

The background of the slide is a complex, abstract composition. It features a dark, muted purple or brownish background. Overlaid on this are several geometric and data-like elements: a network of thin, light-colored lines forming a mesh or web-like structure; numerous small, green and blue dots scattered across the field; and a prominent, lighter-colored, semi-transparent geometric shape (a large triangle or polygon) that serves as a backdrop for the title. In the bottom left corner, there is a small, square inset image showing a cluster of orange and red dots on a light background, with a horizontal bar chart overlaid on it.

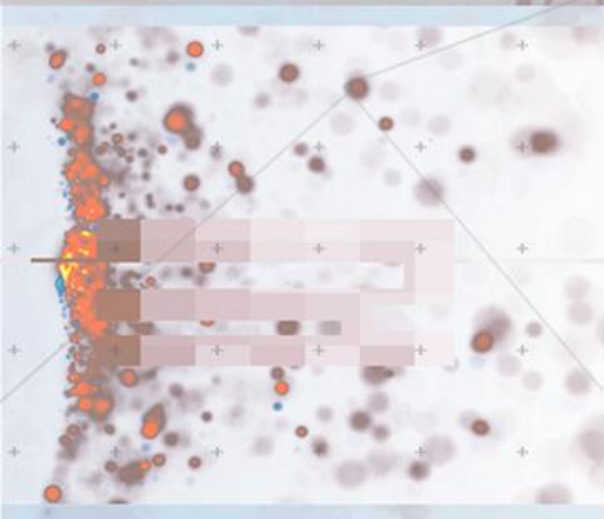
Lecture 3. Efficient Pattern Mining Methods

Lecture 3. Efficient Pattern Mining Methods


- ❑ The Downward Closure Property of Frequent Patterns
- ❑ The Apriori Algorithm
- ❑ Extensions or Improvements of Apriori
- ❑ Mining Frequent Patterns by Exploring Vertical Data Format
- ❑ FPGrowth: A Frequent Pattern-Growth Approach
- ❑ Mining Closed Patterns

The background of the slide is a complex, abstract composition. It features a network of thin, reddish-brown lines connecting various points, some of which are small green dots. This network is overlaid on a light beige background with a subtle grid of small grey plus signs. On the right side, there is a dark, textured area with a pattern of thin, white, intersecting lines. In the bottom right corner, there are some faint, colorful geometric shapes in shades of blue and orange. The overall aesthetic is technical and mathematical.

Session 1: The Downward Closure Property of Frequent Patterns



The Downward Closure Property of Frequent Patterns

- ❑ Observation: From $TDB_1: T_1: \{a_1, \dots, a_{50}\}; T_2: \{a_1, \dots, a_{100}\}$
 - ❑ We get a frequent itemset: (a_1, \dots, a_{50})
 - ❑ Also, its subsets are all frequent: $(a_1), (a_2), \dots, (a_{50}), (a_1, a_2), \dots, (a_1, \dots, a_{49}), \dots$
 - ❑ There must be some hidden relationships among frequent patterns!
- ❑ The **downward closure (also called “Apriori”)** property of frequent patterns
 - ❑ If **{beer, diaper, nuts}** is frequent, so is **{beer, diaper}**
 - ❑ Every transaction containing {beer, diaper, nuts} also contains {beer, diaper}
 - ❑ Apriori: Any subset of a frequent itemset must be frequent
- ❑ Efficient mining methodology
 - ❑ If **any subset of an itemset S** is infrequent, then there is no chance for S to be frequent—why do we even have to consider S!?  A sharp knife for pruning!

Apriori Pruning and Scalable Mining Methods

- ❑ Apriori pruning principle: If there is any itemset which is infrequent, its superset should not even be generated! (Agrawal & Srikant @VLDB'94, Mannila, et al. @ KDD' 94)
- ❑ Scalable mining Methods: Three major approaches
 - ❑ Level-wise, join-based approach: Apriori (Agrawal & Srikant@VLDB'94)
 - ❑ Vertical data format approach: Eclat (Zaki, Parthasarathy, Ogihara, Li @KDD'97)
 - ❑ Frequent pattern projection and growth: FPgrowth (Han, Pei, Yin @SIGMOD'00)