

The background of the slide features a complex network graph with numerous nodes and edges, rendered in shades of red, orange, and green. The nodes are represented by small dots, and the edges are thin lines connecting them. The overall aesthetic is technical and data-driven.

Pattern Mining Applications: Mining Quality Phrases from Text Data

Pattern Mining Applications: Mining Quality Phrases from Text Data

- ❑ From Frequent Pattern Mining to Phrase Mining
- ❑ Previous Phrase Mining Methods
- ❑ ToPMine: Phrase Mining without Training Data
- ❑ SegPhrase: Phrase Mining with Tiny Training Sets

Thanks to Ahmed El-Kishky@UIUC, Jialu Liu@UIUC, Jingbo Shang@UIUC, Xiang Ren@UIUC, Chi Wang@MSR and Marina Danilevsky@IBM for their contributions

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 this network are numerous small, green circular dots. In the upper left corner, there is a smaller, semi-transparent inset showing a different pattern of dots, some in orange and some in brown, arranged in a more structured, grid-like fashion. The overall color palette is muted, with earthy tones and a soft, hazy atmosphere.

From Frequent Pattern Mining to Phrase Mining

Why Phrase Mining?

- Unigrams vs. phrases
 - **Unigrams** (single words) are often *ambiguous*
 - Example: “United”: United States? United Airline? United Parcel Service?
 - **Phrase**: A natural, meaningful, *unambiguous* semantic unit
 - Example: “United States” vs. “United Airline”
- Mining semantically meaningful phrases
 - Transform text data from *word granularity* to *phrase granularity*
 - Enhance the power and efficiency at manipulating unstructured data

From Frequent Pattern Mining to Phrase Mining

- General principle
 - Exploit information redundancy and data-driven criteria to determine phrase boundaries and salience
- Methodology: Exploring three ideas
 - Frequent pattern mining and colocation analysis
 - Phrasal segmentation
 - Quality phrase assessment
- Recent developments of phrase mining methods
 - ToPMine: Mining quality phrase without training (A. El-Kishky, et al., 2015)
 - SegPhrase: Mining quality phrase with tiny training sets (J. Liu, et al., 2015)