



FSM-Viewer: A User-Friendly GUI for Frequent Subgraph Pattern



Jalal Khalil², Da Yan¹, Lyuheng Yuan¹, Saugat Adhikari¹, Cheng Long³, Yang Zhou⁴

¹Indiana University Bloomington, ²St. Cloud State University,

³Nanyang Technological University, ⁴Auburn University

The Problem

Given a big graph G , find all subgraphs that appear **frequently** in G

Frequency: Not anti-monotonic

Minimum Image (MNI): $MNI(S_2) = \min(1, 3) = 1$

Fraction-Score: More accurate than MNI

System	1	2	3	4	5	6	7	8	9	10
T-FSM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Scaleline	✓	✓	x	x	x	✓	x	x	x	x
DistGraph	x	x	x	✓	N/A	x	x	x	x	x
Fractal	✓	✓	x	✓	N/A	x	x	x	x	x
Arabesque	✓	x	x	✓	N/A	x	x	x	x	x
Pangolin	✓	x	x	✓	N/A	x	x	x	x	x
Peregrine	✓	✓	x	✓	x	✓	x	x	x	x

T-FSM Algorithm & System Framework

Domain $D(u)$: The set of candidate vertices in G that u can map to.

Valid Domain $D^*(u)$: The subset of $D(u)$ with valid matches.

Pattern Processing:

Initial 2-edge patterns (computed by the worker)

FSM-Explorer System Architecture

End Users

Parallel Computing

Frequent Subgraph Pattern Mining (FSM)

Instance Matching and Examining (by Subgraph Matching)

Example

UK_POI

Parameters

- Support: 4000
- Support Type: OMNI (Fraction-Score)
- Max Node #: 4
- Thread #: 4

Mine → Mining...

1: Food_&_Drink

2: Food_&_Drink

3: Community

0: ATM

Matching... → Mode: Expansion Graph

Node Id: 79303 Label Name: Food_&_Drink Address: Dominos Pizza, London Bow E3 5LX Coordinates: (-0.03456, 51.53215)

Related to: [6917] [48300] [55732] [65609] [83567]

Mode: Union Graph

Backend with T-thinker

Think-Like-A-Task (TLAV) Model: Divide and Conquer, Often Recursive

CCG Computing Community Consortium

(a) Query graph G_q

(b) Data graph G

(c) State Space Search Tree

FSM: Two-Level Task Scheduling

Subgraph Matching: Ullmann's Algorithm with Additional Pruning-Based Optimizations

References

(Closely related to this demo)

- Jalal Khalil et al. **FSM-Explorer: An Interactive Tool for Frequent Subgraph Pattern Mining from a Big Graph**. In *ICDE 2024 Demo*. **THIS DEMO!**
- Lyuheng Yuan et al. **T-FSM: A Task-Based System for Massively Parallel Frequent Subgraph Pattern Mining from a Big Graph**. In *SIGMOD 2023*.
- Lyuheng Yuan et al. **Faster Depth-First Subgraph Matching on GPUs**. In *ICDE 2024*. **THIS CONFERENCE!**
- Lyuheng Yuan et al. **G2-AIMD: A Memory-Efficient Subgraph-Centric Framework for Efficient Subgraph Search on GPUs**. In *ICDE 2024*. **THIS CONFERENCE!**
- Da Yan et al. **G-thinker: A General Distributed Framework for Finding Qualified Subgraphs in a Big Graph with Load Balancing**. *VLDB Journal* 2022.

(Other selected T-thinker publications)

- Da Yan et al. **Systems for Scalable Graph Analytics and Machine Learning: Trends and Methods**. In *IJCAI 2024 Tutorial*. **UPCOMING!**
- Jalal Khalil et al. **Parallel Mining of Large Maximal Quasi-Clique**. *VLDB Journal* 2022.
- Guimu Guo et al. **Scalable Mining of Maximal Quasi-Cliques: An Algorithm-System Codesign Approach**. In *PVLDB* 2022.
- Guimu Guo et al. **Maximal Directed Quasi-Clique Mining**. In *ICDE 2022*.

Acknowledgement & Contact

Supported by DOE ECRP Award 0000274975, NSF OIA-2229394 (RII Track-4), South Big Data Innovation Hub 2022 S.E.E.D.S. Award, and others from NSF.

Contact: yanda@iu.edu (Da Yan)