Raphtory : Streaming Analysis Of Distributed Temporal Graphs

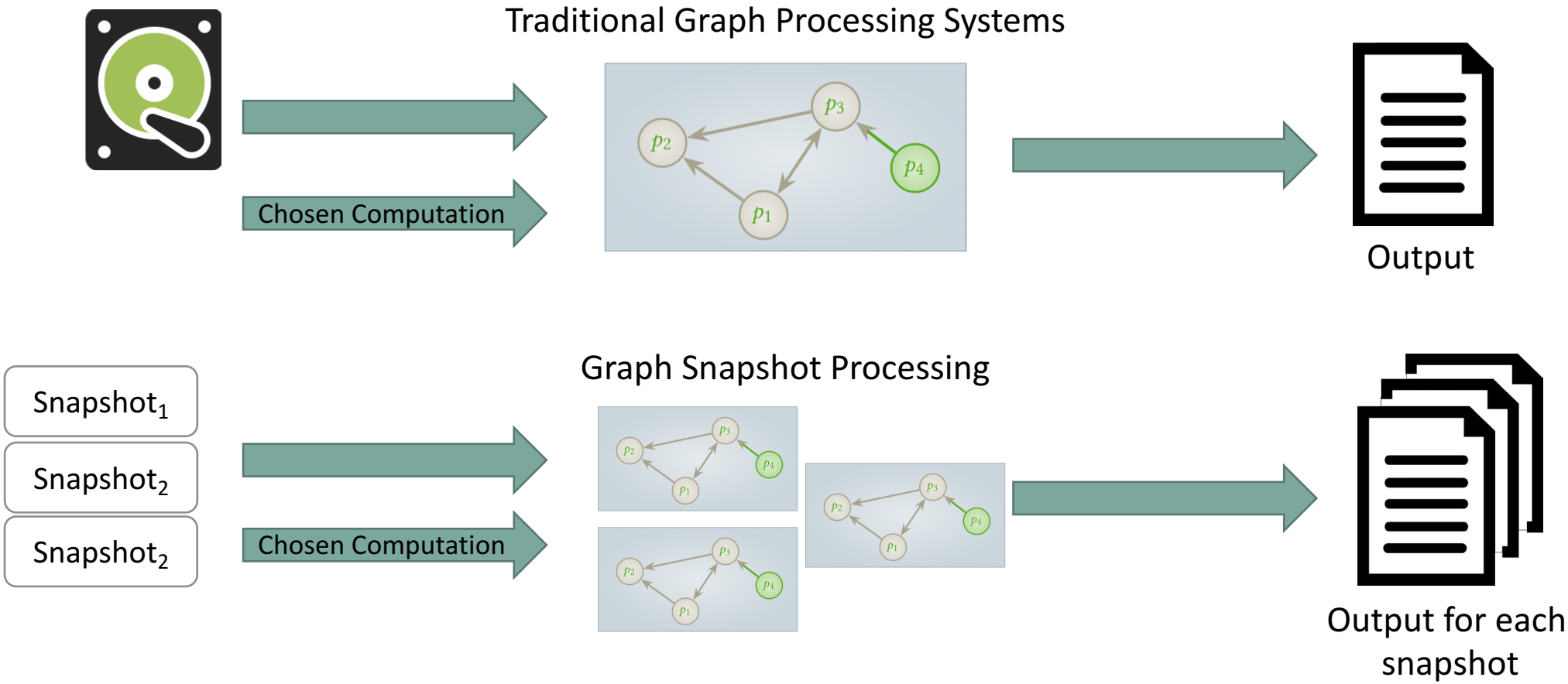
Benjamin Steer, Felix Cuadrado &
Richard G. Clegg

The Alan Turing Institute

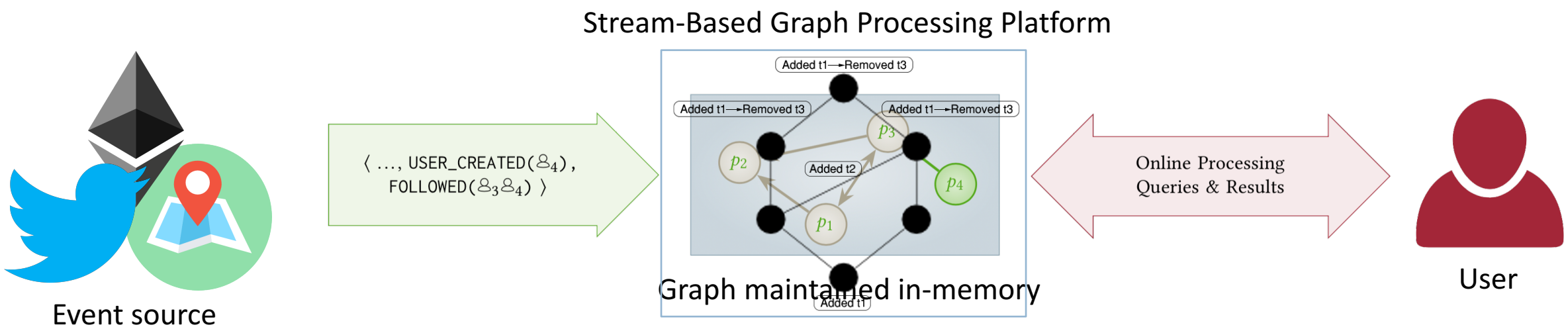
Queen Mary University of London

CHOROGRAPH

Motivation


Traditional Graph Processing Systems
Chosen Computation
Output
Snapshot1
Snapshot2
Snapshot2
Graph Snapshot Processing
Chosen Computation
Output for each snapshot

Motivation

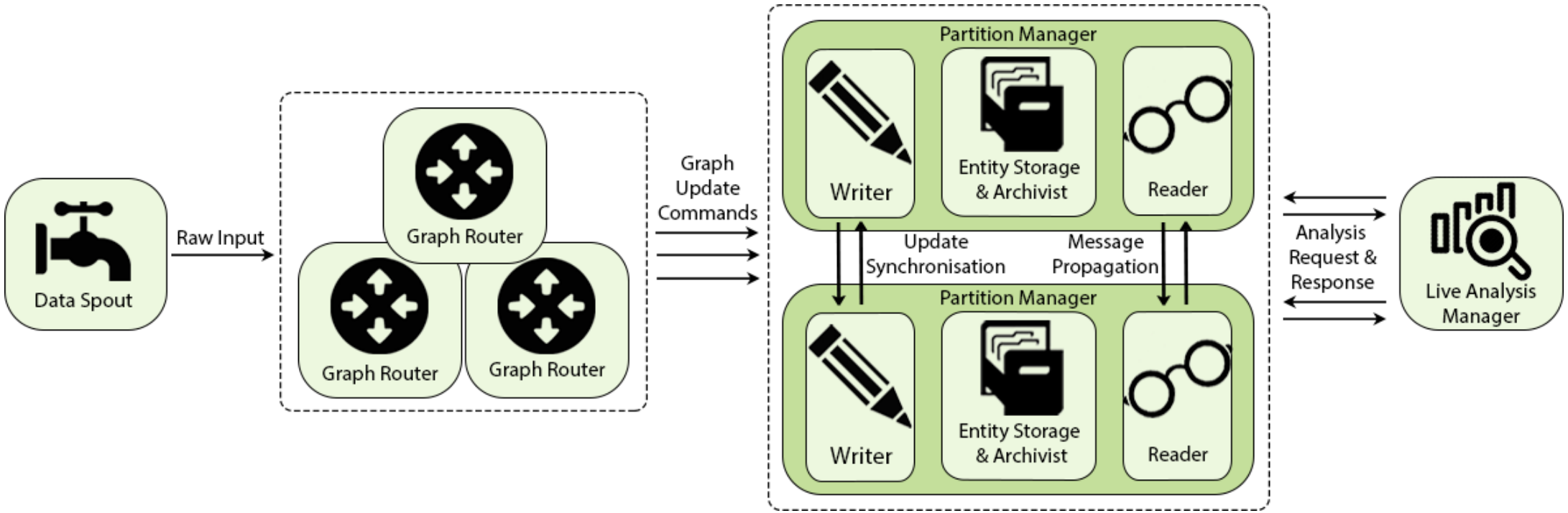


- Analysis on the most recent Graph
- Near real-time updates to metrics
- Compare new updates to previous state
- Temporal graph analysis

Raphtory features

- Temporal Graph Model
  - Formalisation of model and update semantics
- Distributed graph management
  - Stream Ingestion and near real-time maintenance
- Pregel-like temporal Graph Analysis
  - Live, view and temporal range analysis

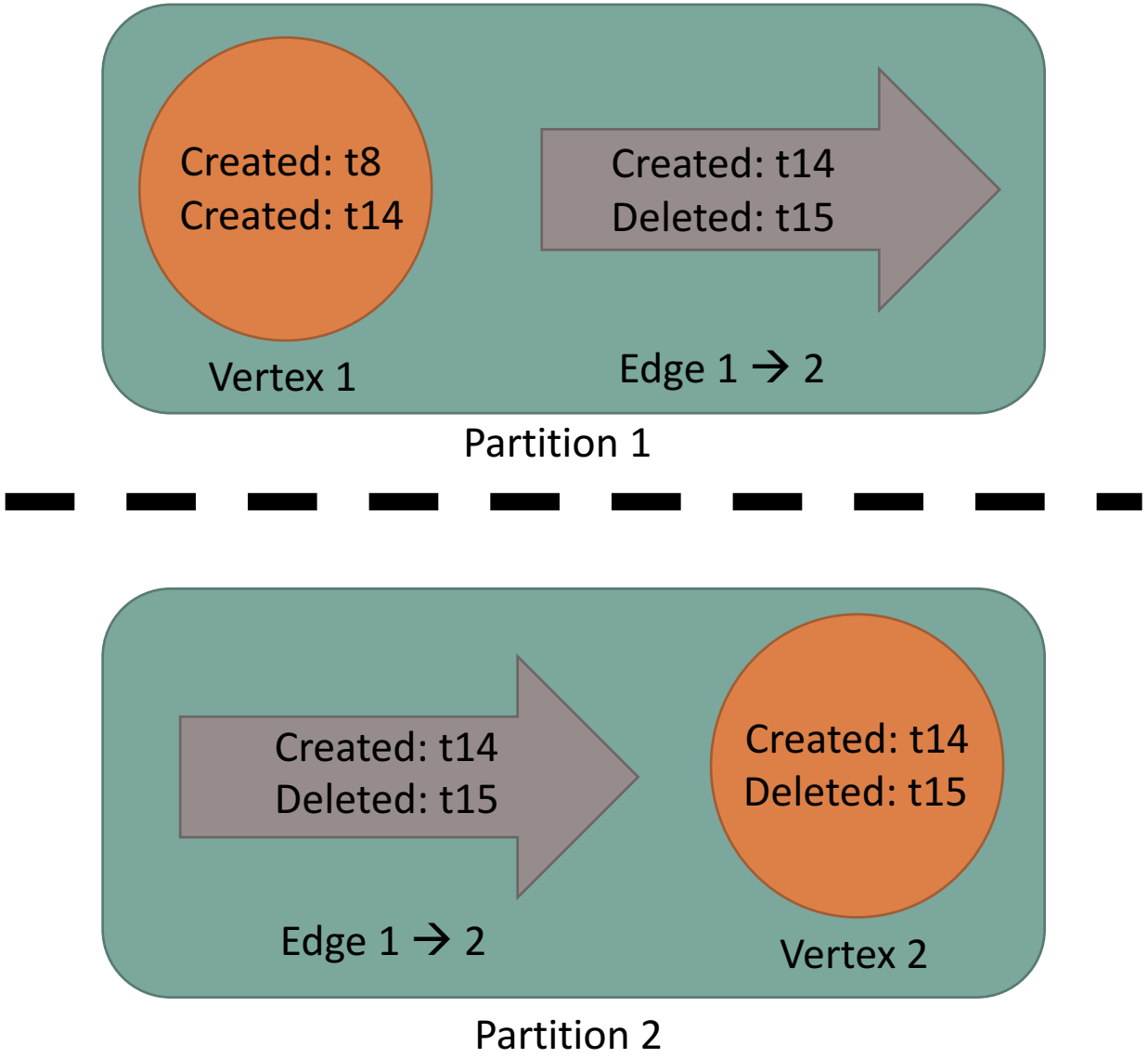
Raphtory Design



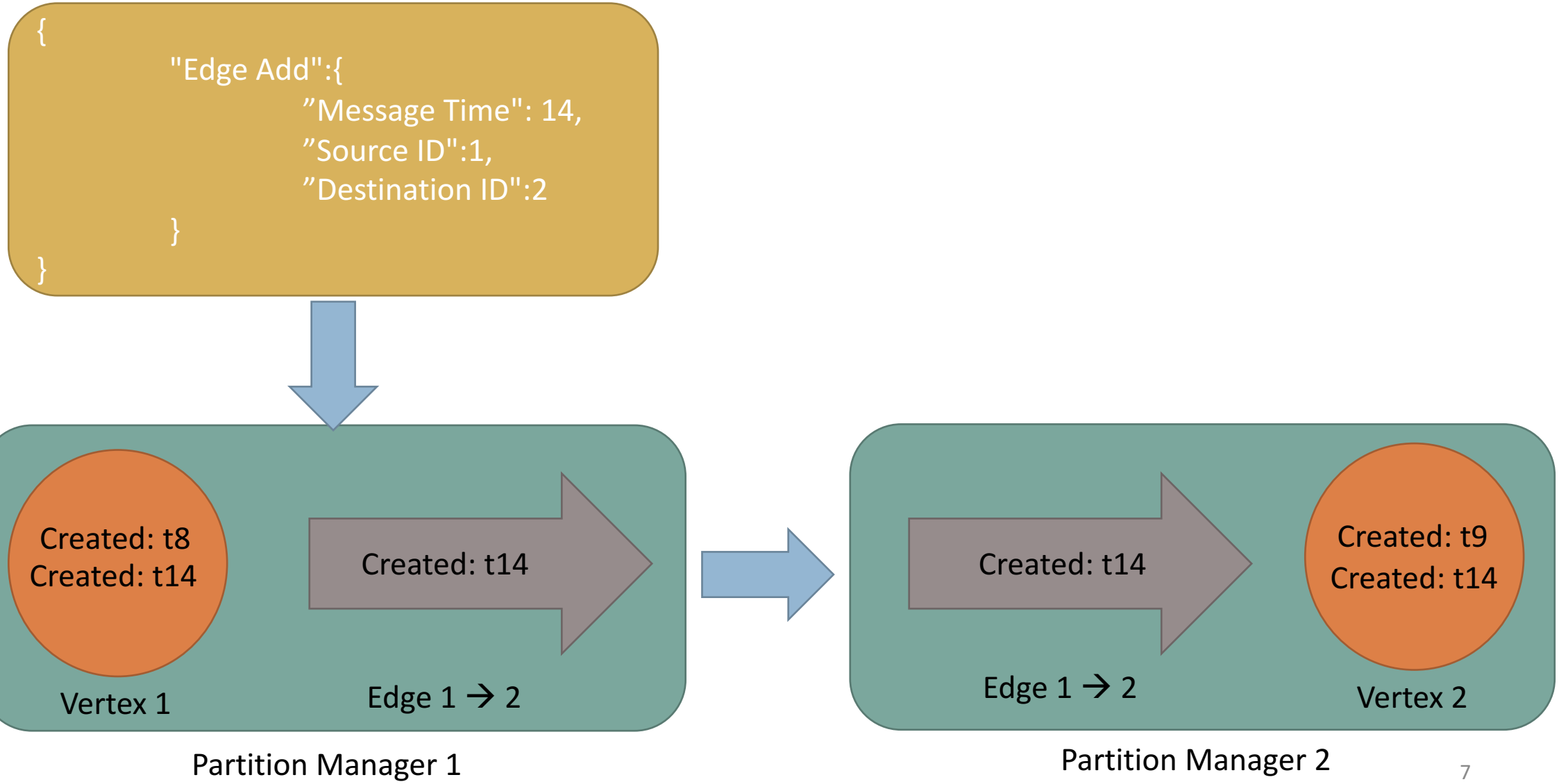
Implemented in Scala using the Akka actor model

[Raphtory: Streaming analysis of distributed temporal graphs, Future Generation Computer Systems 2020, Vol 102, pp 453-464]

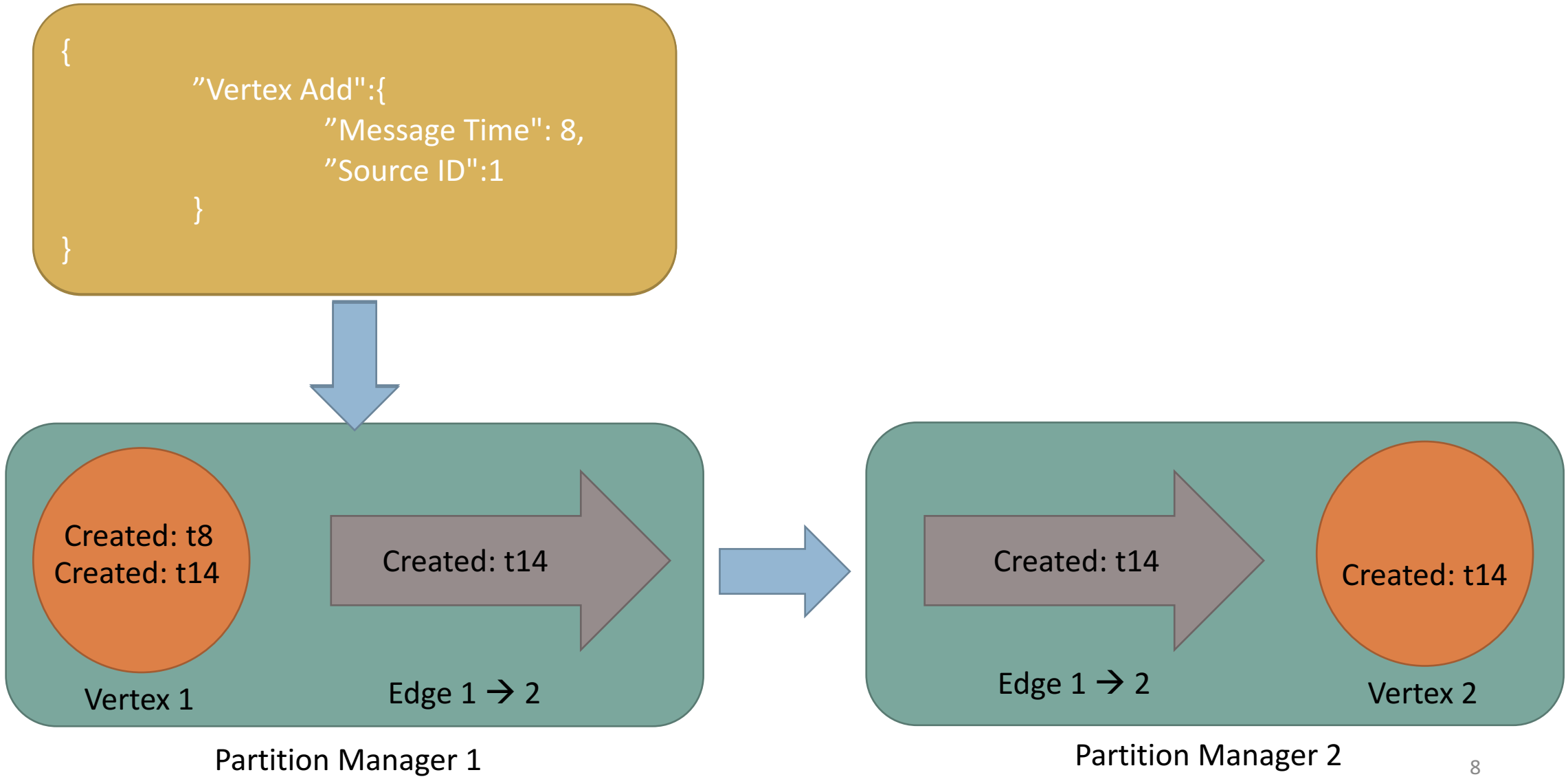
Partition Manger Ingestion



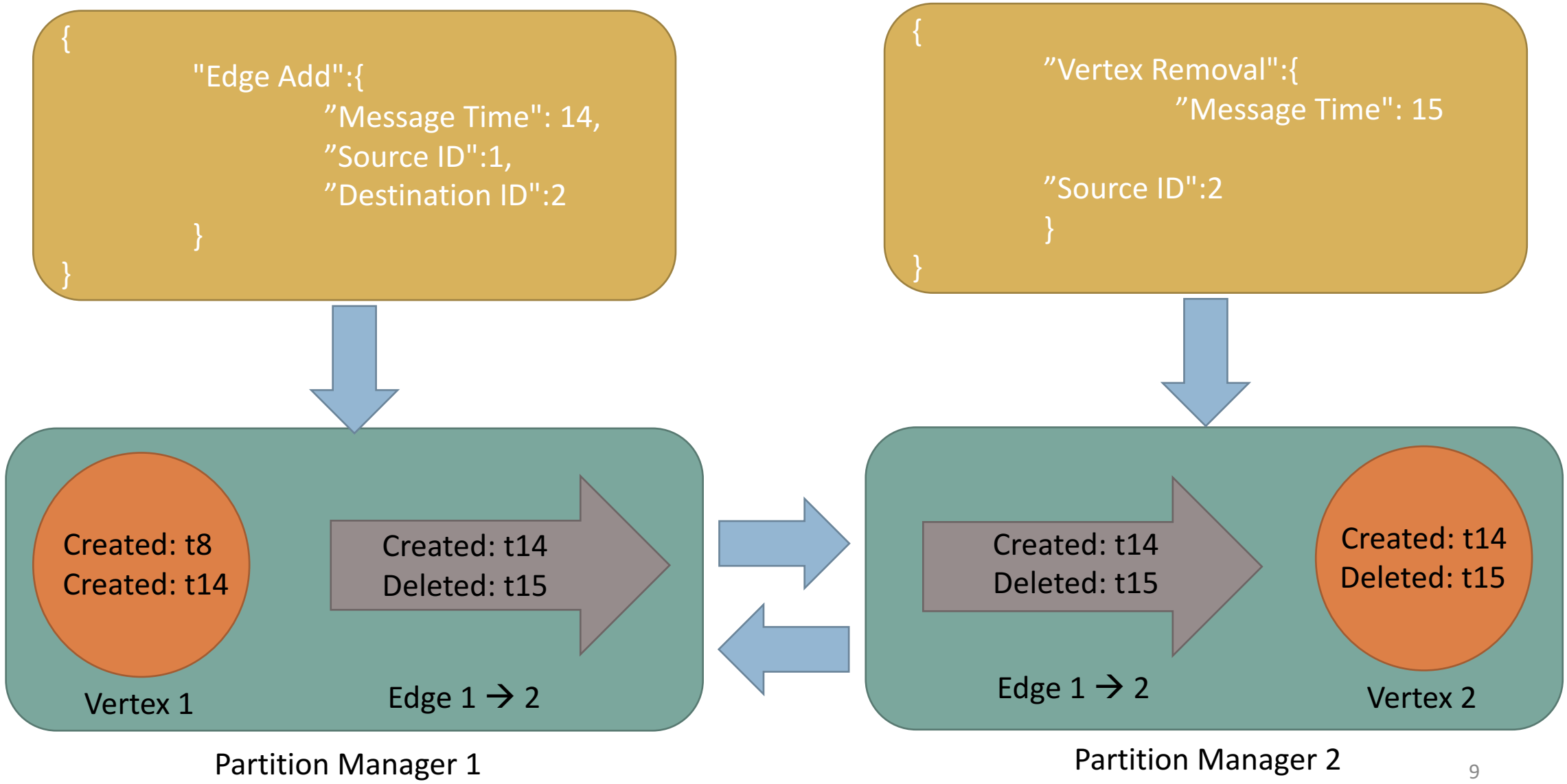
Correct update order



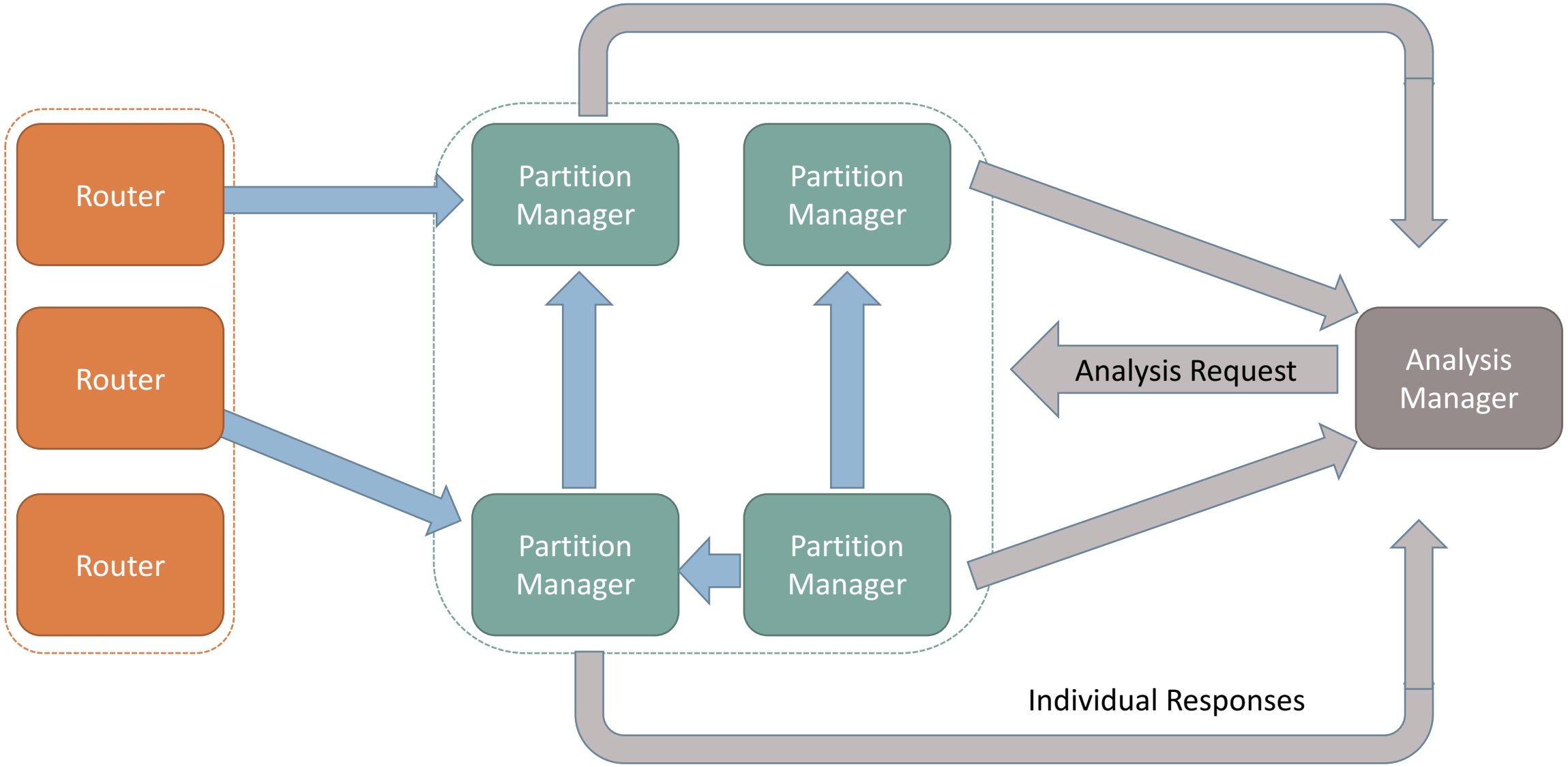
Edge Added Before Vertex



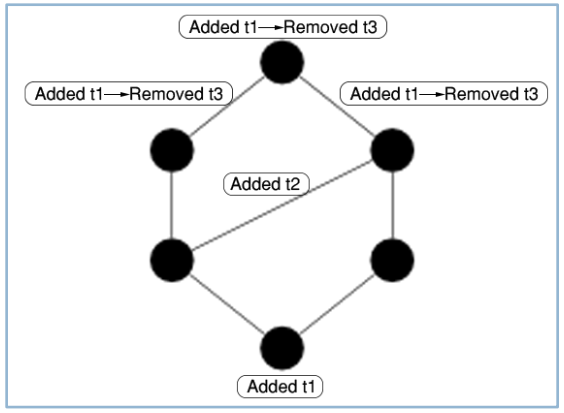
Vertex Deletion Before Edge Addition

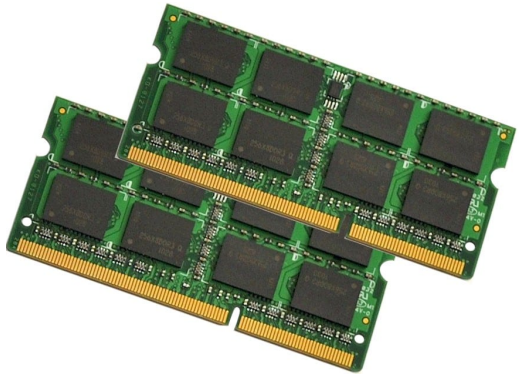


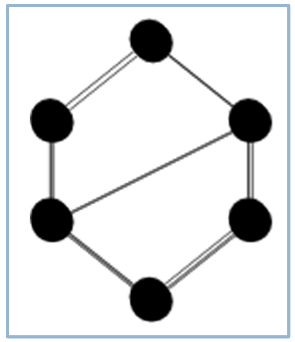
Analysis


Router
Router
Router
Partition Manager
Partition Manager
Partition Manager
Partition Manager
Analysis Request
Analysis Manager
Individual Responses

Live Graph, Views & Snapshots



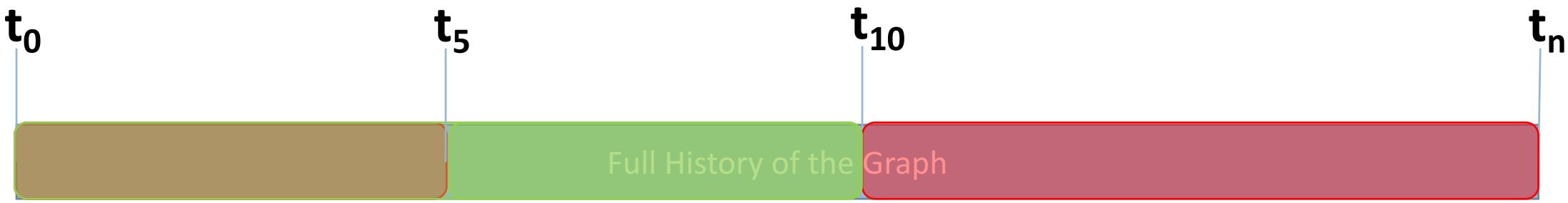




Views & Windowing

Window (Left Hand Filter)

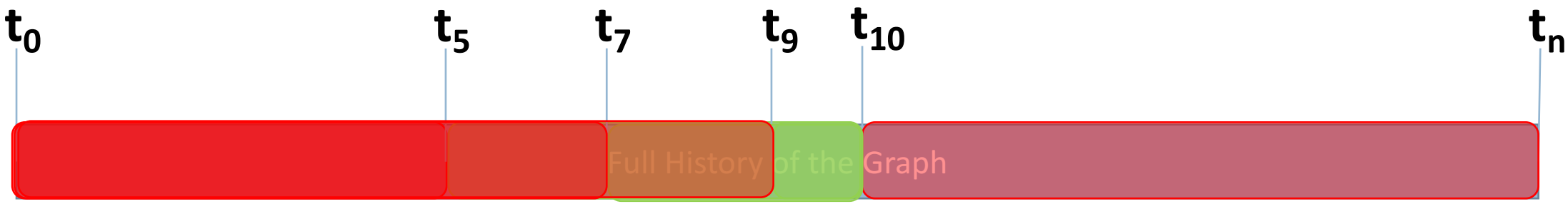
View (Right Hand Filter)



Window Size = 5

Windowing Batches

Batch of Windows
(Decreasing in size)

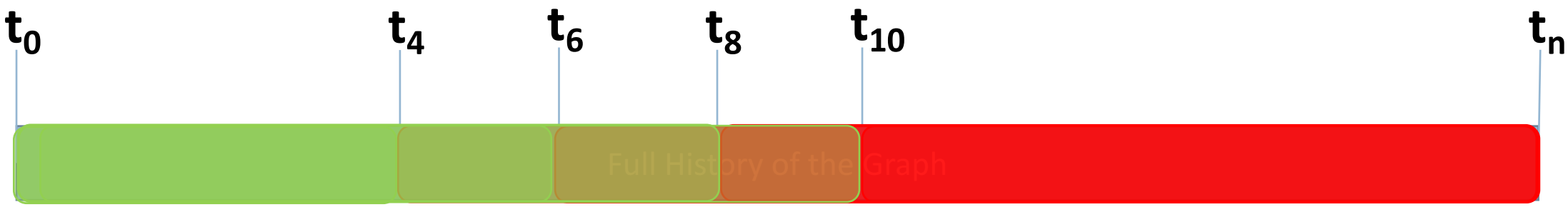


Window Sizes = [5,3,1]

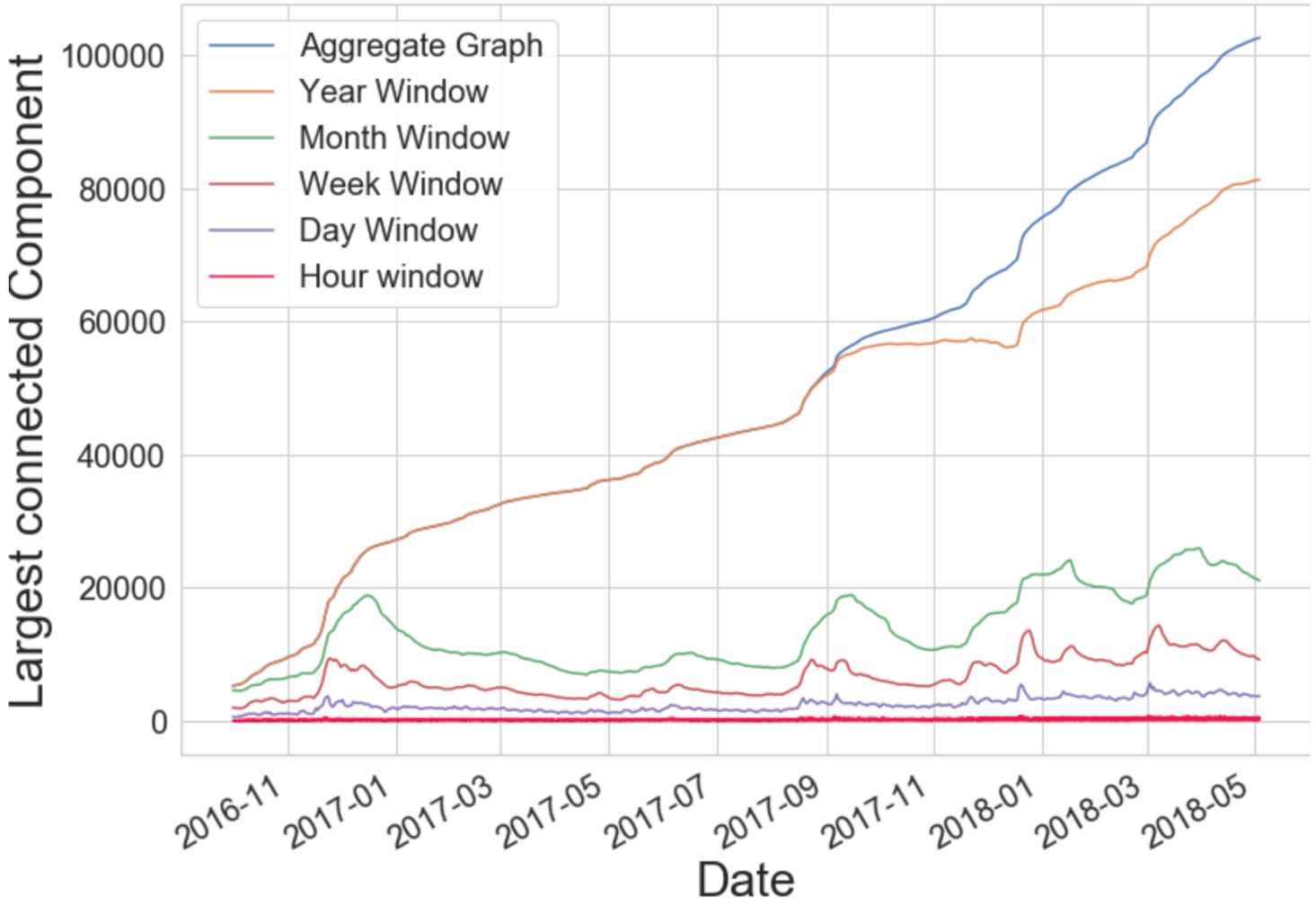
Temporal Range Analysis

Range of Interest = t_4 -> t_{10}

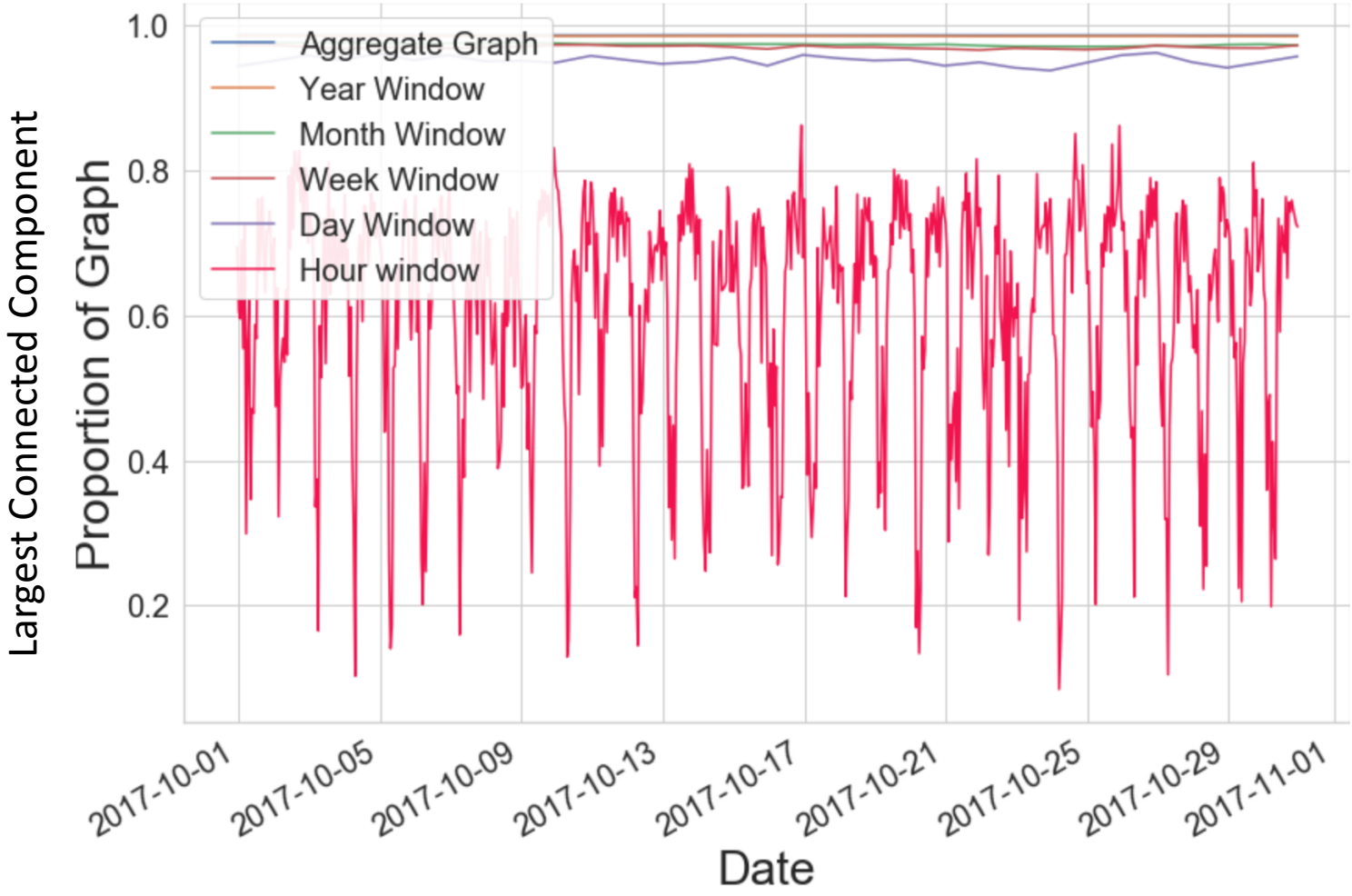
Interval = 2



Gab.ai Connected Components Every Hour Across Lifetime



Gab.ai Connected Components Every Hour Across Lifetime



Using Raphtory

- Available at github: https://github.com/miratepuffin/raphtory
- Includes starting documentation and tutorials
  - Readme goes through a single machine dockerised version that runs connected components over Gab graph.
- Multiple spouts (parsing data from Gab, Twitter, Bitcoin, Ethereum)
- Multiple analysis functions implemented (on views, ranges, window)
  - Connected Components
  - Information Diffusion
  - Top Degree vertex rankings

Future Roadmap and Getting Involved

The Alan Turing Institute

CHOROGRAPH

Drop me a line at

b.a.steer@qmul.ac.uk

Raise PR's/Queries on Git

https://github.com/miratepuffin/raphtory