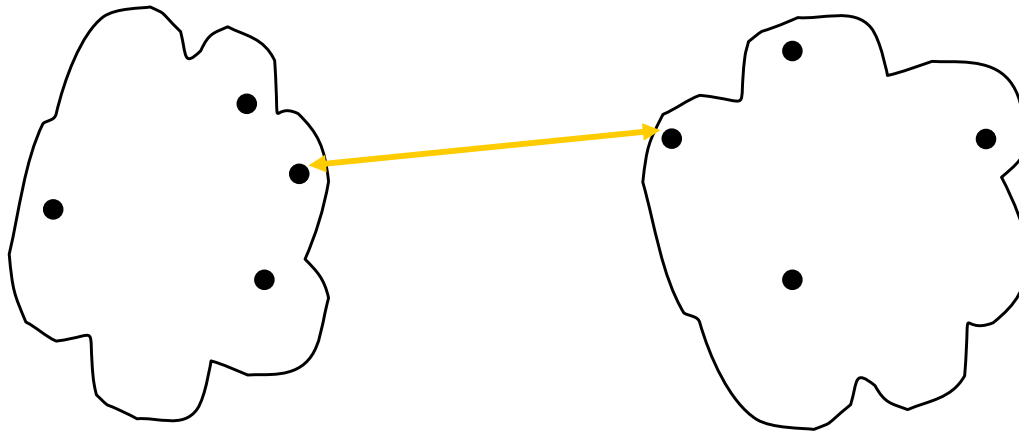


# MIN or Single Link

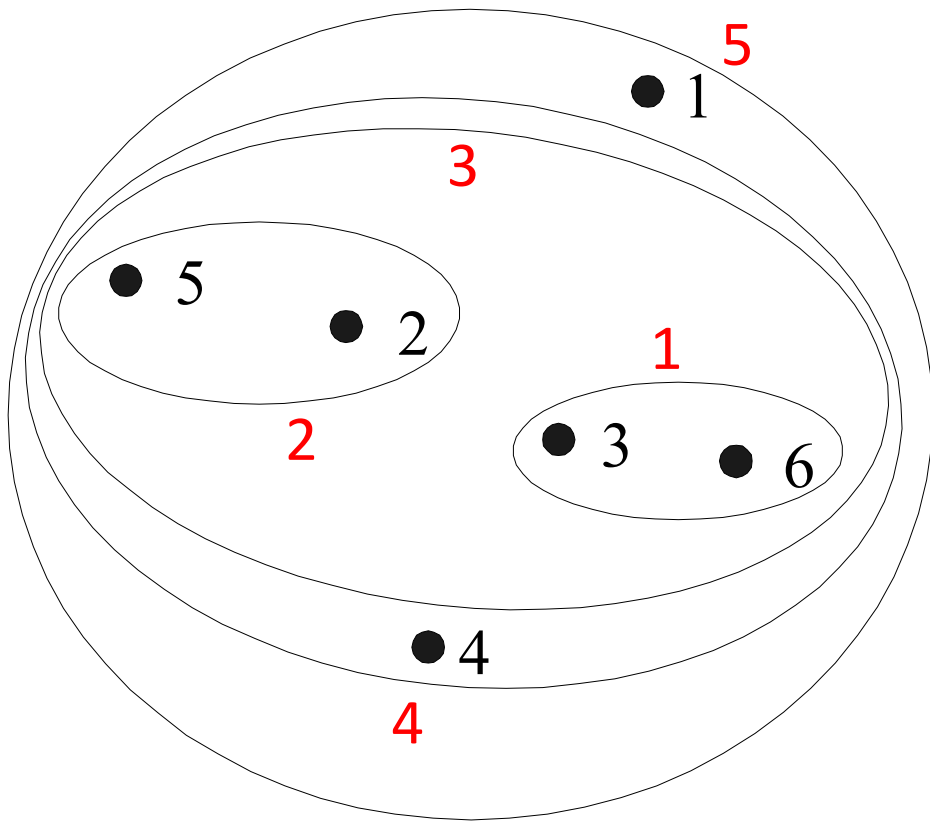
- **Inter-cluster distance**

- The distance between two clusters is represented by the distance of the closest pair of data objects belonging to different clusters.
- Determined by one pair of points, i.e., by one link in the proximity graph

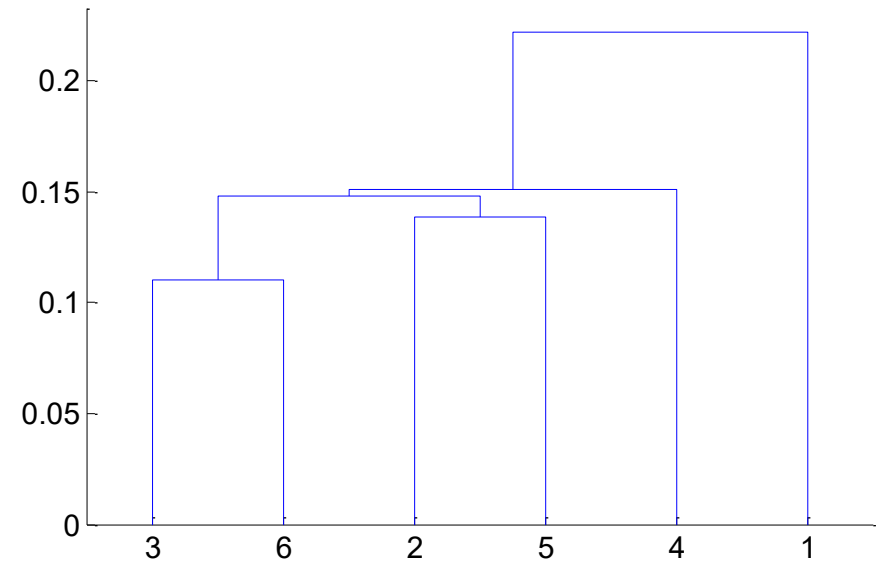


$$d_{\min}(C_i, C_j) = \min_{p \in C_i, q \in C_j} d(p, q)$$

**MIN**

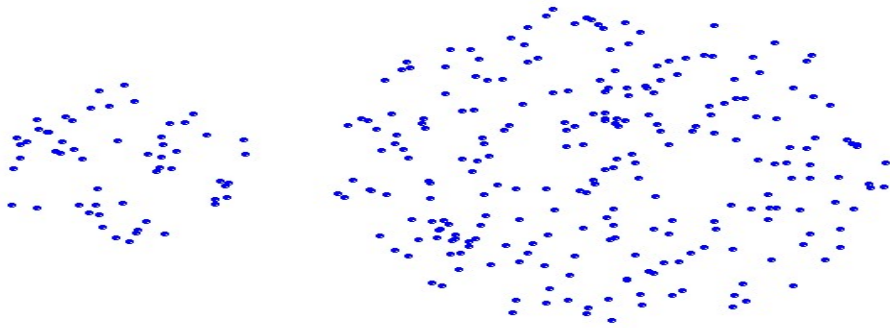


Nested Clusters

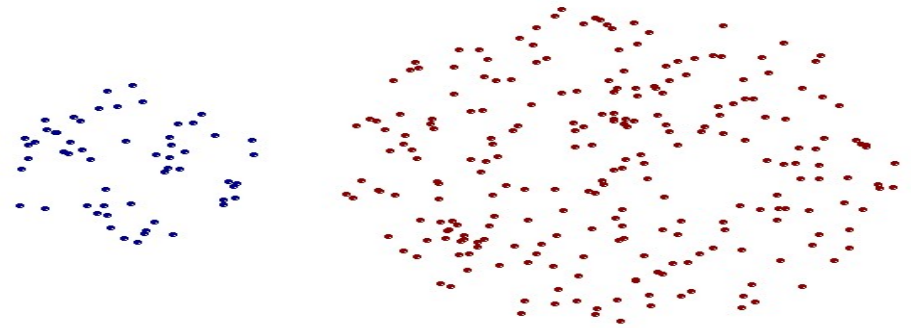


Dendrogram

# Strength of MIN



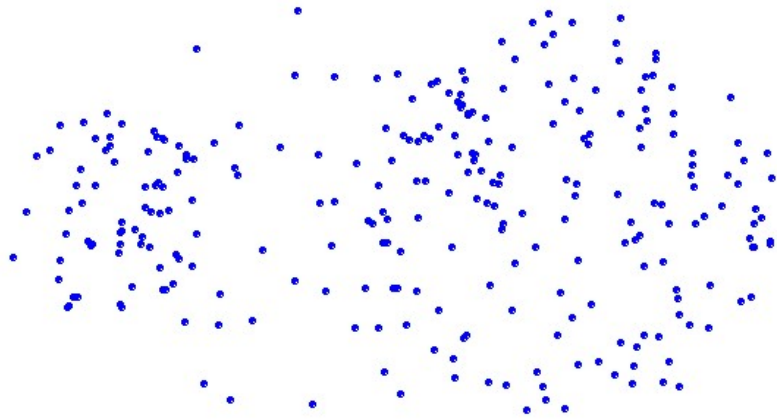
Original Points



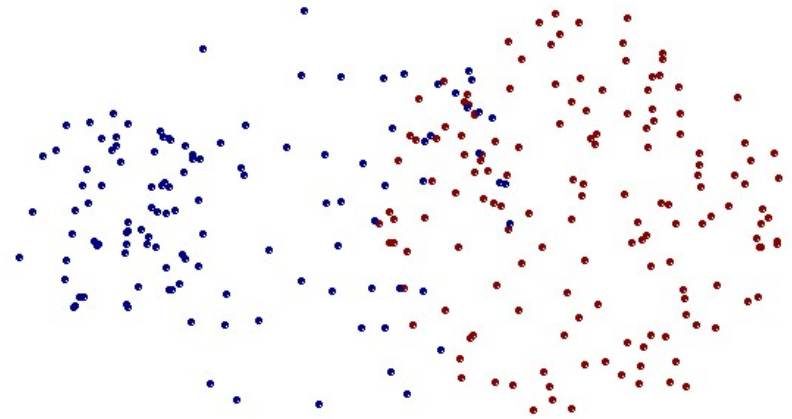
Two Clusters

- Can handle non-elliptical shapes

# Limitations of MIN



Original Points



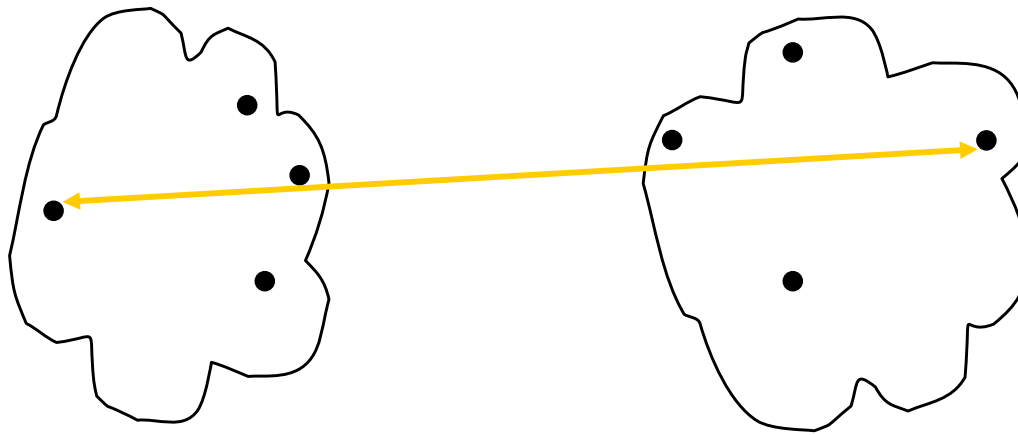
Two Clusters

- Sensitive to noise and outliers

# MAX or Complete Link

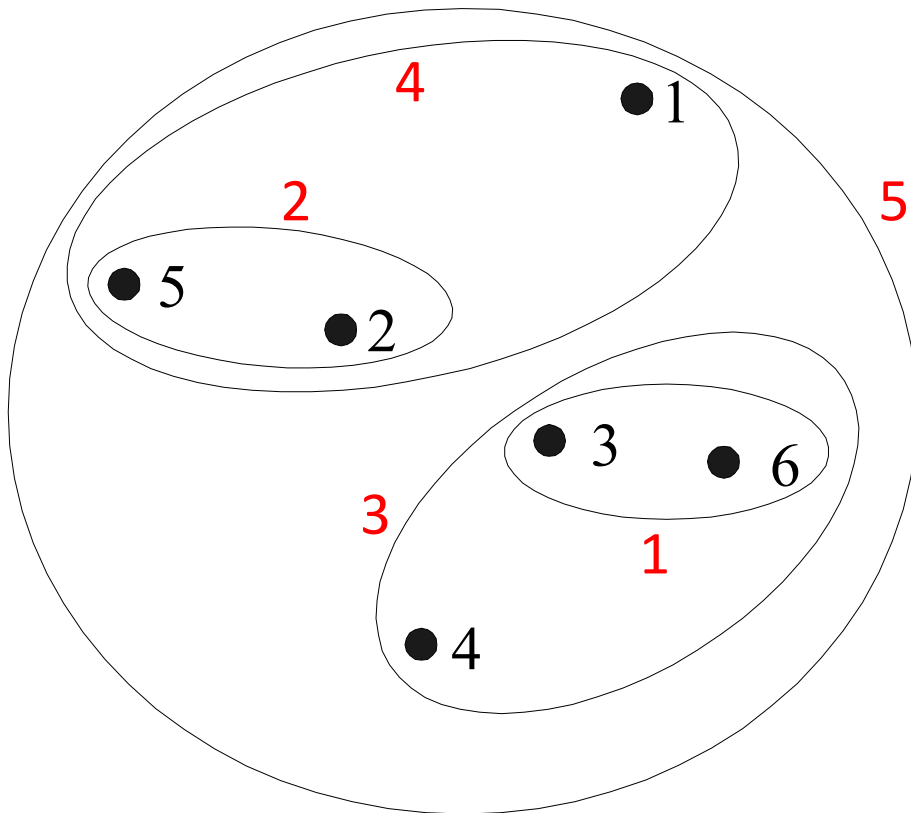
- **Inter-cluster distance**

- The distance between two clusters is represented by the distance of the farthest pair of data objects belonging to different clusters

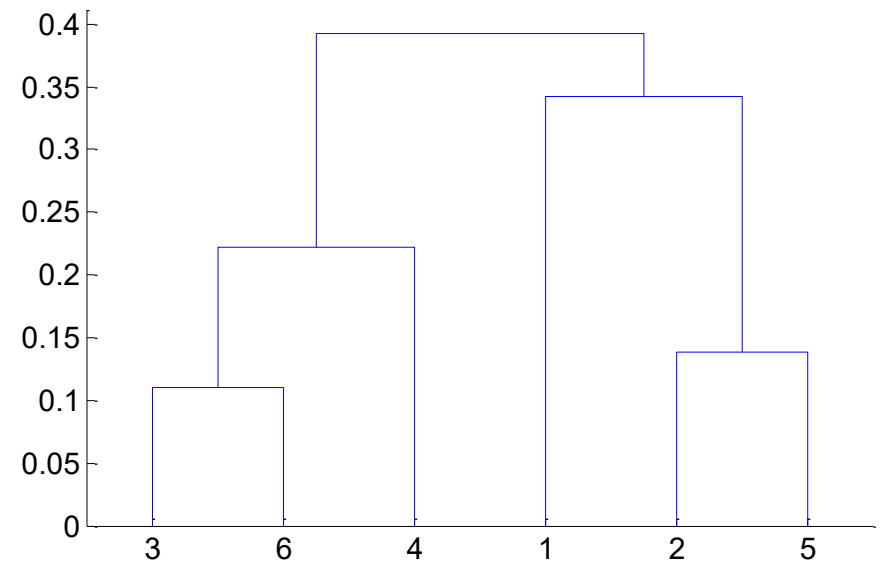


$$d_{\min}(C_i, C_j) = \max_{p \in C_i, q \in C_j} d(p, q)$$

**MAX**

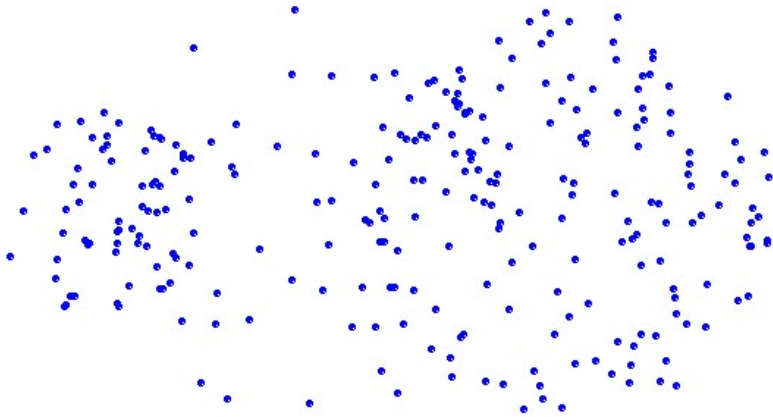


Nested Clusters

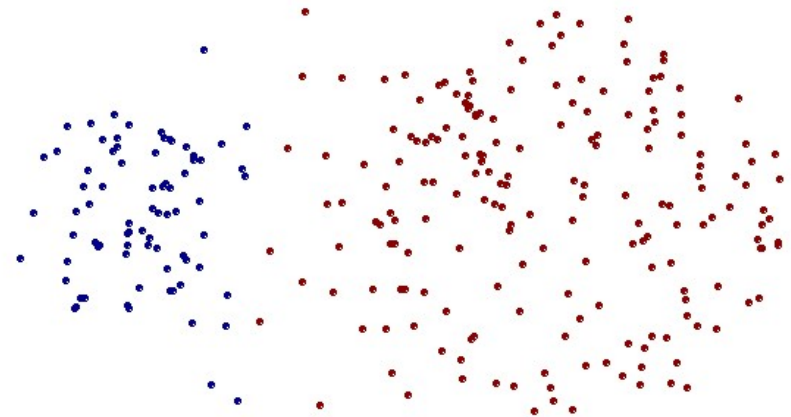


Dendrogram

# Strength of MAX



Original Points

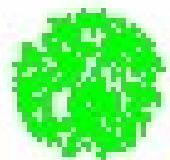
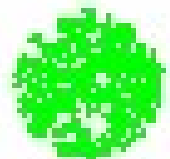
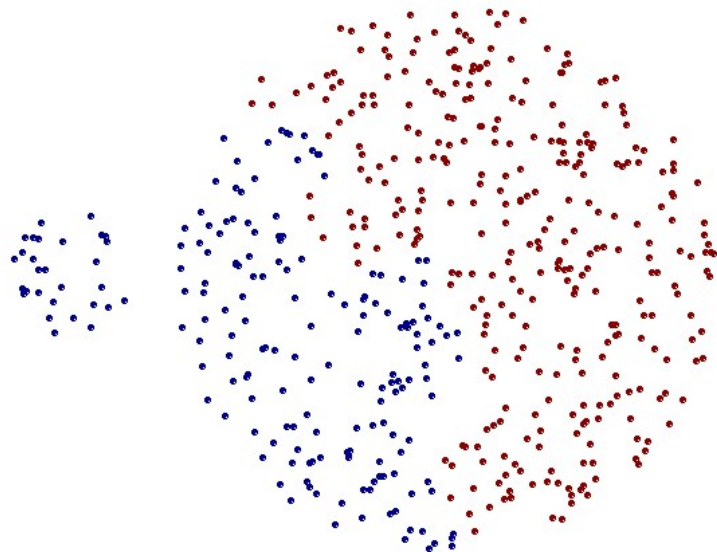
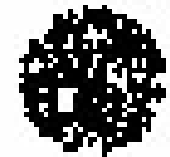
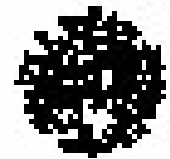
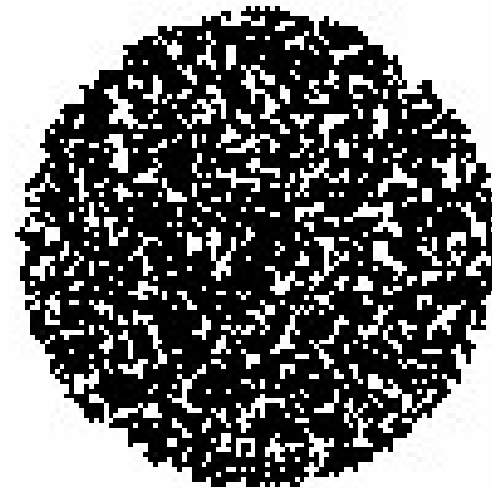
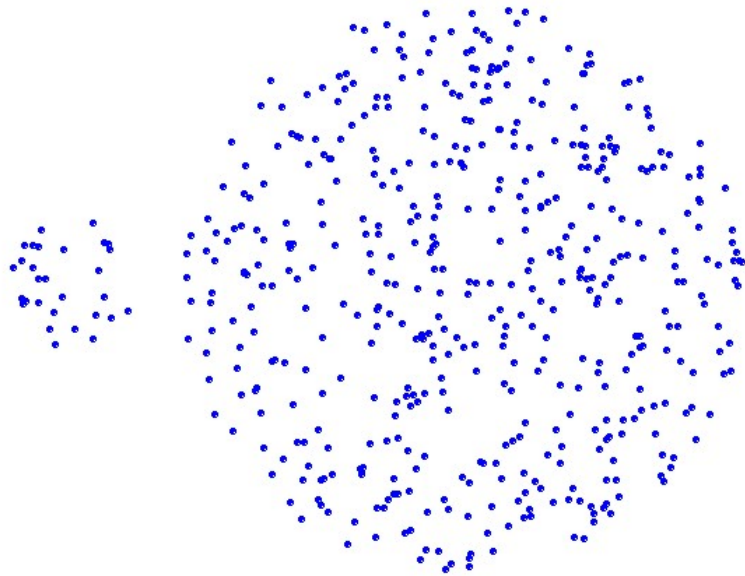


Two Clusters

- Less susceptible to noise and outliers

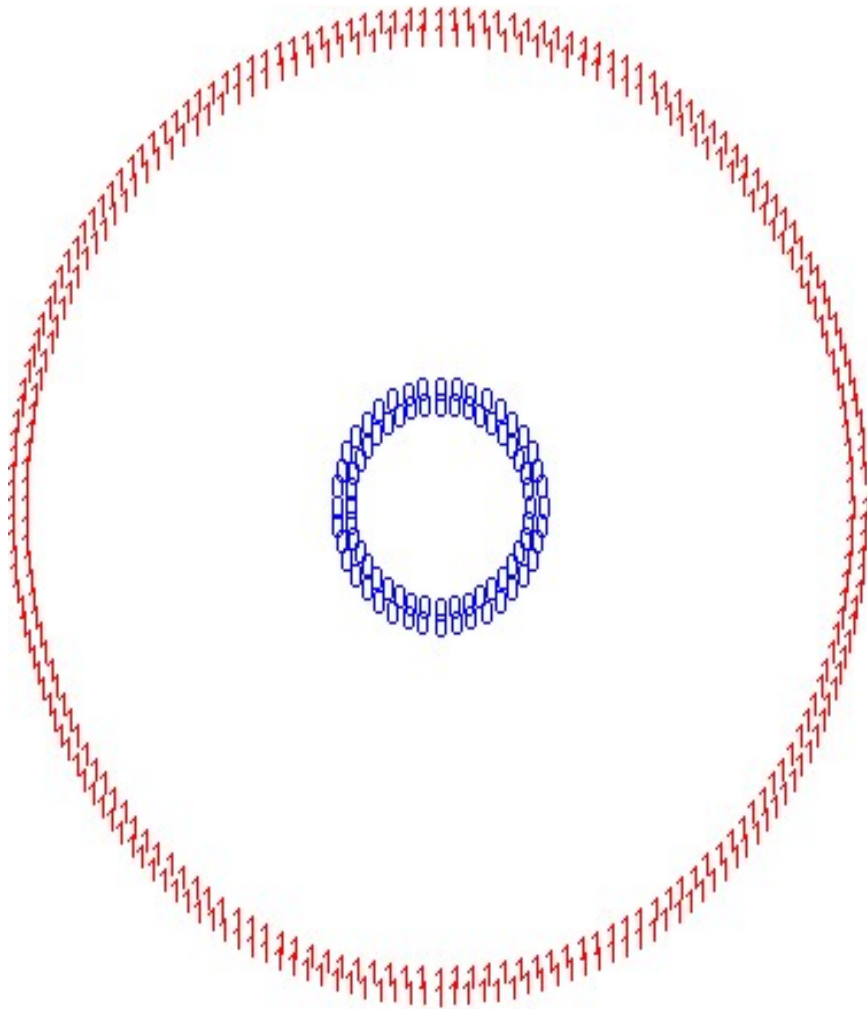
# Limitations of MAX

- Tends to break large clusters

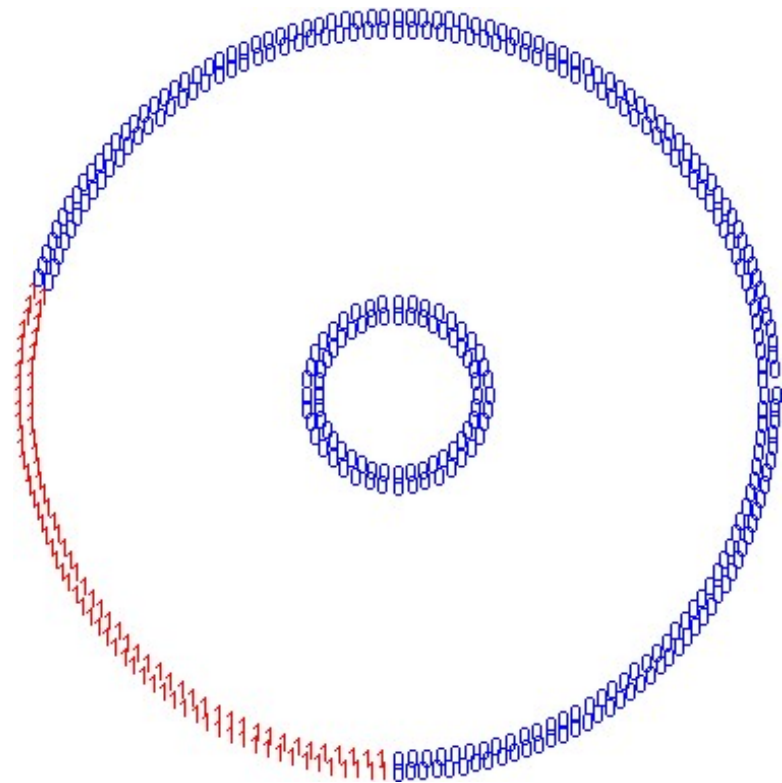




# Limitations of MAX



MIN (2 clusters)



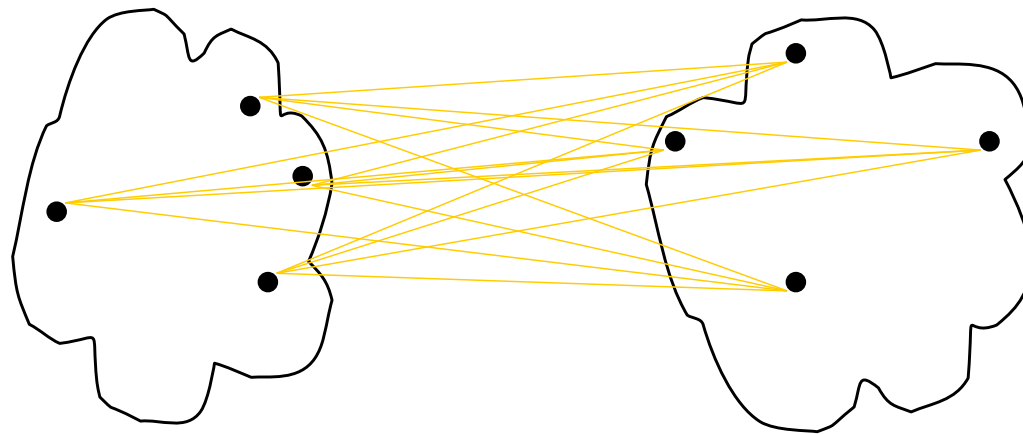
MAX (2 clusters)

- Biased towards globular clusters

# Group Average or Average Link

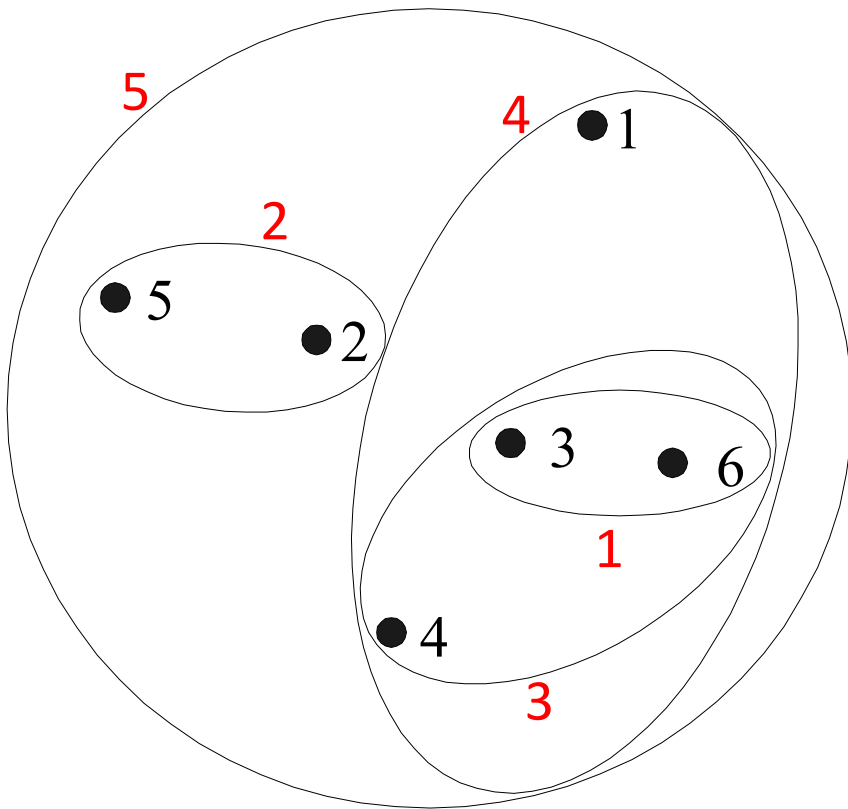
- **Inter-cluster distance**

- The distance between two clusters is represented by the average distance of all pairs of data objects belonging to different clusters
- Determined by all pairs of points in the two clusters

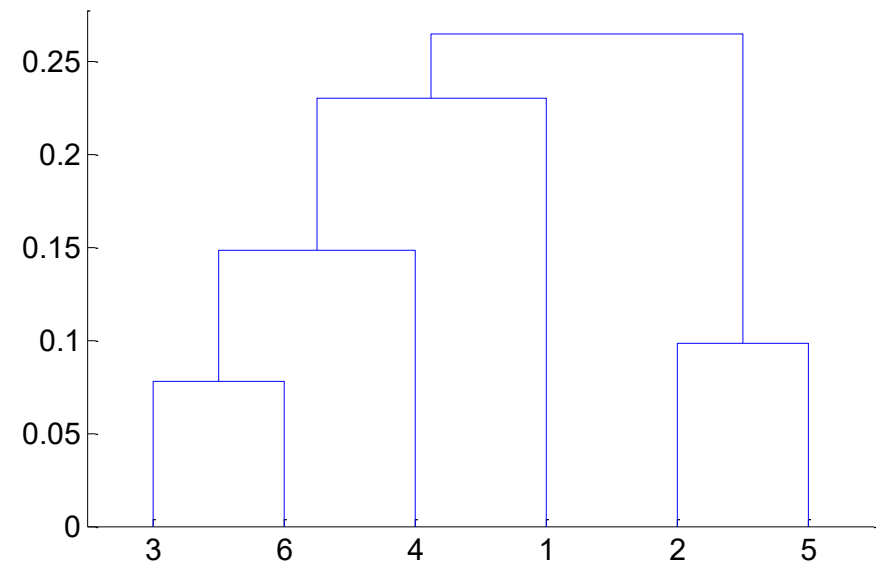


$$d_{\min}(C_i, C_j) = \underset{p \in C_i, q \in C_j}{avg} \ d(p, q)$$

# Group Average



Nested Clusters



Dendrogram

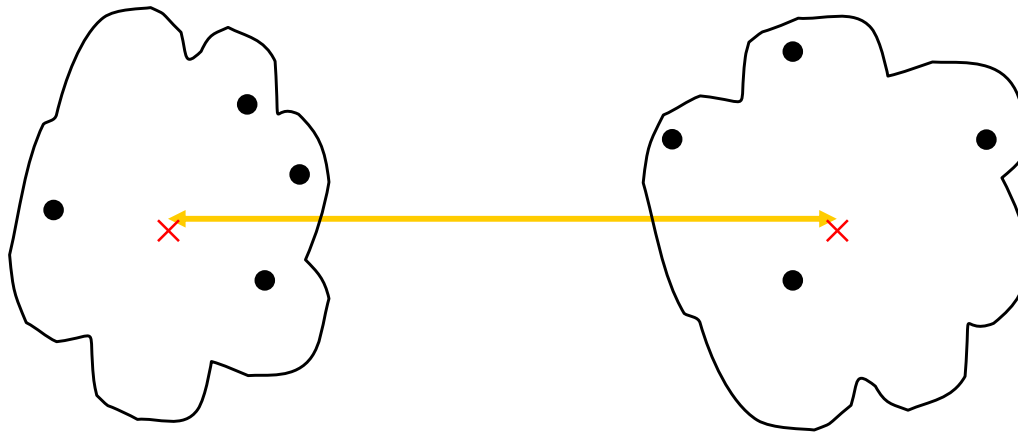
# Group Average

- Compromise between Single and Complete Link
- Strengths
  - Less susceptible to noise and outliers
- Limitations
  - Biased towards globular clusters

# Centroid Distance

- **Inter-cluster distance**

- The distance between two clusters is represented by the distance between the centers of the clusters
- Determined by cluster centroids



$$d_{mean}(C_i, C_j) = d(m_i, m_j)$$