CloseGraph: Mining Closed Graph Patterns

Why Mine Closed Graph Patterns?

- ❑ Challenge: An **n**-edge frequent graph may have 2^n subgraphs

- ❑ Motivation: Explore *closed frequent subgraphs* to handle graph pattern explosion problem

- ❑ A frequent graph G is *closed* if there exists no supergraph of G that carries the same support as G
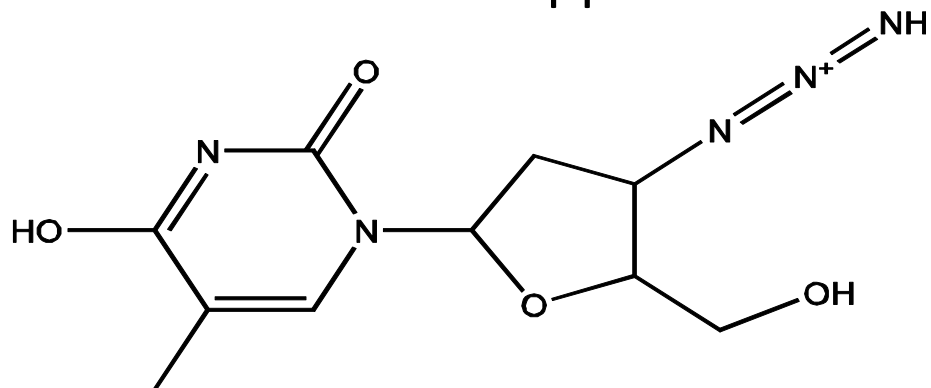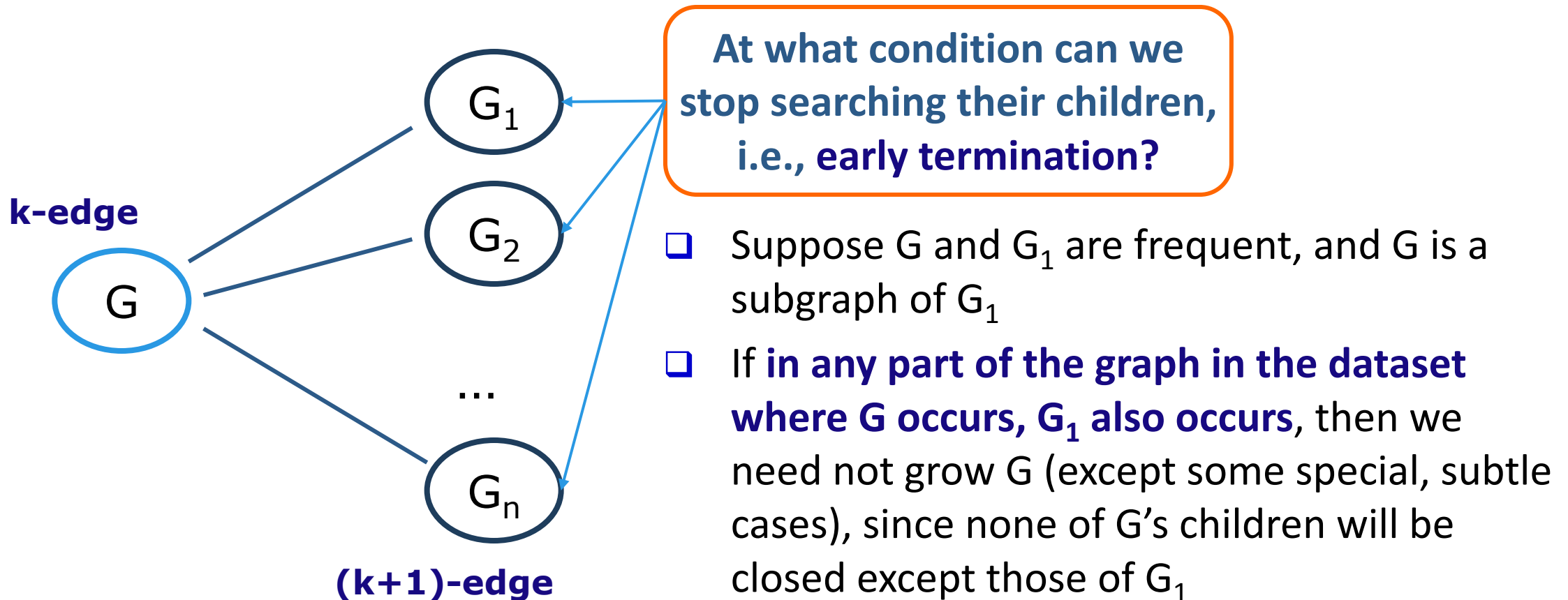

If this subgraph is *closed* in the graph dataset, it implies that none of its frequent super-graphs carries the same support

- ❑ *Lossless compression:* Does not contain non-closed graphs, but still ensures that the mining result is complete

- ❑ Algorithm CloseGraph: Mines closed graph patterns directly

CloseGraph: Directly Mining Closed Graph Patterns

❑ CloseGraph: Mining closed graph patterns by extending gSpan (Yan & Han, KDD'03)

k-edge

G

G_1

G_2

...

G_n

(k+1)-edge

> **At what condition can we stop searching their children, i.e., early termination?**

❑ Suppose G and G_1 are frequent, and G is a subgraph of G_1

❑ If **in any part of the graph in the dataset where G occurs, G_1 also occurs**, then we need not grow G (except some special, subtle cases), since none of G's children will be closed except those of G_1

Experiment and Performance Comparison

❑ The AIDS antiviral screen compound dataset from NCI/NIH

❑ The dataset contains 43,905 chemical compounds

❑ Discovered patterns: The smaller minimum support, the bigger and more interesting subgraph patterns discovered

20%   **10%**   **5%**

# of Patterns: Frequent vs. Closed

Runtime: Frequent vs. Closed