# An Introduction to Number Plate Data and Relevant Research Applications


Pedro Pinto da Silva

Supervisors: Stephen McGough, Matthew Forshaw

EPSRC CDT in Cloud Computing for Big Data, Newcastle University


January 10, 2018

# Outline

1. Context
2. Identifying trips in number plate data
3. Research Applications
4. Future Work

# Context

# The Urban Traffic Management Centre

## UTMC

Responsible for monitoring traffic, road works, incidents, car parks, . . .

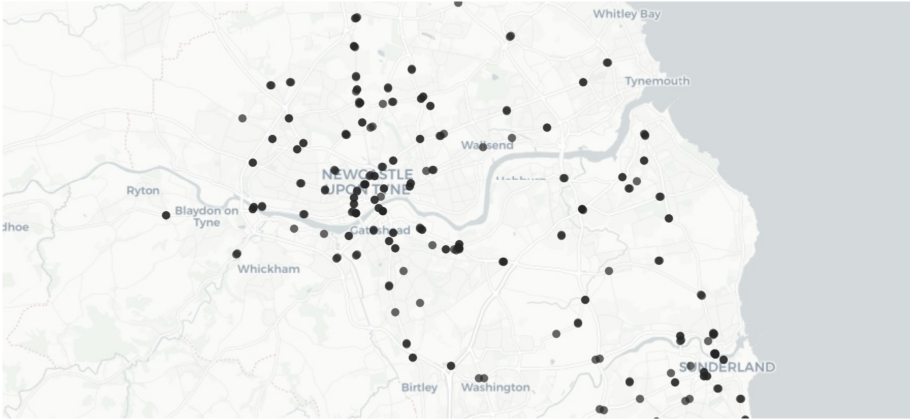



**Figure 1:** Location of automatic number plate recognition (ANPR) cameras in Tyne and Wear.

# The automatic number plate recognition database

## ANPR

The UTMC has kindly provided a subset of the database:

- Number of tables: 82
- Time period: First 6 months of 2017
- Size in disk: $\approx$ 350 GB
- Many tables with static information: e.g Cameras, Nodes, CameraType ...
- 1 major table: Number plate data

# Number plate data

| Tag | Timestamp | Camera Id | Timestamp Error | Tag Confidence | Vehicle Direction |
|---|---|---|---|---|---|
| 12862943 | 2016-02-01 00:00:00 | 18 | 8 | 61 | Towards |
| 169239 | 2016-02-01 00:00:00 | 1031 | 0 | 100 | Away |
| 1862361 | 2016-02-01 00:00:35.080 | 19 | 22 | 83 | Away |

**Table 1:** Sample of number plate data

# Identifying trips in number plate data

# Sighting

## A vehicle observed by a ANPR camera

### Definition

The *ith* sighting of vehicle $k$ can be defined as:

$$sighting_i^k = (camera, time) \tag{1}$$

### Constraints

Consider $S^k$ the set of sightings of vehicle $k$.
Then, $S^k$ is valid if:

$$\cdots < time_{i-1}^k < time_i^k < time_{i+1}^k < \cdots \quad , \forall i \in S^k \tag{2}$$

Assuming that all camera clocks are synchronised.

# Trip

## A sequence of sightings

### Definition

The *uth* trip of vehicle $k$ is a sequence (n-tuple) of sightings of $k$:

$$trip_u^k = \left( sighting_1^k, sighting_2^k, \ldots, sighting_n^k \right) \tag{3}$$

where $n$ is the number of sightings, i.e. length of the trip.

### Constraints

- $n \geq 1$ – The length of a trip has to be equal or greater than 1
- Every two consecutive sightings in a trip occur within a maximum time interval $\tau$.

# Journey time

## A sequence of travel times

### Definition

From the *uth* trip of vehicle $k$, we can calculate the corresponding journey time sequence, of length $n-1$:

$$journey\ time_u^k = \left(time_2^k - time_1^k, \ldots, time_n^k - time_{n-1}^k\right) \tag{4}$$

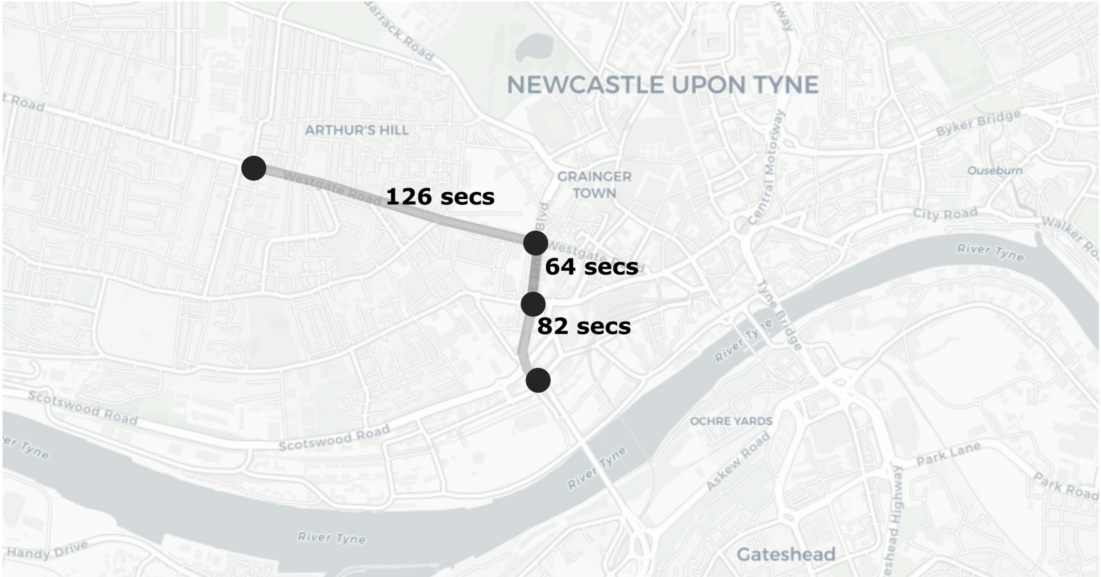
### Valid Trip

A trip of length $\geq 2$ is valid if all partial journey times are lower than a threshold $\tau$:

$$journey\ time_{ui}^k < \tau\ ,\ \forall i \in journey\ time_u^k \tag{5}$$

Note: Realistically there should also be a lower bound for each journey time.

# Example of a trip of length 4



# Origin-destination matrix

Let $P$ be the set of origins and $Q$ the set of destinations:

$$flow_p = \sum_{q \in Q} flow_{pq} \quad (6) \qquad flow_q = \sum_{p \in P} flow_{pq} \quad (7)$$

where $flow_{pq}$ is the number of trips originating at $p$ and terminating at $q$.

$$\text{Trip Matrix} = \begin{bmatrix} flow_{11} & flow_{12} & \dots & flow_{1q} \\ flow_{21} & flow_{22} & \dots & flow_{2q} \\ \dots & \dots & \dots & \dots \\ flow_{p1} & flow_{p2} & \dots & flow_{pq} \end{bmatrix} \quad (8)$$

**Applications** Transportation planning and forecasting.

# Trip data

| Vehicle | Trip | Sighting | Length | Tau | Camera | Journey Time | Trip Id |
|---|---|---|---|---|---|---|---|
| 2362920 | 1 | 1 | 4 | 600 | 1014 | NA | 21 |
| 2362920 | 1 | 2 | 4 | 600 | 1044 | 82 secs | 21 |
| 2362920 | 1 | 3 | 4 | 600 | 35 | 64 secs | 21 |
| 2362920 | 1 | 4 | 4 | 600 | 32 | 126 secs | 21 |

**Table 2:** Sample of trip data

**Additional variables:** Timestamp, Traffic Count, Time Period, Duplicates, Duplicate Time Difference, Starting Camera, Ending Camera

# Challenge 1 – Choosing $\tau$

- Should $\tau$ be fixed or vary over time and edge?
- How do we estimate $\tau$?

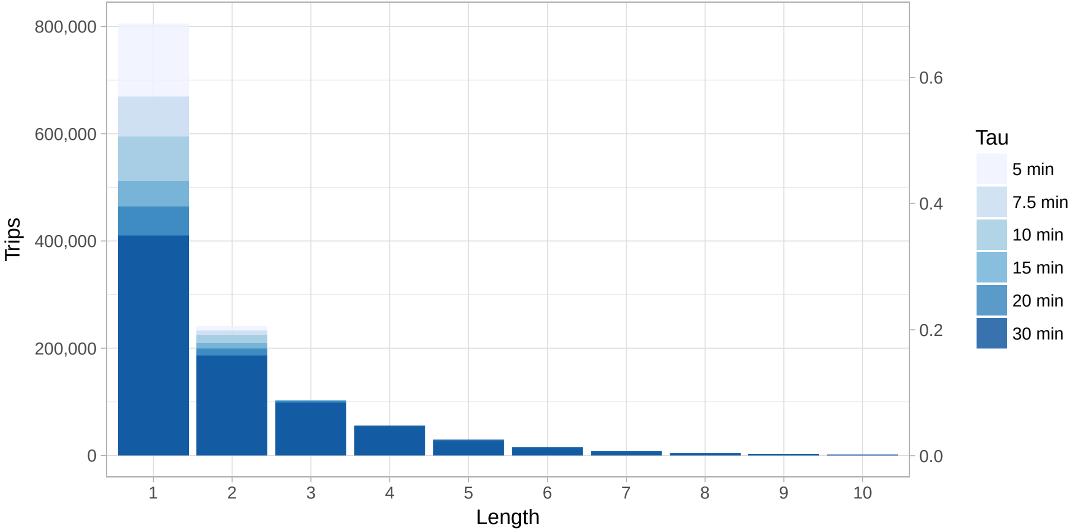
**Simple solution**

Pick a fixed threshold (e.g. 5 min, 10 min, . . . )

**Future work**

- Estimate $\tau$ for each edge in the trip graph as a function of the journey times in that edge.
- Study how varying $\tau$ affects the generation of trip data, and as a consequence, the outcome of the research (e.g. inference on trip matrices, trip graphs, . . . )

# Distribution of trip lengths by $\tau$



# Challenge 2 – Errors in plate scanning

There is a probability (confidence) associated with each sighting.

## Types of errors

- A different plate number was detected instead
- No plate number was detected at all

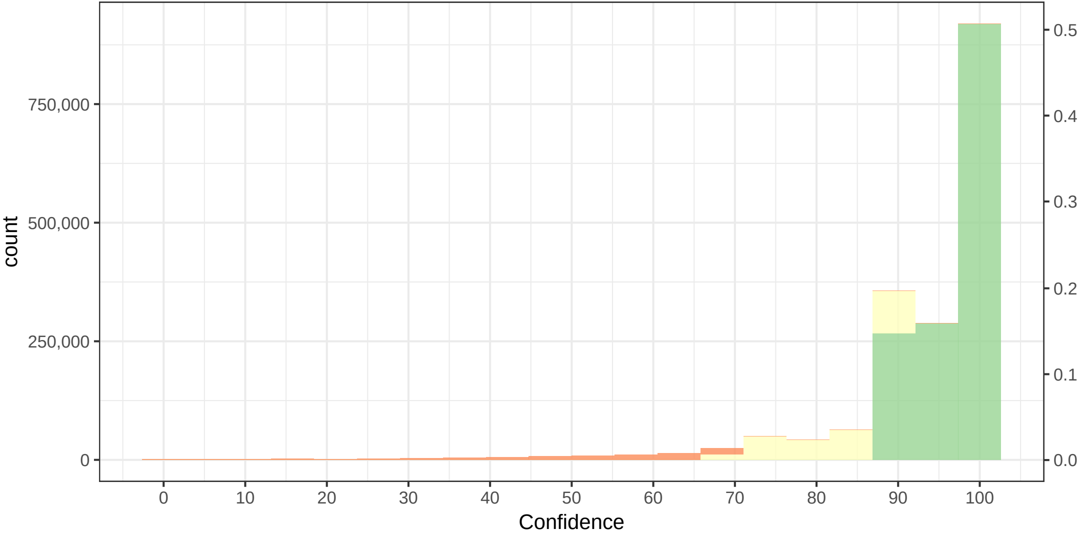
## Simple solution

Filter all sightings with confidence less than e.g. 0.75.

## Future work

- Methods to detect and address the two types of errors
- Model the probability that a trip is missing sightings

# Distribution of plate scan confidence



# Challenge 3 – Duplicate Scannings

A vehicle can be scanned by the same camera multiple times in a single trip.

**Simple solution**

For each trip:

- A sighting is a duplicate if the previous sighting occurs at the same location.

$(1044, 32, 32, 35)$ $(1044, 32, 35, 32)$

$\approx 5\%$of sightings are duplicates

**Future Work**

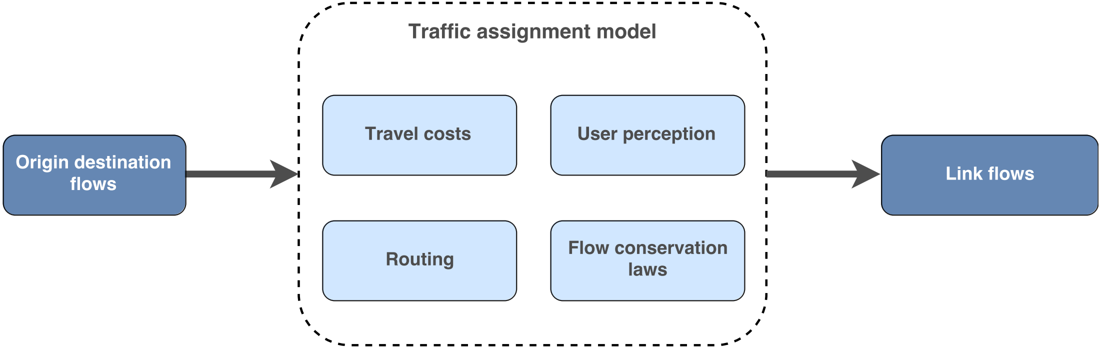
- A sighting is duplicated if there is a previous sighting at the same location and there isn't a significant time interval between them.

# Research Applications

# Application 1 – Transportation Modelling

## Traffic Assignment Problem

How do od-flows disaggretate among the road network?

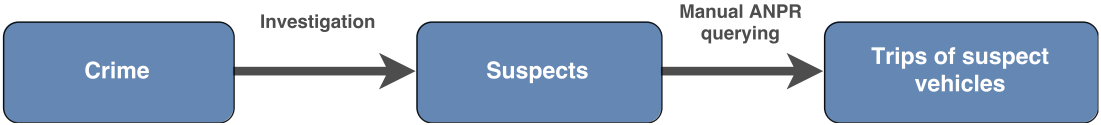



**Applications** Traffic forecasting, transportation planning, traffic simulators

# Application 2 – Abnormal Vehicle Behavior Detection

Law enforcement ocasionally uses the ANPR database to link vehicles to crimes.

## Current workflow

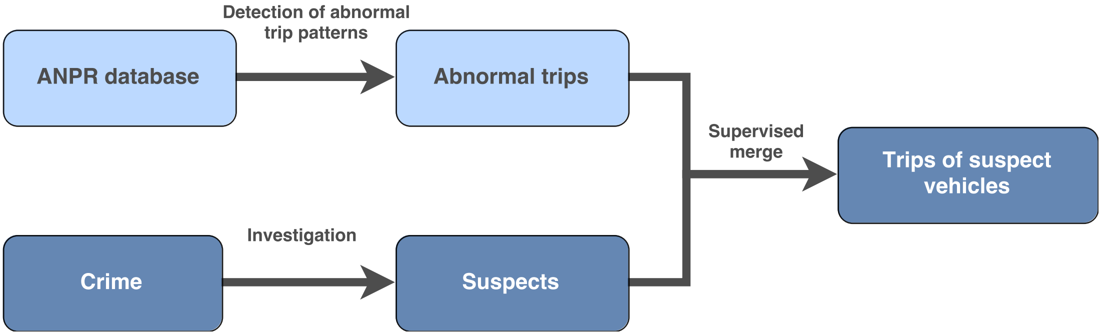



## Limitations

- Lack of technical knowledge
- Time-expensive manual procedure

# Application 2 – Abnormal Vehicle Behavior Detection

## Improved workflow



# Future Work

# Scaling up

Newcastle University

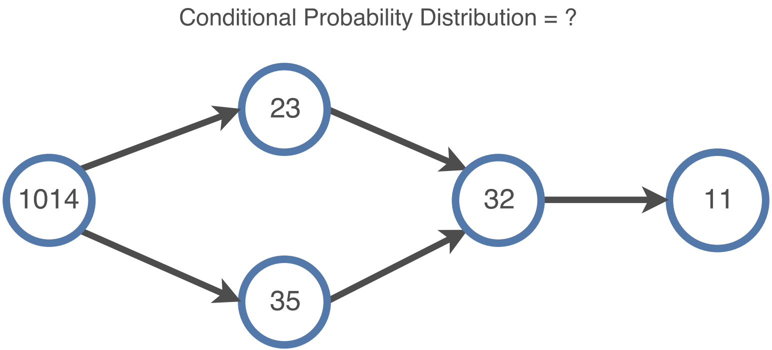
Digital Institute

openstack®

APACHE Spark™

# Probabilistic queries on trip graphs

Which route is likelier to be faster?



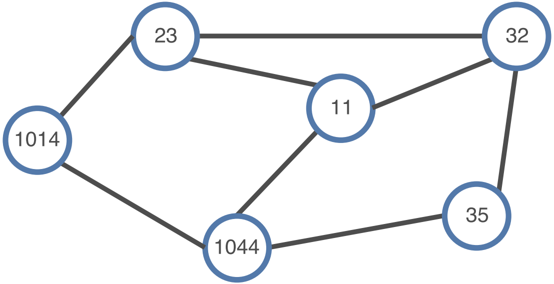

**Applications** Variable message signs with real time route info. and recommendations

# Thank you

A special thank you to Phil Blythe in Civil Engineering and Ray King from the Tyne and Wear Urban Traffic and Management Centre (UTMC) for providing guidance and a copy of the Automatic Number Plate Recognition (ANPR) database.

# Trip graph

Consider the mapping of trips of length $\geq 2$ to trip-pairs of length $= 2$.
Let $G(V, E)$ be the trip graph. $V$ is the set of cameras and $E$ is the set of trip-pairs.




**Weights:** Distribution of journey times, $flow_{pq}$, routing info, probabilities, ...

**Trips** are walks on the trip graph.

# Considerations

## Choosing $\tau$

- Should $\tau$ be fixed or vary over time and camera-pair?
- Estimating $\tau$ from observed journey times

## Errors in number plate recognition

- How to detect and address errors?
- What is the probability that a trip is missing sightings?

## Duplicate sightings

- A sighting is a duplicate if the previous sighting occurs at the same location.
- What about trips containing cycles?

# [EXTRA][DRAFT] More on trip graphs..

- How does the structure and properties of trip graphs vary over time (node connectivity, centrality, ...)?
- How to incorporate/address the time component on trip graphs? For instance, on bayesian networks. Do these methods capture the time-variance of trip choices?
- How good is the distribution of cameras on the road network?
- Does traffic congestion explain particular routing choices made by vehicles?
- Estimate the Markov matrix from an observed set of finite markov chains (a.k.a random walks)
- Compute a trip graph per hour/day/week/month. Compute properties/structure (connectivity, centrality). Cluster the graphs based on these metrics.