

The background of the slide is a complex, abstract composition. It features a network of thin, light-colored lines forming a web-like structure. Overlaid on this are various geometric shapes, including triangles and polygons, some of which are filled with a light blue or orange color. There are also numerous small, dark blue and green dots scattered across the background, some of which are connected by lines, suggesting a graph structure. The overall color palette is muted, with a lot of light blue, grey, and white, accented with small amounts of orange and green.

# 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  $D_g = \{G_i \mid g \subseteq G_i, G_i \in D\}$

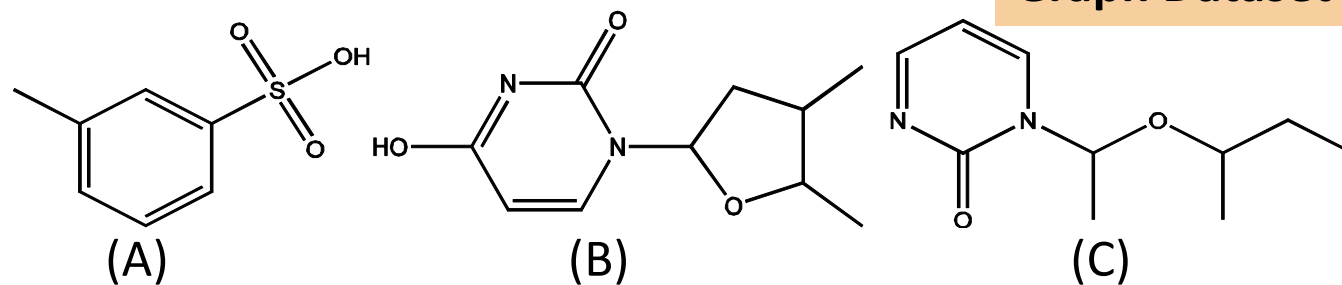
- $\text{support}(g) = |D_g| / |D|$

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

- Ex.: Chemical structures

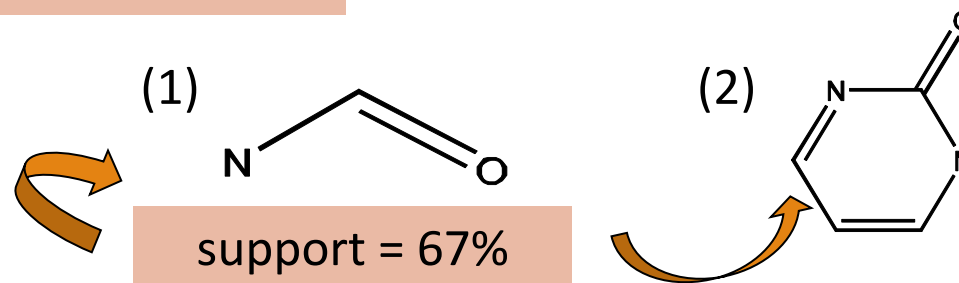
- Alternative:

- 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, 2002])
- ❑ Support calculation
  - ❑ Store embeddings (e.g., GASTON [Nijssen & Kok, 2004], FFSM [Huan, Wang, & Prins, 2003], MoFa [Borgelt & Berthold, ICDM'02])
- ❑ Order of pattern discovery
  - ❑ Path  $\rightarrow$  tree  $\rightarrow$  graph (e.g., GASTON [Nijssen & Kok, 2004])