

Dataset Classic3

On a les métriques suivantes pour 3 clusters :

Kmeans++ :

	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0522	0.0561	0.0782	0.0054
ARI	0.0154	0.0202	0.0680	0.0055

Kmediods :

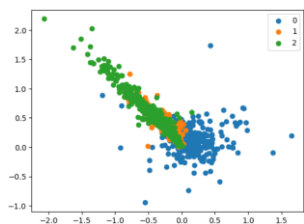
	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0032	0.0604	0.0267	0.0059
ARI	0.0031	0.0321	0.0184	0.0061

Spherical Kmeans :

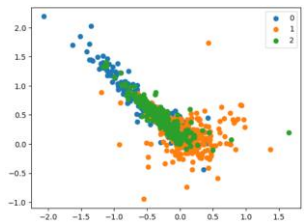
	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.3202	0.0785	0.0893	0.0963
ARI	0.3744	0.0568	0.0793	0.0416

Clustering Hiérarchique :

	Espace d'origine	ACP	t-SNE	UMAP
Critère	Ward	Ward	Ward	Ward
NMI	0.0348	0.0566	0.1169	0.0082
ARI	0.0166	0.0166	0.1068	0.0095



Spherical Kmeans (espace d'origine)



Véritable clusters

Classic4

On a les métriques suivantes pour 4 clusters :

Kmeans++ :

	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0724	0.0430	0.1332	0.0578
ARI	0.01262	0.0213	0.1170	0.0386

Kmediods :

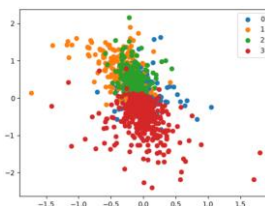
	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0527	0.0513	0.1336	0.0734
ARI	0.0302	0.0274	0.1147	0.0531

Spherical Kmeans :

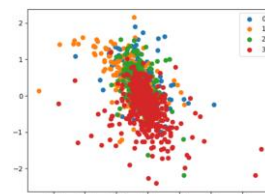
	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.2616	0.0801	0.1336	0.1161
ARI	0.2841	0.0603	0.1144	0.1438

Clustering Hiérarchique :

	Espace d'origine	ACP	t-SNE	UMAP
Critère	Ward	Complete	Complete	Ward
NMI	0.0733	0.0422	0.1168	0.0583
ARI	0.0209	0.0093	0.0991	0.0331



Spherical Kmeans (espace d'origine)



Véritables clusters

Dataset BBC

On a les métriques suivantes pour 5 clusters :

Kmeans++ :

	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0523	0.0523	0.0493	0.0499
ARI	-0.0814	-0.0810	0.0393	0.0288

Kmediods :

	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0495	0.0499	0.0493	0.0491
ARI	0.0008	-0.0333	0.0509	0.0386

Spherical Kmeans :

	Espace d'origine	ACP	t-SNE	UMAP
NMI	0.0113	0.0463	0.0493	0.0425
ARI	-0.0107	-0.0597	0.0489	-0.0440

Clustering Hiérarchique :

	Espace d'origine	ACP	t-SNE	UMAP
Critère	Ward	Ward	Complete	Ward
NMI	0.0526	0.0504	0.0493	0.0495
ARI	-0.0721	-0.0312	0.0367	0.0096

Conclusion

De manière générale les différents clusterings donnent de très mauvais résultats pour chaque dataset.

Spherical Kmeans semble être la méthode de clustering la plus performante (sauf pour le dataset BBC).

De même t-SNE semble être la méthode de réduction de dimension la plus performante.