

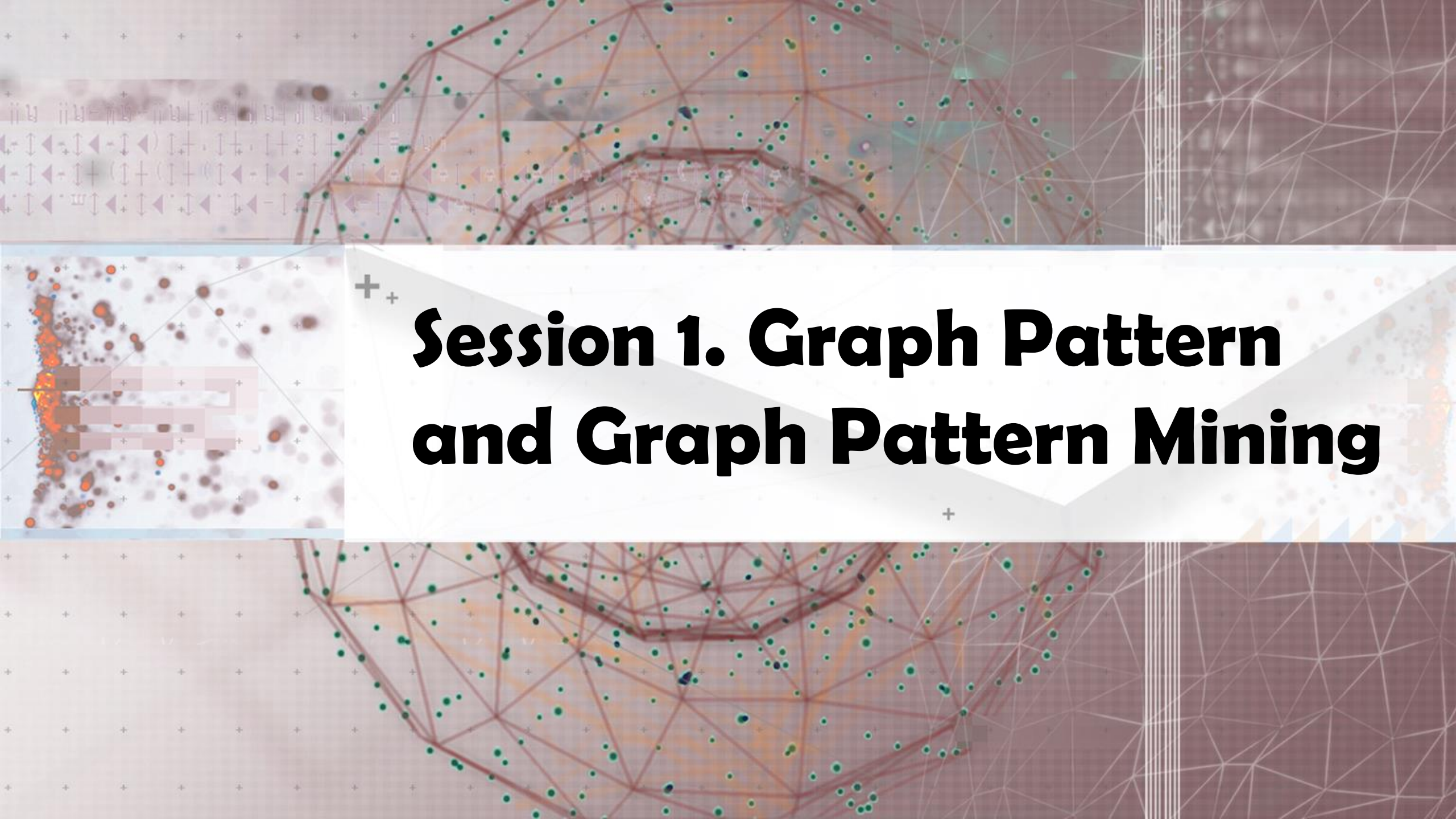
The background of the slide is a complex, abstract composition. It features a network graph with numerous green nodes and red edges, overlaid on a grid of small white plus signs. The background is divided into several geometric sections by white lines, including a large white triangle on the right and a smaller one on the left. The overall color palette is muted, with shades of brown, grey, and white.

Lecture 8. Graph Pattern Mining

Lecture 8. Graph Pattern Mining

- ❑ Graph Pattern and Graph Pattern Mining
- ❑ Apriori-Based Graph Pattern Mining Methods
- ❑ gSpan: A Pattern-Growth-Based Method
- ❑ CloseGraph: Mining Closed Graph Patterns
- ❑ Graph Pattern Mining Application: Graph Indexing
- ❑ Mining Top-K Large Structural Patterns in a Massive Network

Thanks to Xifeng Yan@UCSB and Feida Zhu@SMU.SG for their contributions

The background of the slide is a collage of various data visualization elements. It includes several network graphs with nodes and edges in different colors (red, green, blue). There are also scatter plots with colored dots, a heatmap with a grid of colored squares, and various mathematical symbols and formulas scattered throughout. The overall aesthetic is technical and data-driven.

Session 1. Graph Pattern and Graph Pattern Mining

Frequent (Sub)Graph Patterns

- Given a labeled graph dataset $D = \{G_1, G_2, \dots, G_n\}$, the supporting graph set of a subgraph g is $D_g = \{G_i \mid g \subseteq G_i, G_i \in D\}$.

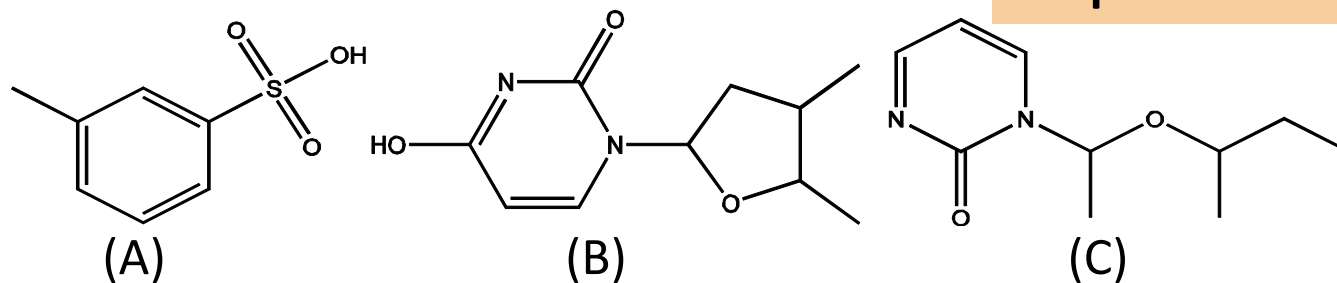
- $\text{support}(g) = |D_g| / |D|$

- A (sub)graph g is **frequent** if $\text{support}(g) \geq \text{min_sup}$

- Ex.: Chemical structures

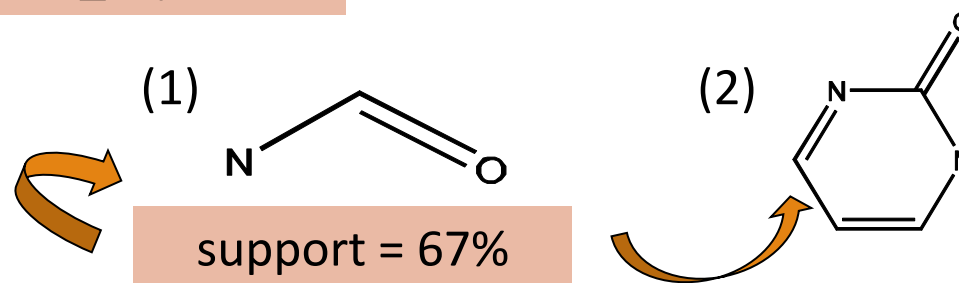
- Alternative:

- Mining frequent subgraph patterns from a single large graph or network



min_sup = 2

Frequent Graph Patterns



Applications of Graph Pattern Mining

- ❑ Bioinformatics
 - ❑ Gene networks, protein interactions, metabolic pathways
- ❑ Chem-informatics: Mining chemical compound structures
- ❑ Social networks, web communities, tweets, ...
- ❑ Cell phone networks, computer networks, ...
- ❑ Web graphs, XML structures, semantic Web, information networks
- ❑ Software engineering: program execution flow analysis
- ❑ Building blocks for graph classification, clustering, compression, comparison, and correlation analysis
- ❑ Graph indexing and graph similarity search

Graph Pattern Mining Algorithms: Different Methodologies

- ❑ Generation of candidate subgraphs
 - ❑ Apriori vs. pattern growth (e.g., FSG vs. gSpan)
- ❑ Search order
 - ❑ Breadth vs. depth
- ❑ Elimination of duplicate subgraphs
 - ❑ Passive vs. active (e.g., gSpan (Yan&Han'02))
- ❑ Support calculation
 - ❑ Store embeddings (e.g., GASTON (Nijssen&Kok'04, FFSM (Huan, et al.'03), MoFa (Borgelt and Berthold ICDM'02))
- ❑ Order of pattern discovery
 - ❑ Path \rightarrow tree \rightarrow graph (e.g., GASTON (Nijssen&Kok'04))