

The background of the slide is a complex, abstract composition. It features a network of thin, reddish-brown lines connecting various points, creating a web-like structure. These points are represented by small, multi-colored dots in shades of green, blue, and yellow. The background is divided into several geometric sections by light gray lines. On the left side, there is a vertical strip with a grid of small, light gray plus signs. In the center, there is a large, white, angular shape that serves as a backdrop for the title. To the left of this shape, there is a smaller, rectangular inset showing a cluster of orange and red dots, with a horizontal band of pink and white squares overlaid on it. The overall aesthetic is technical and modern, suggesting themes of data science, network theory, or pattern recognition.

Lecture 4. Pattern Evaluation

Lecture 4. Pattern Evaluation

- ❑ Interestingness Measures in Pattern Mining
- ❑ Interestingness Measures: Lift and χ^2
- ❑ Null-Invariant Measures
- ❑ Comparison of Interestingness Measures

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 points and patterns. In the upper left, there are rows of small, light-colored symbols that look like mathematical or logical notations. In the lower left, there is a rectangular inset showing a cluster of orange and red dots, possibly representing a specific dataset or a visualization of a network component. The overall color palette is muted, with shades of brown, grey, and white, accented by the colors of the data points.

Session 1: Limitation of the Support-Confidence Framework

How to Judge if a Rule/Pattern Is Interesting?

- ❑ Pattern-mining will generate a large set of patterns/rules
 - ❑ Not all the generated patterns/rules are interesting
- ❑ Interestingness measures: Objective vs. subjective
 - ❑ Objective interestingness measures
 - ❑ Support, confidence, correlation, ...
 - ❑ Subjective interestingness measures: One man's trash could be another man's treasure
 - ❑ Query-based: Relevant to a user's particular request
 - ❑ Against one's knowledge-base: unexpected, freshness, timeliness
 - ❑ Visualization tools: Multi-dimensional, interactive examination

Limitation of the Support-Confidence Framework

- Are s and c interesting in association rules: " $A \Rightarrow B$ " [s, c]? **Be careful!**
- Example: Suppose one school may have the following statistics on # of students who may play basketball and/or eat cereal:

	play-basketball	not play-basketball	sum (row)
eat-cereal	400	350	750
not eat-cereal	200	50	250
sum(col.)	600	400	1000

2-way contingency table

- Association rule mining may generate the following:
 - $play\text{-}basketball \Rightarrow eat\text{-}cereal$ [40%, 66.7%] (higher s & c)
- But this strong association rule is misleading: The overall % of students eating cereal is 75% > 66.7%, a more telling rule:
 - $\neg play\text{-}basketball \Rightarrow eat\text{-}cereal$ [35%, 87.5%] (high s & c)