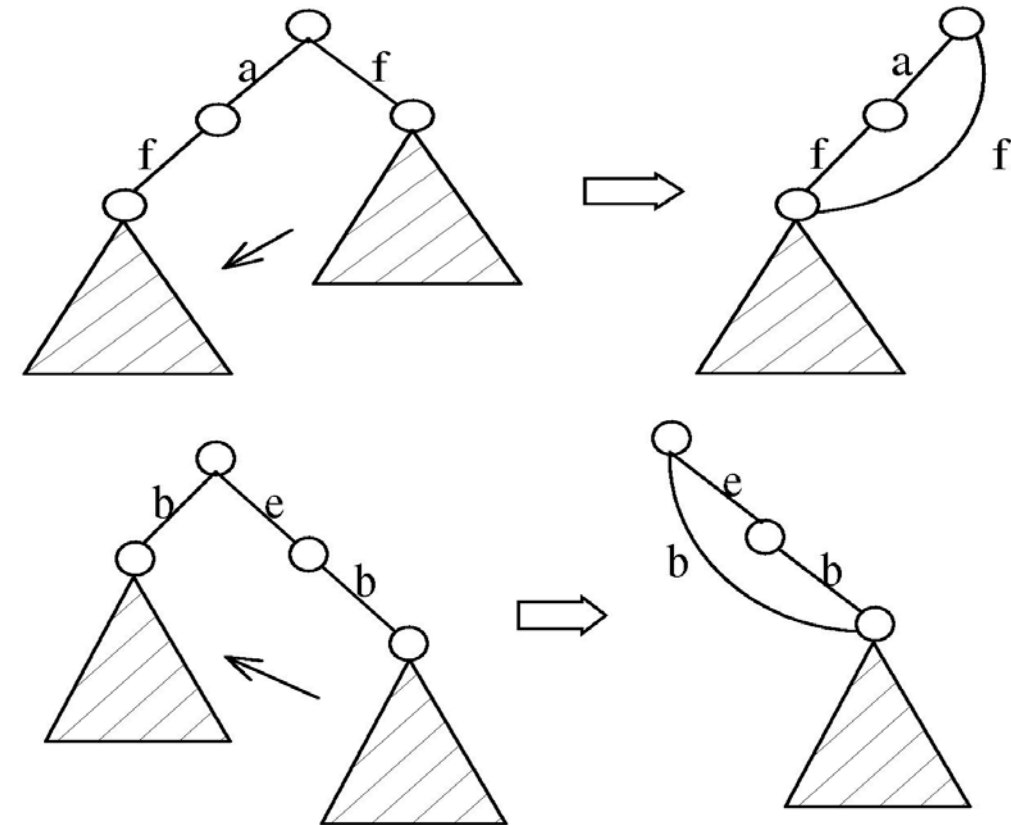


The background features a complex network of red lines connecting green dots, overlaid on a light blue grid. A large, semi-transparent white banner with a grey border is positioned across the center, containing the title text. The banner has small grey plus signs at its corners.

CloSpan: Mining Closed Sequential Patterns

CloSpan: Mining Closed Sequential Patterns

- ❑ A **closed sequential pattern** s : There exists no superpattern s' such that $s' \supset s$, and s' and s have the same support
- ❑ Which ones are closed? $\langle abc \rangle: 20$, $\langle abcd \rangle: 20$, $\langle abcde \rangle: 15$
- ❑ Why directly mine closed sequential patterns?
 - ❑ Reduce # of (redundant) patterns
 - ❑ Attain the same expressive power
- ❑ Property P_1 : If $s \supset s_1$, s is closed iff two project DBs have the same size
- ❑ Explore **Backward Subpattern** and **Backward Superpattern** pruning to prune redundant search space
- ❑ Greatly enhances efficiency (Yan, et al., SDM'03)





Summary

Summary: Sequential Pattern Mining

- ❑ Concepts of Sequential Pattern Mining
- ❑ Sequential Pattern Mining Algorithms
 - ❑ **GSP** (Generalized Sequential Patterns)
 - ❑ Vertical Format-Based Mining: **SPADE**
 - ❑ Pattern-Growth Methods: **PrefixSpan**
- ❑ Mining Closed Sequential Patterns: **CloSpan**

Recommended Readings

- ❑ R. Srikant and R. Agrawal, “Mining sequential patterns: Generalizations and performance improvements”, EDBT’96
- ❑ M. Zaki, “SPADE: An Efficient Algorithm for Mining Frequent Sequences”, Machine Learning, 2001
- ❑ J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach", IEEE TKDE, 16(10), 2004
- ❑ X. Yan, J. Han, and R. Afshar, “CloSpan: Mining Closed Sequential Patterns in Large Datasets”, SDM'03