

# Incremental Frequent Subgraph Mining on Large Evolving Graphs

Ehab Abdelhamid

Imperial College London

Mustafa Canim

IBM

Mohammad Sadoghi

University of California, Davis

Bishwaranjan Bhattacharjee

IBM

Yuan-Chi Chang

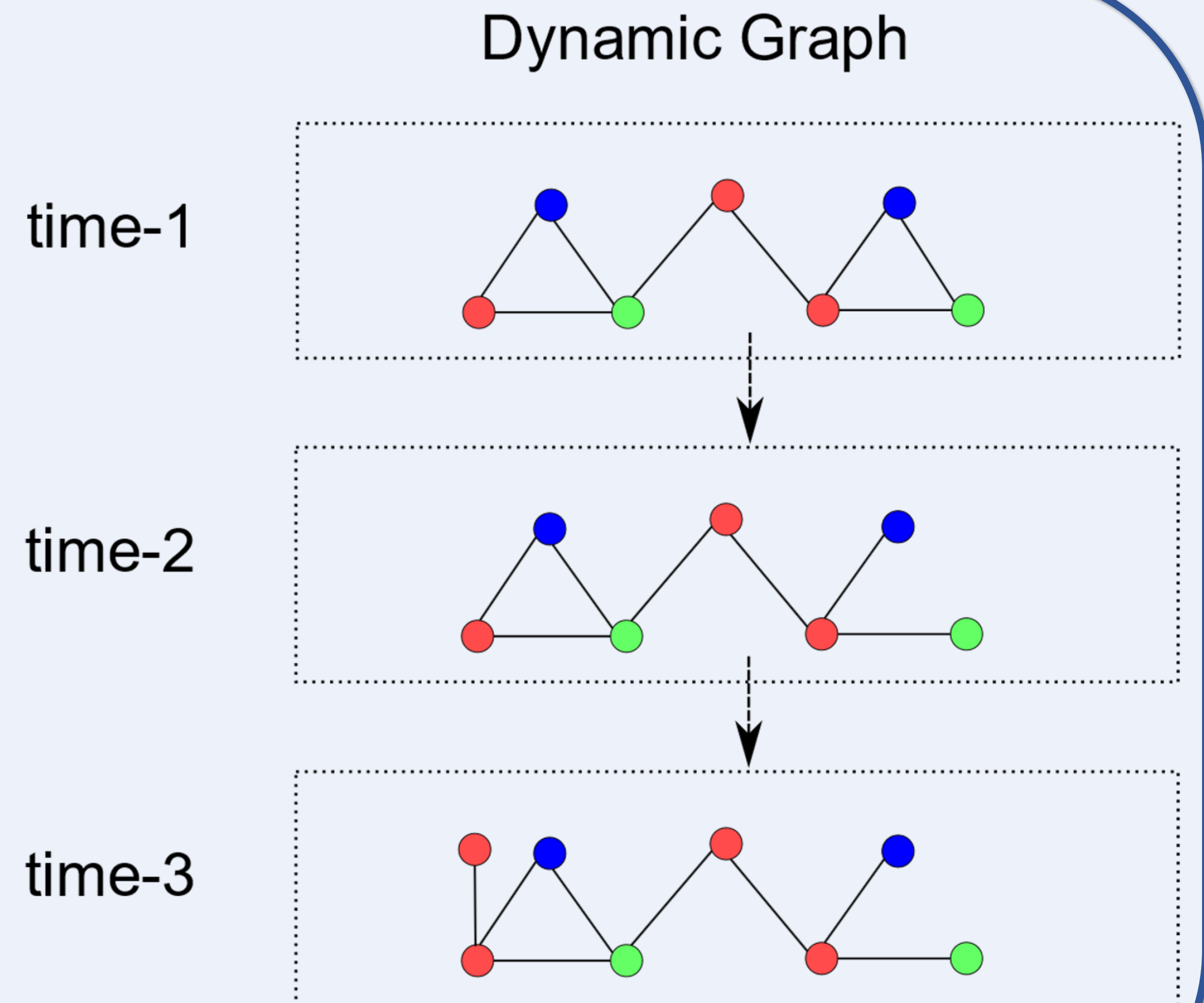
IBM

Panos Kalnis

King Abdullah University of Science and Technology

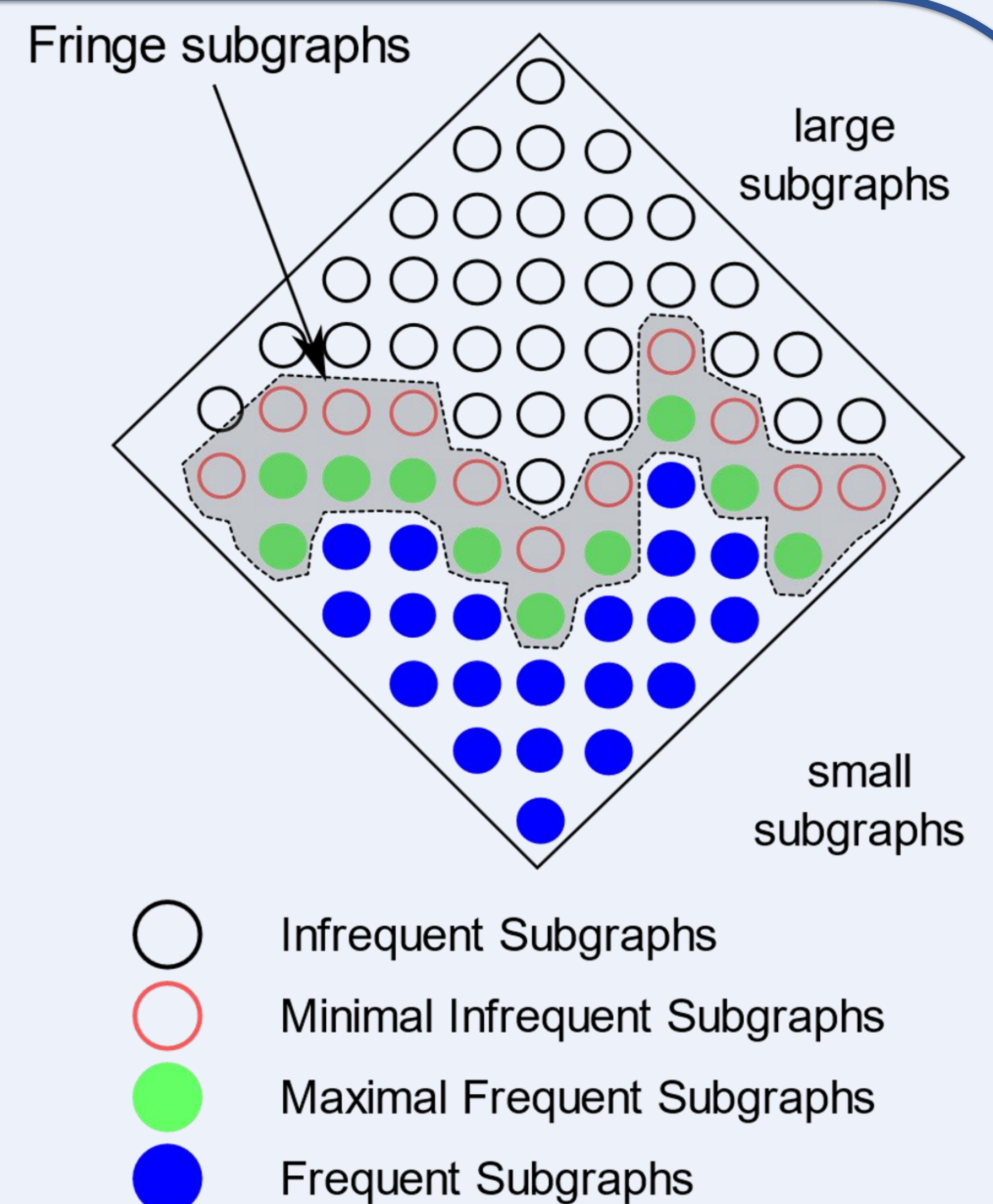
## Motivation

- Graphs are used to represent information of several applications in multidisciplinary domains.
- Frequent subgraph mining (FSM) is the task of finding all subgraphs that appear frequently inside a graph.
- Most research efforts on FSM focused on static graphs (i.e., graphs that never change over time).
- FSM on dynamic graphs is essential for applications such as security and social network analysis.
- Using static graph mining techniques for dynamic graphs is impractical due to their costly overhead.
- We propose **IncGM+** as a solution for dynamic FSM.



## Our Approach

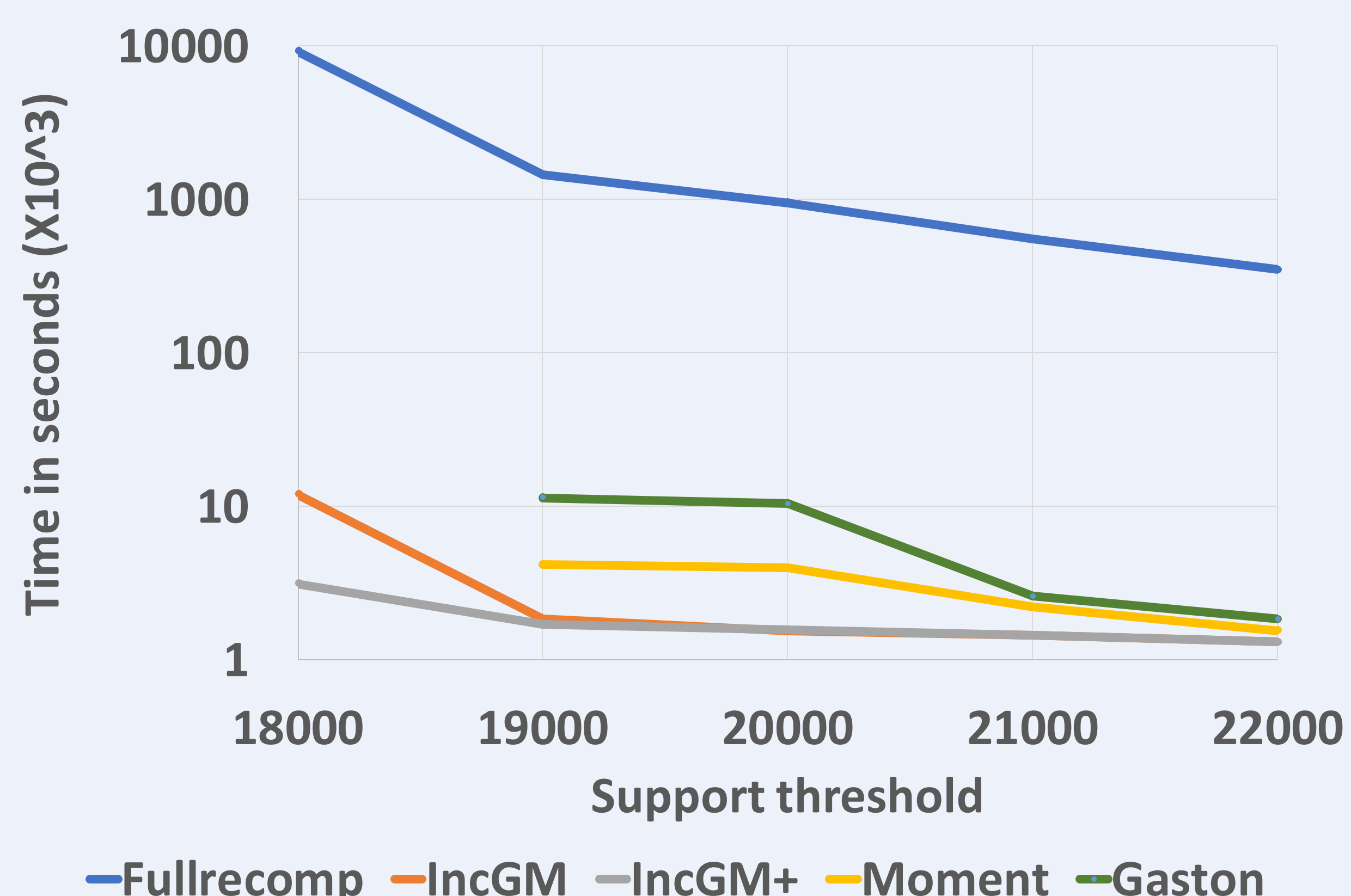
- Maintain a set of carefully selected subgraphs (the fringe subgraphs) and continuously update them as needed.
- Fringe subgraphs are representative subgraphs of the search space and are the most sensitive to dynamic graph updates.
- When a change happens to the input graph, only the fringe subgraphs need to be evaluated. Other subgraphs are evaluated only when needed.
- Refrain from repeatedly searching for occurrences of fringe subgraphs by storing a small number of their occurrences. Consequently, consecutive evaluations are optimized.
- Support batch updates to speed-up overall processing.



## Evaluation

2.9M nodes and 13.9M edges

Patents



11.3M nodes and 85.3M edges

Twitter

