

The background of the slide is a collage of various data visualizations. It includes network graphs with nodes and edges in red, green, and blue. There are also scatter plots with colored dots, a heatmap with a grid of small squares, and a large, faint, stylized letter 'A' in the center. The overall color palette is muted, with a lot of greys, browns, and soft blues.

gSpan: A Pattern Growth Approach

Pattern-Growth Approach

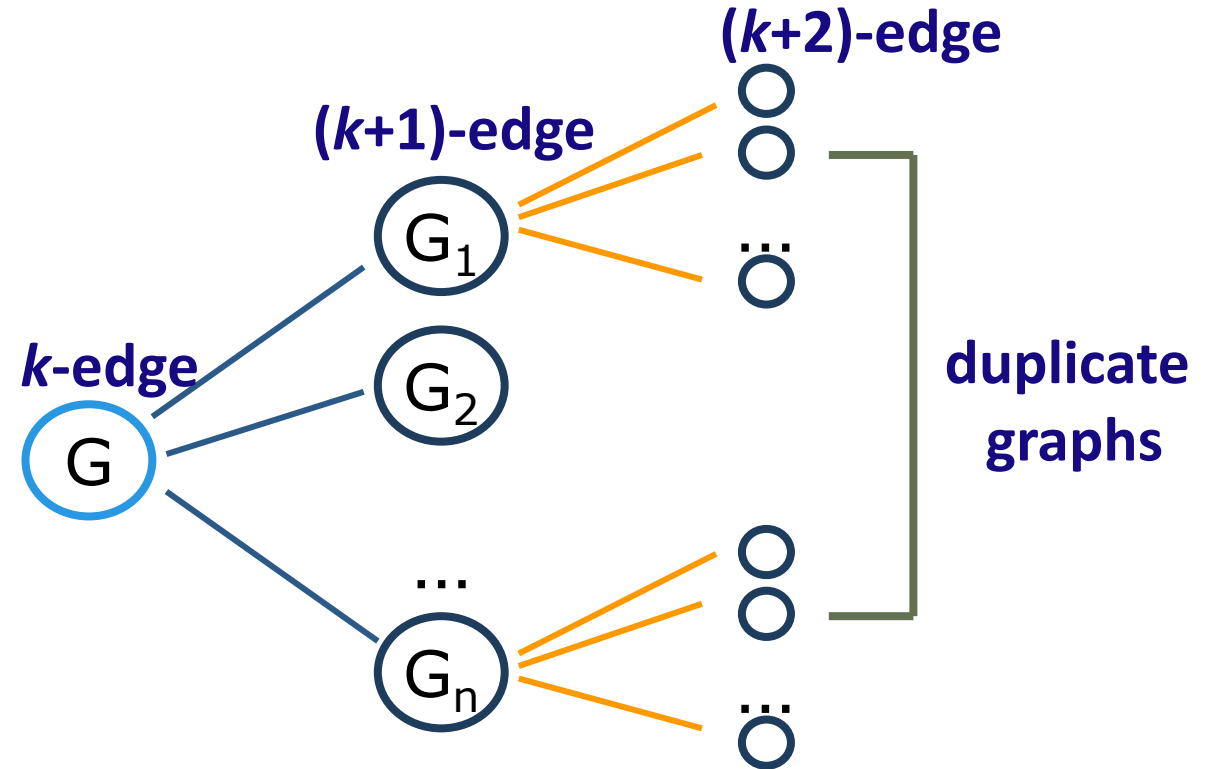
- Depth-first growth of subgraphs from k -edge to $(k+1)$ -edge, then $(k+2)$ -edge subgraphs

- Major challenge

- Generating many duplicate subgraphs

- Major idea to solve the problem

- Define an order to generate subgraphs
 - DFS spanning tree: Flatten a graph into a sequence using depth-first search
 - gSpan (Yan & Han, ICDM'02)



gSPAN: Graph Pattern Growth in Order

- ❑ **Right-most path extension** in subgraph pattern growth
 - ❑ Right-most path: The path from root to the right-most leaf (choose the vertex with the smallest index at each step)
 - ❑ Reduce generation of duplicate subgraphs
- ❑ **Completeness:** The enumeration of graphs using right-most path extension is complete
- ❑ DFS code: Flatten a graph into a sequence using depth-first search

