

Présentation pour la réunion 3

1. K-means pour données géographiques :

- On a exécuté l'algorithme de k-means sur les données géographiques des 4 villes: Paris, Lille, Nantes, Lyon.
- La fonction utilisée pour le clustering est la distance Euclidienne.
- L'évaluation du résultat d'un clustering, c'est-à-dire la mesure de la "qualité" des clusters obtenus, s'effectue selon 2 critères :

Compacité $D(C_k)$: qui représente la distance maximale entre deux RRH du même cluster.
Que l'on souhaite diminuer.

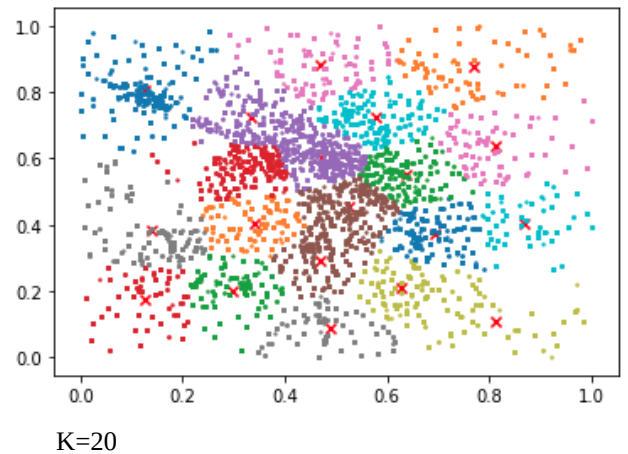
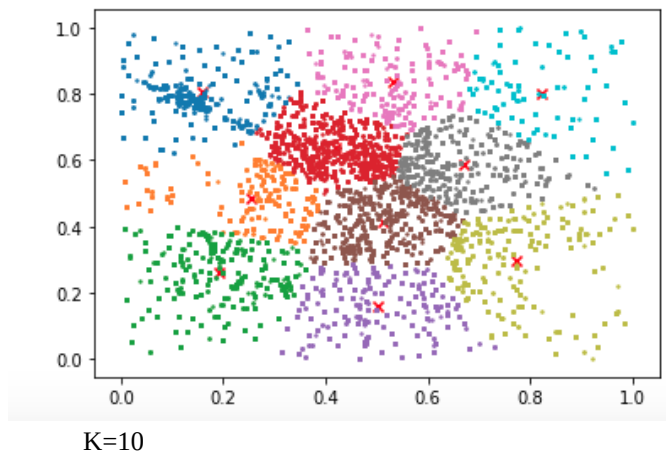
Séparabilité $S(P)$: qui représente la distance minimale entre les centroïdes des clusters.
Que l'on souhaite augmenter.

L'une des mesures utilisées pour évaluer le résultat d'un clustering est l'index de DUNN :

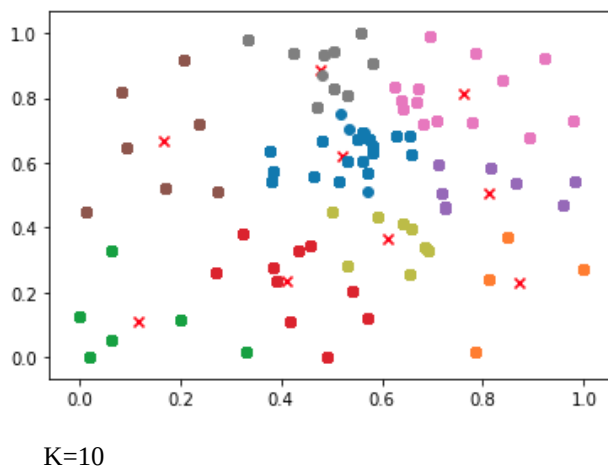
$$I_{DNN} = D(P)/S(P) \quad \text{avec } D(P) = \max D(C_k)$$

Les résultats obtenus:

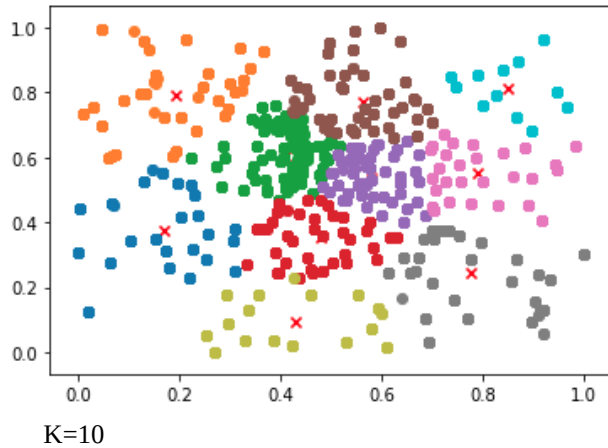
- PARIS



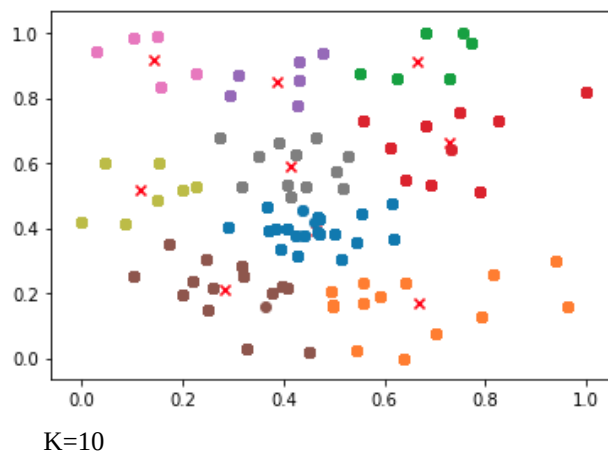
- LILLE



- LYON



- NANTES



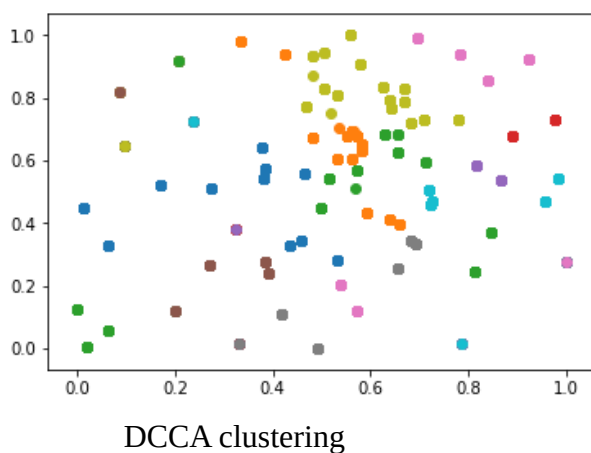
2. Amélioration de l'algorithme DCCA:

2.1. Application de l'algorithme DCCA classique sur les données Orange :

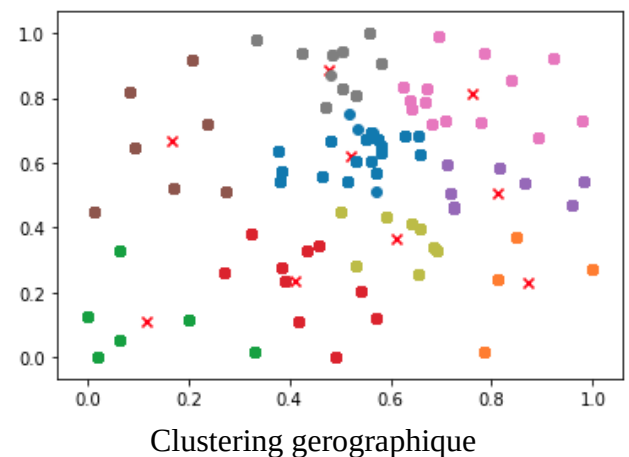
- On construit la matrice F du trafic prévu à partir d'un dataset LTE open source.
- On applique l'algorithme sur le dataset de la ville de Lille.
- Paramètres : taille du set = 1394 RRHs, Taux=0.14

Le résultat :

On obtient un $k=193$ parmi le 1394 clusters initiaux.



VS



On evalue la complementarité des clusters :

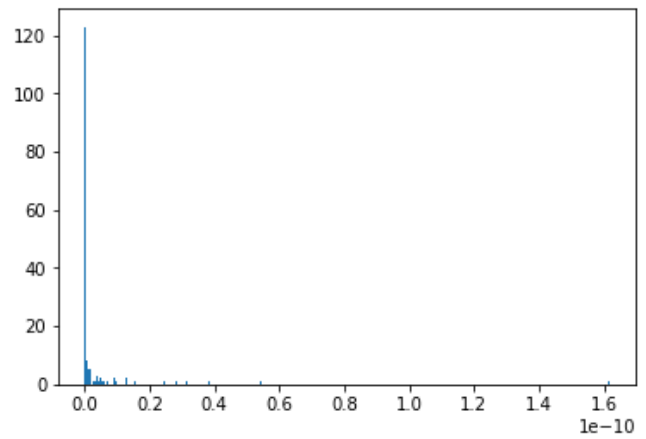
K= 181

max : 1.6166226327266596e-10

min : 7.038546156777134e-47

moyenne : 5.008697417416619e-12

mediane : 9.312759831320452e-14



2.2. Amélioration :

-On modifie le calcul de complémentarité pour qu'il prend en considération tout les rrhs du cluster. Après exécution on obtient :

K = 526

max : 1.6166226327266596e-10

min : 2.2661567582640103e-22

moyenne : 1.6583114805454124e-12

mediane : 1.9427385832877595e-19