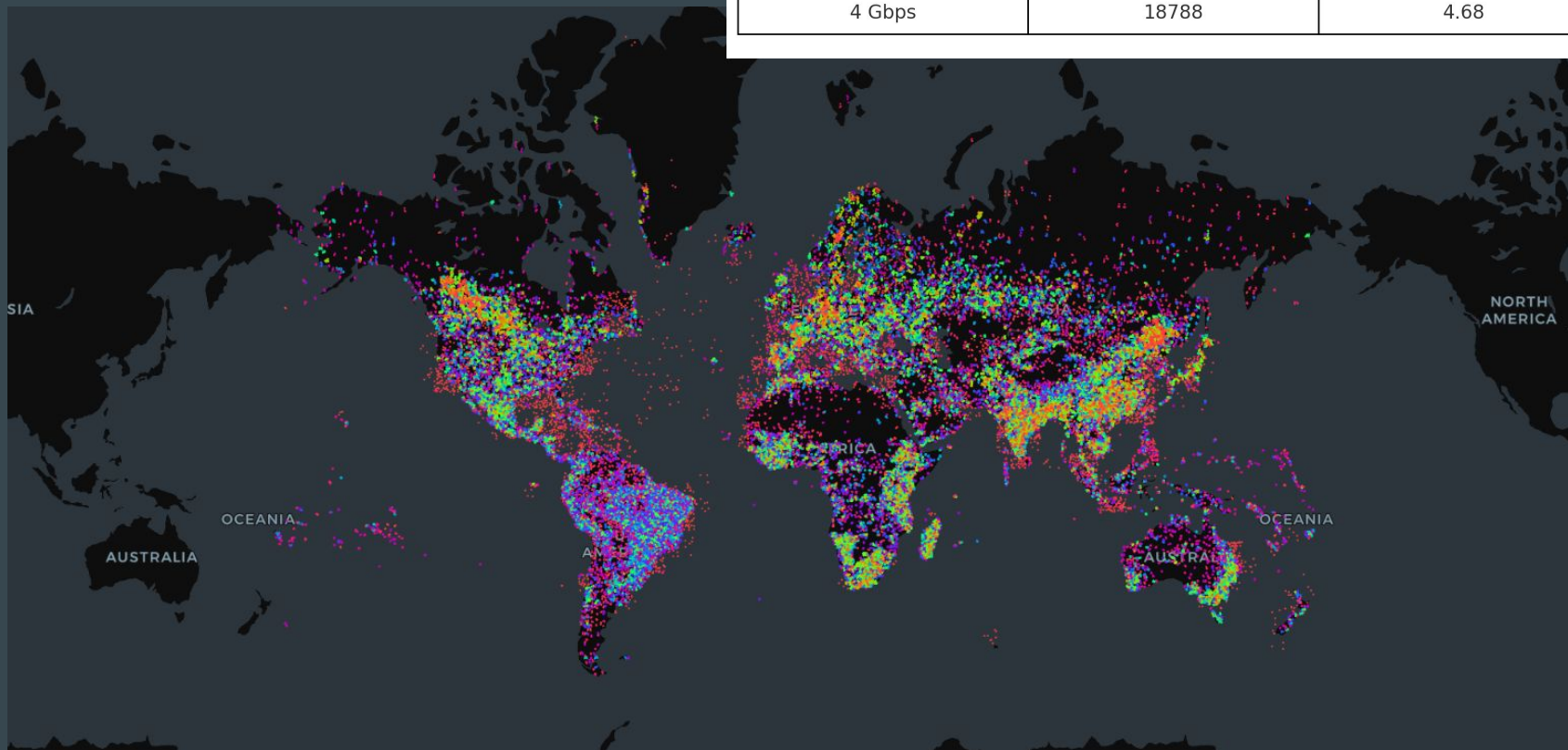
# Algorithme DBSCAN - Etat de l'art

- Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu -> A density-based algorithm for discovering clusters in large spatial databases with noise [lien](#)
- Ahmad M. Bakr, Naia M. Ghanem, Mohamed A. Ismail -> Efficient incremental density-based algorithm for clustering large datasets [lien](#)

# Démonstration + Visualisation des résultats

# Algorithme Ball Tree

Débit	Nb de clusters	Nb d'utilisateurs par cluster
1 Gbps	40408	2.18
2 Gbps	24482	3.59
4 Gbps	18788	4.68



# Algorithme Ball Tree - 1 Gbps





# Algorithme Ball Tree - 2 Gbps

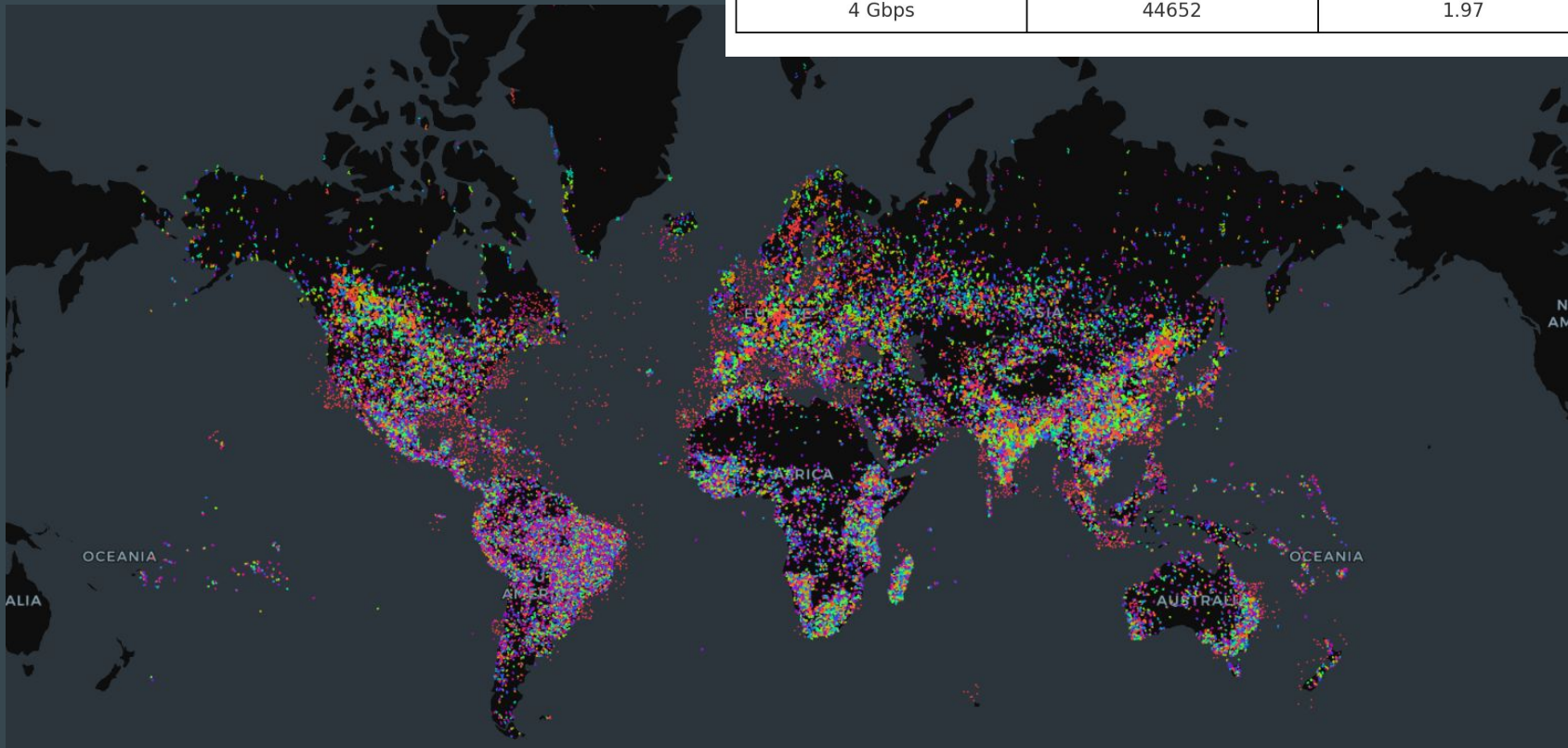


# Algorithme Ball Tree - 4 Gbps



# Algorithme DBSCAN

Débit	Nb de clusters	Nb d'utilisateurs par cluster
1 Gbps	46979	1.87
2 Gbps	44802	1.96
4 Gbps	44652	1.97





# Algorithme DBSCAN - 1 Gbps





# Algorithme DBSCAN - 2 Gbps



# Algorithme de Ball Tree - 4 Gbps



# Conclusion des algorithmes

# Conclusion des algorithmes

## Ball Tree

- Bonne rapidité
- Nombre de clusters largement inférieur à DBSCAN

## DBSCAN

- Variation très faible du nombre de clusters lorsque le débit maximal autorisé change
- Points de bruits sont considérés comme des clusters individuels

**Remise en question**

# Remise en question

Faut-il vraiment chercher à minimiser le nombre de clusters ?

- Perte de redondance
- Perte de ressources / qualité de service
- Moins bonne adaptabilité

Mais

- Réduction du nombre de satellites nécessaires
- Utilisation la plus efficace possible de la taille des clusters

**Autres pistes**

# Algorithme K-Means

- Algorithme des K plus proches voisins
- Très connu
- Clustering uniquement par la distance
- Peu adapté au problème : l'algorithme prends en paramètre le nombre de clusters souhaités, or c'est l'objectif du projet

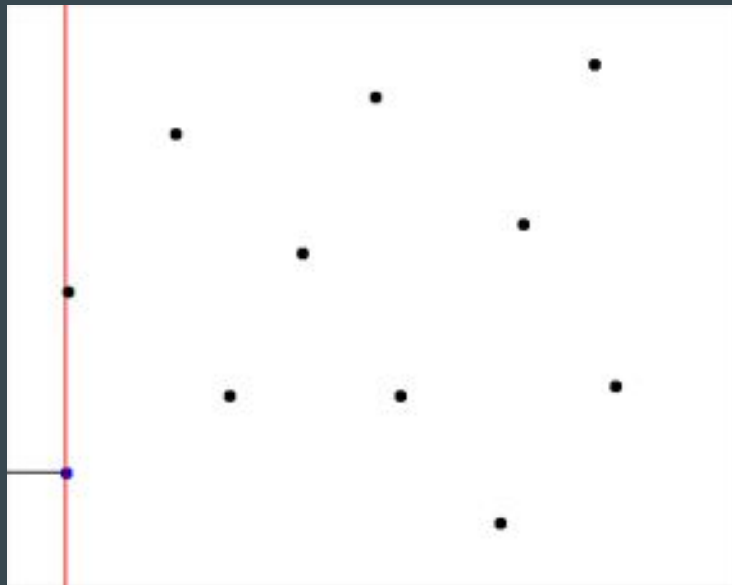


# Algorithme de Sweep

- Balayage des points en partant d'un centre arbitraire (méthode de balayage variable)
- Ajout des points à un cluster si les contraintes sont respectées (distance minimale)

## Propriétés de l'algorithme

- Algorithme itératif
- Insensible aux formes “anormales”  $\neq$  k-means



# Algorithme Sweep Line - Etat de l'art

- Krista Rizman Žalik, Borut Žalik -> A sweep-line algorithm for spatial clustering [lien](#)
- Krista Rizman Žalik, Borut Žalik, Niko Lukac -> Sweep-Hyperplane Clustering Algorithm Using Dynamic Model [lien](#)

