The Role of Graph Databases in Geomatics

Hung Cao, Ikechukwu Maduako, Emerson Cavalheri, Ryan Brideau, Monica Wachowicz

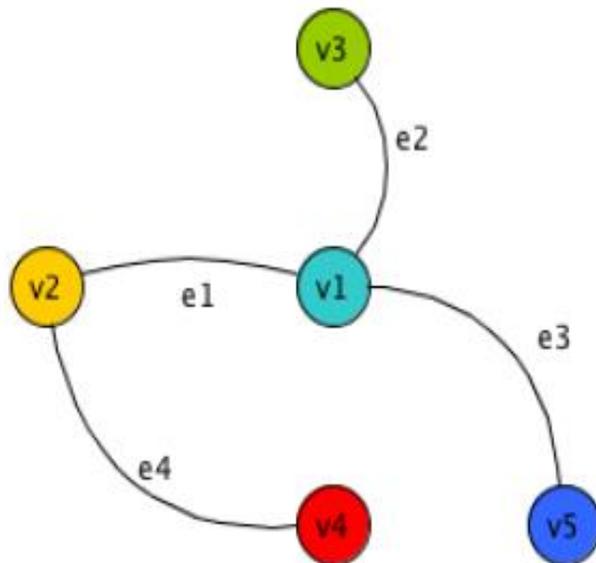{hcao3, imaduako, e.cavalheri, Ryan.Brideau, monicaw}@unb.ca

People in Motion Lab, University of New Brunswick
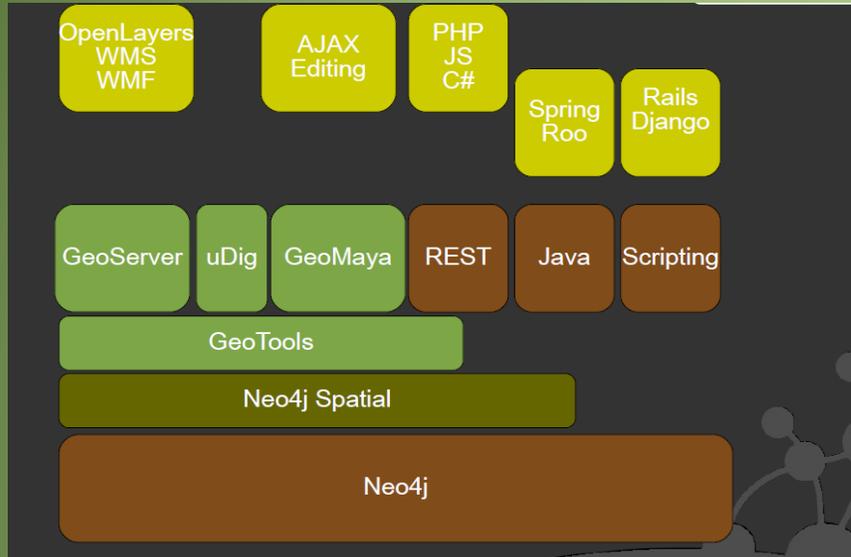
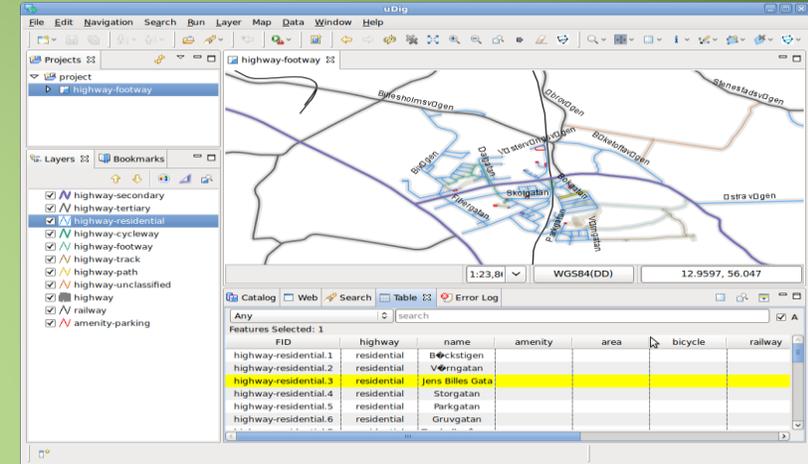# Introduction

What is a graph?

$$G = (V, E)$$

What is a Graph Database?

- It is a database that stores data in a graph.

- It is a data structure that is capable of representing any kind of data for storage and accessibility.

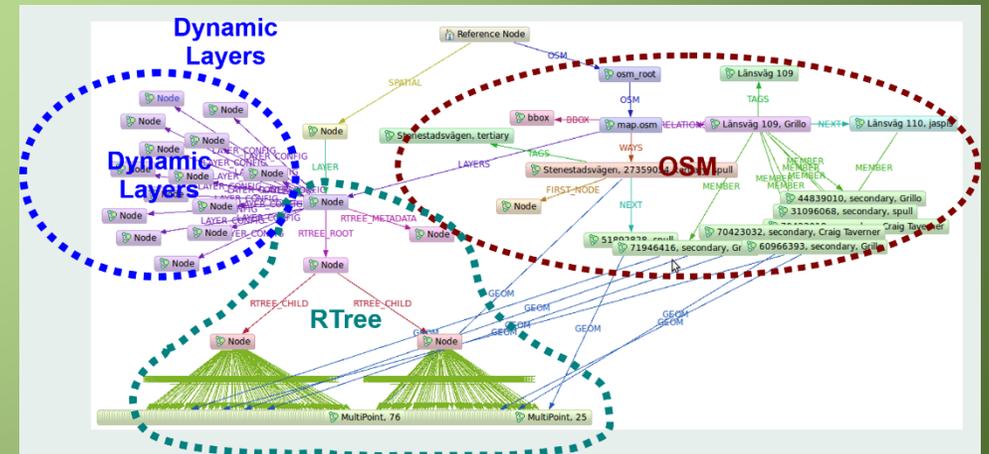# Graph Database in Geomatics

Neo4J Spatial Stack

Neo4j OSM in uDig

Cell Network Analysis

OpenStreetMap Structure in Neo4J

From GPS coordinates to nodes in a graph

Legend

vlr is the vehilce location reports table, where you see vlr, this data comes from this table.

route_id is the id of the route, r is referencing the route table

ta is transit authority and is used in front of some column names, such as ta_route_id that I renamed to rtarouteid and ta_vehicle_id that I renamed to vlrtavehicleid

br is the bid_routes table, brtripid is the trip_id column from the bid_routes table.

tservetime is from the trips table from a column named ta_service_time_id. tstart and tfinish are the start and finish times of each trip
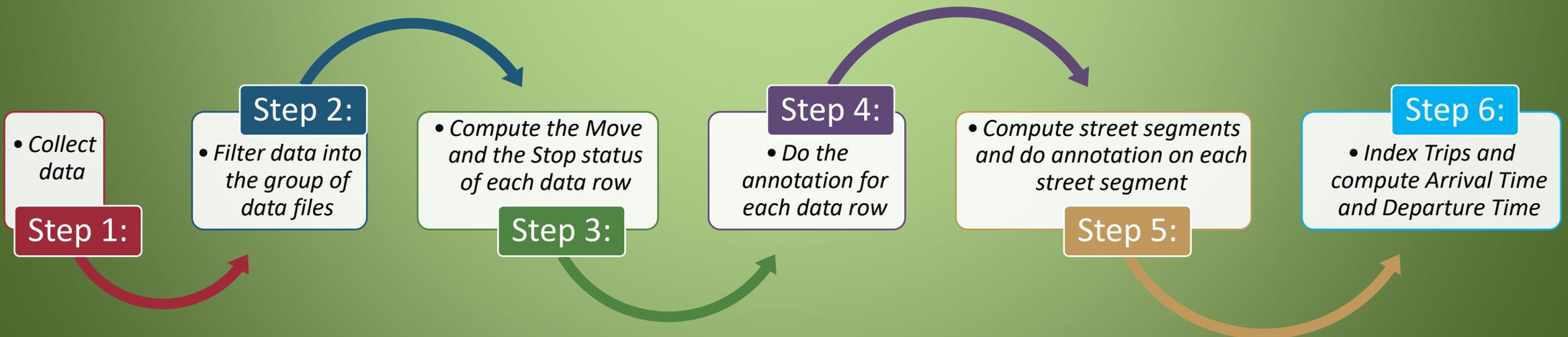
va is from the vehicle_assignment table that is used to assign vehicles to routes

b is the bids table and it contains a description of the bid, which is the route number and route name in the data below
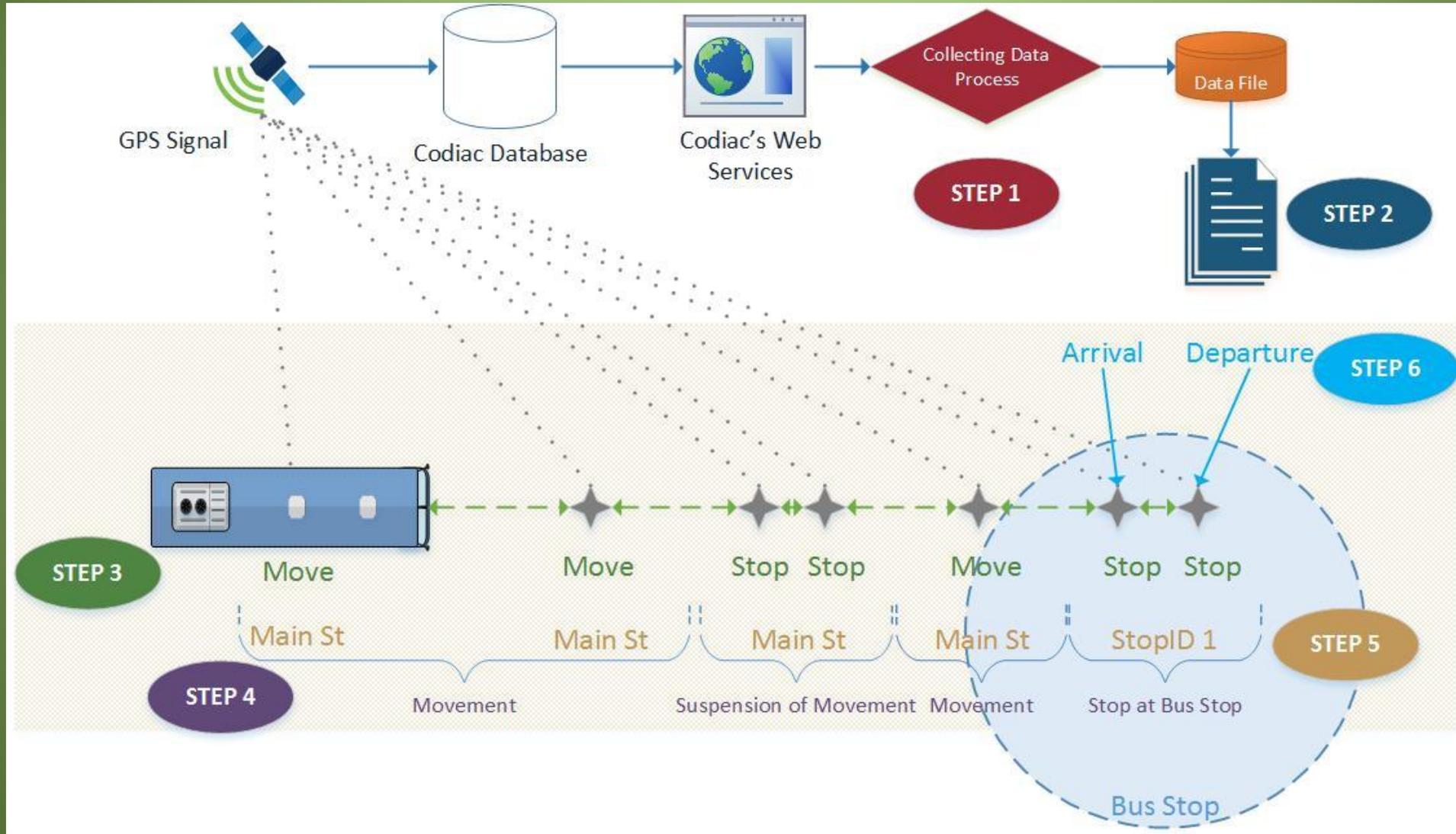
ta_vehicle_id is the descriptive name of the bus, such as com-bus-701 (City of Moncton (com)-bus-701

vlrid	vlr.route_id	route name	rtarouteid	routename	brtripid	tservetime	ttatripid	tstart	tfinish	vlr.Latitude	vlr.longitude	vabidid	bdescription	vlrupdate	vlrvehicleid	vlrtavehicleid
37	21	50S	50s	Red Line (Skip U de M)	44	SU	50S-11	13:00:00	13:30:00	46.0958300000	-64.7590300000	76	50s Skips Univeristy	2016-07-17 12:59:53	37	com-bus-720
3	2	51	51	Green Line	123	SU	51-76	13:00:00	13:45:00	46.0865000000	-64.7848400000	53	51 Green Line A (Sunday)	2016-07-17 12:59:57	3	com-bus-601
4	2	51	51	Green Line	122	SU	51-75	12:30:00	13:15:00	46.0996600000	-64.8062900000	5	51 Green Line B (Sunday)	2016-07-17 12:59:55	4	com-bus-602
2	3	52	52	Blue Line	209	SU	52-73	13:00:00	13:30:00	46.0861500000	-64.7855500000	9	52 Blue Line A (Sunday)	2016-07-17 12:...	2	com-bus-600

- Data was provided by Codiac Transit Moncton

- There are 17 fields in each row of the data. E.g.: TripID, RouteID, Longitude, Latitude, Time Stamp

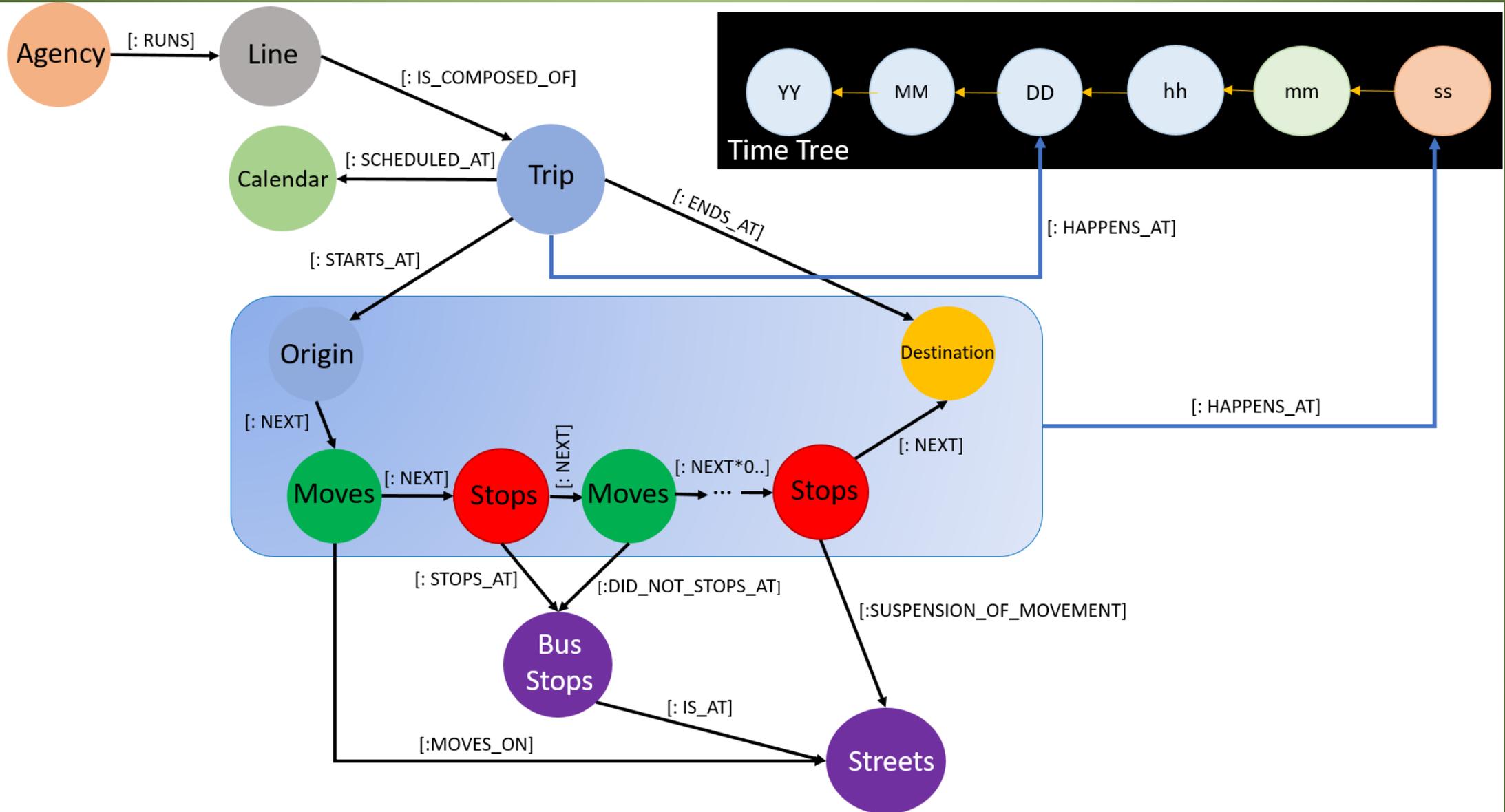- Each row is collected every 5 seconds.

Step 1:
- *Collect data*

Step 2:
- *Filter data into the group of data files*

Step 3:
- *Compute the Move and the Stop status of each data row*

Step 4:
- *Do the annotation for each data row*

Step 5:
- *Compute street segments and do annotation on each street segment*

Step 6:
- *Index Trips and compute Arrival Time and Departure Time*

From GPS coordinates to nodes in a graph

Outcomes

Time-Varying Graph Data Model

System Configuration & Dataset

Graph Database System : Neo4J 3.01

Development Language : Cypher & Python
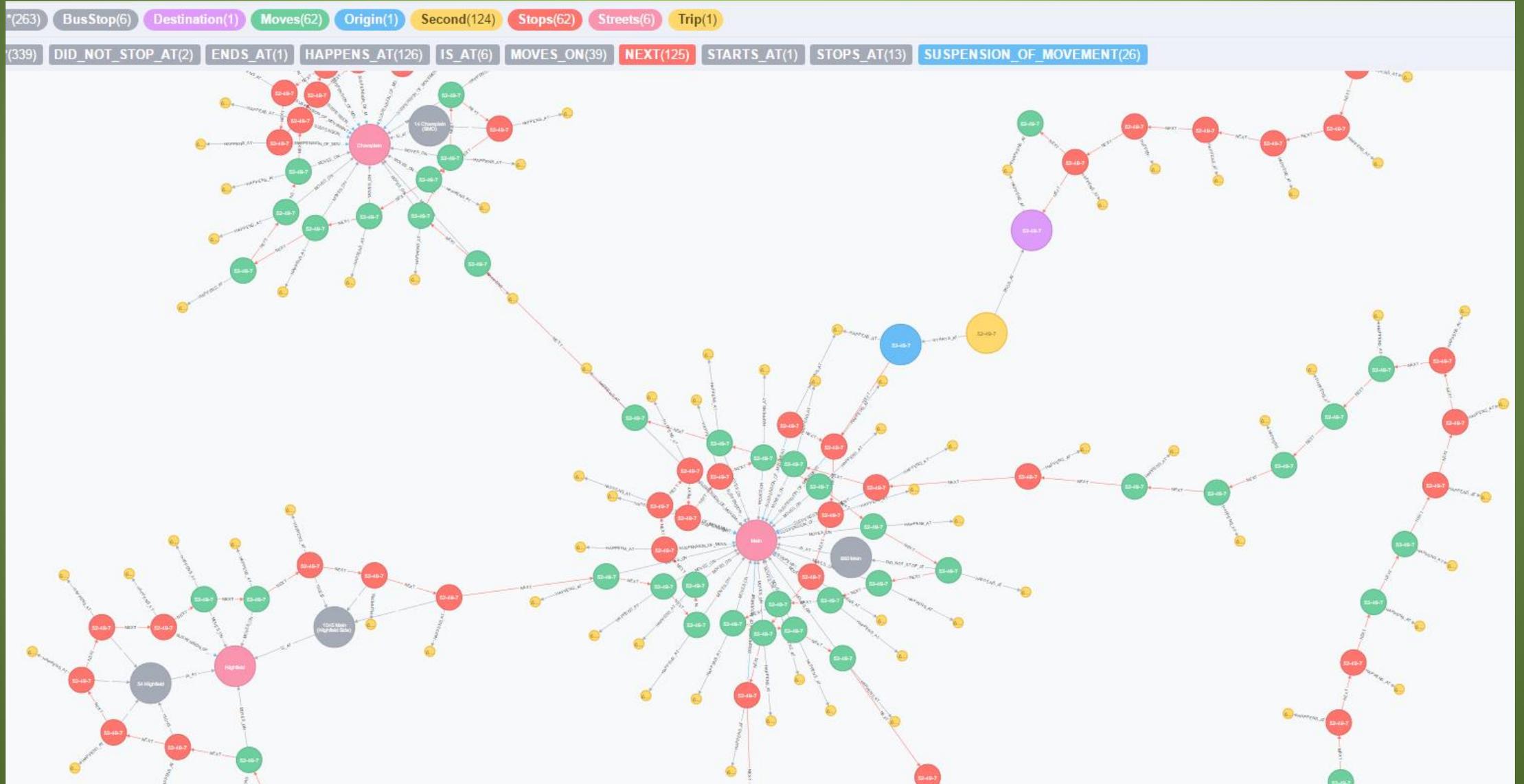
Machine: 3GHz, 32GB Memory & 3TB Disk

Database composition size:

2 weeks of data / 30 buses routes
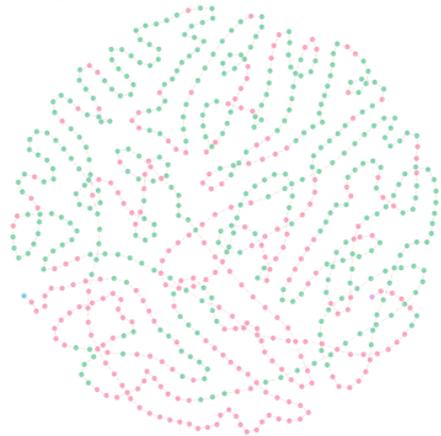Approximately 1 million nodes

4.5 million directed weighted edges

Trip Connectivity
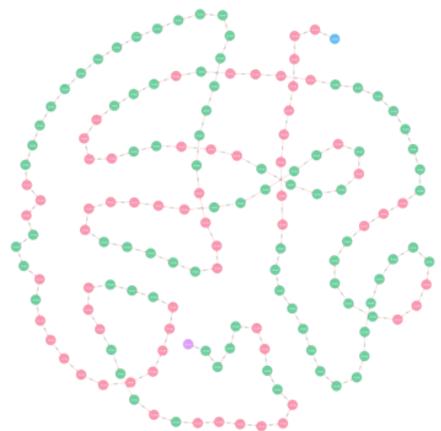
Longest and shortest path at peak hours

8am

Longest path – 81-3-9
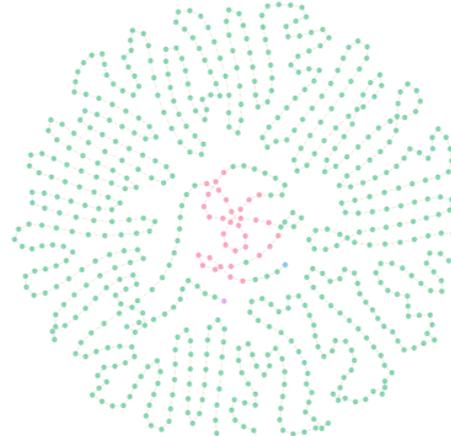
Stats:
#nodes: 631
#edges: 630
#duration: 59min
#distance: 24 km

Shortest path – 51-9-8

Stats:
#nodes: 154
#edges: 153
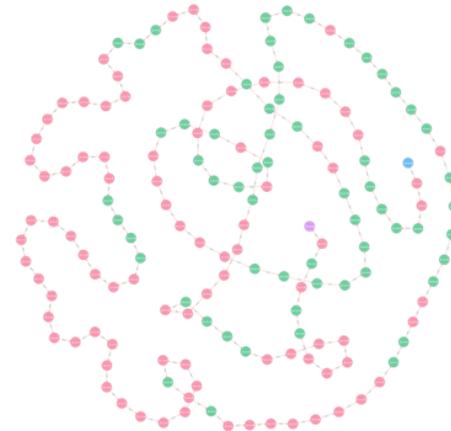#duration: 45min
#distance: 7.6 km

4pm

Longest path – 65-21-9

Stats:
#nodes: 638
#edges: 637
#duration: 44min
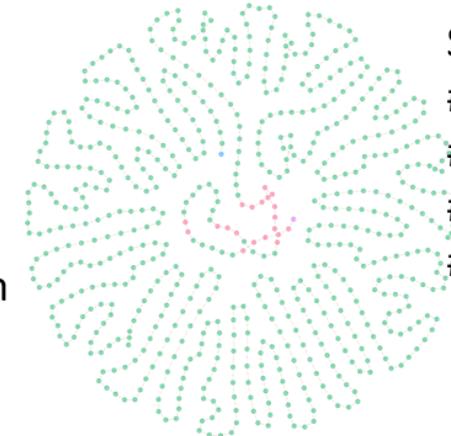#distance: 65.2 km

Shortest path – 52-44-10

Stats:
#nodes: 156
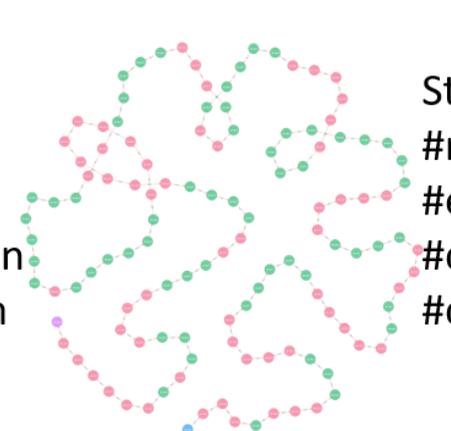#edges: 155
#duration: 18 min
#distance: 3.4km

6pm

Longest path – 65-24-9

Stats:
#nodes: 650
#edges: 649
#duration: 30min
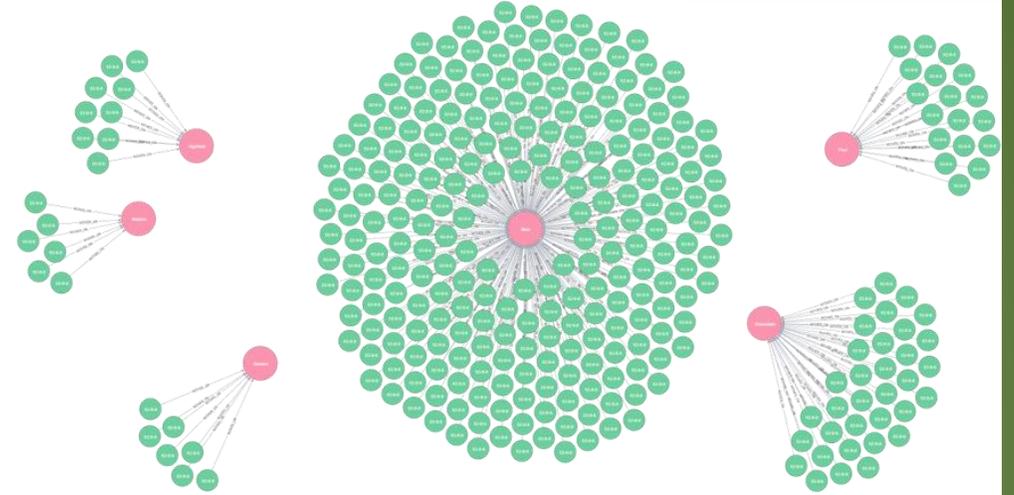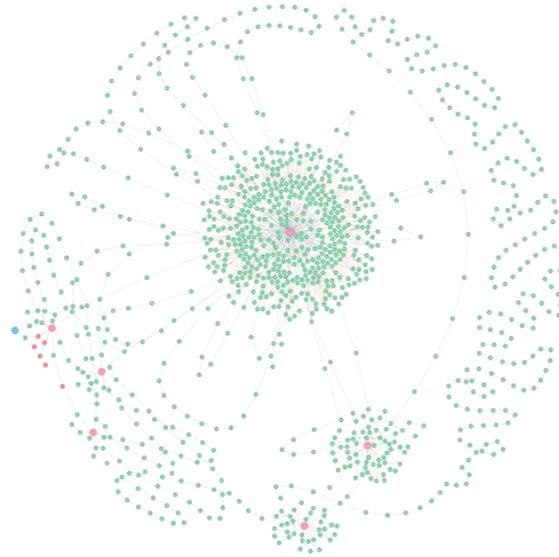#distance: 13.5 km

Shortest path – 52-49-7

Stats:
#nodes: 125
#edges: 124
#duration: 14min
#distance: 4.1 km
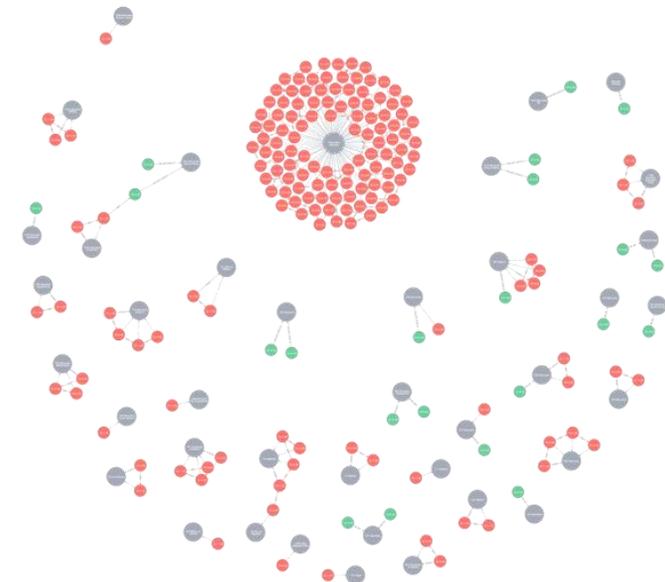
Degree of Centrality

Busiest streets

	A	B
1	Street	degree
2	Main	50779
3	Mountain	40447
4	Plaza	21141
5	Wheeler	12693
6	Champlain	10101
7	St George	9441
8	Elmwood	8756
9	Paul	8449
10	Weldon	8394

Busiest bus stops

	A	B
1	character	degree
2	Plaza Blvd (Walmart)	16835
3	Champlain Place	16110
4	1111 Main	14192
5	54 Highfield	1585
6	77 Weldon	1537
7	Riverview Place	1129
8	655 Main (City Hall)	989
9	1045 Main (Highfield Side	982
10	353 St George at Weldon	966
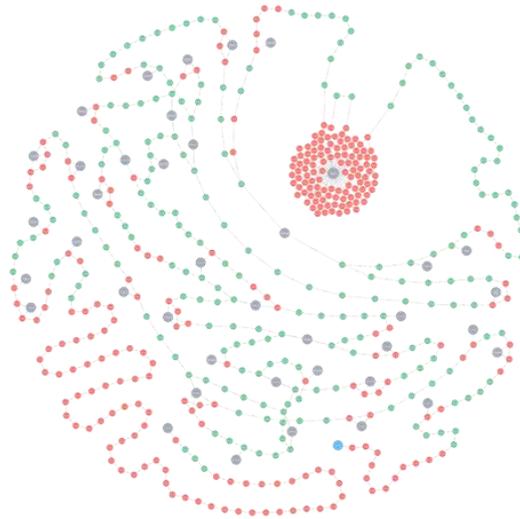
Degree of Centrality by Bus Lines

Data Visualization

How Arrival Time Varies With Time of Day

Conclusions

- Intuitive for data representation

- Reliable with ACID transactions

- Fast processing using a custom disk-based, native storage engine

- Scalable up to several billions of nodes/relationships/properties

- No standard graph query language

People In Motion Lab

www.people-in-motion-lab.org