Hierarchical vs. Partitional Clustering

- A distinction among different types of clusterings is whether the set of clusters is nested or unnested.
- A **partitional clustering** a simply a division of the set of data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset.
- A **hierarchical clustering** is a set of nested clusters that are organized as a tree.

Why of Hierarchical Clustering?

1. It does not assume a particular value of k, as needed by k-means clustering.
2. The generated tree may correspond to a meaningful taxonomy.
3. Only a distance or "proximity" matrix is needed to compute the hierarchical clustering.

Hierarchical Clustering

Two types of algorithms:

1. **Agglomerative ("Bottom-up")**
   Start with the points as individual clusters and, at each step, merge the closest pair of clusters.

2. **Divisive ("Top-down")**
   Start with one, all-inclusive cluster and, at each step, split a cluster until only singleton clusters of individual points remain.

Basic Agglomerative Clustering

Basic agglomerative hierarchical clustering algorithm.

1: Compute the proximity matrix, if necessary.
2: **repeat**
3:     Merge the closest two clusters.
4:     Update the proximity matrix to reflect the proximity between the new cluster and the original clusters.
5: **until** Only one cluster remains.

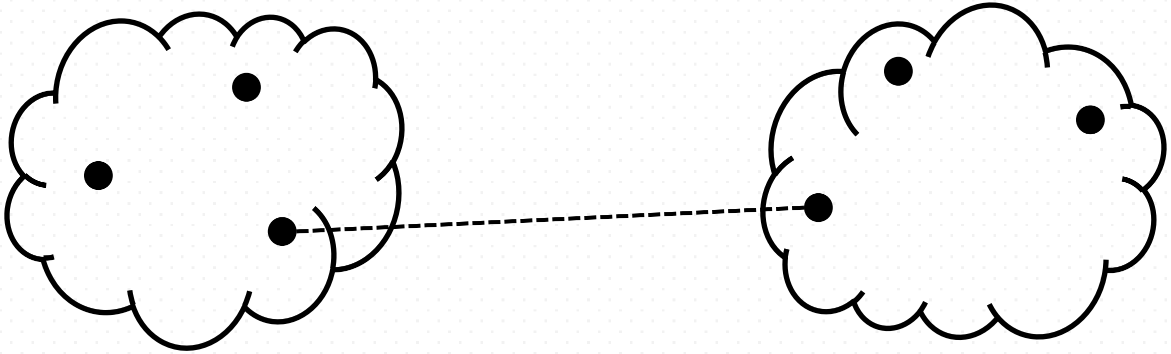
Defining Proximity Between Clusters

- The key operation of basic agglomerative clustering is the computation of the proximity between two clusters.
- The definition of cluster proximity differentiates the various agglomerative hierarchical techniques.
- MAX (single link), MIN (complete link), and group average are **graph-based proximities**.
- Ward's method is a **prototype-based proximity**.

MIN (Single Link) Proximity

Defines cluster proximity as the shortest distance between two points, x and y, that are in different clusters, A and B:

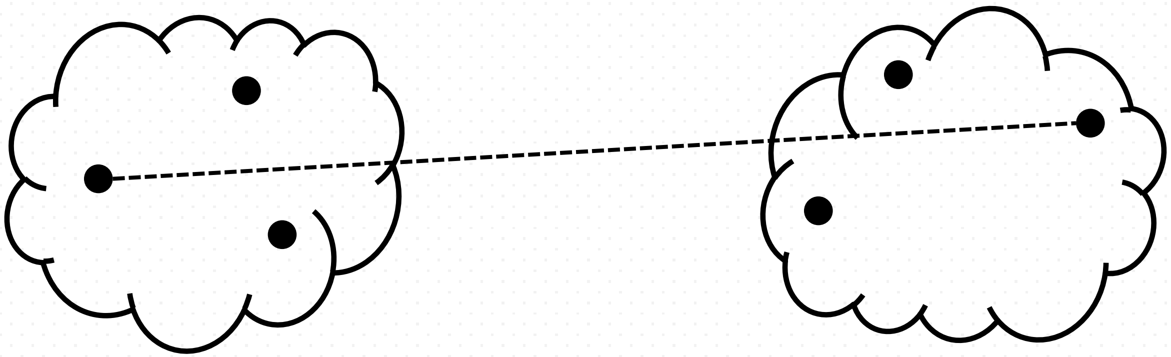
$$d(A,B) = \min_{x \in A, y \in B} d(x - y)$$



MAX (Complete Link) Proximity

Defines cluster proximity as the furthest distance between two points, x and y, that are in different clusters, A and B:

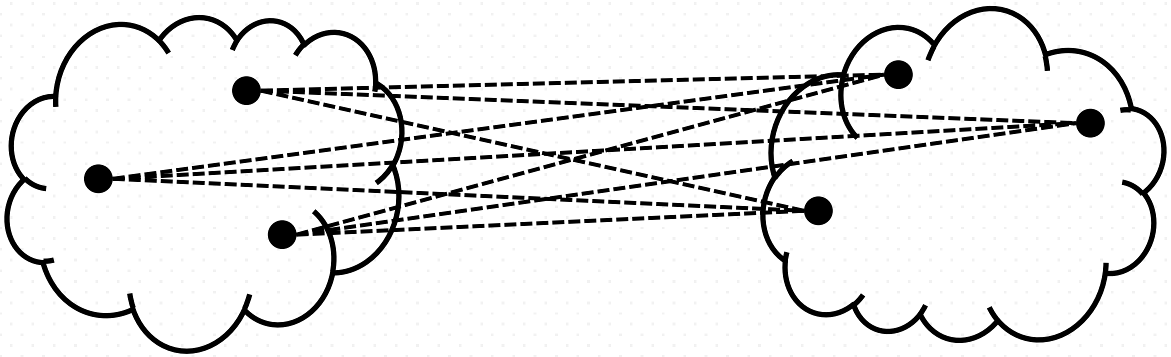
$$d(A, B) = \max_{x \in A, y \in B} d(x - y)$$



Group Average Proximity

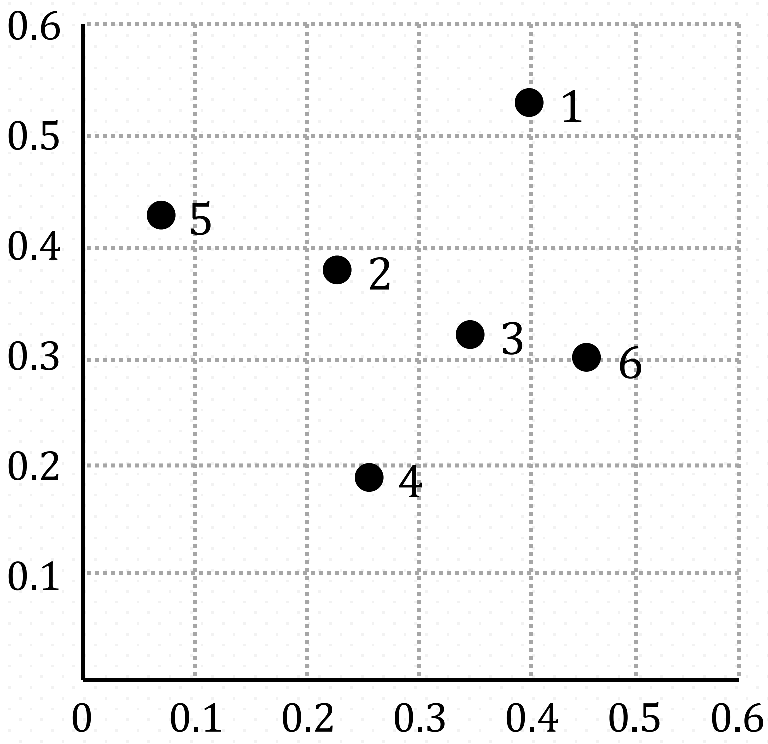
Defines cluster proximity as the average distance between two points, x and y, that are in different clusters, A and B (number of points in cluster j is n_j):

$$d(A, B) = \sum_{x \in A, y \in B} d(x - y)/n_A n_B$$



Example Data for Clustering

Set of 6 Two-Dimensional Points

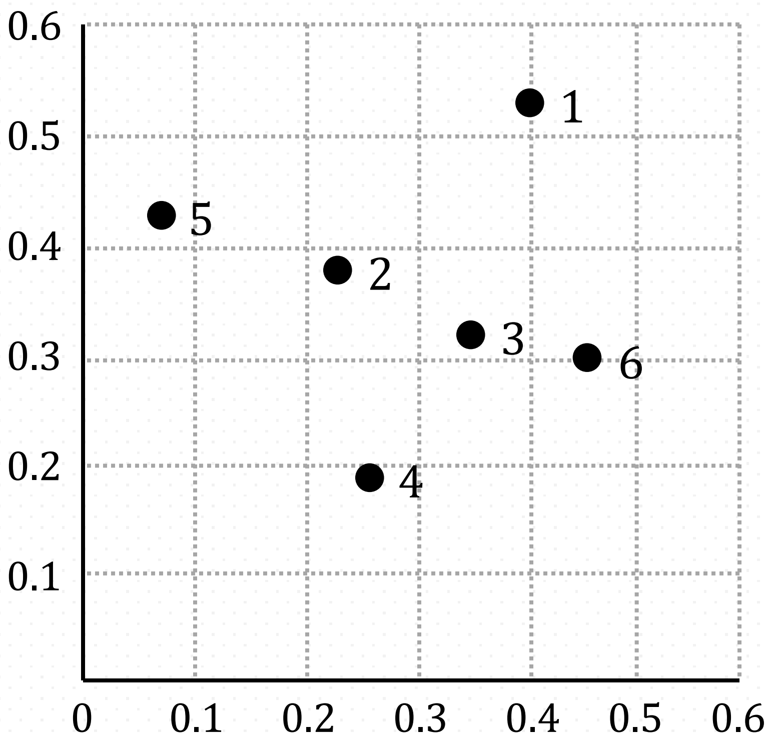


xy Coordinates of 6 Points

Point	*x* Coordinate	*y* Coordinate
p1	0.40	0.53
p2	0.22	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.30

Example Data for Clustering

Set of 6 Two-Dimensional Points

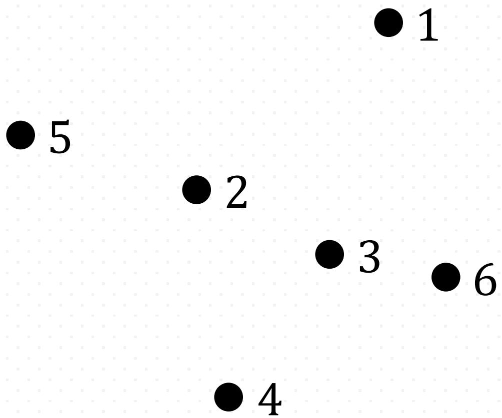


Euclidean Distance Matrix for 6 Points

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

Example of Single Link Clustering

Nested Cluster Diagram

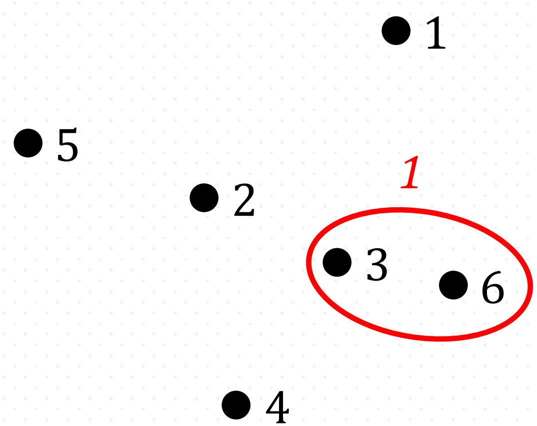


Single Link Distance Matrix

	1	2	3	4	5	6
1	0	0.24	0.22	0.37	0.34	0.23
2		0	0.15	0.20	0.14	0.25
3			0	0.15	0.28	0.11
4				0	0.29	0.22
5					0	0.39
6						0

Example of Single Link Clustering

Nested Cluster Diagram



Single Link Distance Matrix

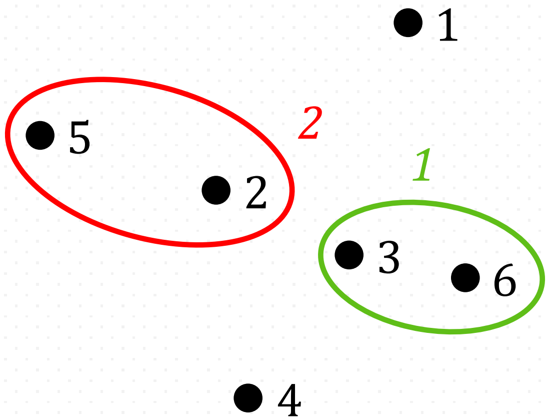
	1	2	3	4	5	6
1	0	0.24	*0.22*	0.37	0.34	0.23
2		0	*0.15*	0.20	0.14	0.25
3			0	*0.15*	*0.28*	0.11
4				0	0.29	0.22
5					0	0.39
6						0

1

Points 3 and 6 have the smallest single link proximity distance. Merge these points into one cluster and update the distances to this new cluster. For example, the distance from point 1 to this cluster is 0.22 (the distance to point 3).

Example of Single Link Clustering

Nested Cluster Diagram



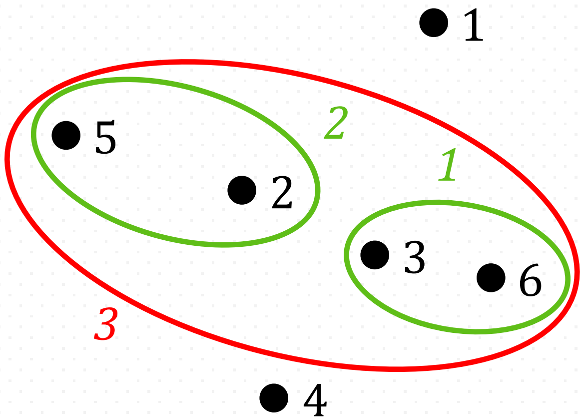
Single Link Distance Matrix

	1	2	4	5	3,6
1	0	*<u>0.24</u>*	0.37	<u>0.34</u>	0.22
2		0	*<u>0.20</u>*	0.14	0.15
4			0	<u>0.29</u>	0.15
5				0	0.28
3,6					0

2 Points 2 and5 have the smallest single link proximity distance. Merge these points into one cluster and update the distances to this new cluster.

Example of Single Link Clustering

Nested Cluster Diagram



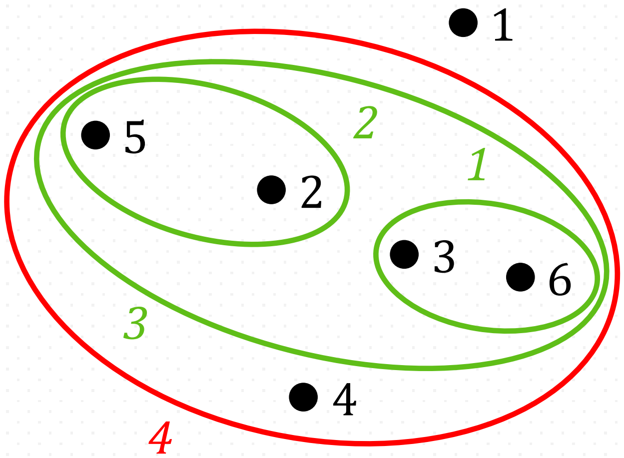
Single Link Distance Matrix

	1	4	2,5	3,6
1	0	0.37	0.24	*0.22*
4		0	0.20	*0.15*
2,5			0	0.15
3,6				0

3 And iterate…

Example of Single Link Clustering

Nested Cluster Diagram



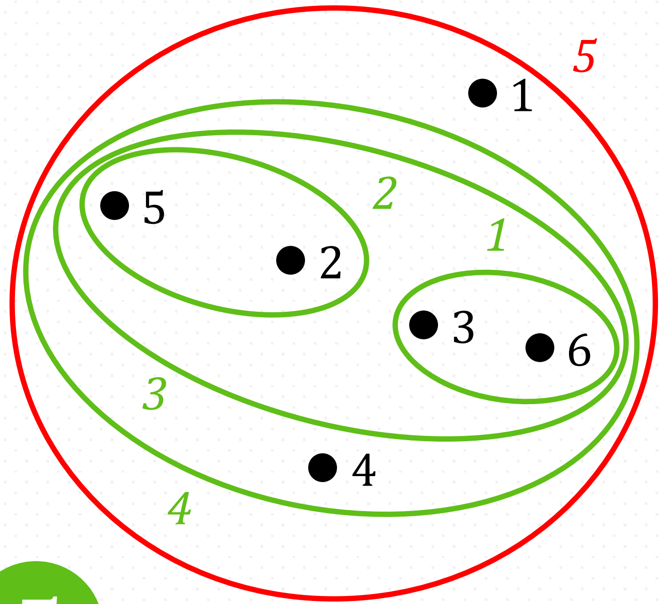
Single Link Distance Matrix

	1	4	2,5,3,6
1	0	0.37	*0.22*
4		0	0.15
2,5,3,6			0

4 And iterate…

Example of Single Link Clustering

Nested Cluster Diagram



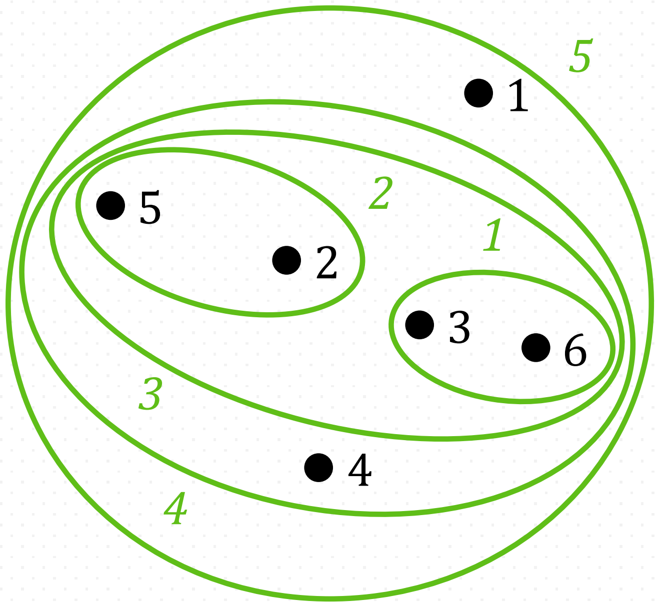
Single Link Distance Matrix

	1	4,2,5,3,6
1	0	0.22
2,5,3,6		0

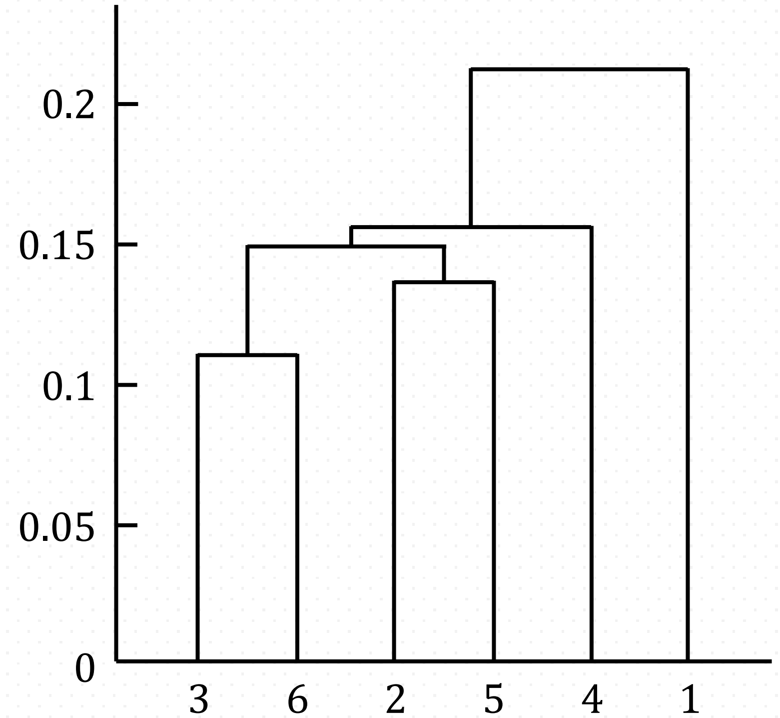
5 And iterate until there is one all-inclusive cluster.

Example of Single Link Clustering

Nested Cluster Diagram

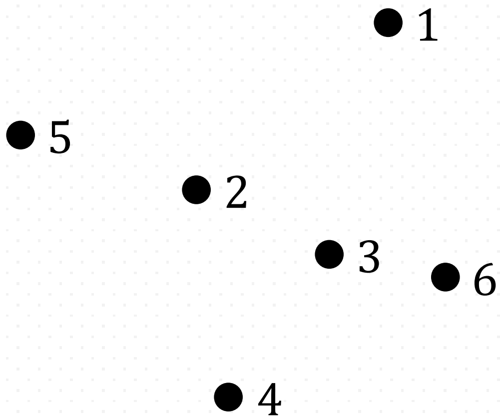


Hierarchical Tree Diagram



Example of Complete Link Clustering

Nested Cluster Diagram

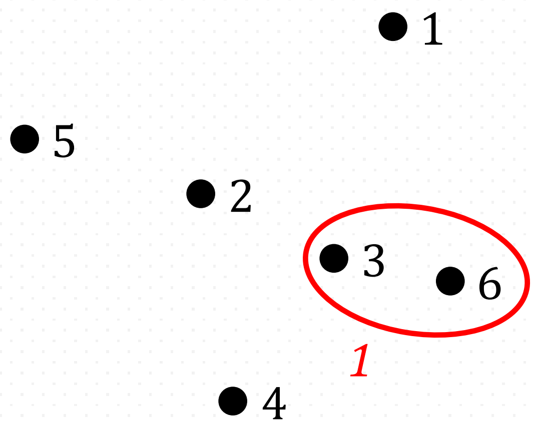


Single Link Distance Matrix

	1	2	3	4	5	6
1	0	0.24	0.22	0.37	0.34	0.23
2		0	0.15	0.20	0.14	0.25
3			0	0.15	0.28	0.11
4				0	0.29	0.22
5					0	0.39
6						0

Example of Complete Link Clustering

Nested Cluster Diagram



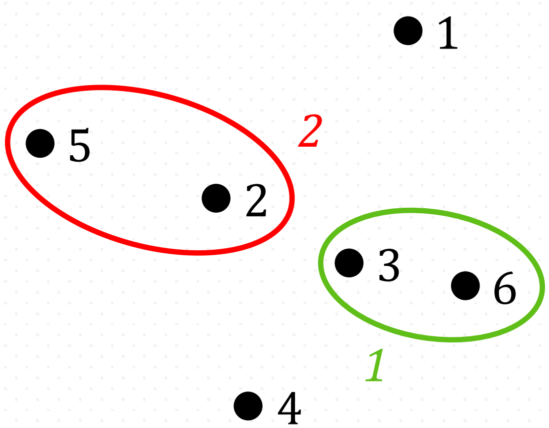
Complete Link Distance Matrix

	1	2	3	4	5	6
1	0	0.24	0.22	0.37	0.34	*0.23*
2		0	0.15	0.20	0.14	*0.25*
3			0	0.15	0.28	0.11
4				0	0.29	*0.22*
5					0	*0.39*
6						0

1 Points 3 and 6 have the smallest complete link proximity distance. Merge these points into one cluster and update the distances to this new cluster. For example, the distance from point 1 to this cluster is 0.23 (the distance to point 6).

Example of Complete Link Clustering

Nested Cluster Diagram



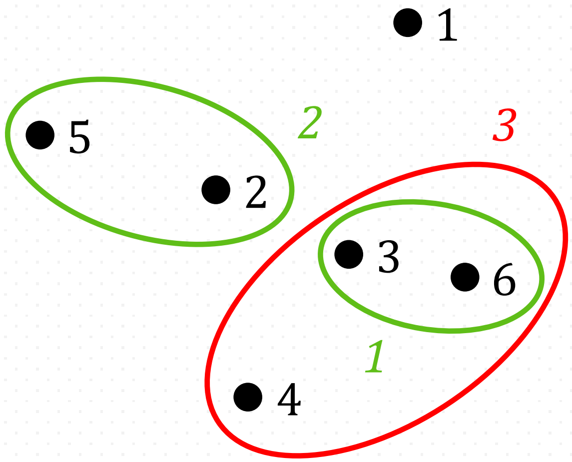
Complete Link Distance Matrix

	1	2	4	5	3,6
1	0	0.24	0.37	*0.34*	0.23
2		0	0.20	0.14	0.25
4			0	*0.29*	0.22
5				0	0.39
3,6					0

2 Points 2 and 5 have the smallest complete link proximity distance. Merge these points into one cluster and update the distances to this new cluster.

Example of Complete Link Clustering

Nested Cluster Diagram



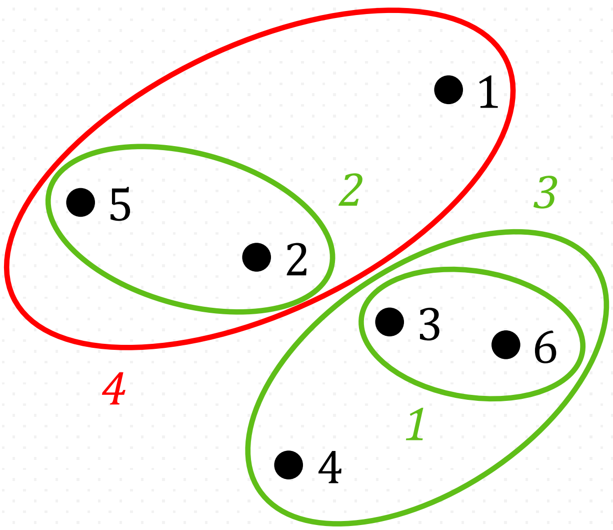
Complete Link Distance Matrix

	1	4	2,5	3,6
1	0	*0.37*	0.34	0.23
4		0	0.29	0.22
2,5			0	*0.39*
3,6				0

3 And iterate…

Example of Complete Link Clustering

Nested Cluster Diagram



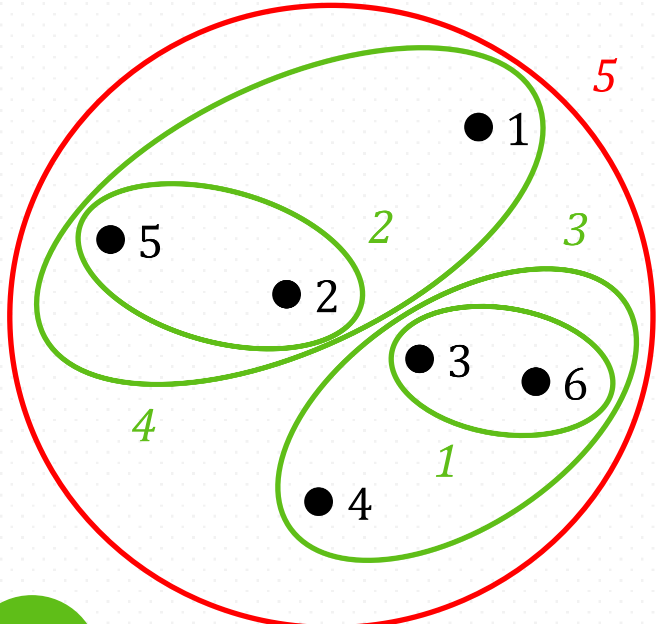
Complete Link Distance Matrix

	1	2,5	4,3,6
1	0	0.34	<u>0.37</u>
2,5		0	*<u>0.39</u>*
4,3,6			0

4 And iterate…

Example of Complete Link Clustering

Nested Cluster Diagram



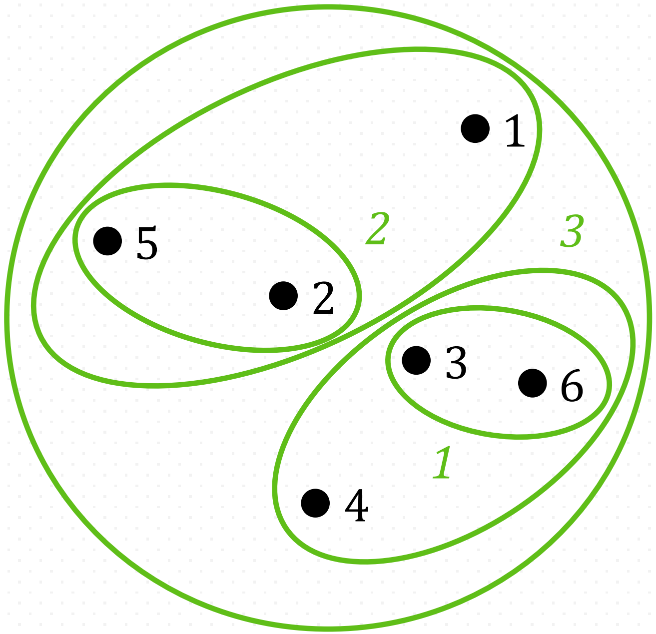
Complete Link Distance Matrix

	1,2,5	4,3,6
1,2,5	0	0.39
4,3,6		0

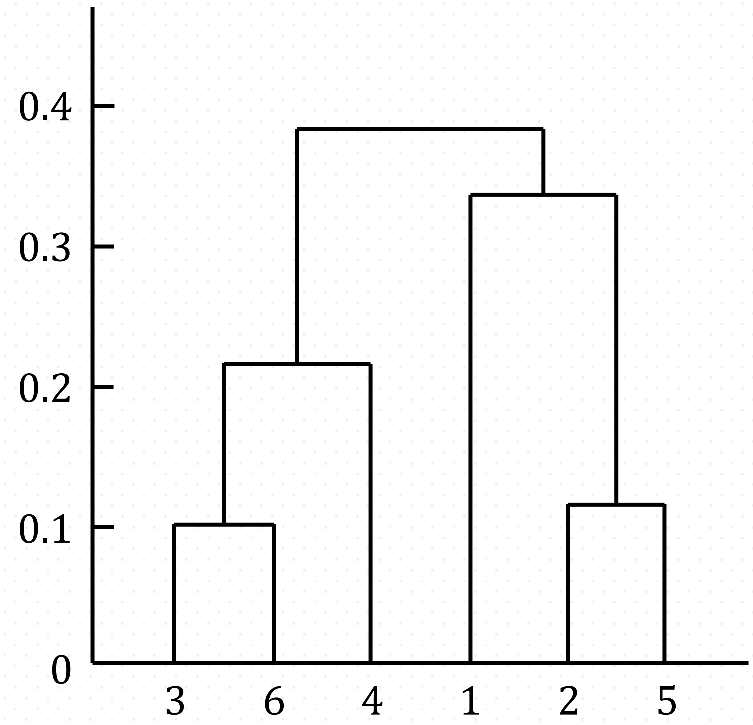
5 And iterate until there is one all-inclusive cluster.

Example of Complete Link Clustering

Nested Cluster Diagram

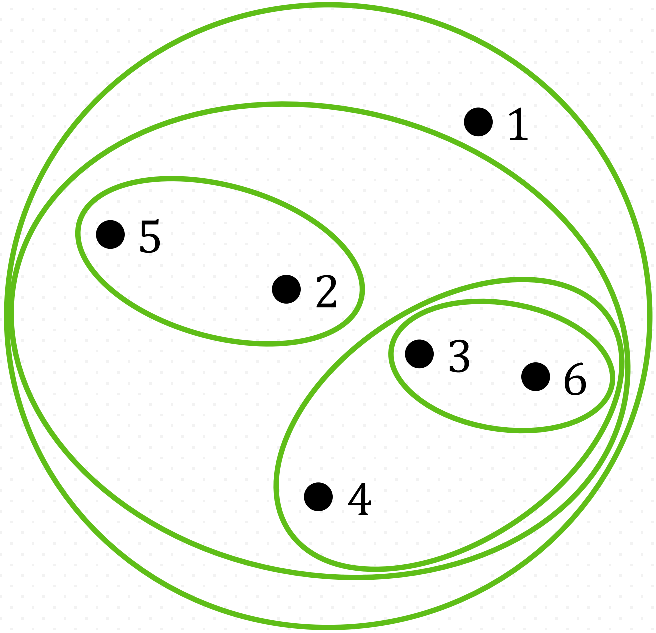


Hierarchical Tree Diagram

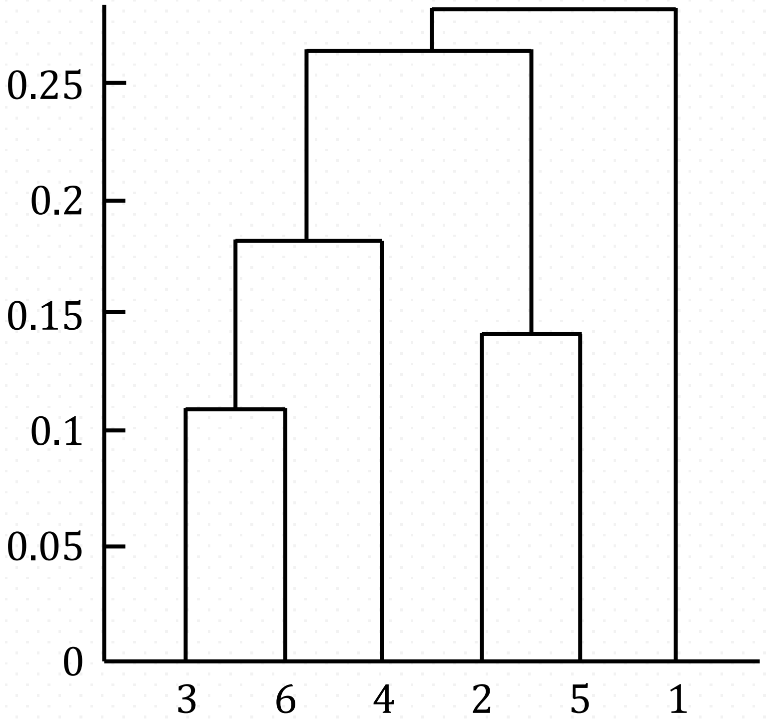


Example of Group Average Clustering

Nested Cluster Diagram



Hierarchical Tree Diagram



Discussion of Proximity Methods

- **Single link** is “chain-like” and good at handling non-elliptical shapes, but is sensitive to outliers.
- **Complete link** is less susceptible to noise and outliers, but can break large clusters and favors globular shapes.
- **Group average** is an intermediate approach between the single and complete link approaches.

Ward's Method

Assumes that a cluster is represented by its centroid, and measures the proximity between two clusters in terms of the increase in sum of the squared error (SSE) that results from merging the two clusters:

$$d(A,B) = SSE_{A\cup B} - SSE_A - SSE_B$$

where A and B are clusters.

Note that for hierarchical clustering, the SSE starts at 0.

Discussion of Hierarchical Clustering

- Useful if the underlying application has a taxonomy.
- Agglomerative hierarchical clustering algorithms are expensive in terms of their computational and storage requirements.
- Merges are final and cannot be undone at a later time, preventing global optimization and causing trouble for noisy, high-dimensional data.

And now...

Lets see some hierarchical clustering!