Agglomerative Clustering: Average vs. Centroid Links

- ❑ Agglomerative clustering with **average link**
  - ❑ **Average link**: The average distance between an element in one cluster and an element in the other (i.e., all pairs in two clusters)
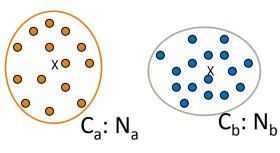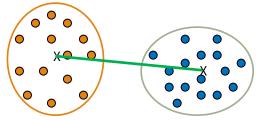   - ❑ Expensive to compute
- ❑ Agglomerative clustering with **centroid link**
  - ❑ **Centroid link**: The distance between the centroids of two clusters
- ❑ **Group Averaged Agglomerative Clustering (GAAC)**
  - ❑ Let two clusters C_a and C_b be merged into $C_{a \cup b}$. The new centroid is:
    - ❑ N_a is the cardinality of cluster C_a, and c_a is the centroid of C_a
    $$c_{a \cup b} = \frac{N_a c_a + N_b c_b}{N_a + N_b}$$
  - ❑ The similarity measure for GAAC is the average of their distances
- ❑ Agglomerative clustering with **Ward's criterion**

  - ❑ **Ward's criterion:** The increase in the value of the SSE criterion for the clustering obtained by merging them into $C_a \cup C_b$: $$W(C_{a \cup b}, c_{a \cup b}) - W(C, c) = \frac{N_a N_b}{N_a + N_b} d(c_a, c_b)$$

C_a: N_a C_b: N_b