Abbiamo considerato lo stesso sottoinsieme di attributi numerici utilizzati in dbscan prendendo in considerazione tutti i record del dataset. I dendrograms mostrano sull’asse delle ascisse tutti i punti, a loro volta raggruppati per agevolarne la lettura, e sulle ordinate la distanza tra i cluster.

A seconda dei metodi utilizzati, il clustering gerarchico restituisce risultati migliori rispetto all’approccio density based.

Come prevedibile, il metodo single link è quello da cui abbiamo ottenuto risultati peggiori. I clusters tendono ad agglomerarsi molto rapidamente ad una distanza molto ravvicinata. Questo comportamento, come ci ha confermato anche dbscan, è dovuto alla continuità dei record ed alla loro vicinanza. Entrambe le metriche, euclidean e codine, sono insoddisfacenti.

(sx single link euclidean - DX SINGLE LINK COSINE)

Differentemente dal precedente, i metodi Ward, complete e average restituiscono risultati apprezzabili. Il primo di questi, che tende a minimizzare la varianza all’interno dei cluster, sembra emulare i risultati di kmean con la formazione di 3 cluster di differente dimensione. Il metodo complete ed average sembrano eguagliarsi, entrambi partizionano il dataset in 6 cluster di dimensione molto simile. 