

# Probabilistic Hierarchical Clustering

---

- ❑ Algorithmic hierarchical clustering
  - ❑ Nontrivial to choose a good distance measure
  - ❑ Hard to handle missing attribute values
  - ❑ Optimization goal not clear: heuristic, local search
- ❑ Probabilistic hierarchical clustering
  - ❑ Use probabilistic models to measure distances between clusters
  - ❑ Generative model: Regard the set of data objects to be clustered as a sample of the underlying data generation mechanism to be analyzed
  - ❑ Easy to understand, same efficiency as algorithmic agglomerative clustering method, can handle partially observed data
- ❑ In practice, assume the generative models adopt common distribution functions, e.g., Gaussian distribution or Bernoulli distribution, governed by parameters