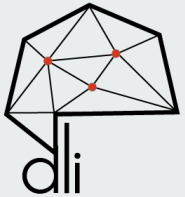
Clustering with D-Wave

Quantum implementation on D-Wave

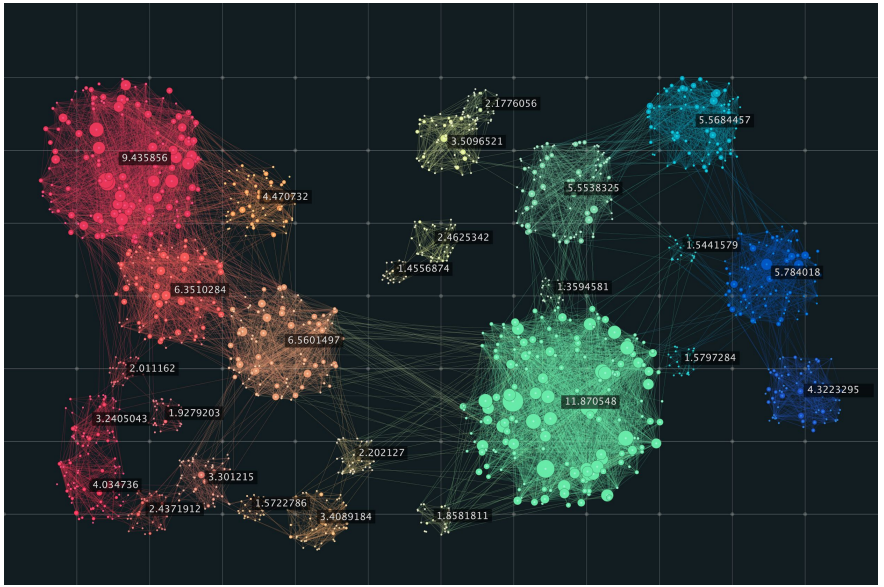


Summary

1. What is Clustering
2. Hierarchical vs Non Hierarchical
3. Similarity
4. Clustering algorithms
5. Clustering with D-Wave

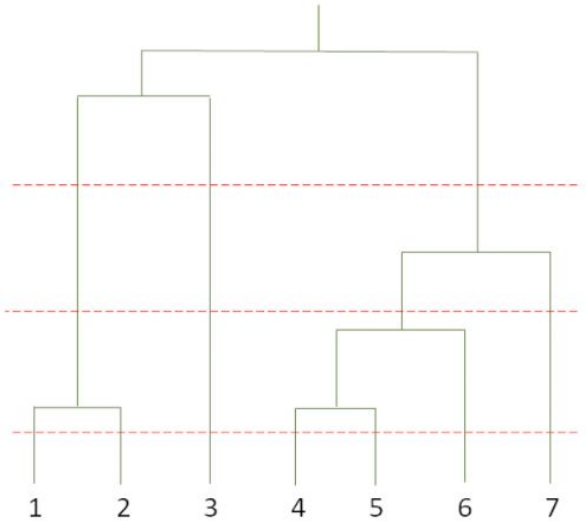
What is Clustering

- Unsupervised learning
- Requires data, but no labels
- Detect patterns in
  - Group of patients in a clinical trial
  - Profiles in a complex interactions system
  - Regions of images
- Useful when you want to understand what the data are really trying to tell you

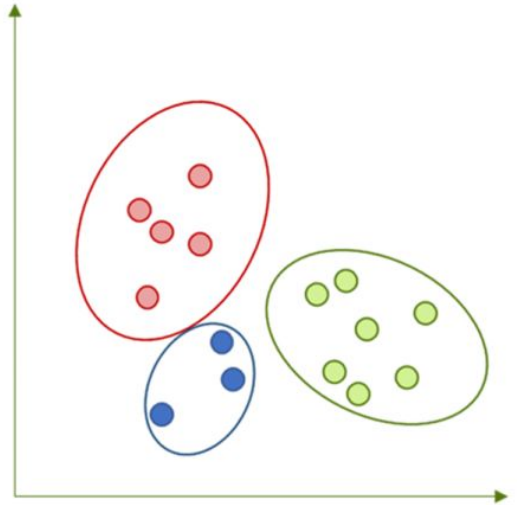


Basic types

Hierarchical



Non-hierarchical



Concept of Similarity

- To cluster you need to define a distance metric
- It depends strictly on the type of data you are looking at
- Examples:

Euclidean Distance

$$\text{dist}(\vec{x}, \vec{y}) = ||\vec{x} - \vec{y}||_2^2$$

Cosine Similarity

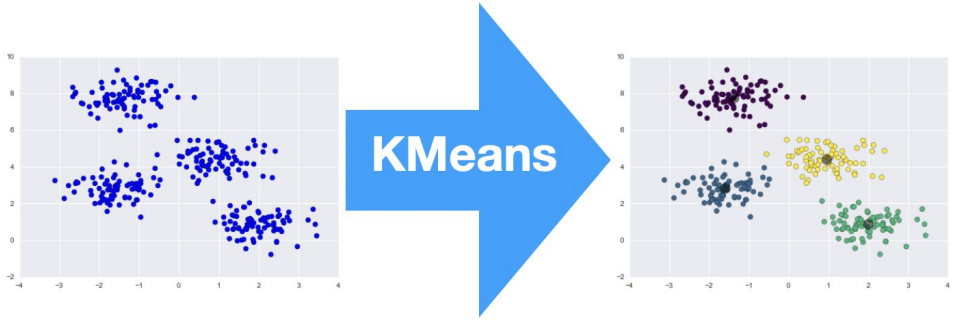
$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|}$$

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

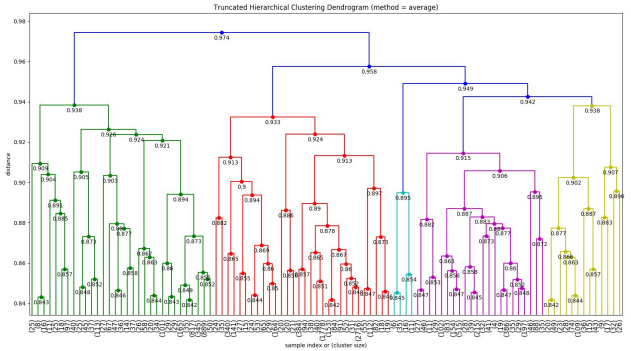
$$\|\vec{b}\| = \sqrt{b_1^2 + b_2^2 + b_3^2 + \cdots + b_n^2}$$

Clustering Algorithms

- Partition algorithms:
  - K-Means
  - Mixture Models
  - Spectral Clustering



- Hierarchical algorithms:
  - Top down → divisive
  - Bottom up → agglomerative



Clustering with D-Wave

We must create a Binary Quadratic Model for Clustering:

- Each data point can only be a part of one cluster
- Data points that are close together should be a part of the same cluster
- Data points that are far apart should be in different clusters

This is a Hard Clustering problem

D-Wave Implementation - Some Details

General formalization

- In this example we consider only 3 clusters: blue, red and green
- Every qbit represent whether or not that node is in a specific cluster
- Every data point have 3 qbits associated to it

Key aspects:

- Minimizing BQM means minimizing distance → Ensure all close points are in the same cluster
- The distance is Cosine Similarity for this example

References

- https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/#:~:text=Clustering%20is%20the%20task%20of,and%20assign%20them%20into%20clusters.
- https://cloud.dwavesys.com/leap/example-details/242592132