

The background of the slide is a complex, abstract composition. It features a network of thin, reddish-brown lines forming a web-like structure. Scattered throughout are numerous small, colored dots in shades of green, blue, and orange. On the left side, there is a vertical strip with a grid of small, light-colored squares. In the center, a large, white, V-shaped or chevron-like graphic element points downwards, serving as a backdrop for the title. The overall aesthetic is technical and data-driven.

Mining Compressed Patterns

Mining Compressed Patterns

Pat-ID	Item-Sets	Support
P1	{38,16,18,12}	205227
P2	{38,16,18,12,17}	205211
P3	{39,38,16,18,12,17}	101758
P4	{39,16,18,12,17}	161563
P5	{39,16,18,12}	161576

- ❑ Closed patterns
 - ❑ P1, P2, P3, P4, P5
 - ❑ Emphasizes too much on support
 - ❑ There is no compression
- ❑ Max-patterns
 - ❑ P3: information loss
- ❑ Desired output (a good balance):
 - ❑ P2, P3, P4

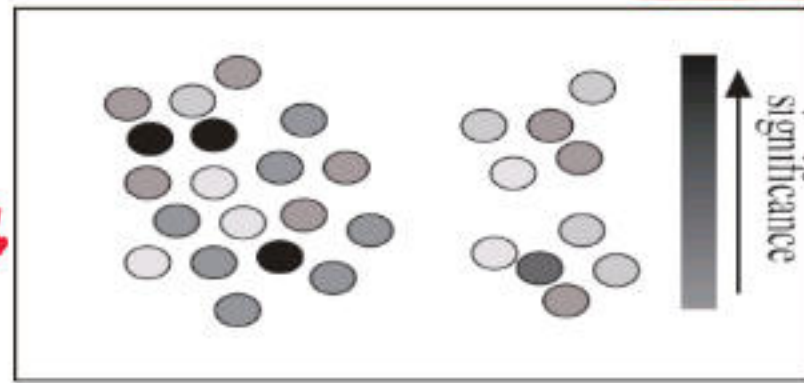
- ❑ Why mining compressed patterns?
 - ❑ Too many scattered patterns but not so meaningful
- ❑ Pattern distance measure

$$\star Dist(P_1, P_2) = 1 - \frac{|T(P_1) \cap T(P_2)|}{|T(P_1) \cup T(P_2)|}$$
- ❑ δ -clustering: For each pattern P, find all patterns which can be expressed by P and whose distance to P is within δ (δ -cover)
- ❑ All patterns in the cluster can be represented by P
- ❑ Method for efficient, direct mining of compressed frequent patterns (e.g., D. Xin, J. Han, X. Yan, H. Cheng, "On Compressing Frequent Patterns", Knowledge and Data Engineering, 60:5-29, 2007)

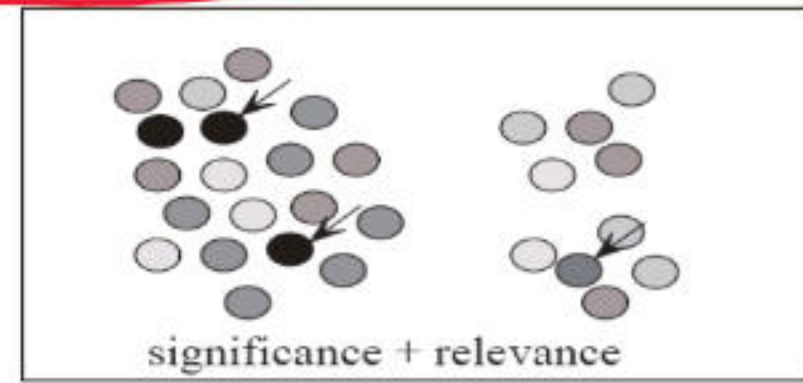
Redundancy-Aware Top-k Patterns

- Desired patterns: high significance & low redundancy

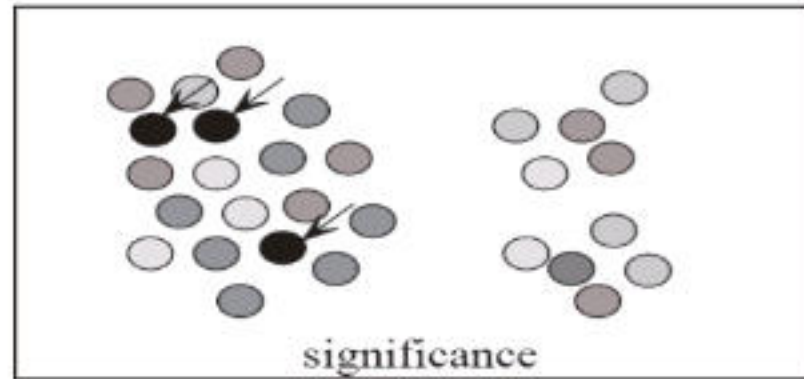
more
darker: \checkmark significant
lighter: less significant



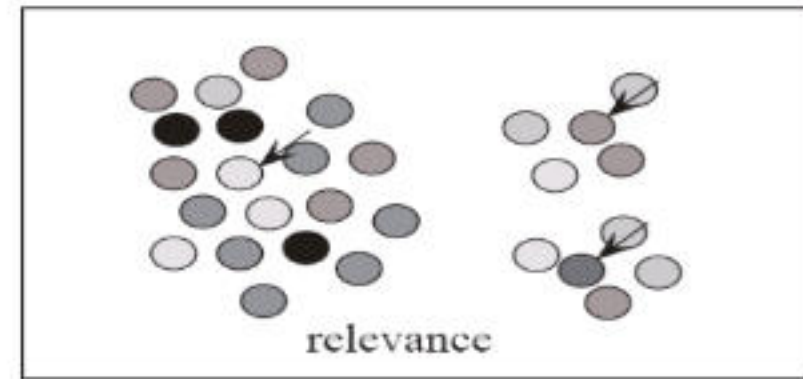
(a) a set of patterns



(b) redundancy-aware top-k

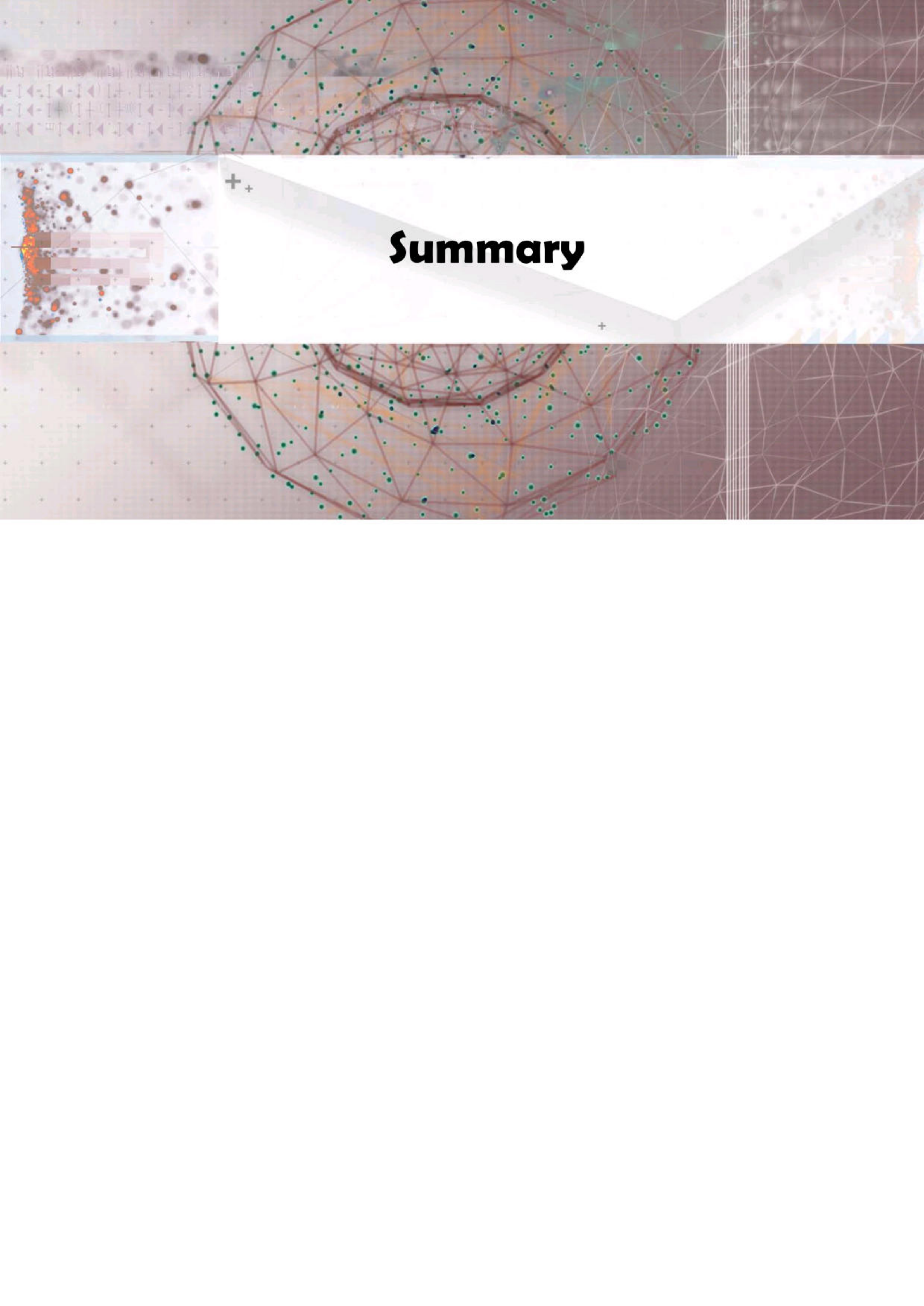


(c) traditional top-k



(d) summarization

- Method: Use MMS (Maximal Marginal Significance) for measuring the combined significance of a pattern set
- Xin et al., Extracting Redundancy-Aware Top-K Patterns, KDD'06



Summary

Summary: Mining Diverse Patterns

- Efficient methods have been developed for mining various kinds of patterns
 - Mining Multiple-Level Associations
 - Mining Multi-Dimensional Associations
 - Mining Quantitative Associations
 - Mining Negative Correlations
 - Mining Compressed and Redundancy-Aware Patterns

Recommended Readings

- ❑ R. Srikant and R. Agrawal, “Mining generalized association rules”, VLDB'95
- ❑ Y. Aumann and Y. Lindell, “A Statistical Theory for Quantitative Association Rules”, KDD'99
- ❑ K. Wang, Y. He, J. Han, “Pushing Support Constraints Into Association Rules Mining”, IEEE Trans. Knowledge and Data Eng. 15(3): 642-658, 2003
- ❑ D. Xin, J. Han, X. Yan and H. Cheng, "On Compressing Frequent Patterns", Knowledge and Data Engineering, 60(1): 5-29, 2007
- ❑ D. Xin, H. Cheng, X. Yan, and J. Han, "Extracting Redundancy-Aware Top-K Patterns", KDD'06
- ❑ J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent Pattern Mining: Current Status and Future Directions", Data Mining and Knowledge Discovery, 15(1): 55-86, 2007