Raphtory: Streaming Analysis Of Distributed Temporal Graphs

Benjamin Steer, Felix Cuadrado & Richard G. Clegg
Motivation

Traditional Graph Processing Systems

Graph Snapshot Processing

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

Partition Manger Ingestion

Partition 1

Vertex 1
Created: t8
Created: t14

Edge 1 → 2
Created: t14
Deleted: t15

Partition 2

Edge 1 → 2
Created: t14
Deleted: t15

Vertex 2
Created: t14
Deleted: t15
Correct update order

{ "Edge Add": { "Message Time": 14, "Source ID": 1, "Destination ID": 2 } }

Partition Manager 1

Created: t8
Created: t14
Vertex 1

Created: t14
Edge 1 → 2

Partition Manager 2

Created: t14
Edge 1 → 2

Created: t9
Created: t14
Vertex 2
Edge Added Before Vertex

{ "Vertex Add": { "Message Time": 8, "Source ID": 1 } }

Created: t8
Created: t14
Vertex 1

Created: t14
Edge 1 → 2

Created: t14
Edge 1 → 2
Vertex 2

Partition Manager 1

Partition Manager 2
Vertex Deletion Before Edge Addition

\{ 
  "Edge Add":{ 
    "Message Time": 14, 
    "Source ID":1, 
    "Destination ID":2 
  } 
\} 

\{ 
  "Vertex Removal":{ 
    "Message Time": 15 
    "Source ID":2 
  } 
\} 

Partition Manager 1

Vertex 1
Created: t8
Created: t14

Edge 1 → 2
Created: t14
Deleted: t15

Partition Manager 2

Vertex 2
Created: t14
Created: t14

Edge 1 → 2
Created: t14
Deleted: t15
Deleted: t15
Live Graph, Views & Snapshots
Views & Windowing

Window (Left Hand Filter)  View (Right Hand Filter)

$t_0$  $t_5$  $t_{10}$  $t_n$

Full History of the Graph

Window Size = 5
Windowing Batches

Batch of Windows (Decreasing in size)

Window Sizes = [5,3,1]
Temporal Range Analysis

Range of Interest = $t_4 \rightarrow t_{10}$
Interval = 2
Gab.ai Connected Components
Every Hour Across Lifetime

![Graph showing connected components over time with different windows (Aggregate, Year, Month, Week, Day, Hour) and their respective growth trends over a period from 2016 to 2018.](image)
Gab.ai Connected Components Every Hour Across Lifetime
Using Raphtory

- Available at github: [https://github.com/miratepuffin/raphtory](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

Drop me a line at
b.a.steer@qmul.ac.uk

Raise PR’s/Queries on Git
https://github.com/miratepuffin/raphtory