

The background of the slide is a collage of various data visualization elements. It includes several network graphs with nodes and edges in different colors (red, green, blue, orange). There are also scatter plots with points of various colors (orange, red, blue, green) and some heatmaps. The overall aesthetic is technical and data-driven.

Session 8: Probabilistic Hierarchical Clustering

Question

- ❑ Which of the following statements about probabilistic hierarchical clustering is FALSE:
 - ❑ When measuring the quality of a clustering, we take the product of the likelihood of each individual cluster.
 - ❑ Two clusters are merged if their distance defined in the probabilistic setting is > 0 .
 - ❑ Probabilistic hierarchical clustering is considered a generative model: the set of data objects to be clustered is assumed to be drawn from some underlying data generation mechanism.
- ❑ Answer: Second option. We only merge two clusters if their distance is < 0 , which results in an increase in the clustering quality.