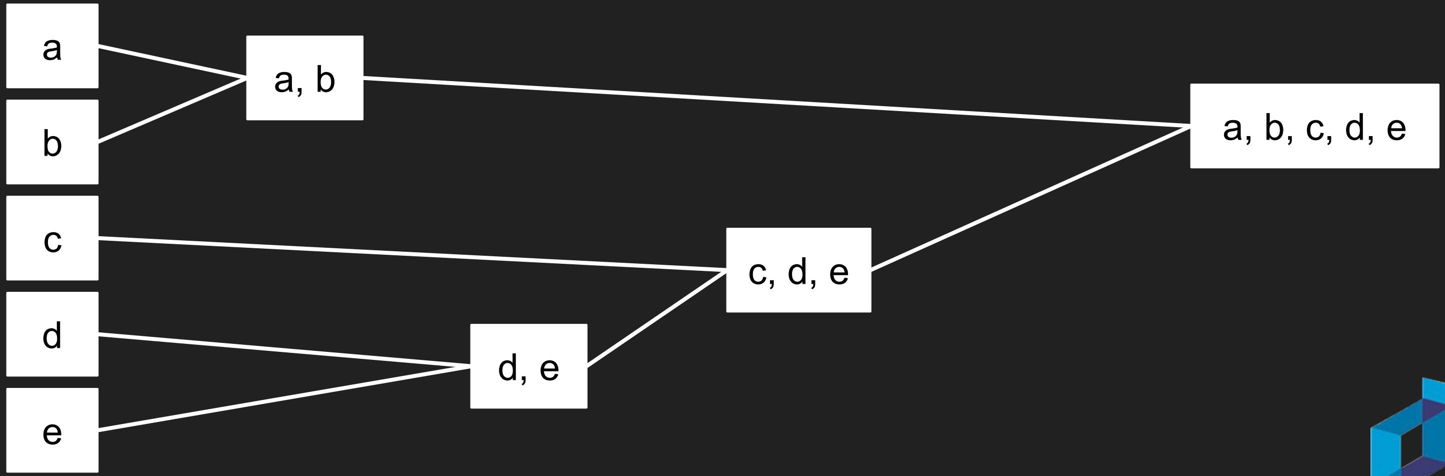
Hierarchical Clustering

Abner Cardoso, Anderson Oliveira, Anderson Uchôa,
Bruno Yusuke, Guilherme Marques and Hugo Villamizar

DEPARTAMENTO
DE INFORMÁTICA
PUC RIO

What is hierarchical clustering?

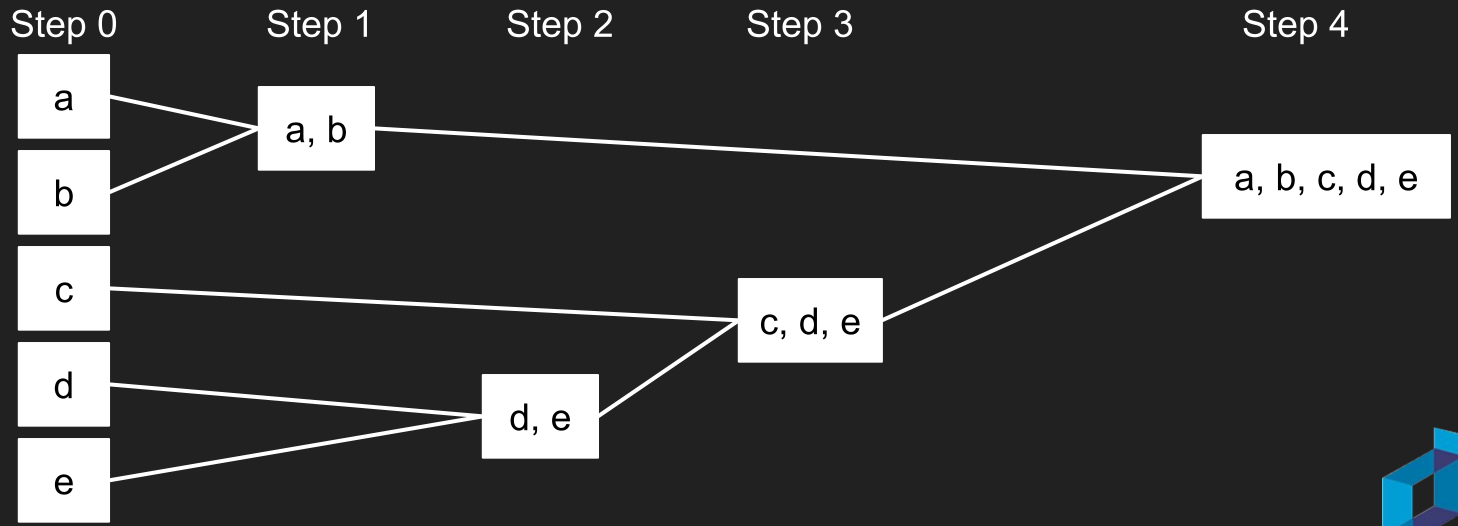
- Is the most common clustering technique
- Produces a hierarchy of nested clusters
- The hierarchy be visualized as a dendrogram: a tree like diagram that records the sequences of merges or splits



DEPARTAMENTO
DE INFORMÁTICA
PUC RIO

Agglomerative approach (aka bottom-up)

- Suppose we have five items, a, b, c, d, and e.
- Initially, we consider one cluster for each item
- Then, at each step we merge together the most similar clusters, until we generate one cluster



DEPARTAMENTO DE INFORMÁTICA PUC RIO

Agglomerative Clustering Algorithm

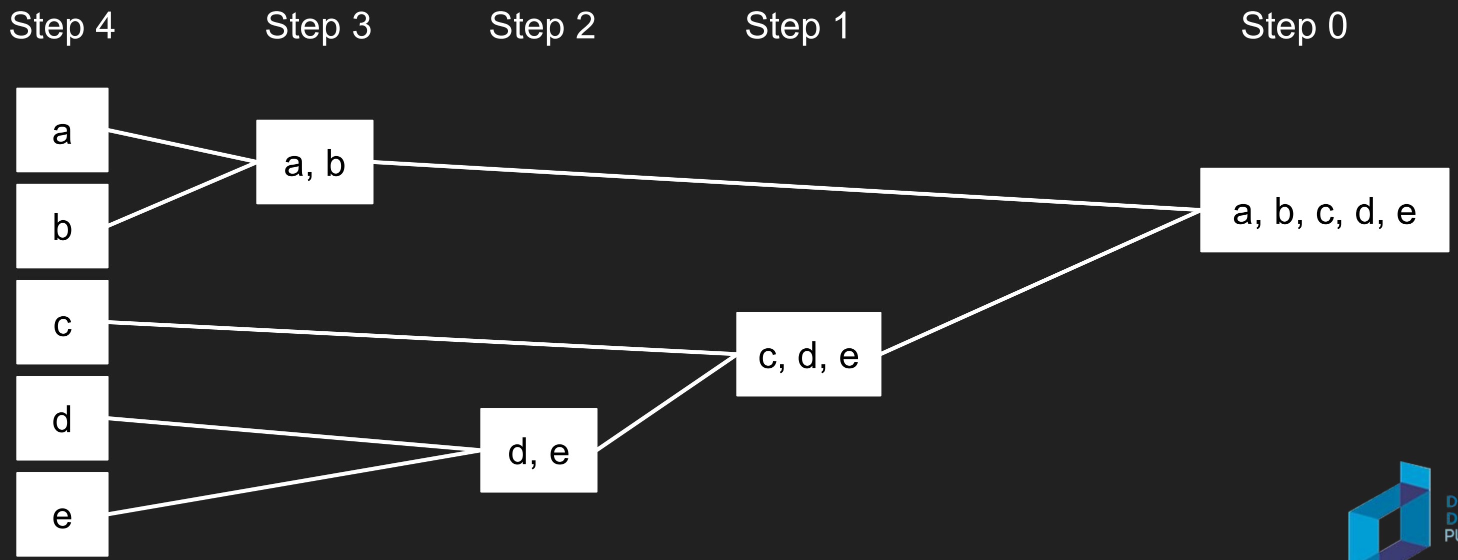
- More popular hierarchical clustering technique

- Compute the proximity matrix
- Let each data point be a cluster
- Repeat
  - Merge the two closest clusters
  - Update the proximity matrix
- Until only a single cluster remains

- Key operation is the computation of the proximity of two clusters
- Different approaches to defining the distance between clusters distinguish the different algorithms

Divisive approach (aka top-down)

- We start from one cluster containing the five elements
- Then, at each step we split one cluster to improve intra cluster similarity, until all the elements are contained in one cluster
- Top-down clustering requires a method for splitting a cluster



DEPARTAMENTO DE INFORMÁTICA PUC RIO

Comparison between the algorithms

Agglomerative	**Divisive**	**Reason**
Less complex	More Complex	We need a flat clustering to split each cluster
Less efficient in some cases	More efficient in some cases	Using an efficient flat algorithm like K-Means, divisive algorithms are linear in the number of patterns and clusters
Less accurate	More accurate	Agglomerative considers local distribution and Divisie considers global distribution

Proximity Matrix

- Stores the distances between each point.

ID	A	B	C
A	0	3	1
B	3	0	5
C	1	5	0

Proximity Matrix

- Stores the distances between each point.
- As clustering proceeds, rows and columns are merged and the distance is updated. (Merge with smaller pairwise distance)

ID	A	B	C
A	0	3	1
B	3	0	5
C	1	5	0

DEPARTAMENTO
DE INFORMÁTICA
PUC RIO

Proximity Matrix

- Stores the distances between each point.
- As clustering proceeds, rows and columns are merged and the distance is updated.
- At each iteration, as merge occurs, the table becomes smaller.

ID	AC	B
AC	0	3
B	3	0

Proximity Matrix

- Stores the distances between each point.
- As clustering proceeds, rows and columns are merged and the distance is updated.
- In the end, with all clusters merged, the distance becomes 0.

ID	ABC
ABC	0

Similarity Matrix

- Stores the similarity between each point. (Values between 0 and 1)
- As clustering proceeds, rows and columns are merged and the similarity is updated. (Merge with most similar pairwise)

ID	A	B	C
A	1	0	0.85
B	0	1	0.5
C	0.85	0.5	1

Metric / Linkage criterias

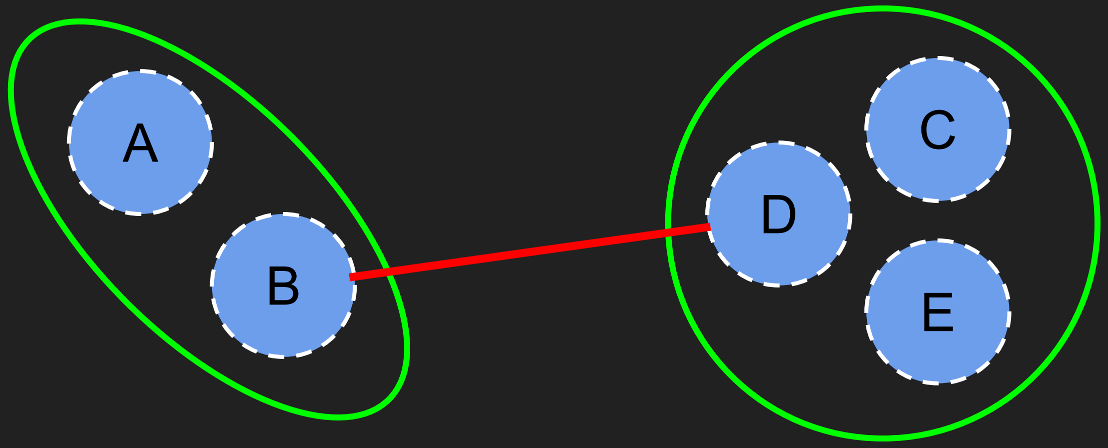
- Which clusters should be combined (for agglomerative)?
- Or where the cluster should be split (for divisive)?
- Using a good Metric (measure of distance between pairs), such as:
  - Euclidean Distance, Manhattan Distance, City Block Distance.
- Choose a Linkage criteria (measure the similarity between pairwise clusters), such as:
  - Ward, Single, Average and Complete Linkage

- Linkage criteria

- Single: Merges clusters based on minimum distance between two points from two different clusters.

- Linkage criteria

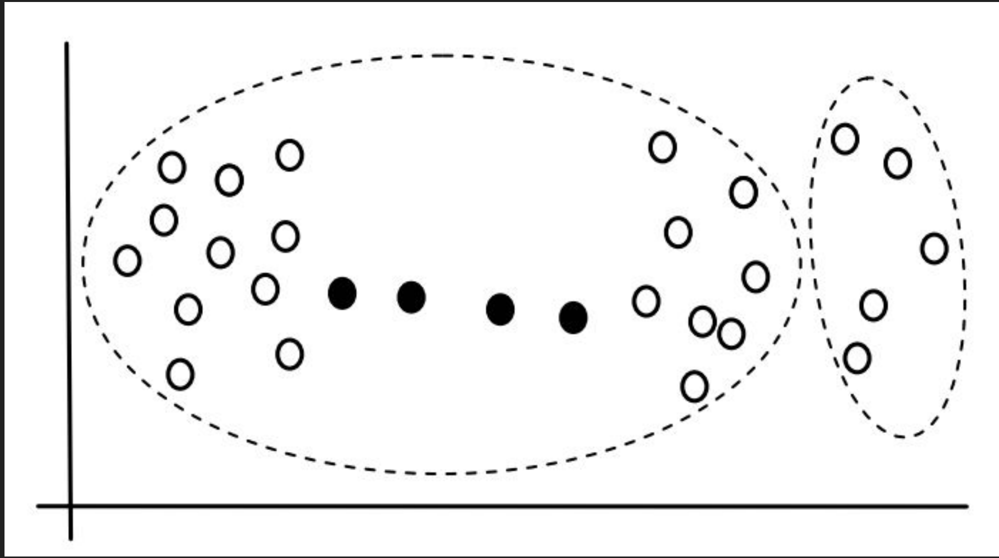
- Single: Merges clusters based on minimum distance between two points from two different clusters.



- Linkage criteria: Single Linkage

- Advantages:
  - Can handle non-elliptical shapes
- Limitation:
  - Sensitive to noise and outliers
  - May cause chaining effect

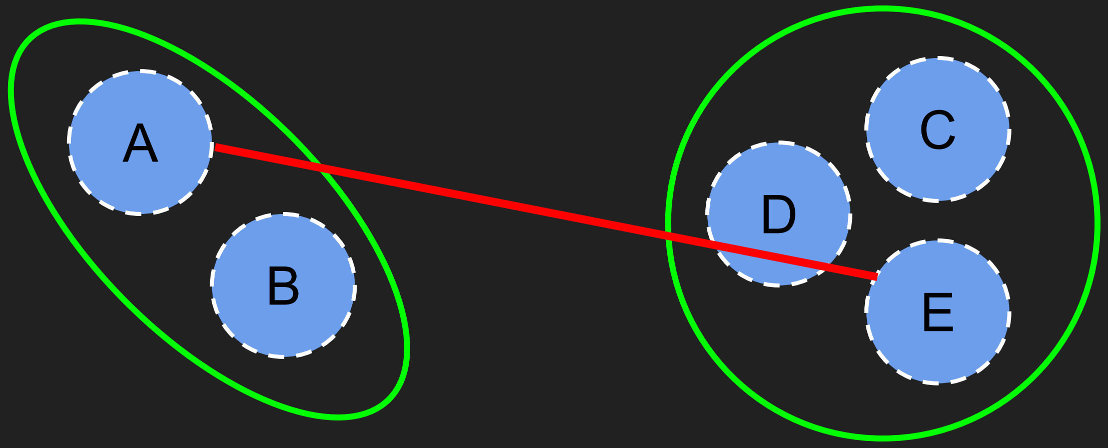
- Linkage criteria: Single Linkage



- Linkage criteria
- Complete: Merges clusters based on maximum distance between two points from two different clusters.

- Linkage criteria

- Complete: Merges clusters based on maximum distance between two points from two different clusters.



- Linkage criteria: Complete Linkage

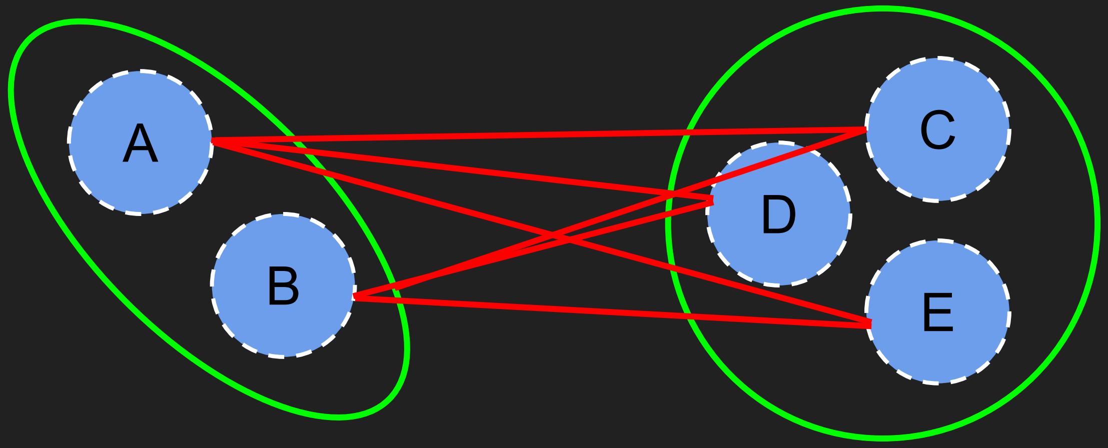
- Advantages:
    - Less susceptible to noise and outliers
- Limitation:
    - Tends to break large clusters

- Linkage criteria

- Average: Merges clusters based on average distance between all points in one group to all points in other group

- Linkage criteria

- Average: Merges clusters based on average distance between all points in one group to all points in other group



- Linkage criteria: Average Linkage

- Advantages:
  - Less susceptible to noise and outliers
- Limitation:
  - Can’t handle complicated forms

- Linkage criteria

- Ward: Merge the pair of clusters that leads to minimum increase in total within-cluster variance.

- Linkage criteria

- Ward: Merge the pair of clusters that leads to minimum increase in total within-cluster variance.

- Linkage criteria: Ward’s Linkage

- Advantages:
  - Less susceptible to noise and outliers
  - Can be used to initialize K-means
- Limitation:
  - Biased towards globular clusters

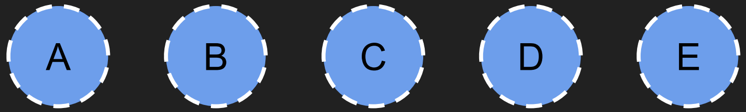
Dendrogram

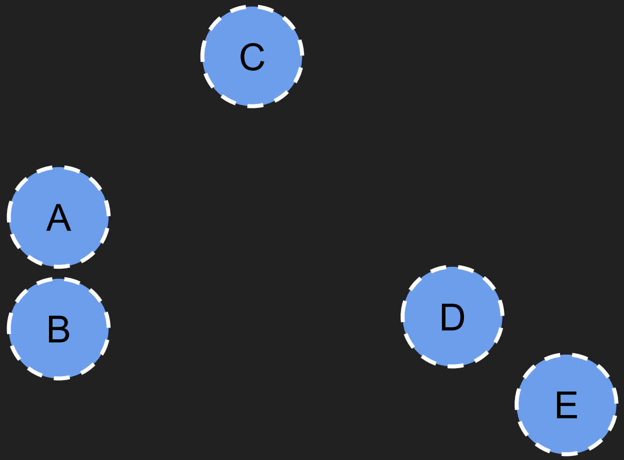
- Diagram representing a tree.

- Shows the hierarchical relationship between the clusters.

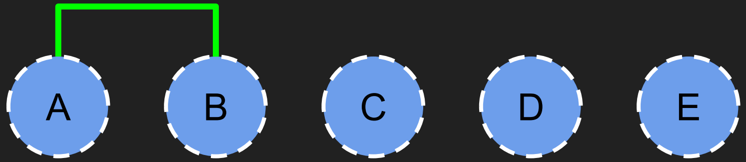
- Dendrogram is the main output of Hierarchical Clustering.

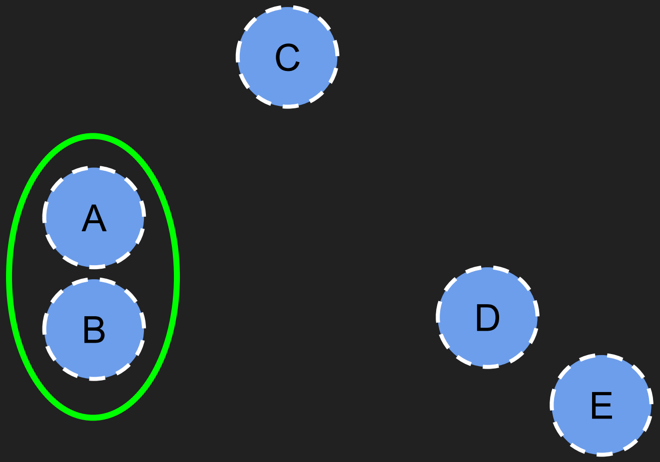
Dendrogram



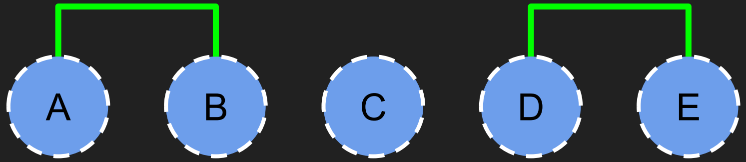


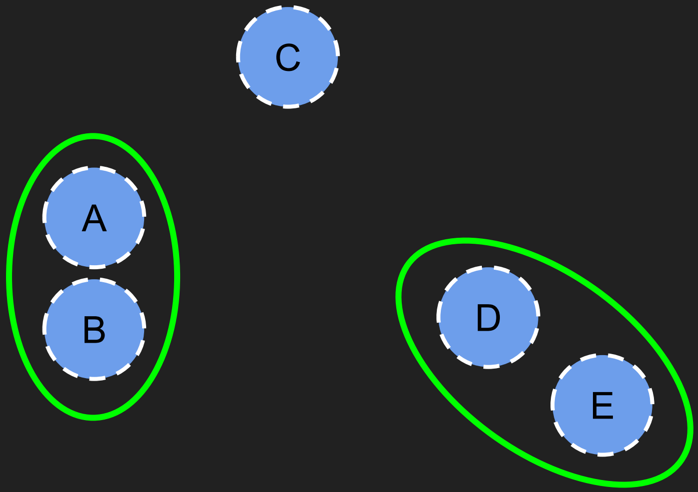
Dendrogram



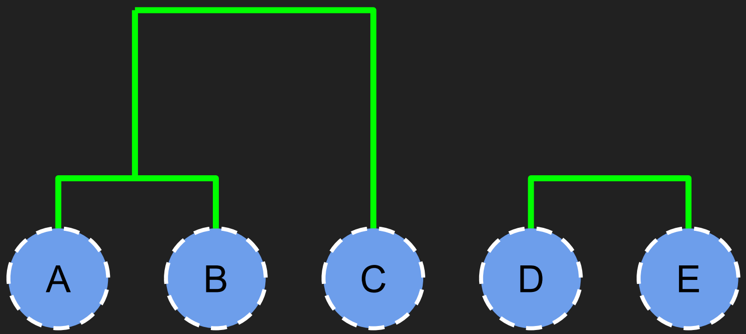


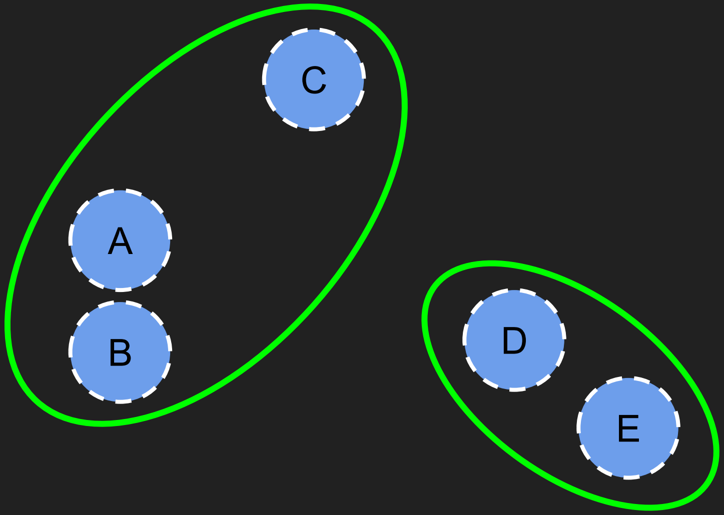
Dendrogram



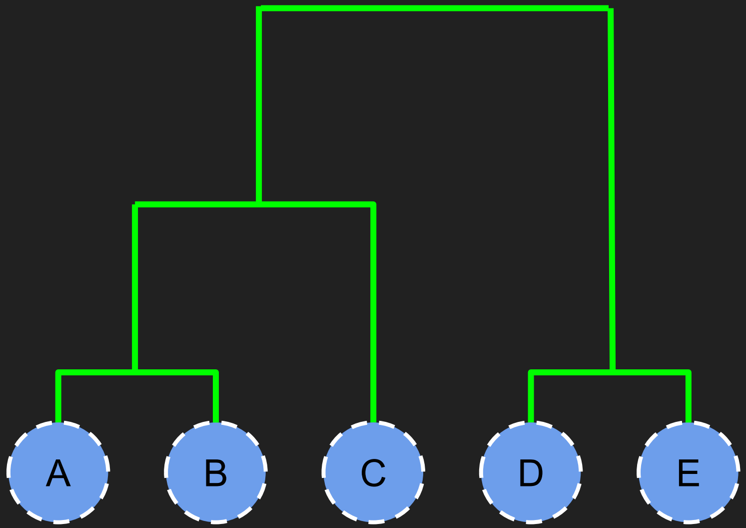


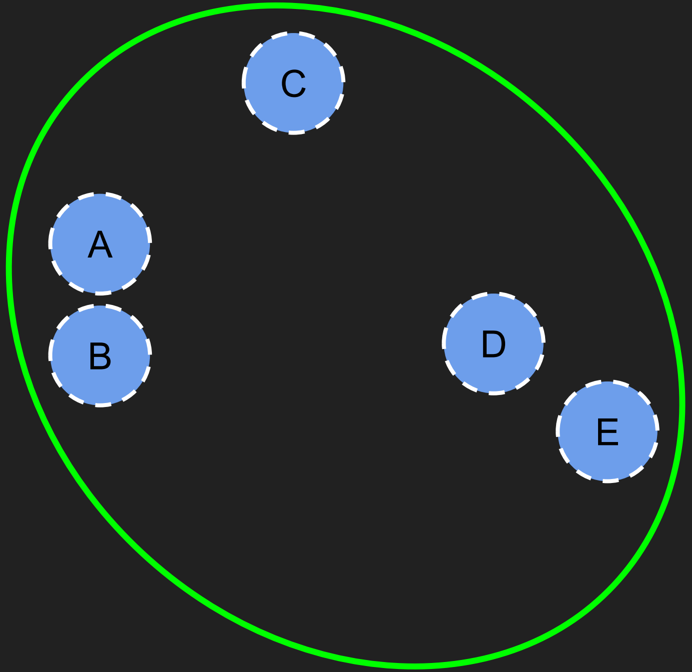
Dendrogram



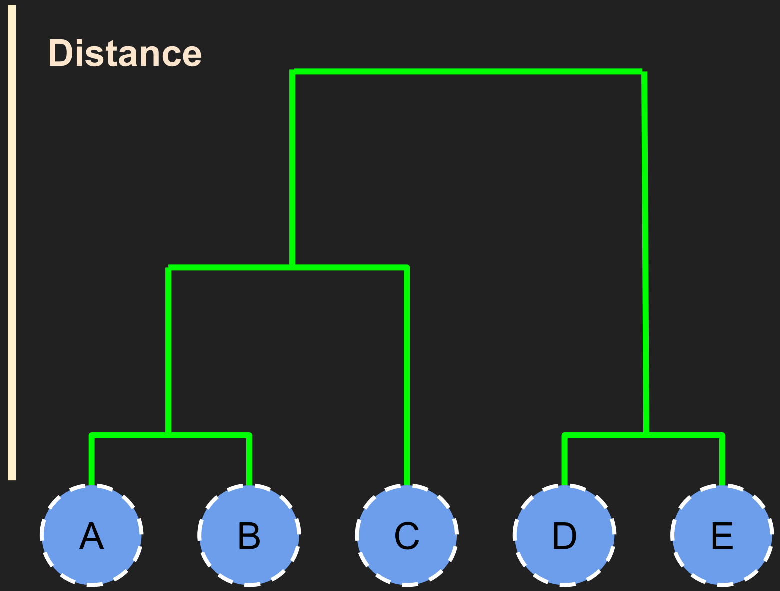


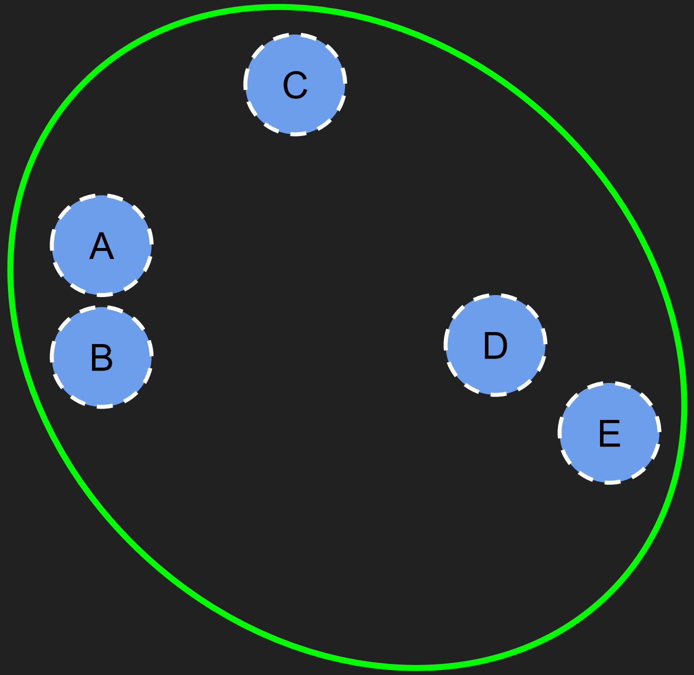
Dendrogram



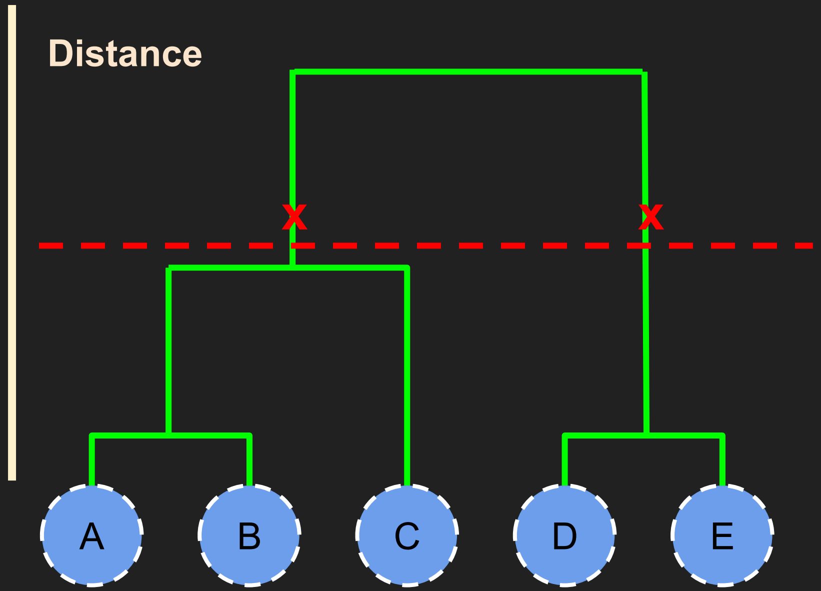


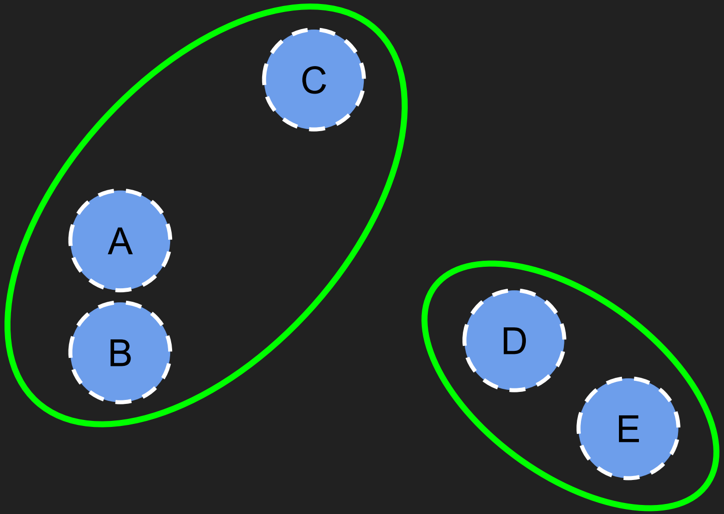
Dendrogram



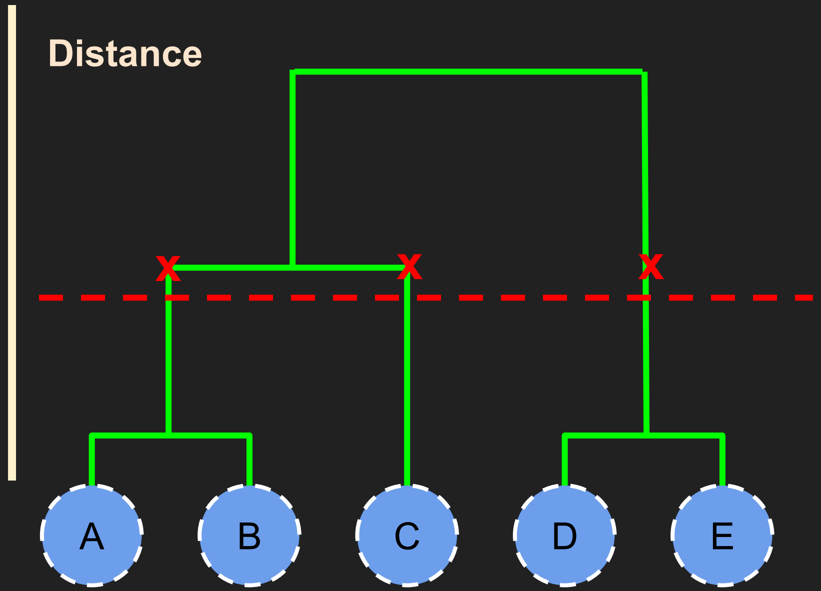


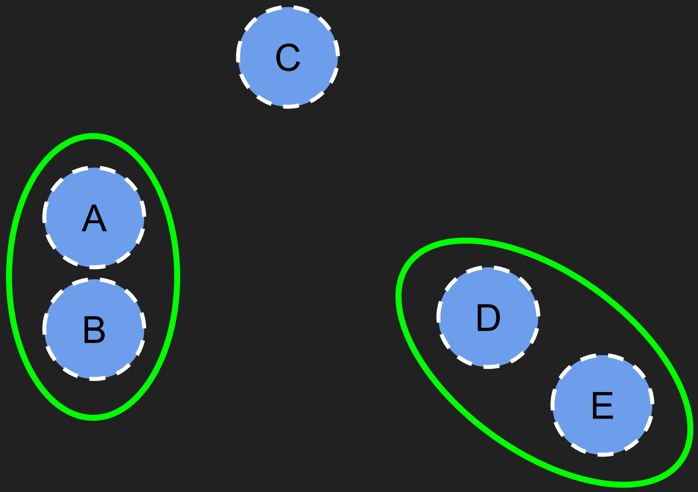
Dendrogram





Dendrogram





Dendrogram

Pros

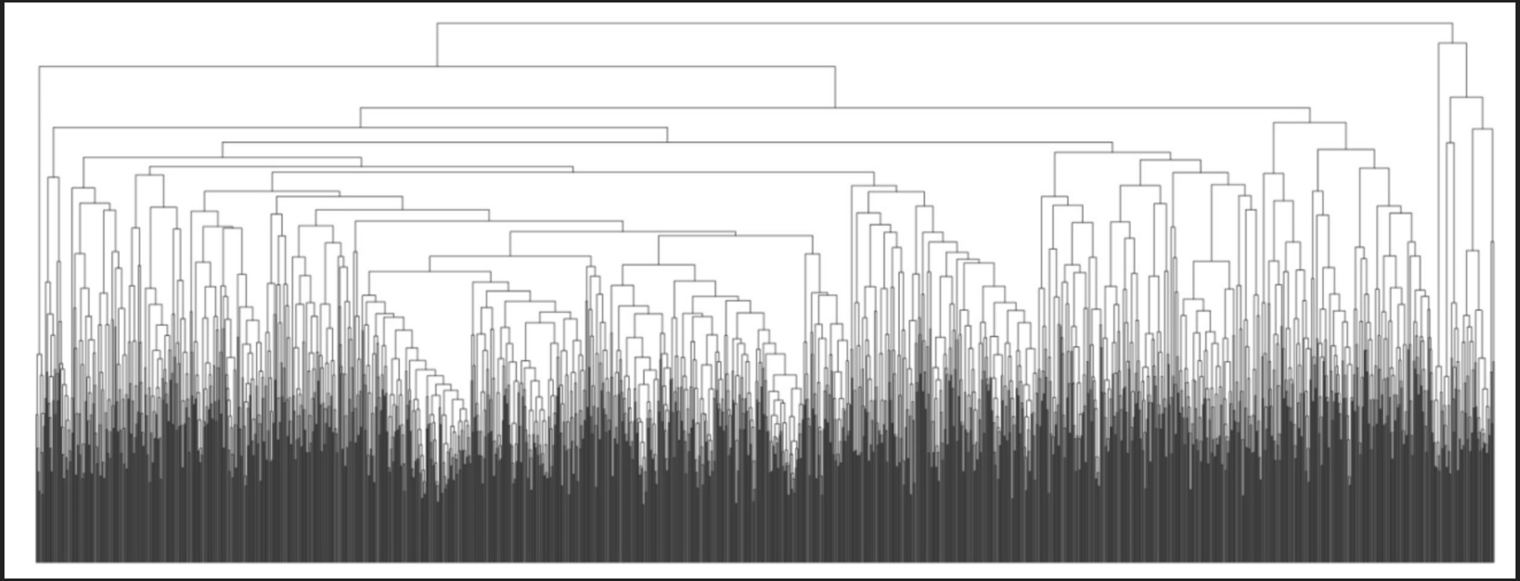
It shows **all the possible linkages** between clusters

General overview of the data

No need to preset the number of clusters

Dendrogram

Cons...



Examples

https://github.com/gpmarques/hierarchical_clustering

Conclusions

- Advantages:
  - Do not have to assume any particular number of clusters
  - Any desired number of clusters can be obtained by ‘cutting’ the dendrogram at the proper level
- Disadvantages:
  - They do not scale well: time complexity of at least O(n2), where n is the number of total objects

References

- http://web.mit.edu/6.S097/www/resources/Hierarchical.pdf
- Slides of Prof. Pier Luca Lanzi
- Finding groups in data an introduction to cluster analysis, book.
- http://www.stat.cmu.edu/~cshalizi/350/lectures/08/lecture-08.pdf

Hierarchical Clustering

Abner Cardoso, Anderson Oliveira, Anderson Uchôa,
Bruno Yusuke, Guilherme Marques and Hugo Villamizar