

The background features a complex network of thin, light-colored lines forming a mesh-like structure. Overlaid on this are various data visualization elements: a grid of small grey plus signs, a series of purple arrows pointing left, and a large, dense network of green and blue dots connected by red lines. In the lower-left corner, there is a rectangular inset showing a scatter plot with orange and blue dots, a red line, and a heatmap with a color gradient from yellow to red.

Session 5. 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)

