

The background of the slide is a complex, abstract composition. It features a network of thin, light-colored lines forming a web-like structure. Overlaid on this are various data visualizations: a grid of small, light-colored dots, a series of vertical bars of varying heights, and a large, irregular shape filled with a dense pattern of small, colored dots (green, blue, and yellow). The overall color palette is muted, with shades of brown, beige, and light blue.

What Is Pattern Discovery? Why Is It Important?

What Is Pattern Discovery?

❑ What are patterns?

- ❑ **Patterns**: A set of items, subsequences, or substructures that occur frequently together (or strongly correlated) in a data set
- ❑ Patterns represent **intrinsic** and **important properties** of datasets

❑ **Pattern discovery**: Uncovering patterns from massive data sets

❑ Motivation examples:

- ❑ What products were often purchased together?
- ❑ What are the subsequent purchases after buying an iPad?
- ❑ What code segments likely contain copy-and-paste bugs?
- ❑ What word sequences likely form phrases in this corpus?

Pattern Discovery: Why Is It Important?

- ❑ Finding **inherent regularities** in a data set
- ❑ **Foundation** for many essential data mining tasks
 - ❑ Association, correlation, and causality analysis
 - ❑ Mining sequential, structural (e.g., sub-graph) patterns
 - ❑ Pattern analysis in spatiotemporal, multimedia, time-series, and stream data
 - ❑ Classification: Discriminative pattern-based analysis
 - ❑ Cluster analysis: Pattern-based subspace clustering
- ❑ Broad applications
 - ❑ Market basket analysis, cross-marketing, catalog design, sale campaign analysis, Web log analysis, biological sequence analysis