



# Module : Data Mining & Texte Mining

## 1<sup>ère</sup> Année Master Big Data & Aide à la Décision

### Semestre 2 / Année 2018/2019 / Feuille de TD N° 4

#### **Exercice 1**

Considérons le dataset suivant :

Pt.	X	Y
P1	2	3
P2	3	1
P3	4	2
P4	11	5
P5	12	4
P6	12	6
P7	7	5
P8	8	4
P9	8	6

Appliquer l'algorithme K-Means en commençant par les centroïdes  $P_3$  et  $P_6$ . Préciser le nombre d'itérations pour la convergence de l'algorithme.

#### **Exercice 2**

Considérons la matrice de similarité suivante :

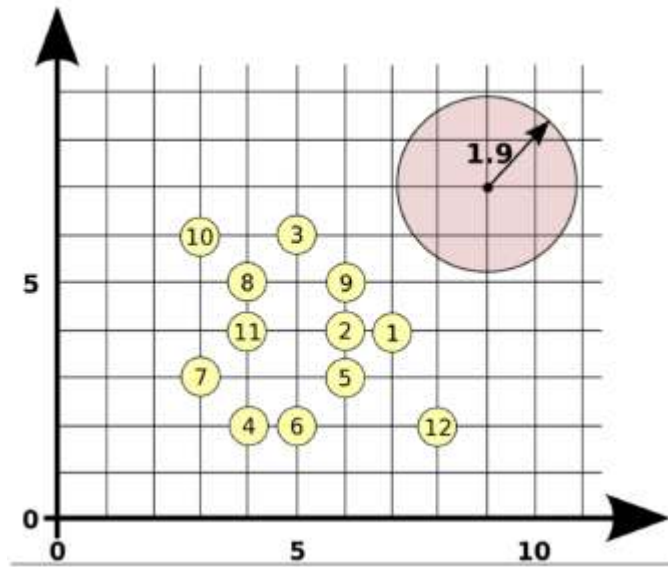
	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

Appliquer l'algorithme AHC :

- (i) en utilisant la méthode du single-linkage.
- (ii) en utilisant la méthode du complete-linkage.

#### **Exercice 3**

Appliquer l'algorithme DBSCAN sur le dataset de la figure suivante :



Prendre  $\varepsilon = 1.9$  et  $\text{MinPts} = 4$ . Indiquer si un point est central, bordure ou bruit et indiquer les clusters obtenus.