CHAMELEON

La particolarità di questo algoritmo è quella di basarsi su una metrica di similarità dinamica.
Due cluster vengono uniti se la loro interconnettività e vicinanza sono alte relativamente alla interconnettività interna e la vicinanza degli oggetti all'interno dei cluster stessi.

GRAPH-BASED

Perché utilizza un graph-partitioning algorithm.

CLUSTER GERARCHICO

Usa un cluster di tipo agglomerativo, combinando sotto-cluster.

K-NN GRAPH

Gli oggetti p e q vengono connessi se q è fra i primi K vicini di p.
Il raggio del vicinato dipende da quando quel punto giace su una regione densa. Più è densa e più i K vicini più vicini stanno in un raggio minore.
Questa densità viene registrata negli archi pesati del grafo, più una regione è densa, più i suoi archi nel grafo avranno peso maggiore. Sono indice di vicinanza.
Il risultato sono cluster molto naturali.

ALGORITMO DI GRAPH-PARTITIONING

Un cluster C viene diviso nei sottocluster C_i e C_j. Questo si fa tagliando gli archi che minimizzino la somma degli dei pesi degli archi da tagliare. L'EDGE CUT: $\delta C\{C_i, C_j\}$ fa si di mantenere l'assoluta interconnettività fra i cluster C_i e C_j.

PASSI CHAMELEON

1) Tutti i punti appartengono a un cluster solo.

2) Selezionare il più grande sotto-cluster per usare l'algoritmo di graph-partitioning per tagliare gli archi e formare a ogni iterazione due cluster diversi.

3) Ripetere finché il più grande cluster ha meno di un numero specificato di vertici.

4) Ora entra in gioco il cluster gerarchico agglomerativo per unire i sottocluster trovati col graph-partitioning, esso si basa sulla misura di interconnettività relativa RI:

$$RI(C_i, C_j) = \frac{|\delta C\{c_i, c_j\}|}{\frac{1}{2}(|\delta C_{C_i}| + |\delta C_{C_j}|)}$$

Dove $|\delta C\{c_i, c_j\}|$ è la somma dei pesi del EDGE CUT e $|\delta C_{C_i}|$ è l'interconnettività interna del singolo cluster (Cioè la somma minima degli archi che si dovrebbero tagliare per dividere C_i in due)

La misura di vicinanza relativa RC è definita:

$$RC(C_i, C_j) = \frac{\bar{S}_{EC\{C_i, C_j\}}}{\frac{|C_i|}{|C_i|+|C_j|}\bar{S}_{EC\{C_i\}} + \frac{|C_j|}{|C_i|+|C_j|}\bar{S}_{EC\{C_j\}}}$$

Dove $\bar{S}_{EC\{C_i, C_j\}}$ è la media dei pesi che connettono i vertici di C_i con quelli di C_j.
Mentre $\bar{S}_{EC\{C_i\}}$ è la media dei pesi che appartengono al più piccolo EDGE CUT di C_i. $|C_i|$ = # punti di C_i.

Merge: Due cluster vengono uniti se $RI(C_i, C_j)$ e $RC(C_i, C_j)$ sono entrambi sopra della threshold specificate a priori (T_{RI} e T_{RC}).

Nota: Quando si hanno più cluster che soddisfano i requisiti di threshold, si fa il merge dei due con $RI(C_i, C_j)$ maggiore.

COMPLESSITÀ $O(n^2)$

PRO: Ottimale nel cercare le forme di cluster più varie, meglio di BIRCH e DBSCAN