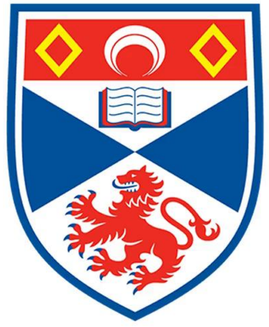
Fully Dynamic Coreset Spectral Clustering

Ben Jourdan, Peter Macgregor, Gregory Schwartzman

ICML 2026

THE UNIVERSITY
of EDINBURGH



University of
St Andrews

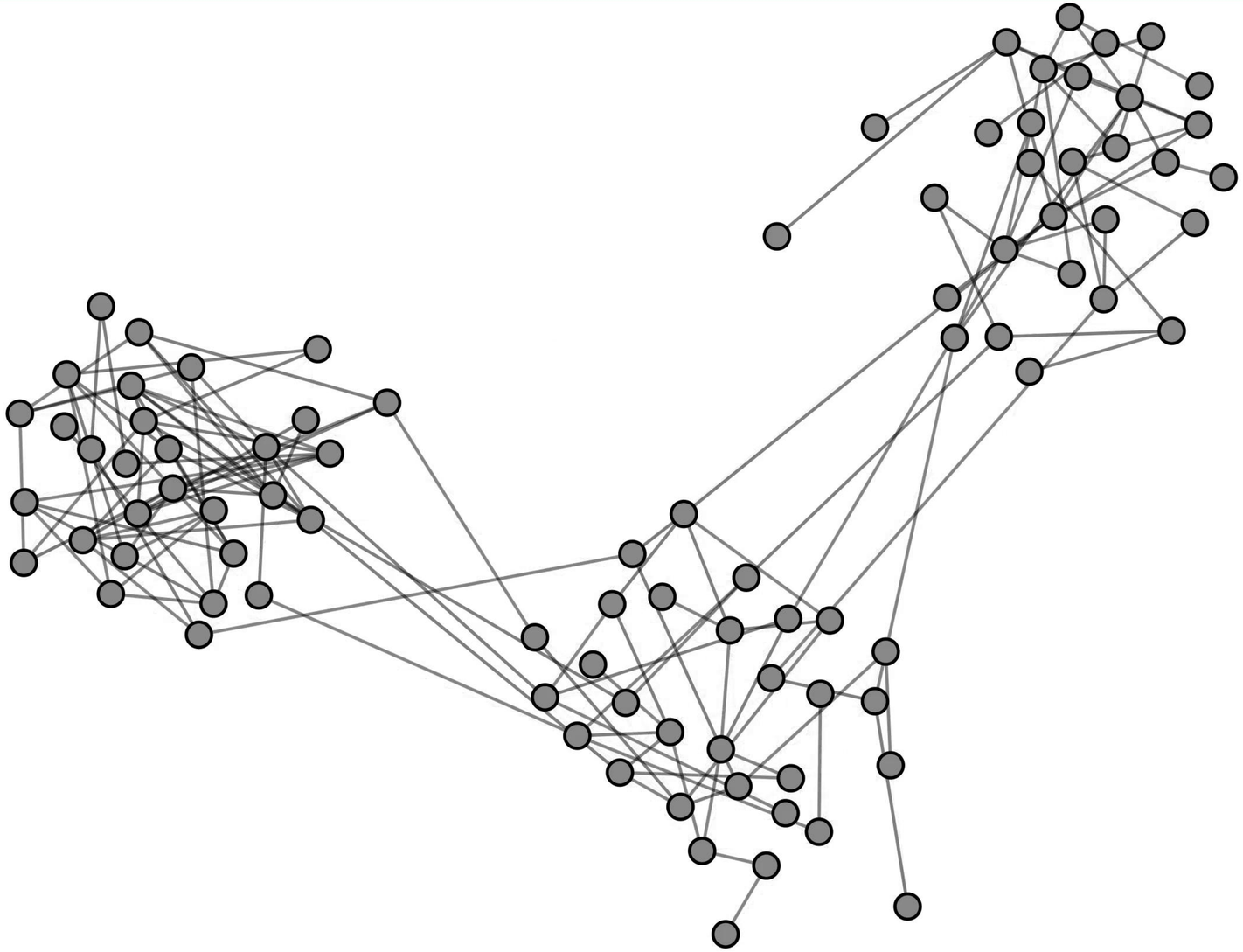
JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY
1990

Motivation

Clustering dynamic graphs

Interface:

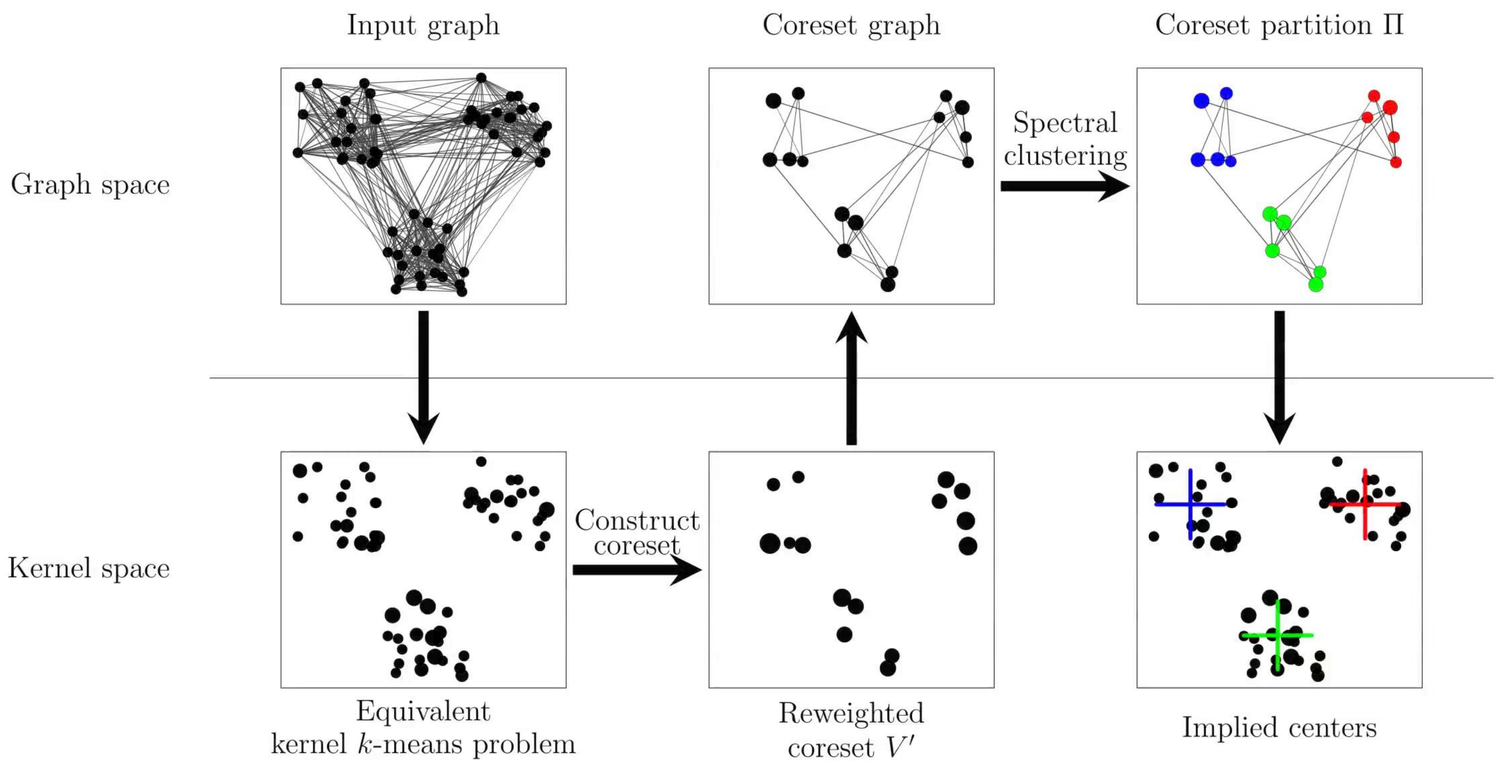
- Edge updates (insert/delete/modify)
- Cluster membership queries



edge updates

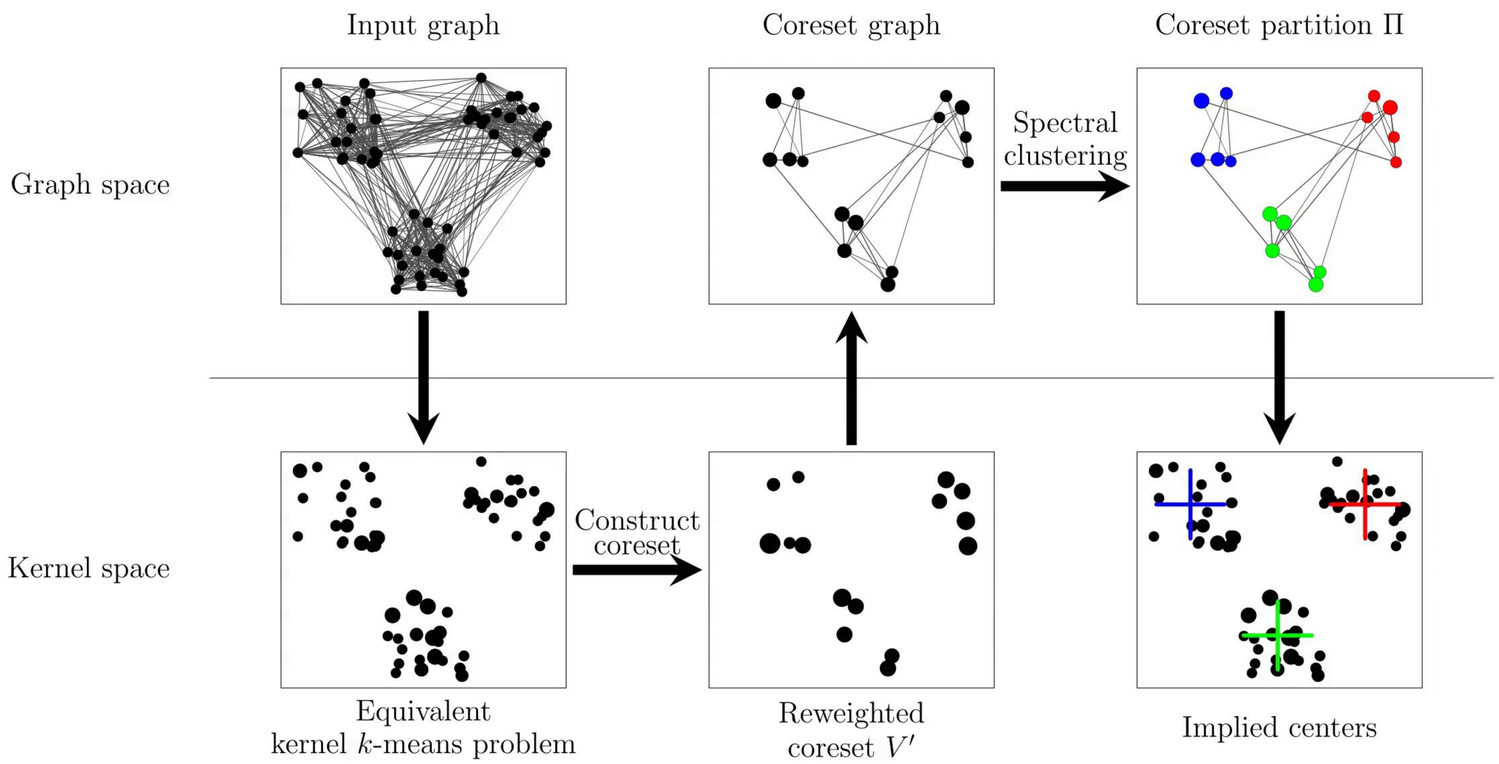
Motivation

Static Coreset Spectral Clustering (Jourdan et. al. 25)



Motivation

Static Coreset Spectral Clustering (Jourdan et. al. 25)

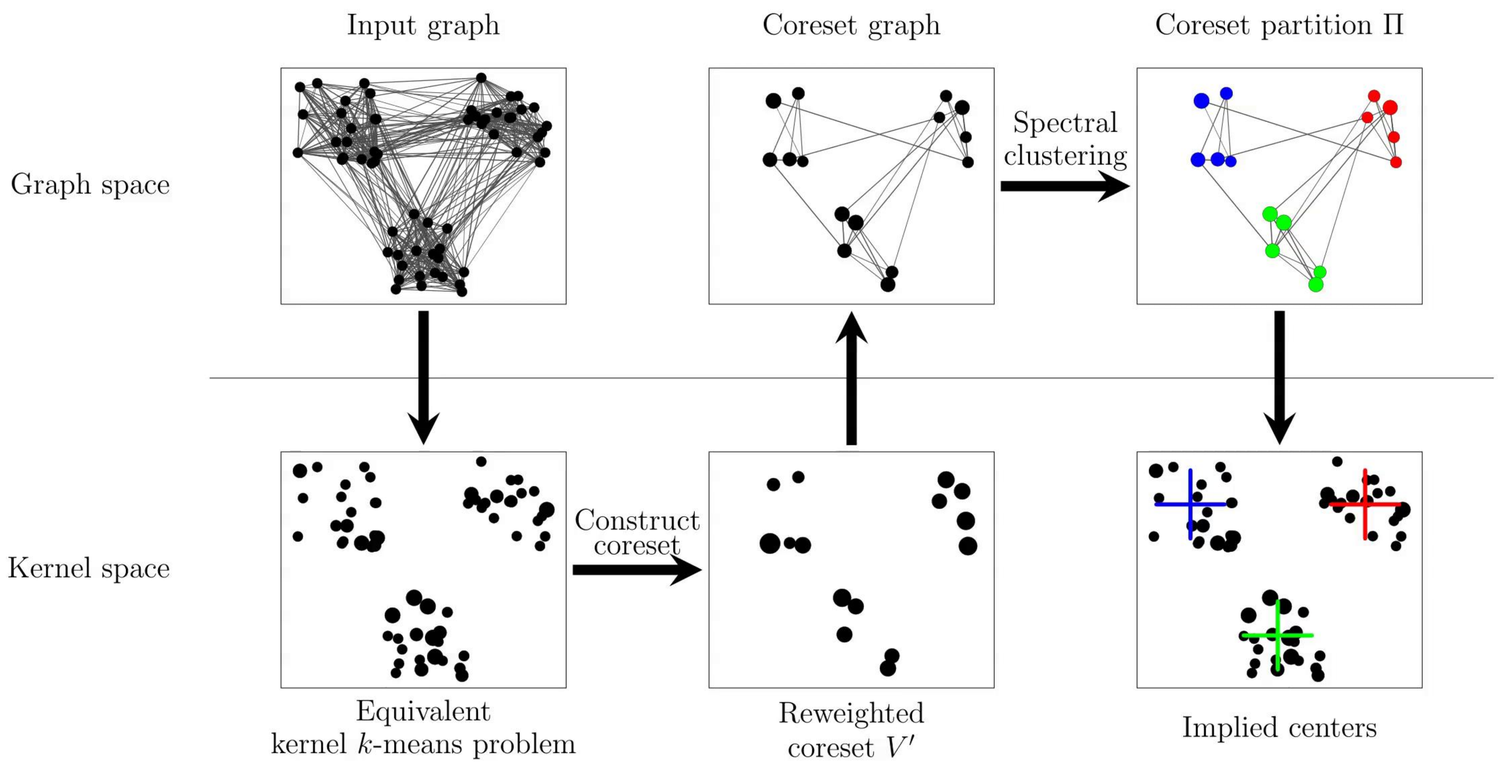


Merge and Reduce

Henzinger & Kale (20) + Jiang et. al. (24) + Jourdan et. al. (25)

Motivation

Static Coreset Spectral Clustering (Jourdan et. al. 25)



Merge and Reduce

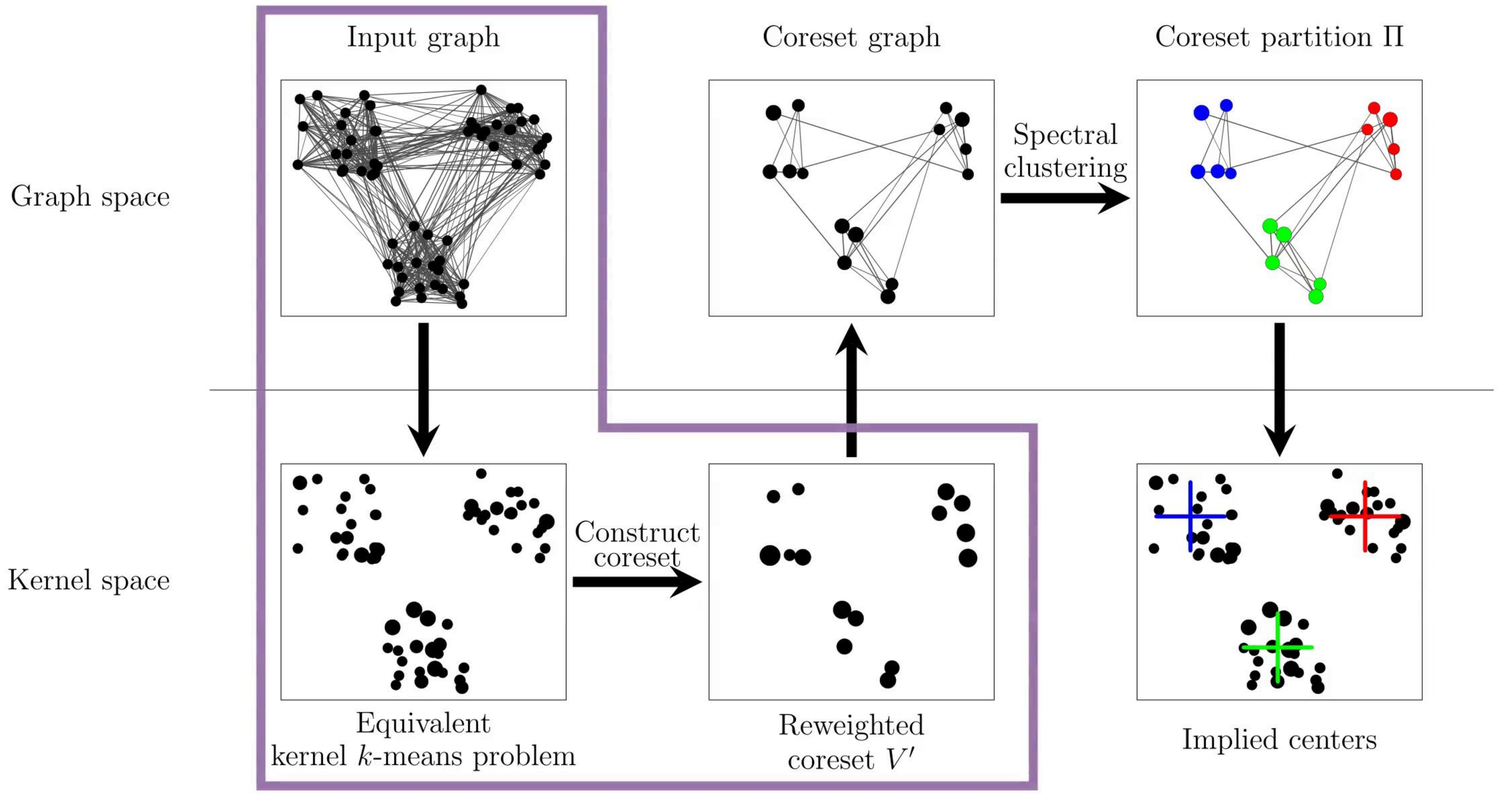
Henzinger & Kale (20) + Jiang et. al. (24) + Jourdan et. al. (25)

Challenge

Can we avoid the merge-and-reduce construction?

Motivation

Static Coreset Spectral Clustering (Jourdan et. al. 25)



Merge and Reduce

Henzinger & Kale (20) + Jiang et. al. (24) + Jourdan et. al. (25)

Challenge

Can we avoid the merge-and-reduce construction?

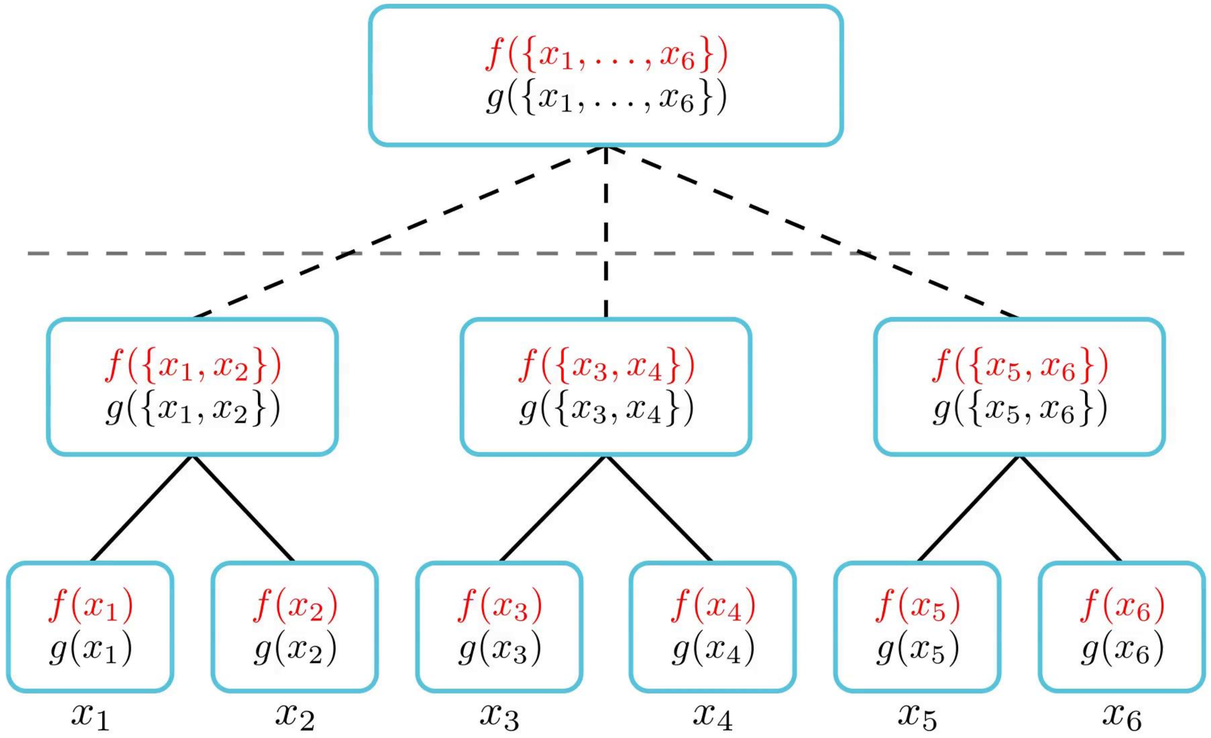
Goal

Efficiently maintain and sample from the coreset sampling distribution of Static CSC.

CSC Sampling Distribution

Static CSC sampling distribution

- Uses a sampling tree
- Exploits input graph sparsity



$$f(S) = \sum_{x \in S} f(x) \qquad g(S) = \sum_{x \in S} g(x)$$

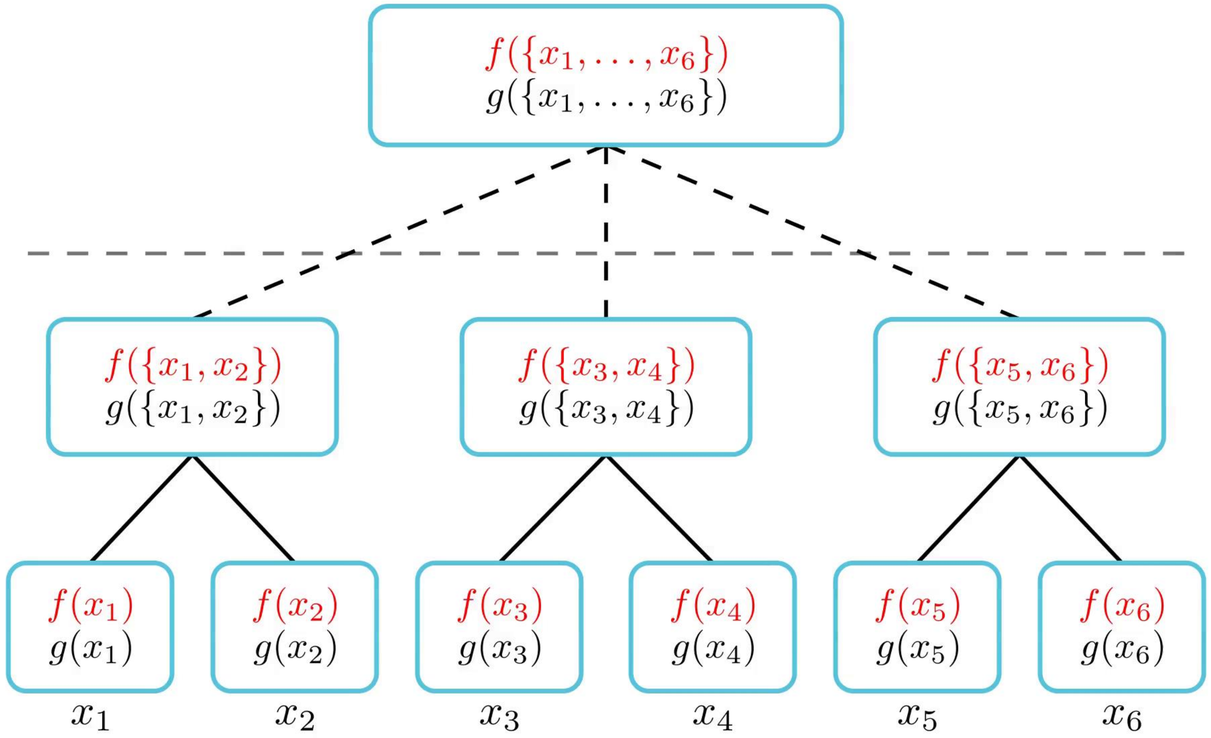
CSC Sampling Distribution

Static CSC sampling distribution

- Uses a sampling tree
- Exploits input graph sparsity

Dynamic issues

Forced sampling tree rebuilds on edge updates and during sampling.



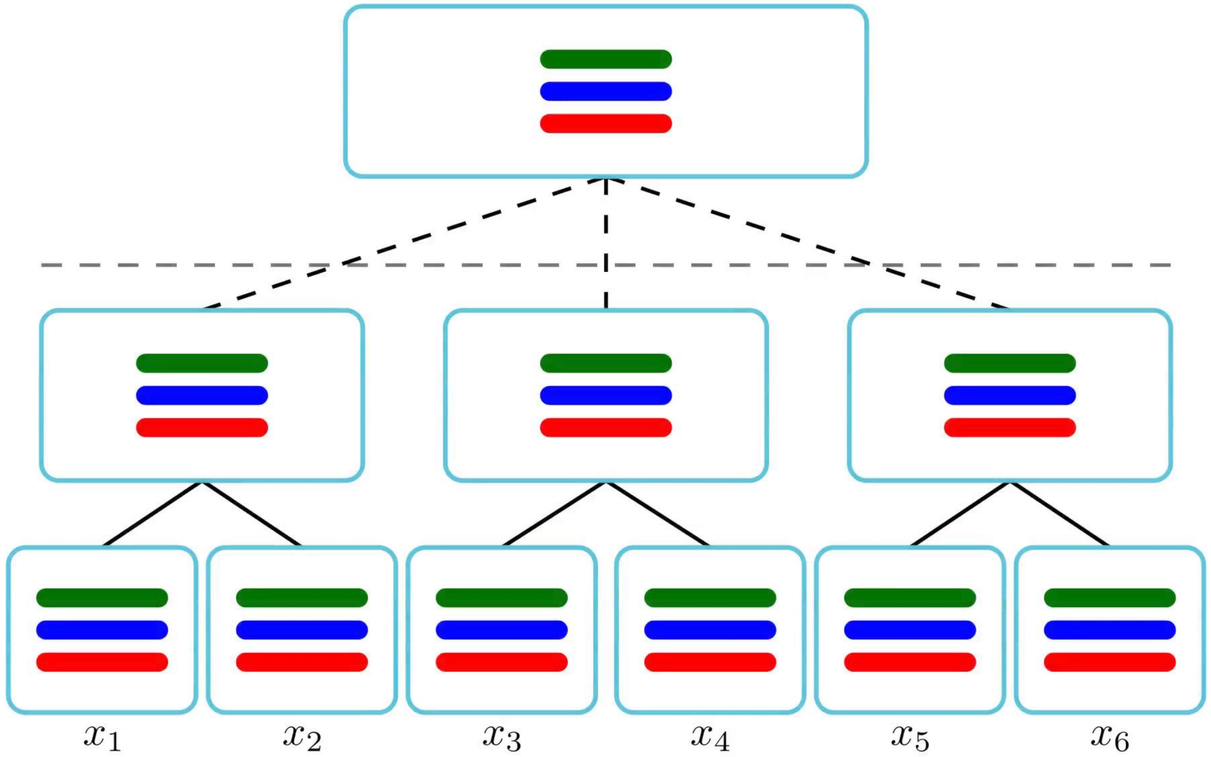
$$f(S) = \sum_{x \in S} f(x) \qquad g(S) = \sum_{x \in S} g(x)$$

Just-In-Time Sampling Trees

Linear decompositions of tree values

We decompose f and g into functions of "linear" terms

- Persistent terms
- Query-time terms
- Timestamps



Just-In-Time Sampling Trees

Linear decompositions of tree values

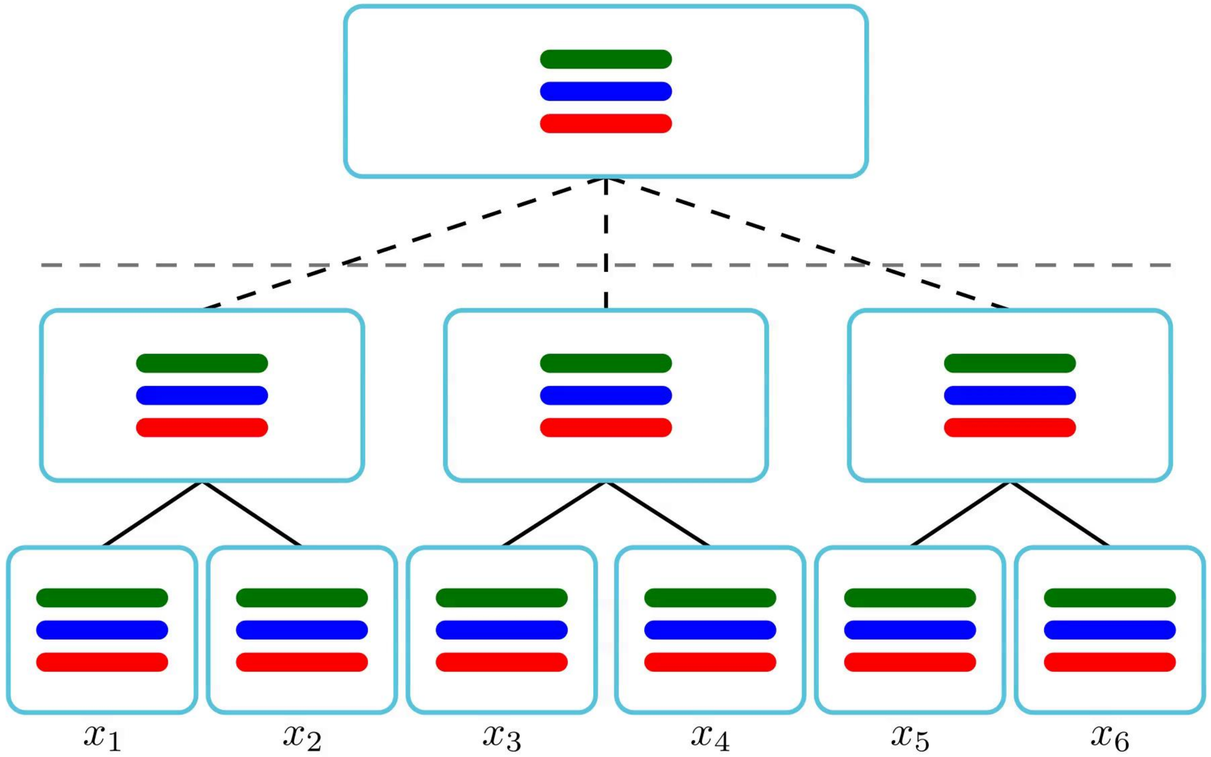
We decompose f and g into functions of "linear" terms

- Persistent terms
- Query-time terms
- Timestamps

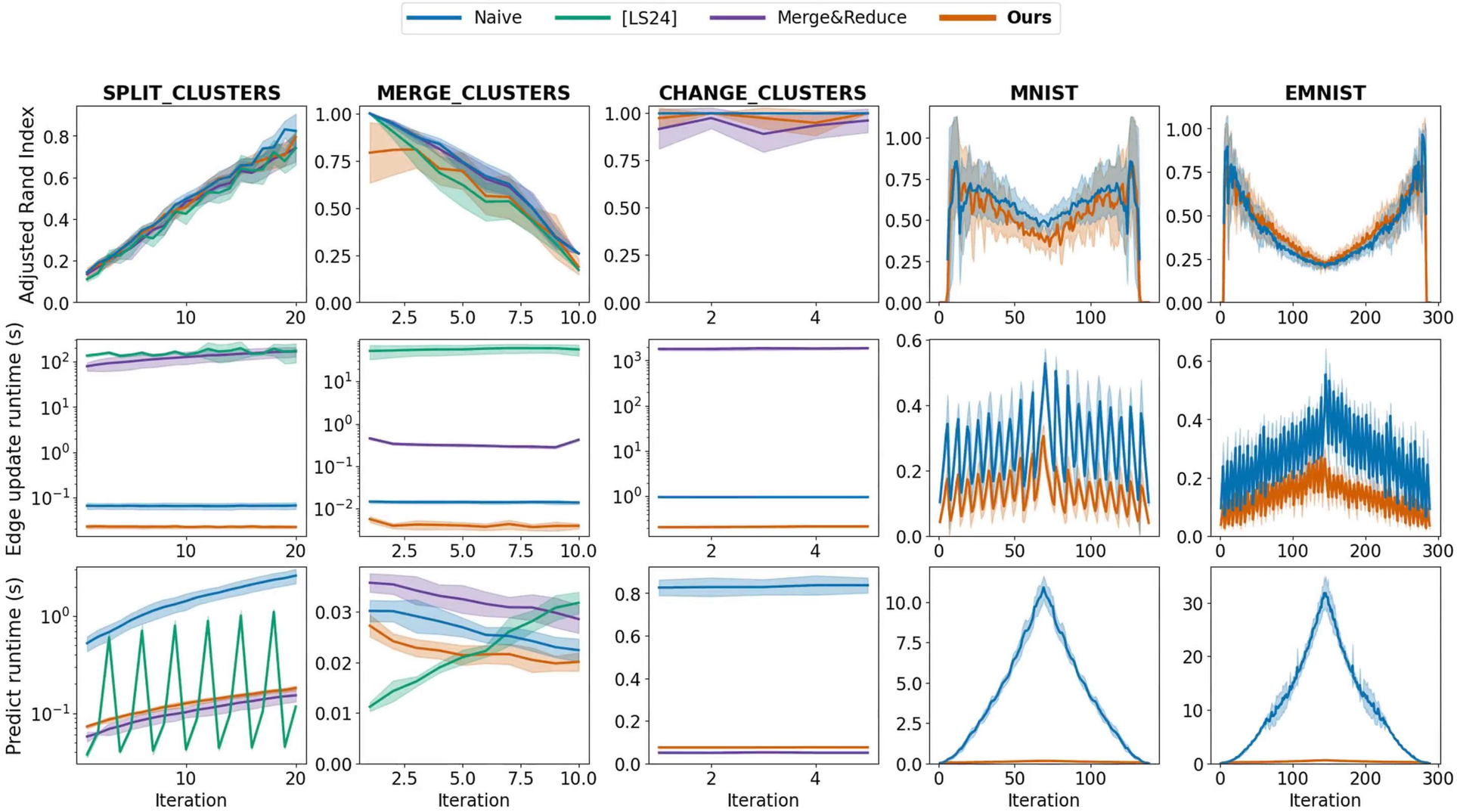
Result

Same distribution but efficient!
Compared to merge-and-reduce:

- Faster edge updates
- Similar query time



Experiments



Mixed workloads: update/query time and ARI across synthetic and real-world dynamic graph sequences.