

New York Taxi Trip Duration

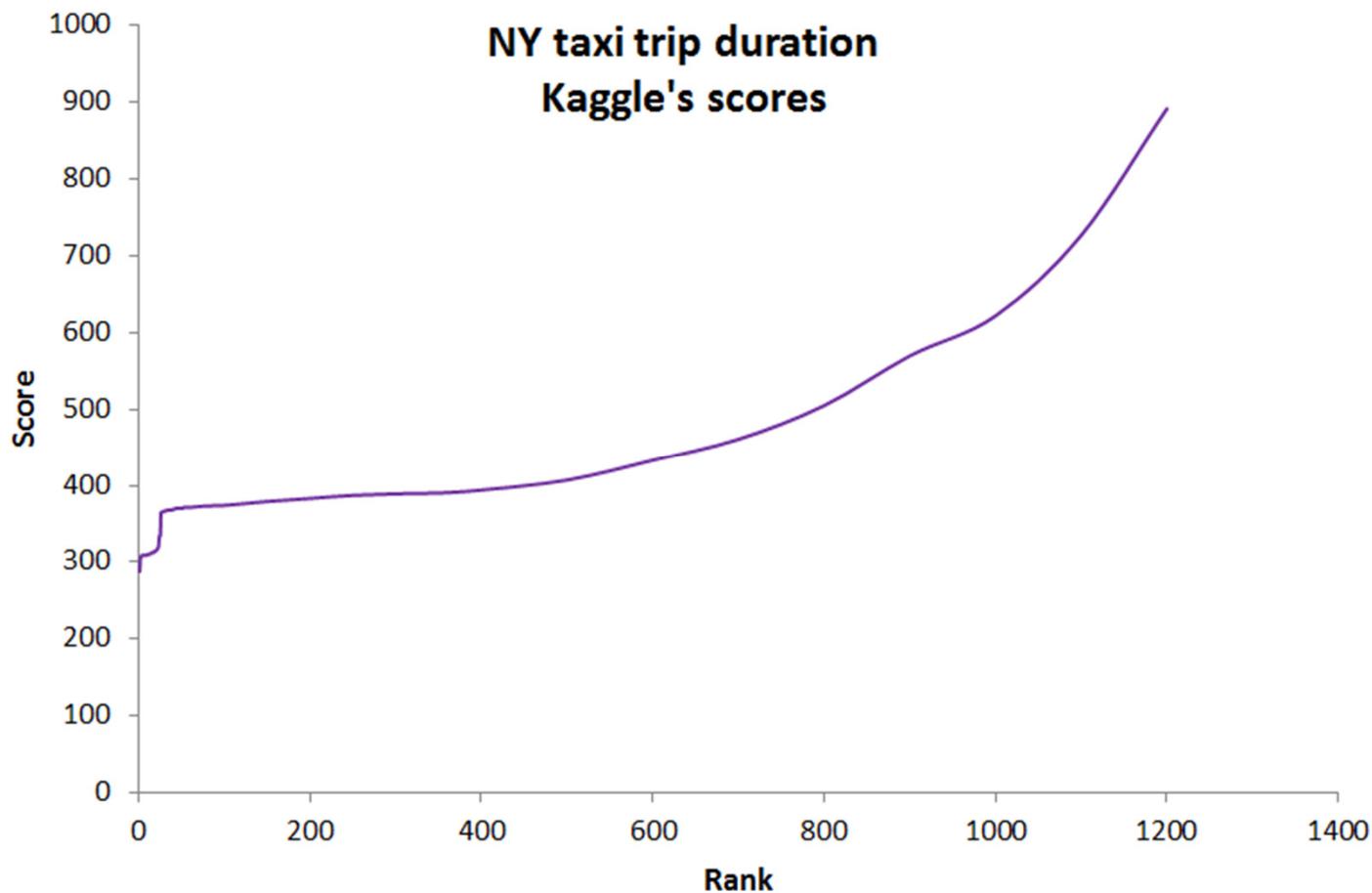
The Graph-Way to reach top 1% solution



Serge NAKACHE
Engineer @ Thales

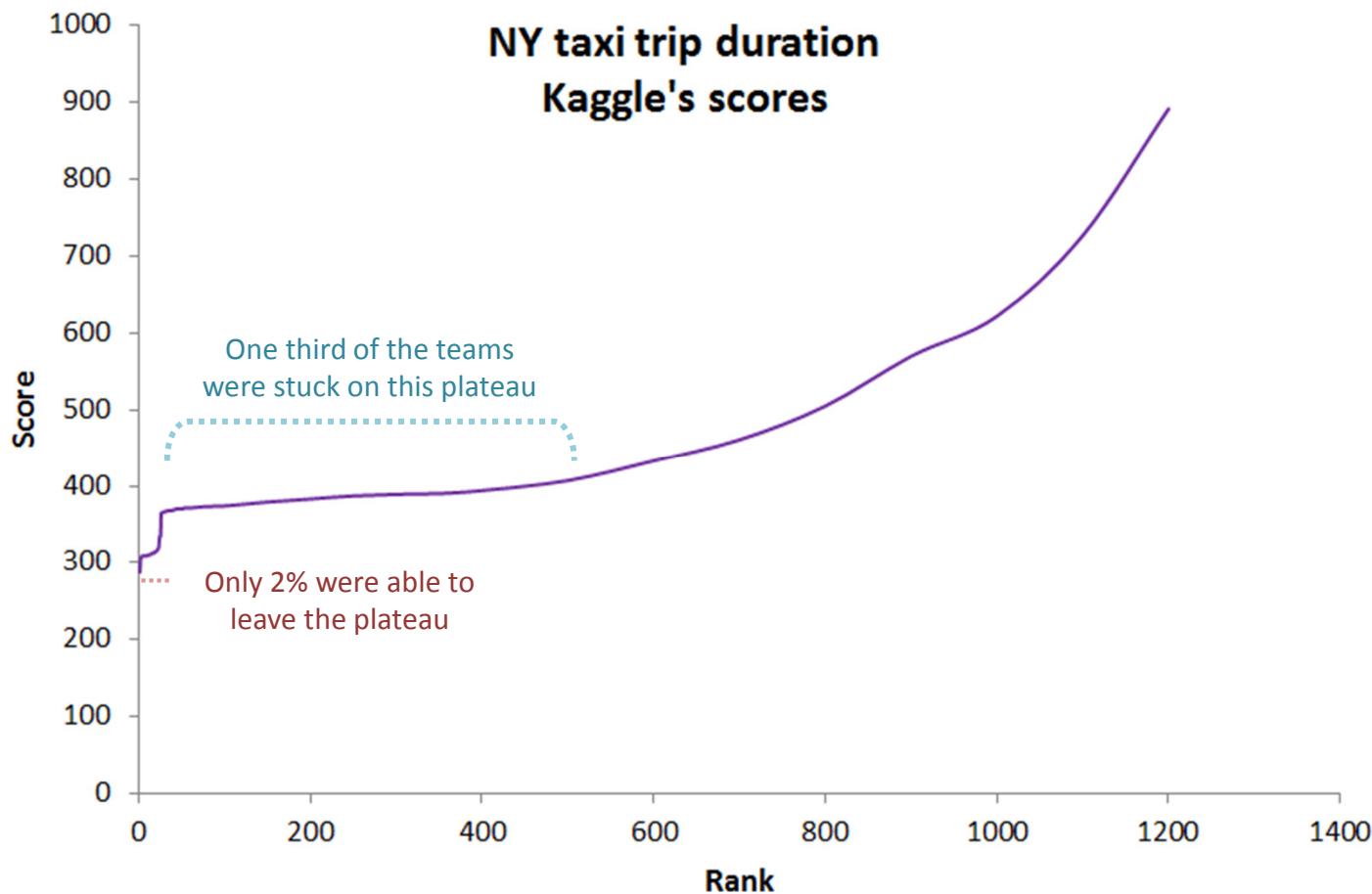
New York Taxi Trip Duration

The Graph-Way to reach top 1% solution



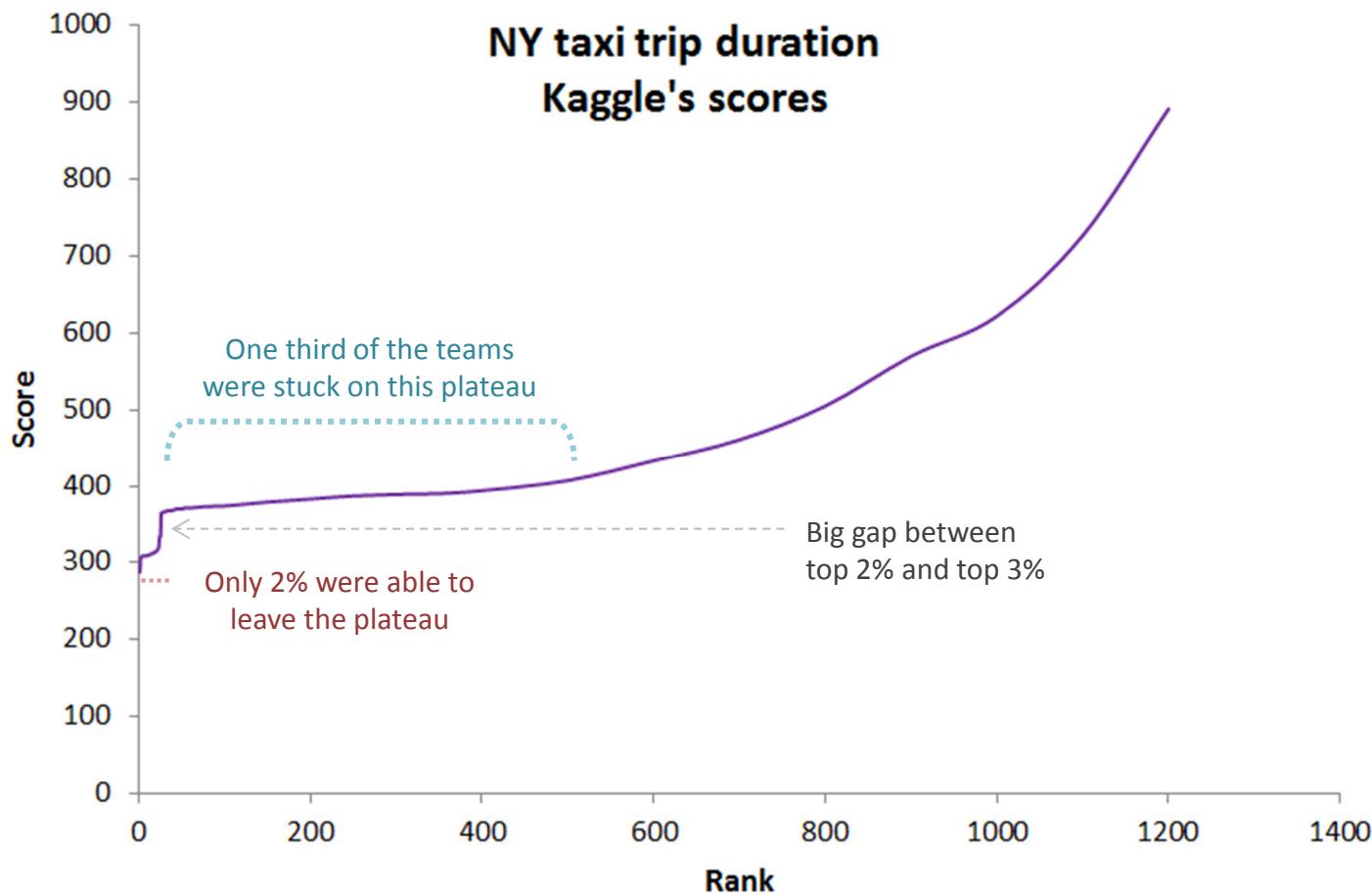
New York Taxi Trip Duration

The Graph-Way to reach top 1% solution



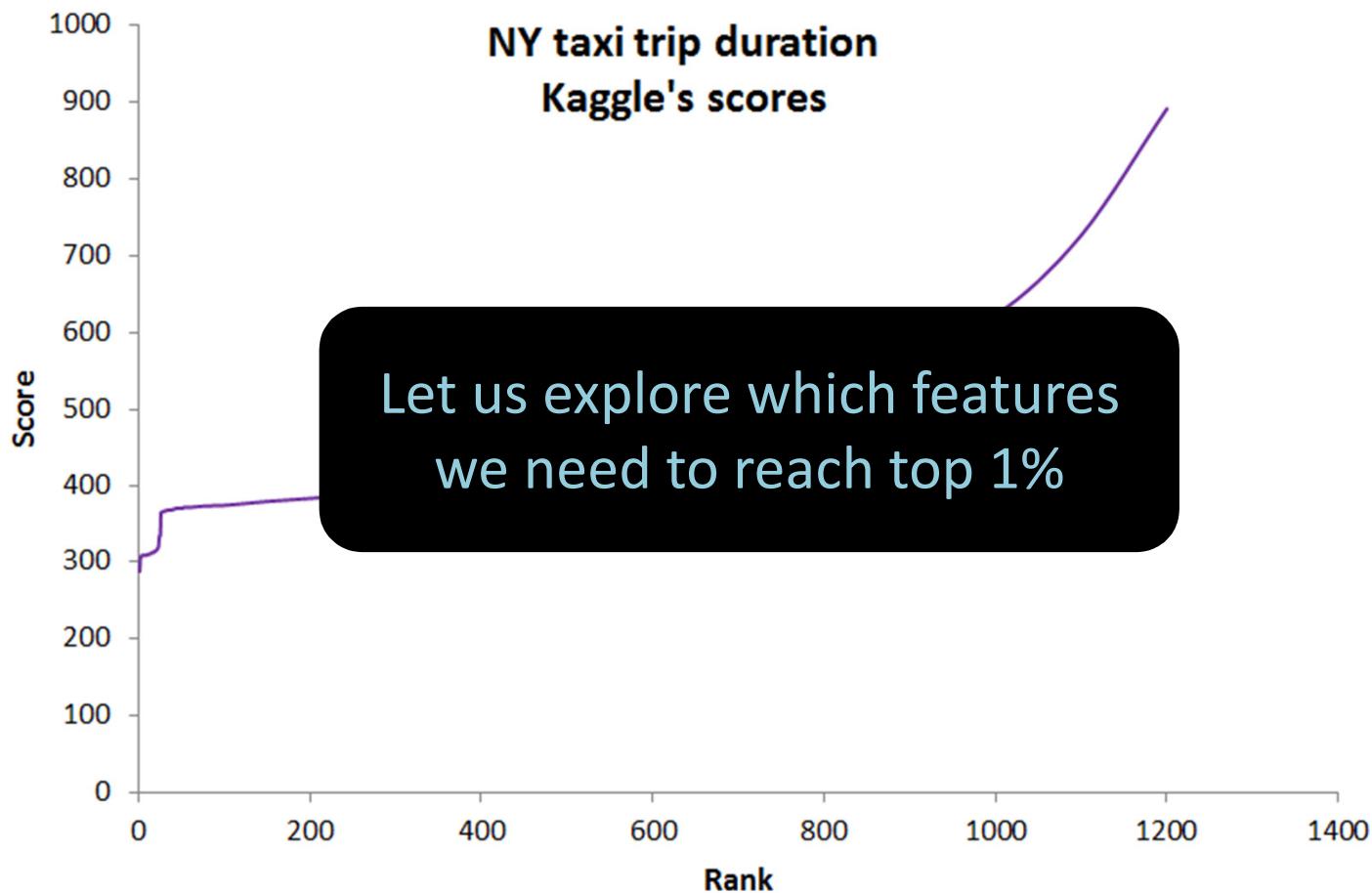
New York Taxi Trip Duration

The Graph-Way to reach top 1% solution



New York Taxi Trip Duration

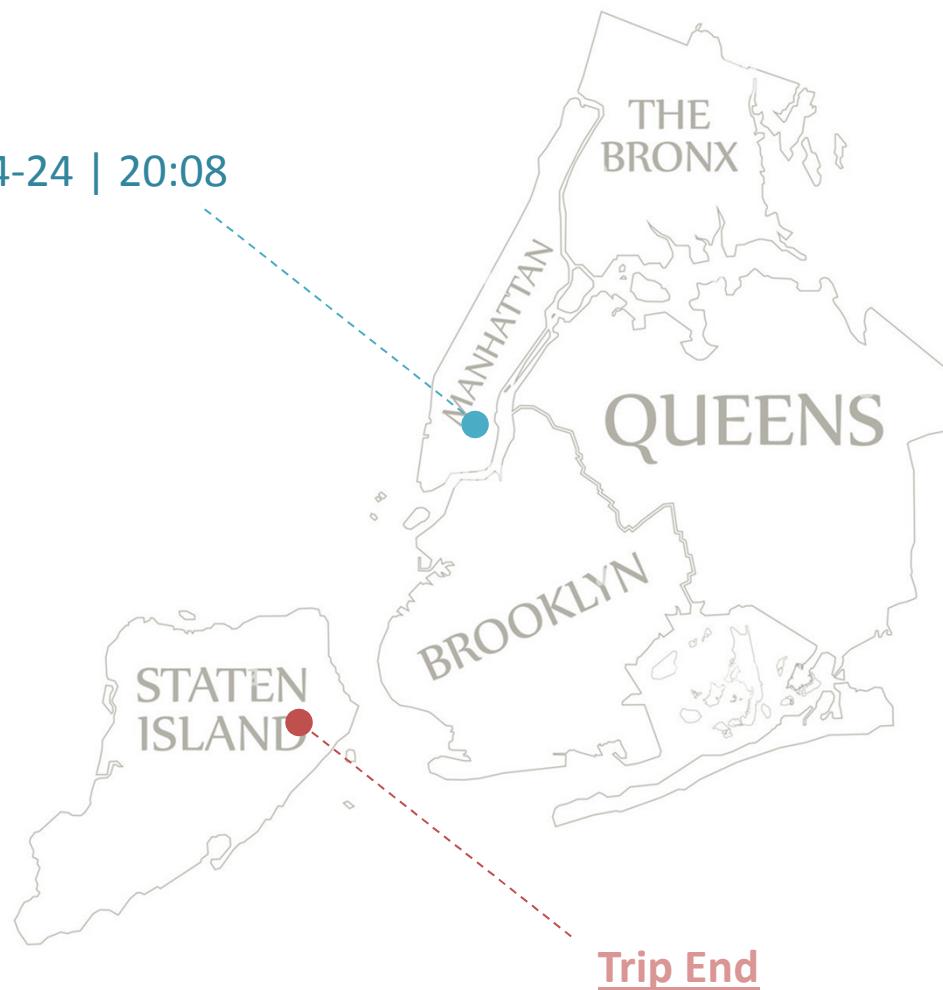
The Graph-Way to reach top 1% solution



Goal

Trip Start

Saturday | 2016-04-24 | 20:08

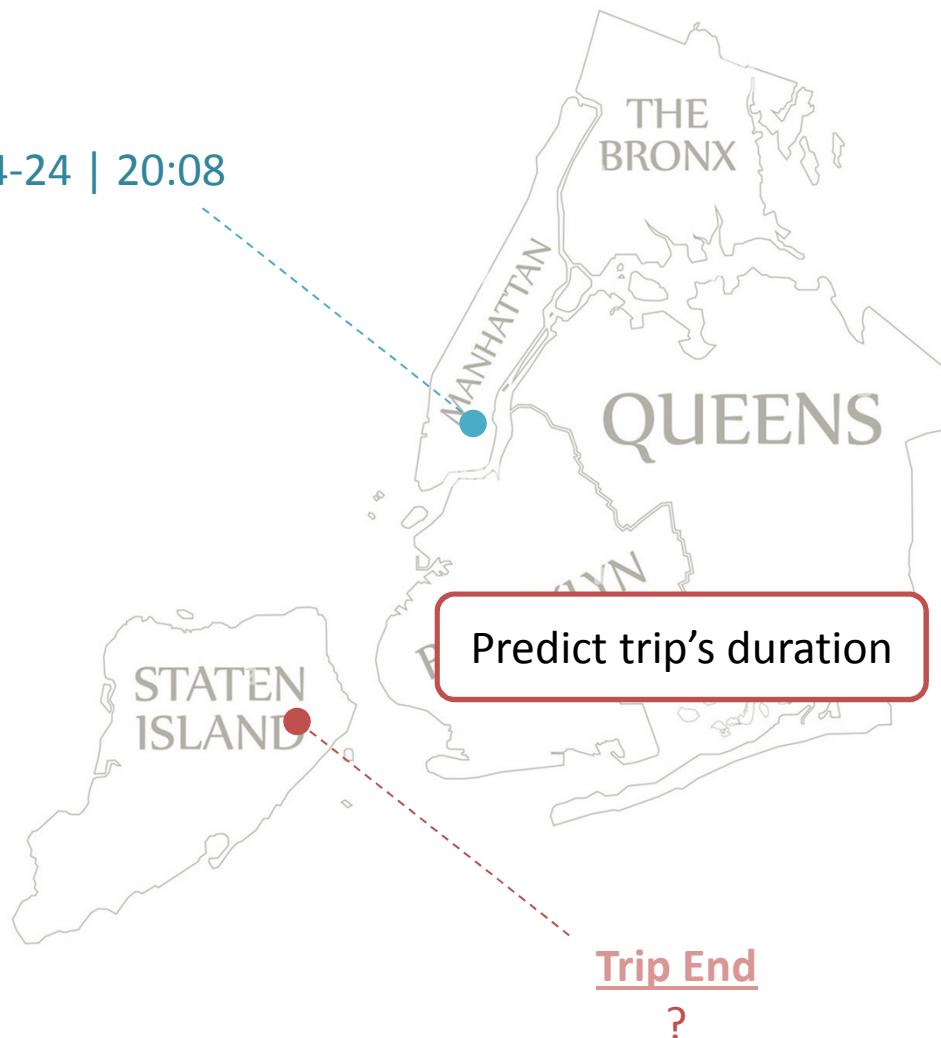


Trip End

Goal

Trip Start

Saturday | 2016-04-24 | 20:08



Goal

Trip Start

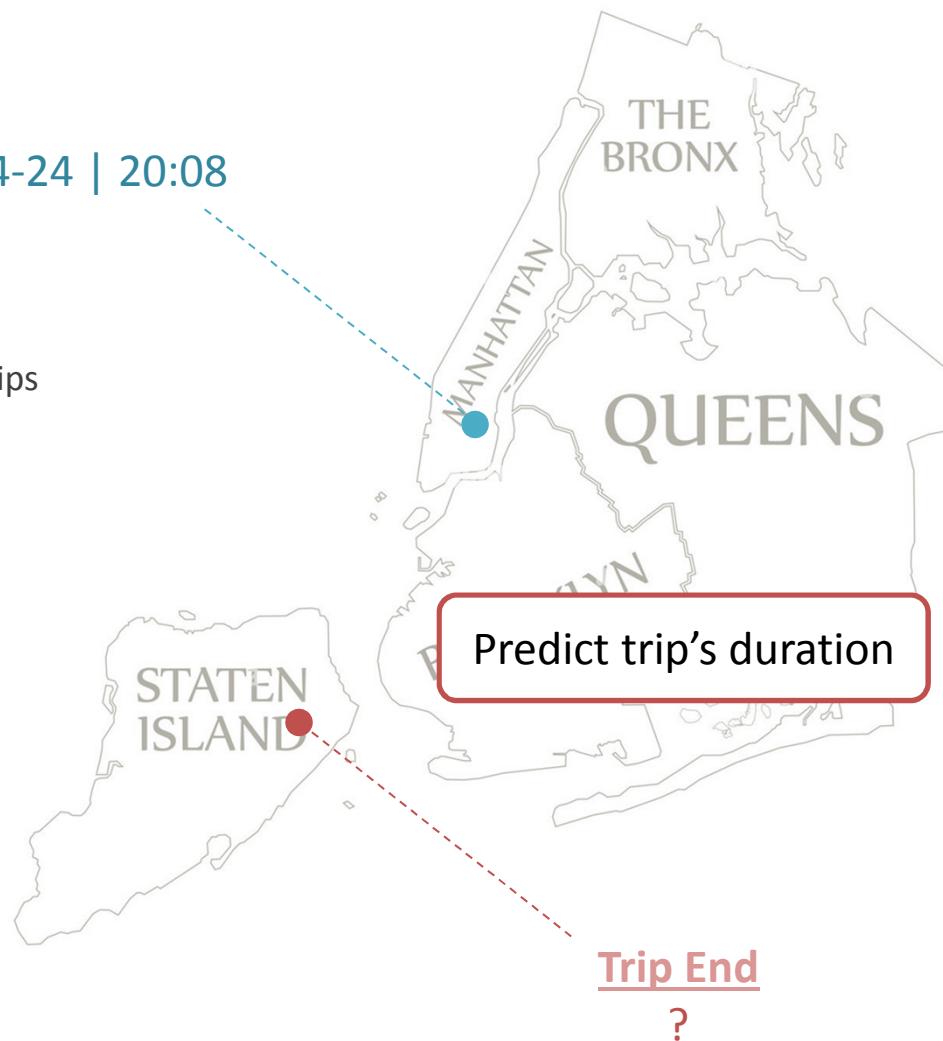
Saturday | 2016-04-24 | 20:08

Training set : 1,458,644 trips

Test set : 625,134 trips

- Pickup coordinates
- Dropoff coordinates
- Start datetime
- End datetime
- Passengers
- Vendor

Features



Classic method used to characterize a trip



Classic method used to characterize a trip

- Start (x,y)
- End (x,y)



Classic method used to characterize a trip

- Start (x,y)
- End (x,y)
- Direct distance



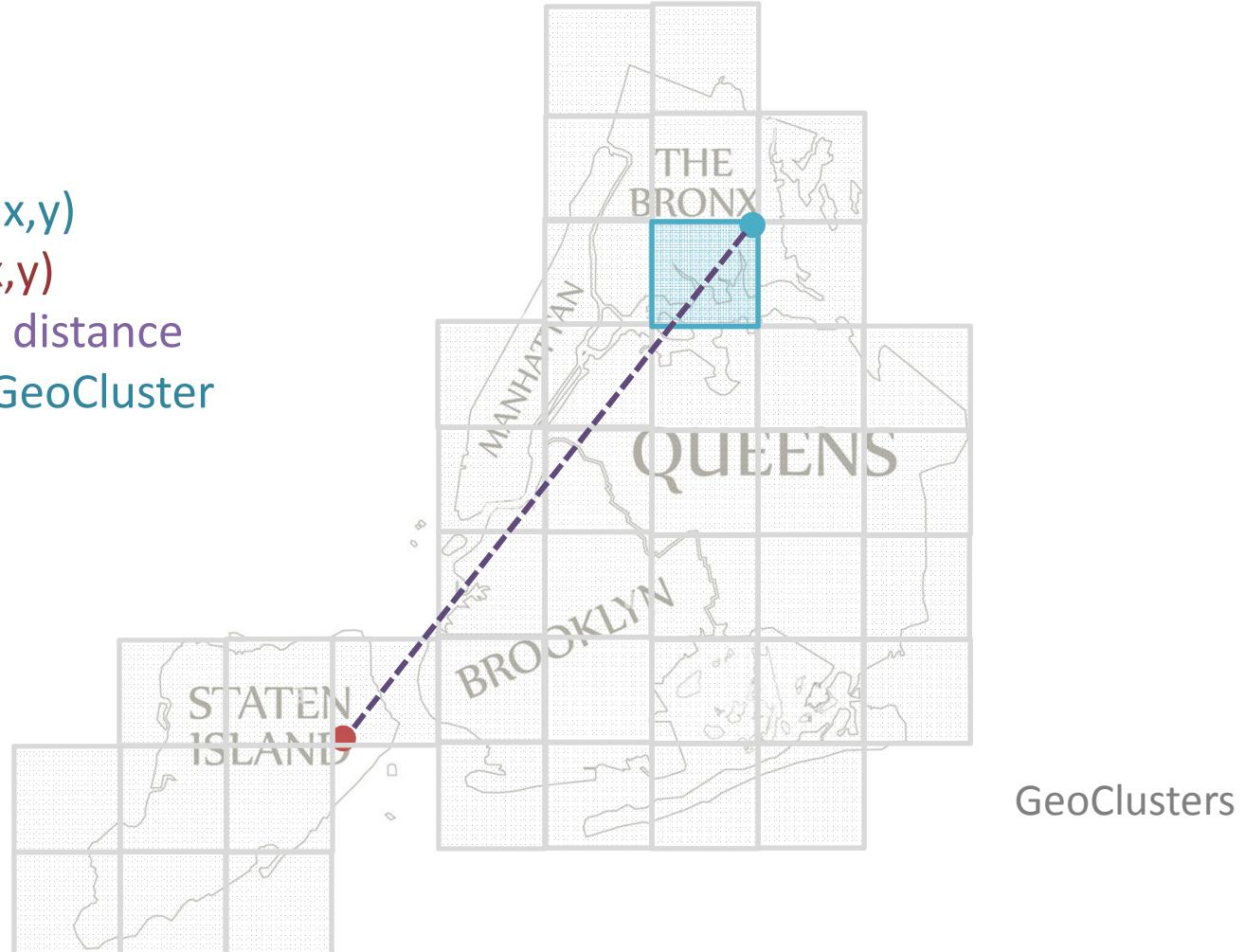
Classic method used to characterize a trip

- Start (x,y)
- End (x,y)
- Direct distance



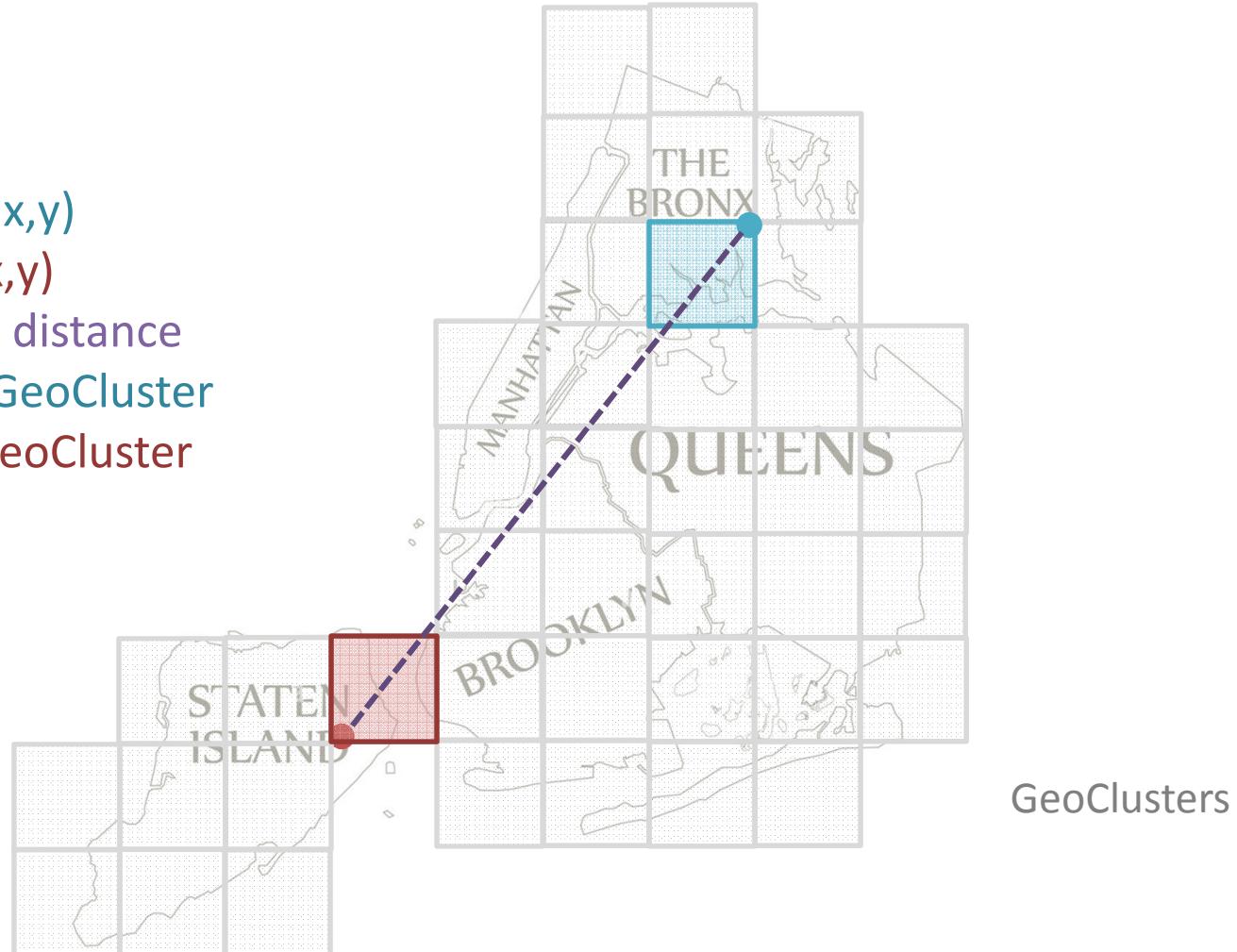
Classic method used to characterize a trip

- Start (x,y)
- End (x,y)
- Direct distance
- Start GeoCluster



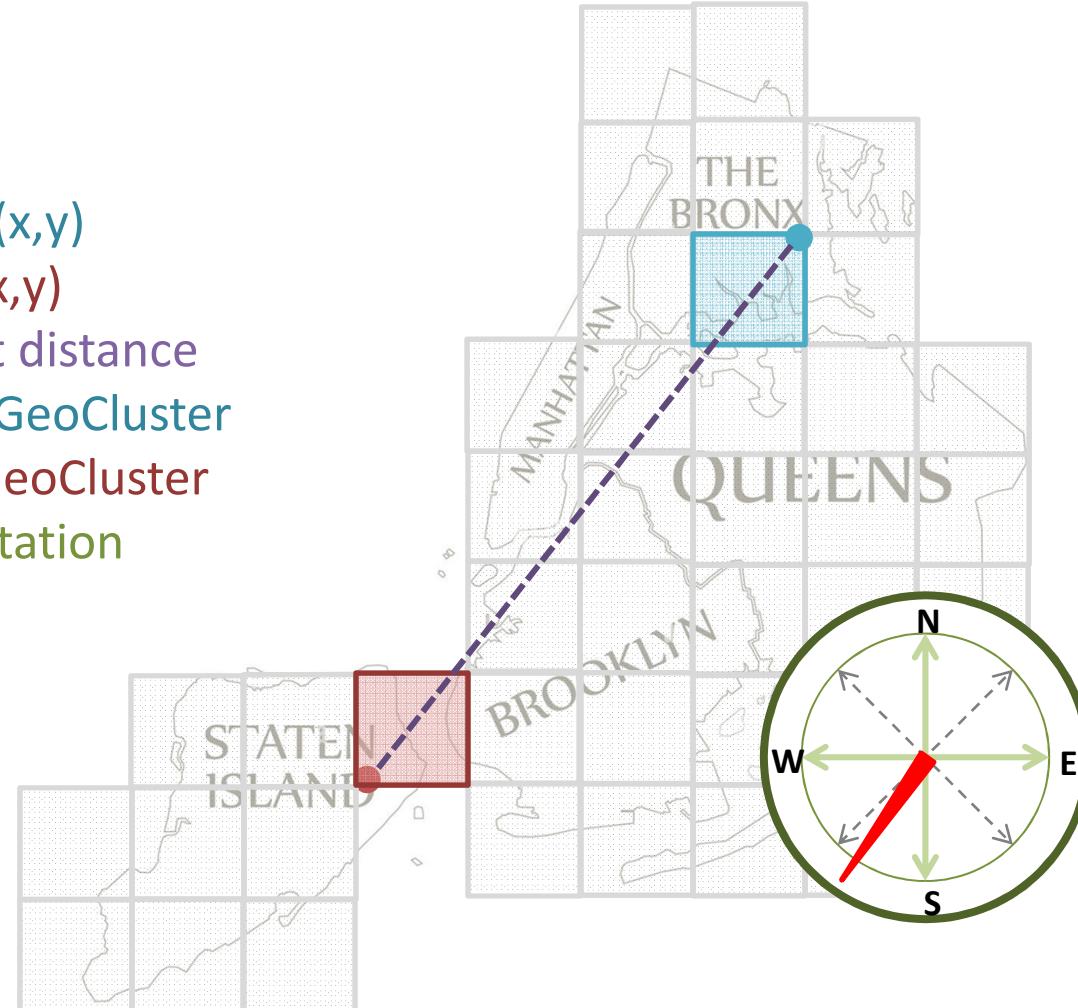
Classic method used to characterize a trip

- Start (x,y)
 - End (x,y)
 - Direct distance
- Pair of GeoClusters {
- Start GeoCluster
 - End GeoCluster



Classic method used to characterize a trip

- Start (x,y)
 - End (x,y)
 - Direct distance
- Pair of GeoClusters {
- Start GeoCluster
 - End GeoCluster
 - Orientation

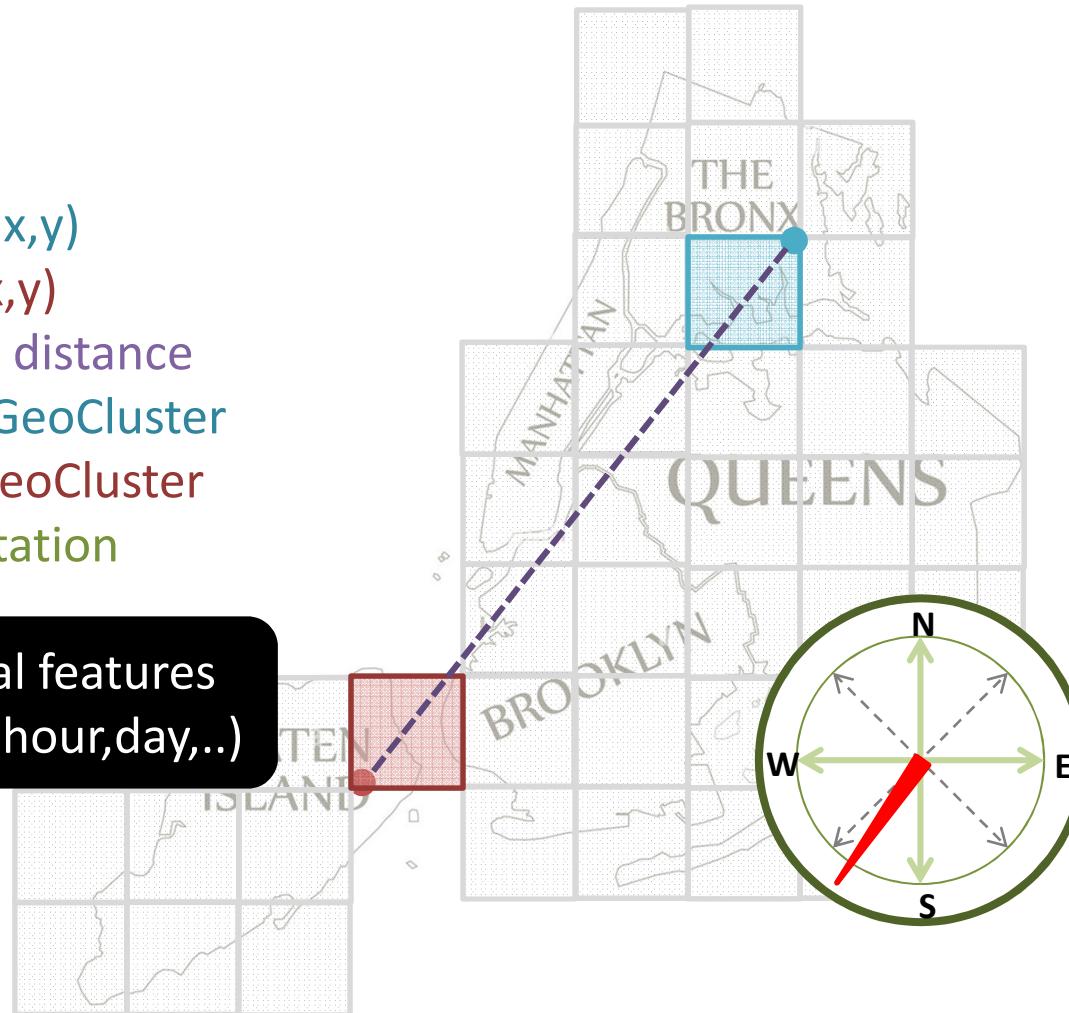


Classic method used to characterize a trip

- Start (x,y)
 - End (x,y)
 - Direct distance
- Pair of GeoClusters {
- Start GeoCluster
 - End GeoCluster
 - Orientation

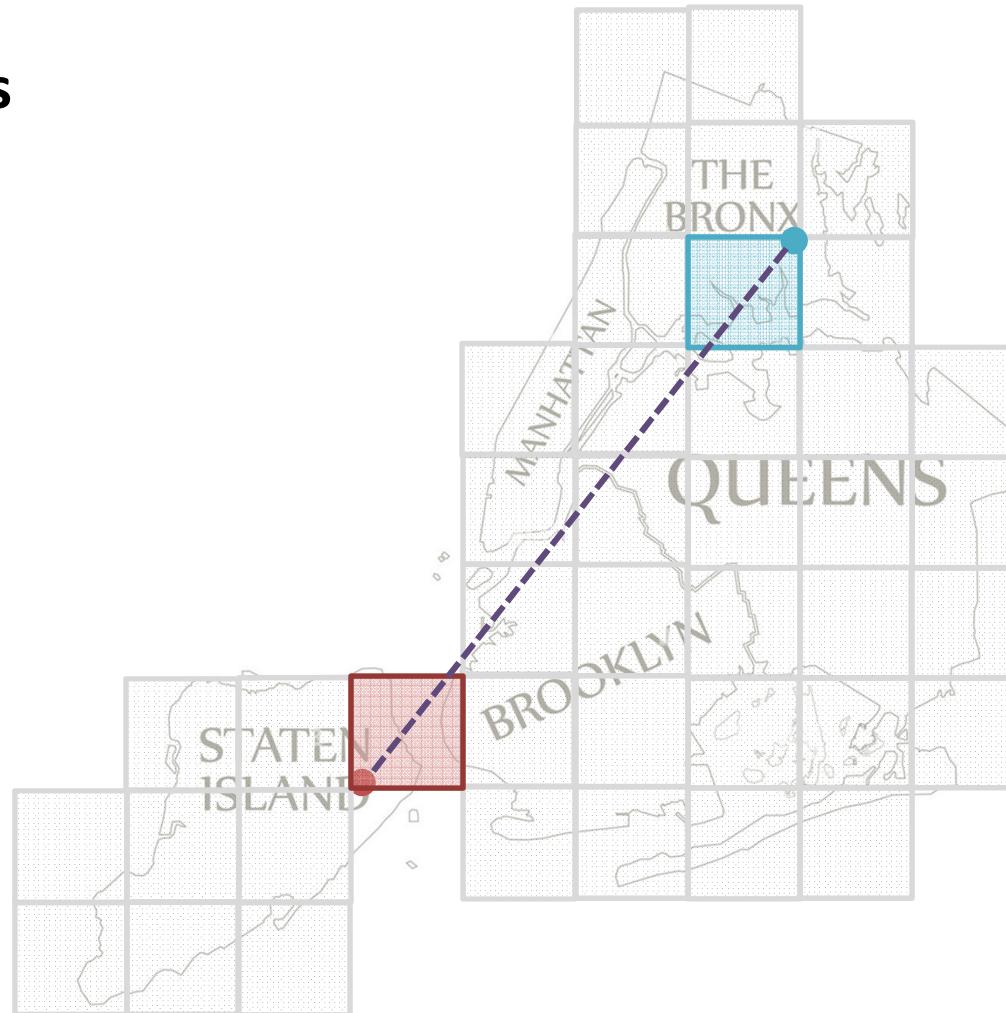


Temporal features
(weather, hour, day, ...)



Classic method used to characterize a trip

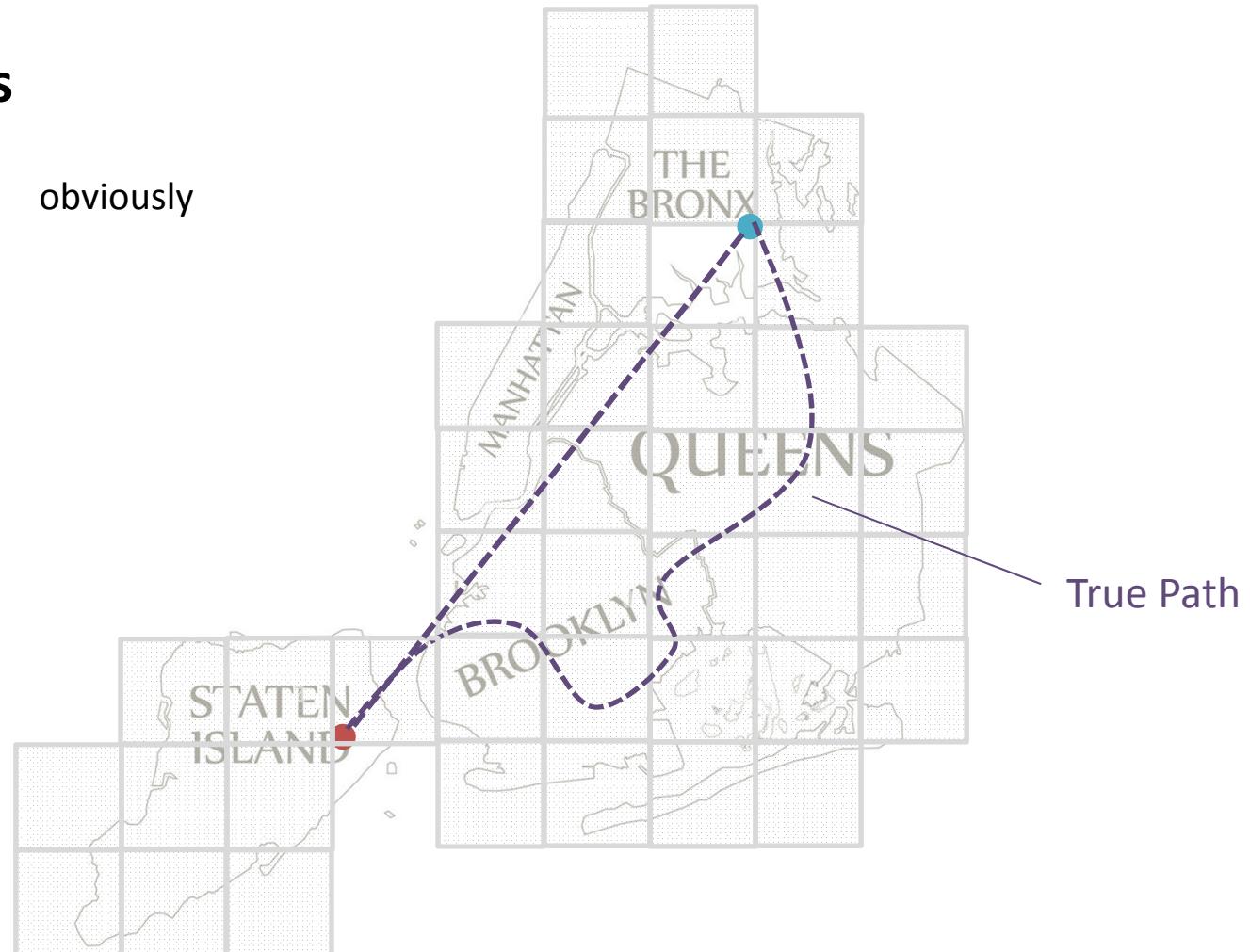
Problems



Classic method used to characterize a trip

Problems

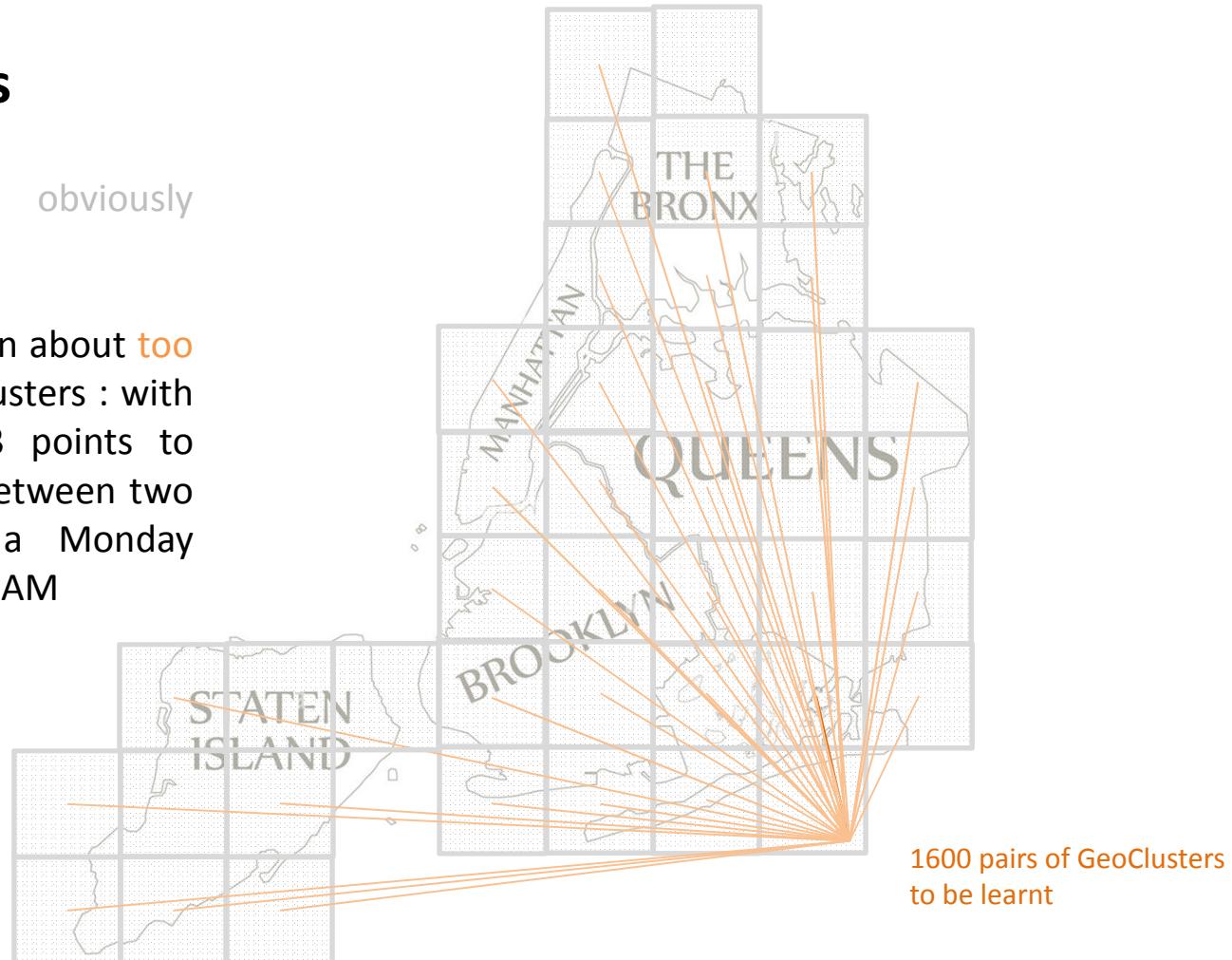
- Direct distance obviously inaccurate



Classic method used to characterize a trip

Problems

- Direct distance obviously inaccurate
- Too little data to learn about too much pairs of geo clusters : with 40 geo clusters, ~3 points to learn trip duration between two random clusters for a Monday between 8 AM and 9 AM



Classic method used to characterize a trip

Problems

- Direct distance obviously inaccurate
- Too little data to learn about too much pairs of geo clusters : with 40 geo clusters, ~3 points to learn trip duration between two random clusters for a Monday between 8 AM and 9 AM
- Close starting points and orientation can lead to very different trips durations



Classic method used to characterize a trip

Problems

- Direct distance obviously inaccurate
- Too little data to learn about too much pairs of geo clusters : with 40 geo clusters, ~3 points to learn trip duration between two random clusters for a Monday between 8 AM and 9 AM
- Close starting points and orientation can lead to very different trips durations



Classic method used to characterize a trip

Problems

- Direct distance obviously inaccurate
- Too little data to learn about too much pairs of geo clusters : with 40 geo clusters, ~3 points to learn trip duration between two random clusters for a Monday between 8 AM and 9 AM
- Close starting points and orientation can lead to very different trips durations

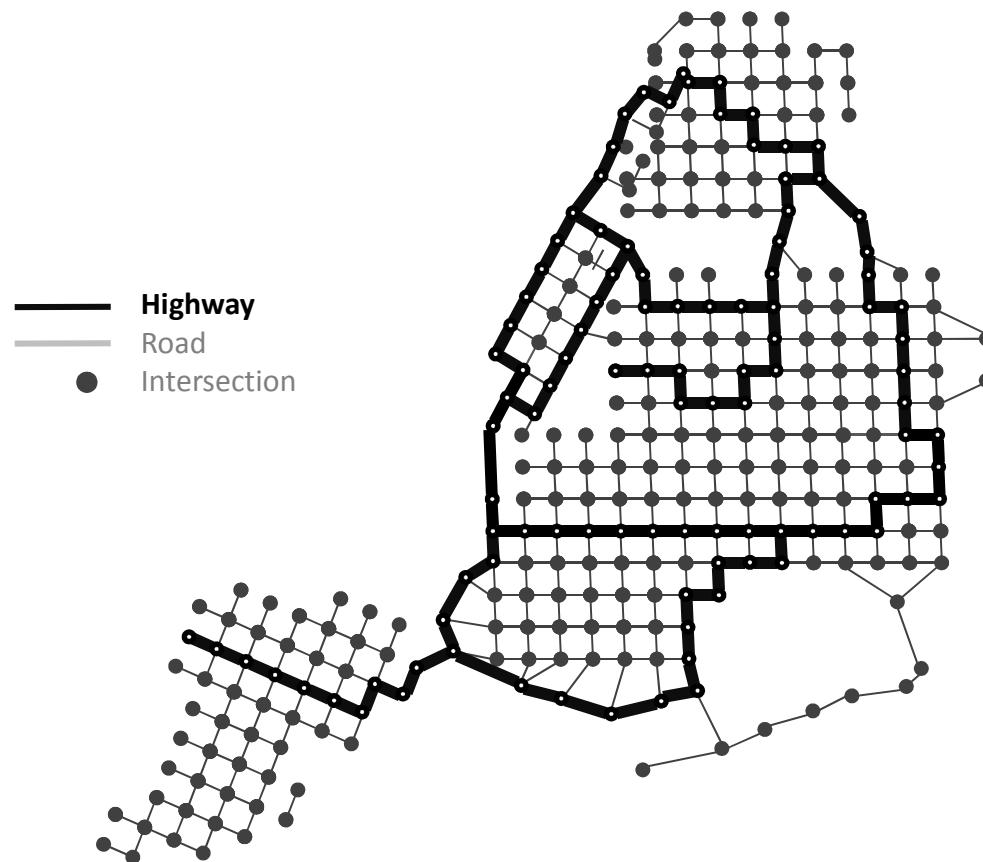


Close pairs
Trip duration x2.5

The Graph Way



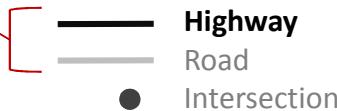
The Graph Way



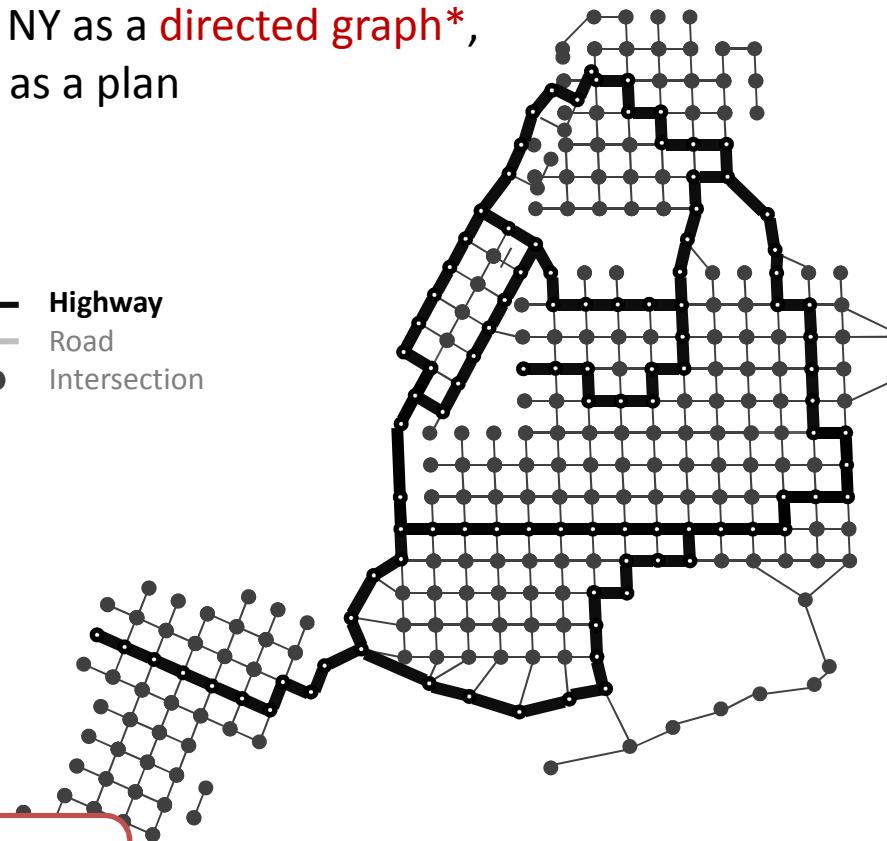
The Graph Way

See NY as a **directed graph***,
Not as a plan

Edge*



Node



* Weight of an edge :
Appr.speed based on heuristics

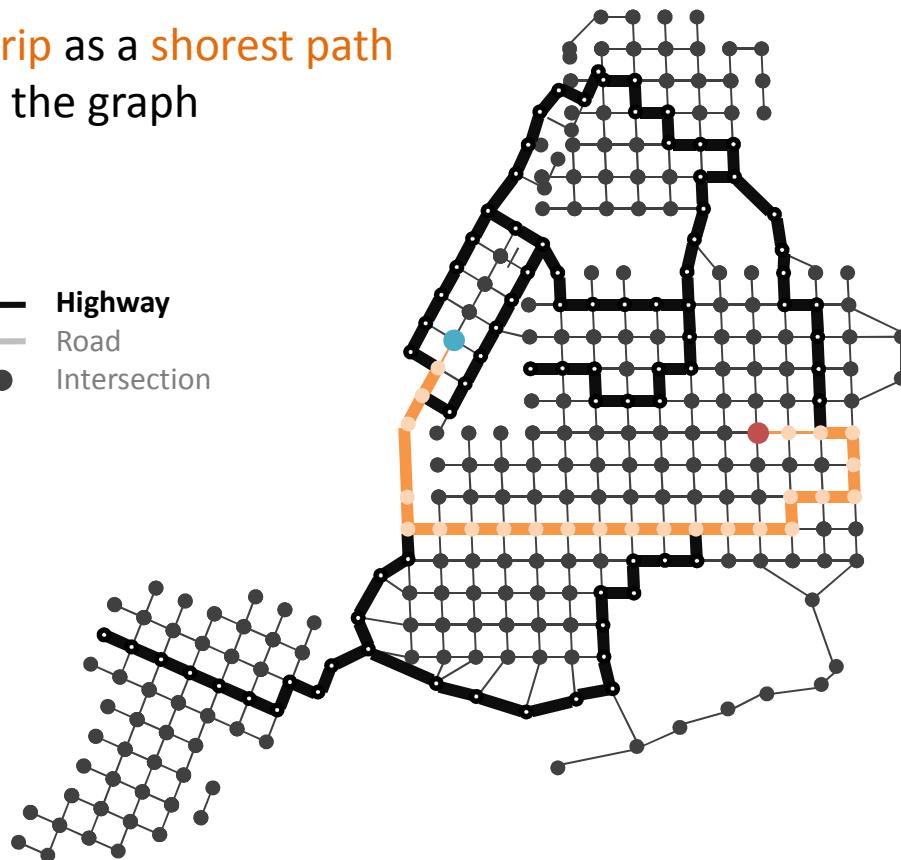
* New York roadnetwork is available
at osmnx.readthedocs.io

- 40,000 nodes
- 130,000 edges

The Graph Way

See a **trip** as a **shortest path**
Within the graph

- Highway
- Road
- Intersection



The Graph Way

See a **trip** as a **shortest path**
Within the graph, which is a sequence of **edges**



The Graph Way

See a **trip** as a **shortest path**
Within the graph, which is a sequence of **edges**



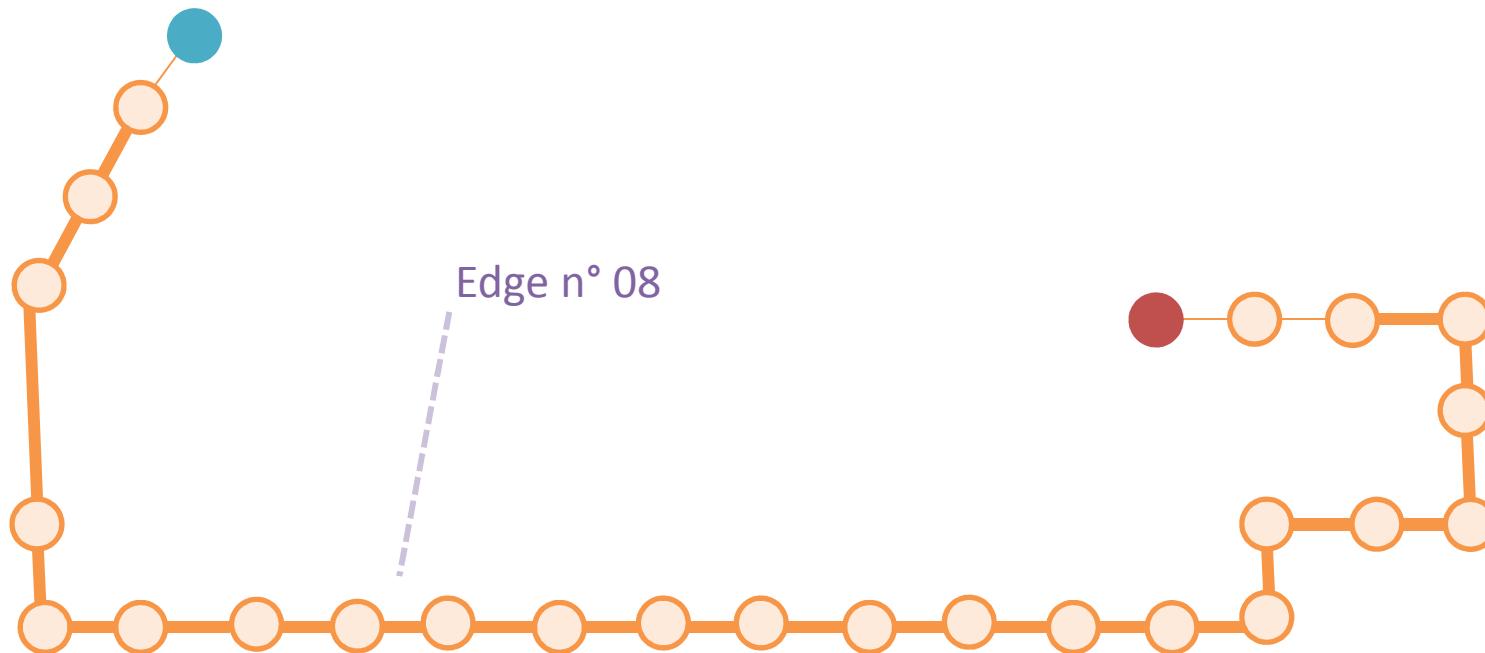
The Graph Way

See a **trip** as a **shortest path**
Within the graph, which is a sequence of **edges**



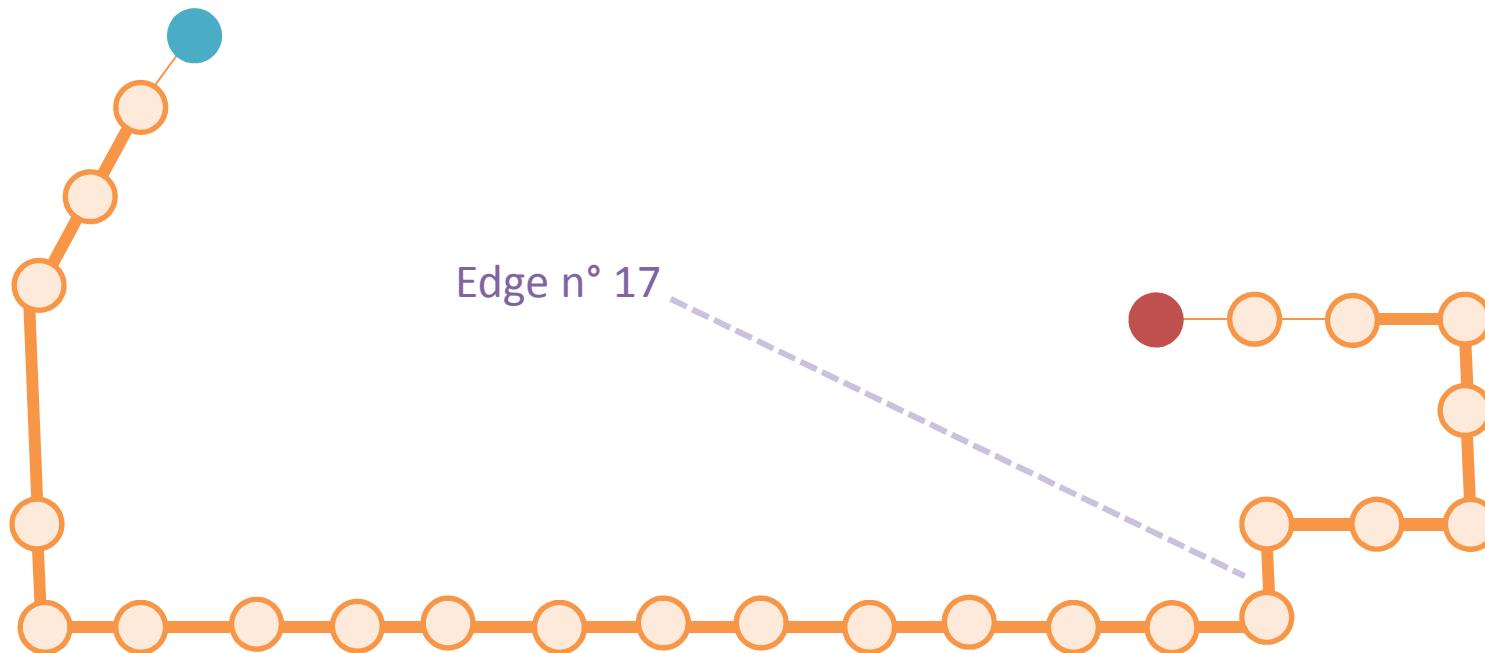
The Graph Way

See a **trip** as a **shortest path**
Within the graph, which is a sequence of **edges**



The Graph Way

See a **trip** as a **shortest path**
Within the graph, which is a sequence of **edges**



The Graph Way

New Goal :



The Graph Way

New Goal :

- ## 1. Predict edge's duration



The Graph Way

New Goal :

1. Predict **edge**'s duration
2. Sum up **shortest path edges**'s duration to predict trip's duration



How to characterize an edge

to predict its duration

1. Road features
2. Socio-economic features
3. Temporal features
4. Graph-based features

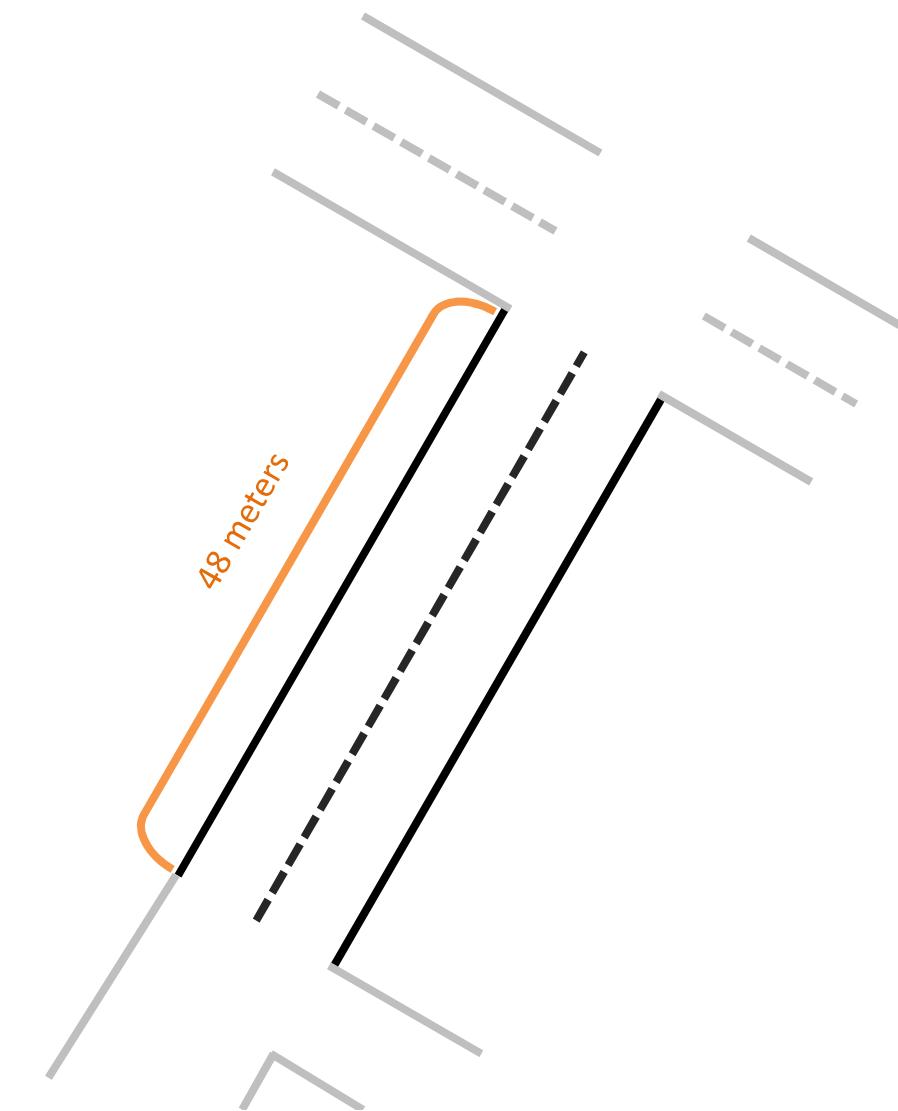


How to characterize an edge

to predict its duration

1. Road features

- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two way edge
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights

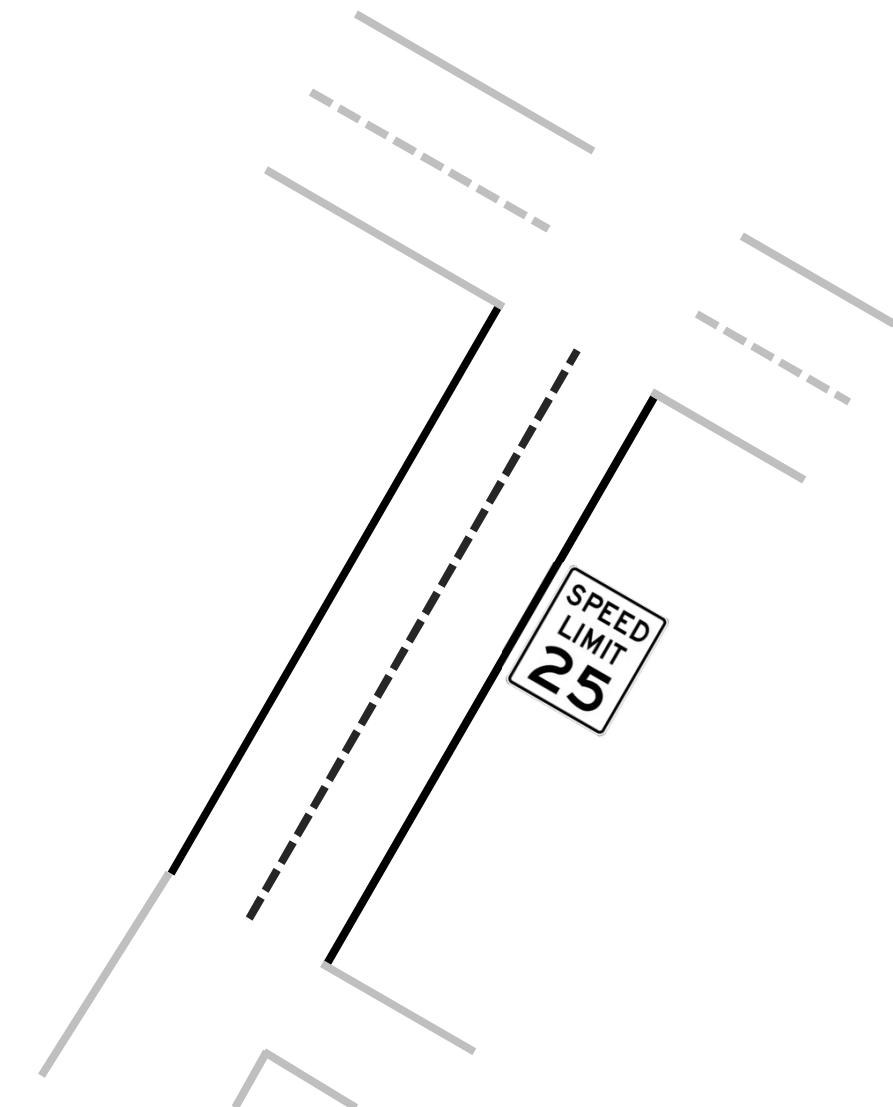


How to characterize an edge

to predict its duration

1. Road features

- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two way edge
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights

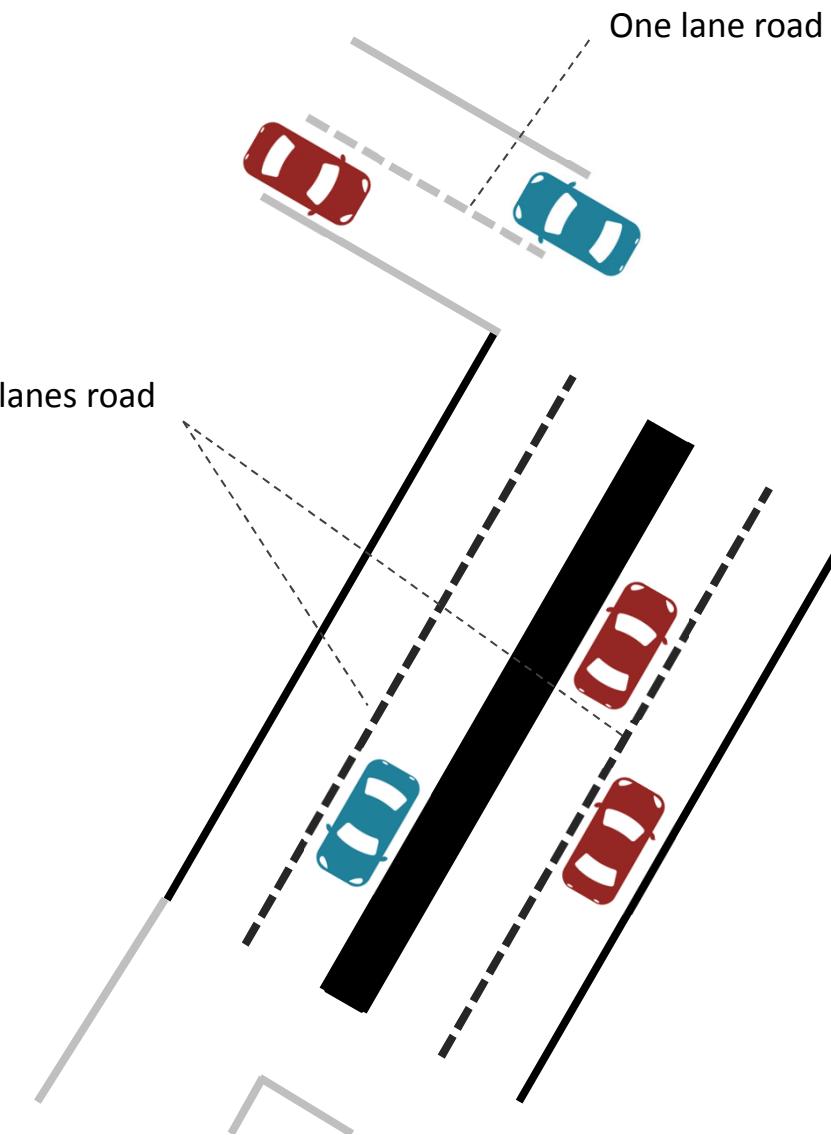


How to characterize an edge

to predict its duration

1. Road features

- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two way edge
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights

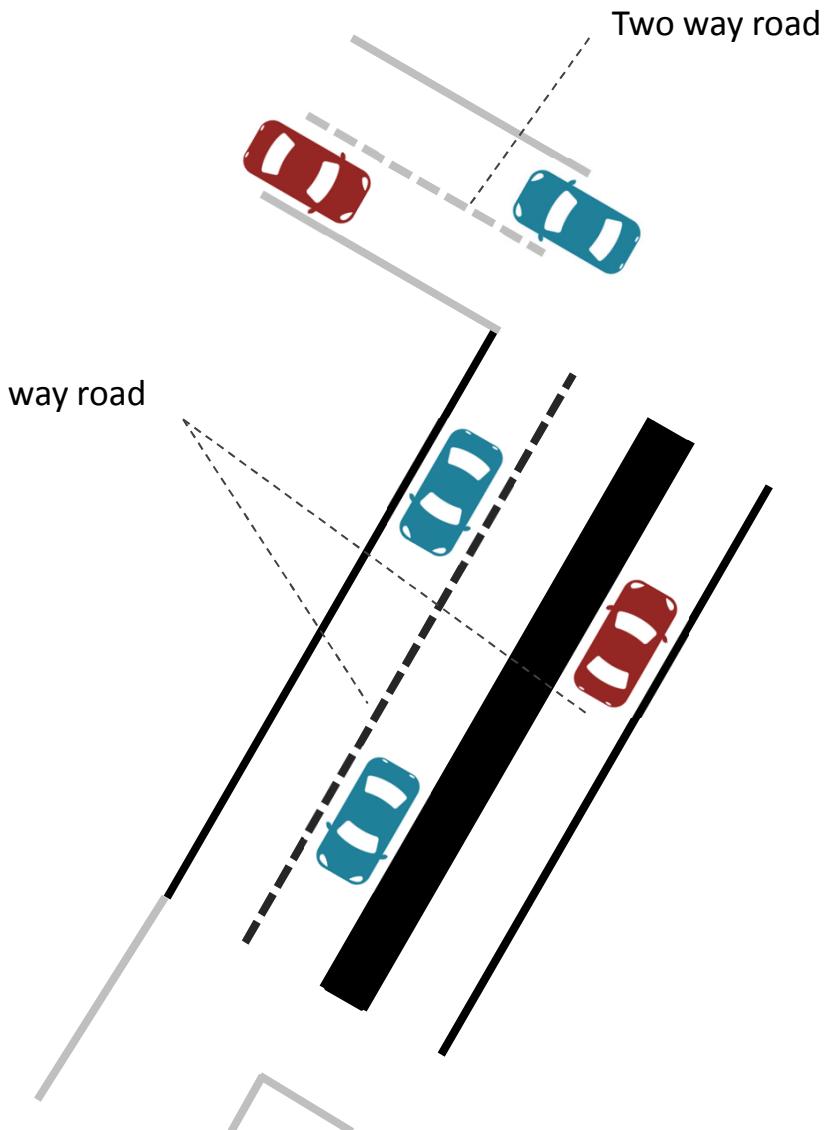


How to characterize an edge

to predict its duration

1. Road features

- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two-way
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights

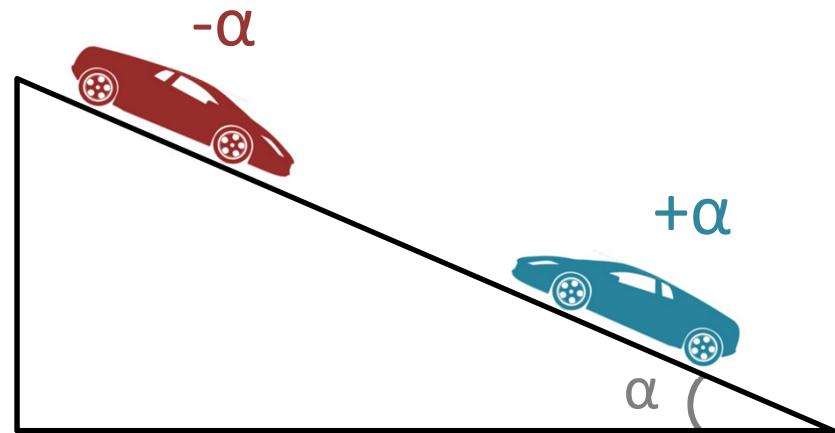


How to characterize an edge

to predict its duration

1. Road features

- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two way
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights

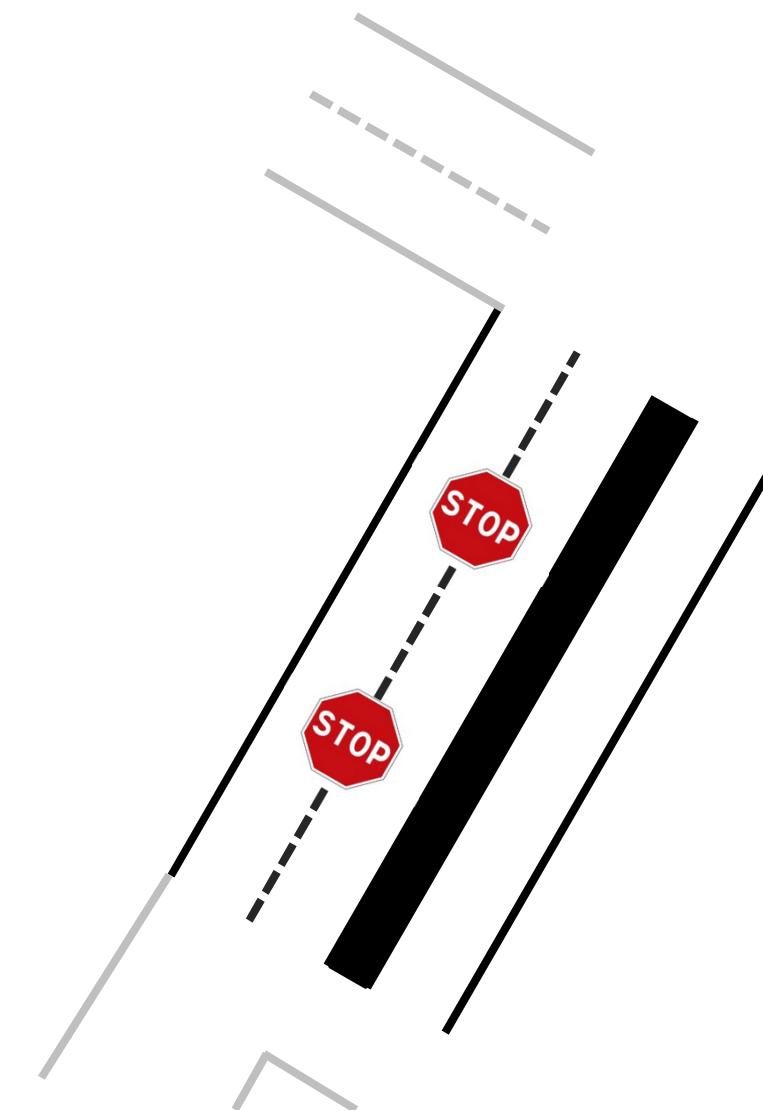


How to characterize an edge

to predict its duration

1. Road features

- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two way edge
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights

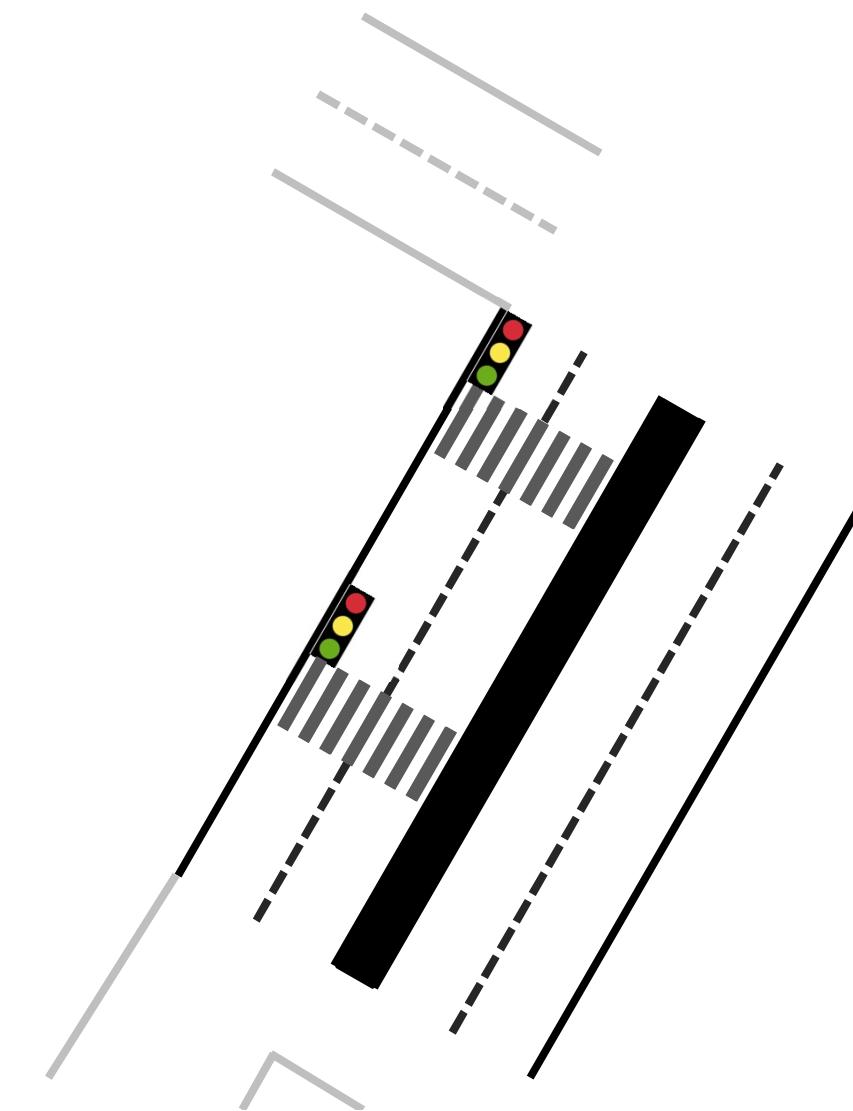


How to characterize an edge

to predict its duration

1. Road features

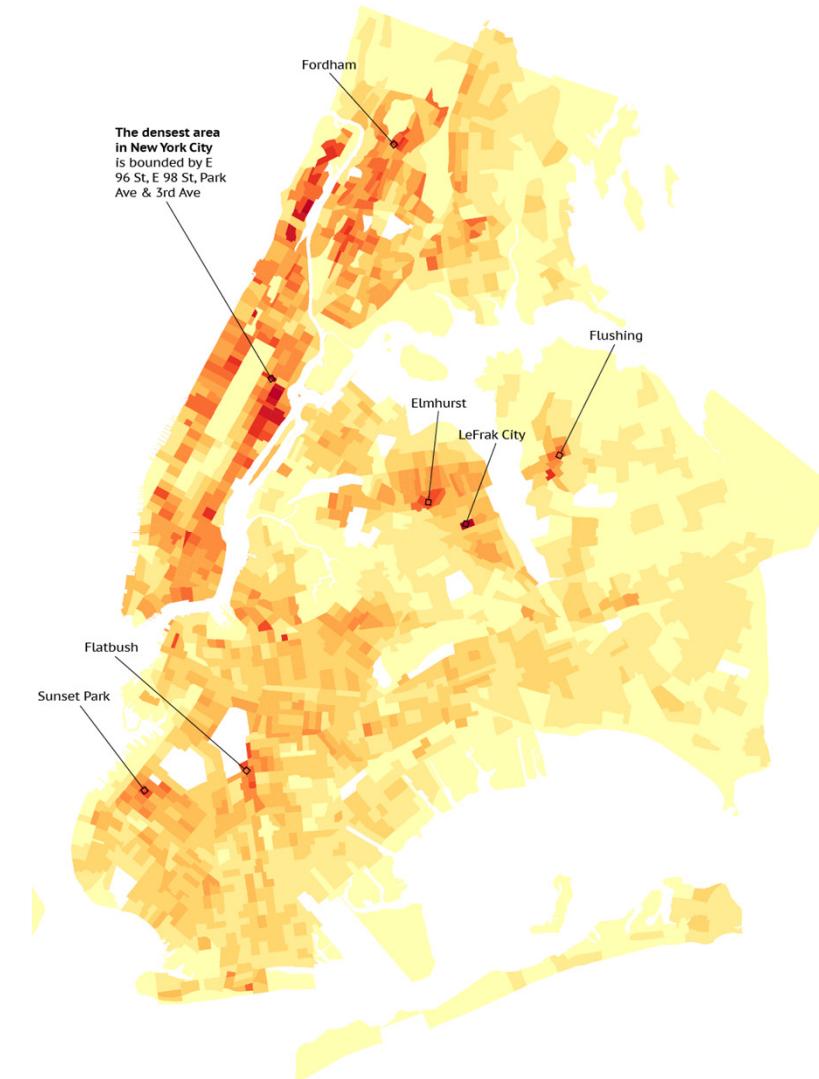
- a. Length
- b. Speed limit
- c. Number of lanes
- d. Two way edge
- e. Elevation gain
- f. Number of stops
- g. Number of traffic lights



How to characterize an edge to predict its duration

1. Socio-economic features

- a. Population density
- b. Landuse
- c. Family composition
- d. Household median income



How to characterize an edge to predict its duration

1. Socio-economic features

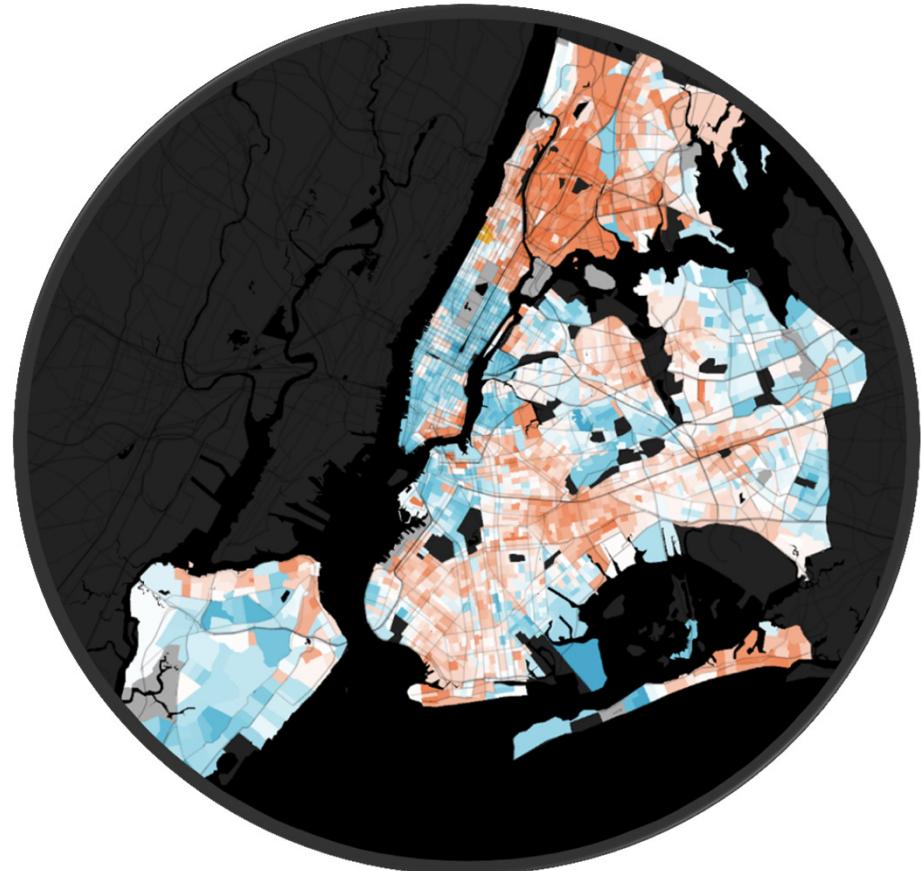
- a. Population density
- b. Landuse
- c. Family composition
- d. Household median income



How to characterize an edge to predict its duration

1. Socio-economic features

- a. Population density
- b. Landuse
- c. Family composition
- d. Household median income



How to characterize an edge

to predict its duration

1. Socio-economic features

- a. Population density
- b. Landuse
- c. Family composition
- d. Household median income



How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

Feature encoding is a
Non-trivial problem!

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
 - b. Day of the week
 - c. Weather
 - d. Luminosity
 - e. Day of the week
 - f. Working day / Week-end / Holiday
 - g. N. days to switch
 - h. N. days from switch
- 86350 s | 23:59:10
00007 s | 00:00:07

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

86350 s | 23:59:10
00007 s | 00:00:07

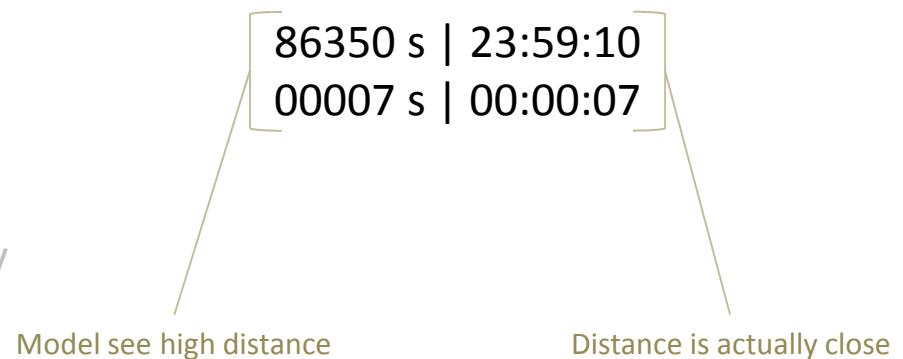
Model see high distance

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
 - b. Day of the week
 - c. Weather
 - d. Luminosity
 - e. Day of the week
 - f. Working day / Week-end / Holiday
 - g. N. days to switch
 - h. N. days from switch
- 86350 s | 23:59:10
00007 s | 00:00:07

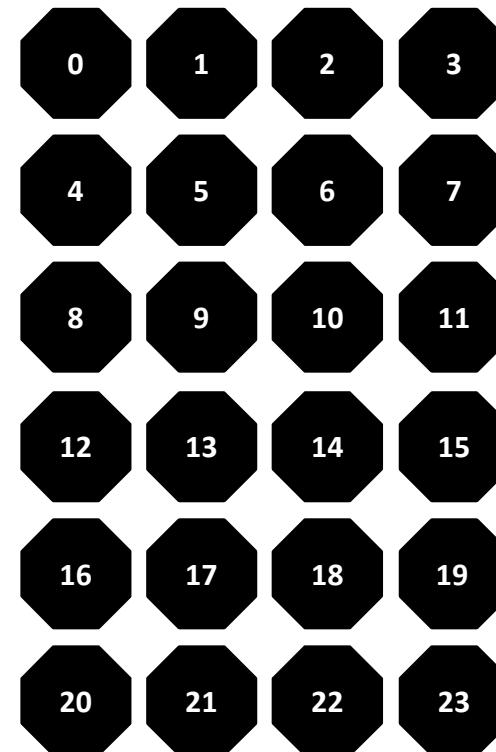
How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

Time as a class

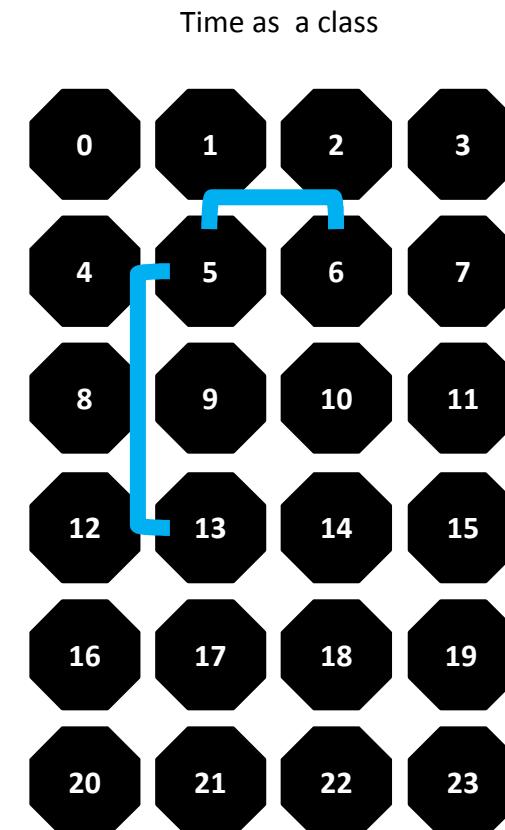


How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



Class 5 AM is **as different** from
Class 6 AM as from Class 1 PM

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

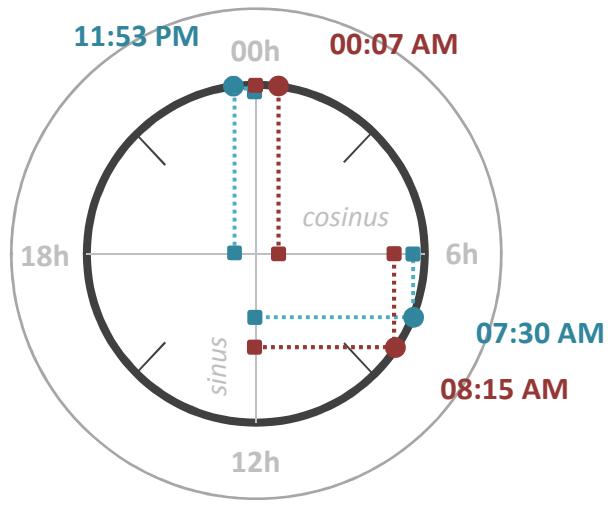


How to characterize an edge to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

Express starting time of the edge
as coordinates on a clock



Closed times leads to
close coordinates

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

But ...

We know first edge
starting time only



How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

But ...

We know first edge
starting time only



To access 3rd edge starting time
We need to predict previous
two edges duration first

How to characterize an edge

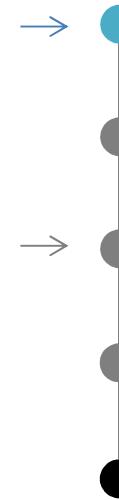
to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

But ...

We know first edge
starting time only



To access 3rd edge starting time
We need to predict previous
two edges duration first



We need to **predict edges's duration sequentially** like we'd do for a RNN

How to characterize an edge

to predict its duration

1. Temporal features

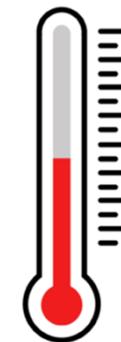
- a. Starting time
 - b. Day of the week
 - c. Weather
 - d. Luminosity
 - e. Day of the week
 - f. Working day / Week-end / Holiday
 - g. N. days to switch
 - h. N. days from switch
- Encode day of the week
Just as starting time
(coordinates on a circle)

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



Temperature



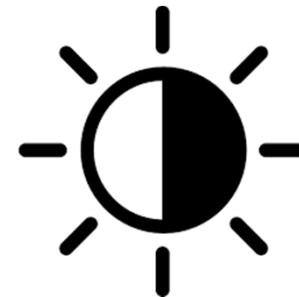
Precipitations

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



Float between 0 and 1
Ex: *twilight 0.35 / night 0 / noon 1*

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

JAN

M	T	W	T	F	S	S
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

Working day
 Week-end
 Holiday

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

JAN

M	T	W	T	F	S	S
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

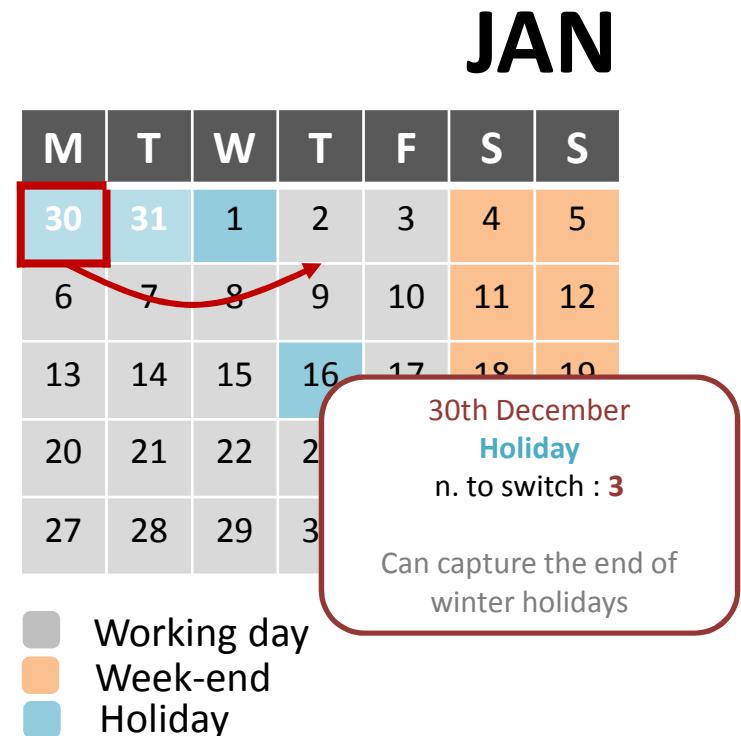
Working day
 Week-end
 Holiday

How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



How to characterize an edge

to predict its duration

1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the week
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch

JAN

M	T	W	T	F	S	S
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

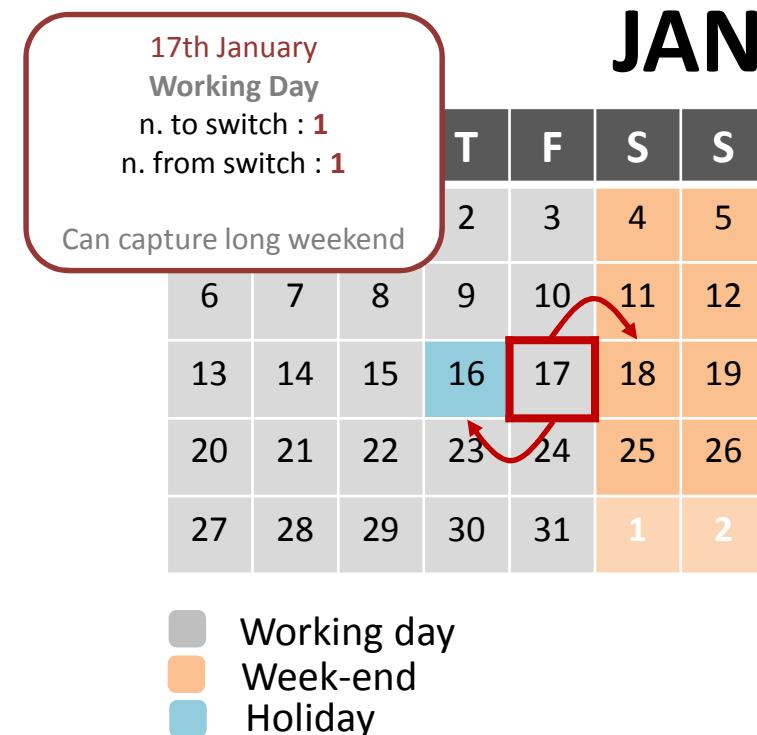
Working day
 Week-end
 Holiday

How to characterize an edge

to predict its duration

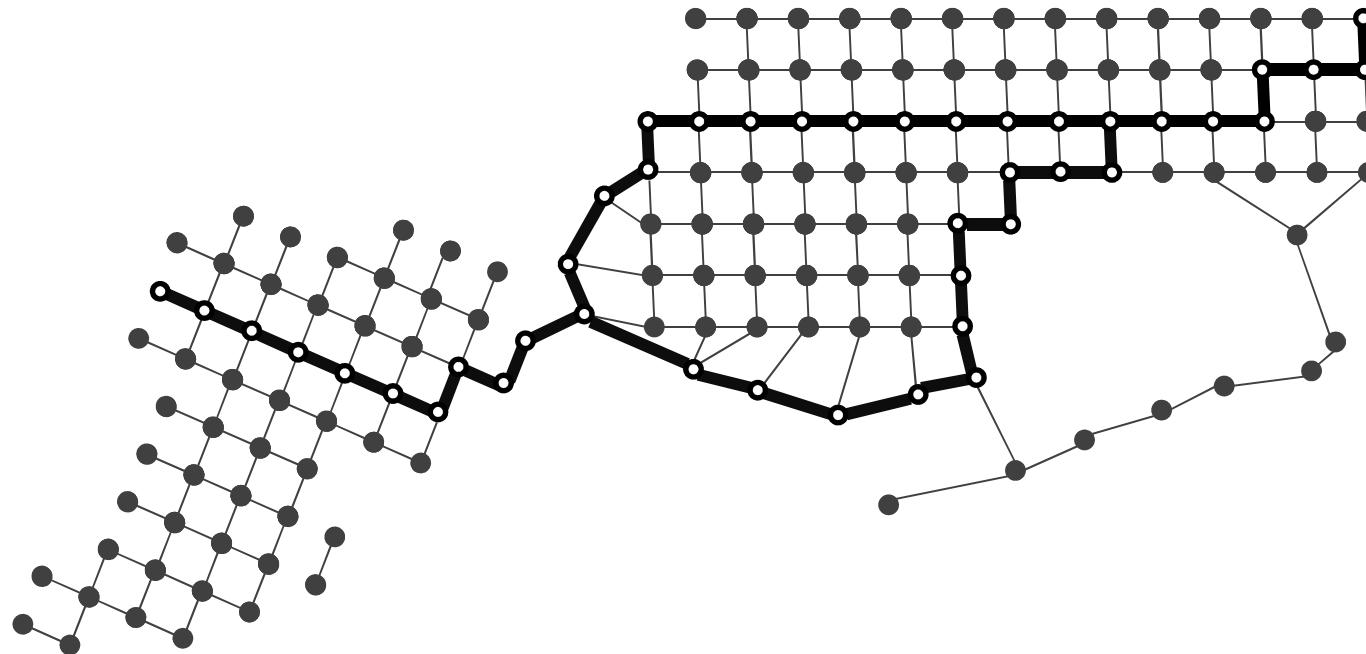
1. Temporal features

- a. Starting time
- b. Day of the week
- c. Weather
- d. Luminosity
- e. Day of the month
- f. Working day / Week-end / Holiday
- g. N. days to switch
- h. N. days from switch



How to characterize an edge to predict its duration (Graph Features)

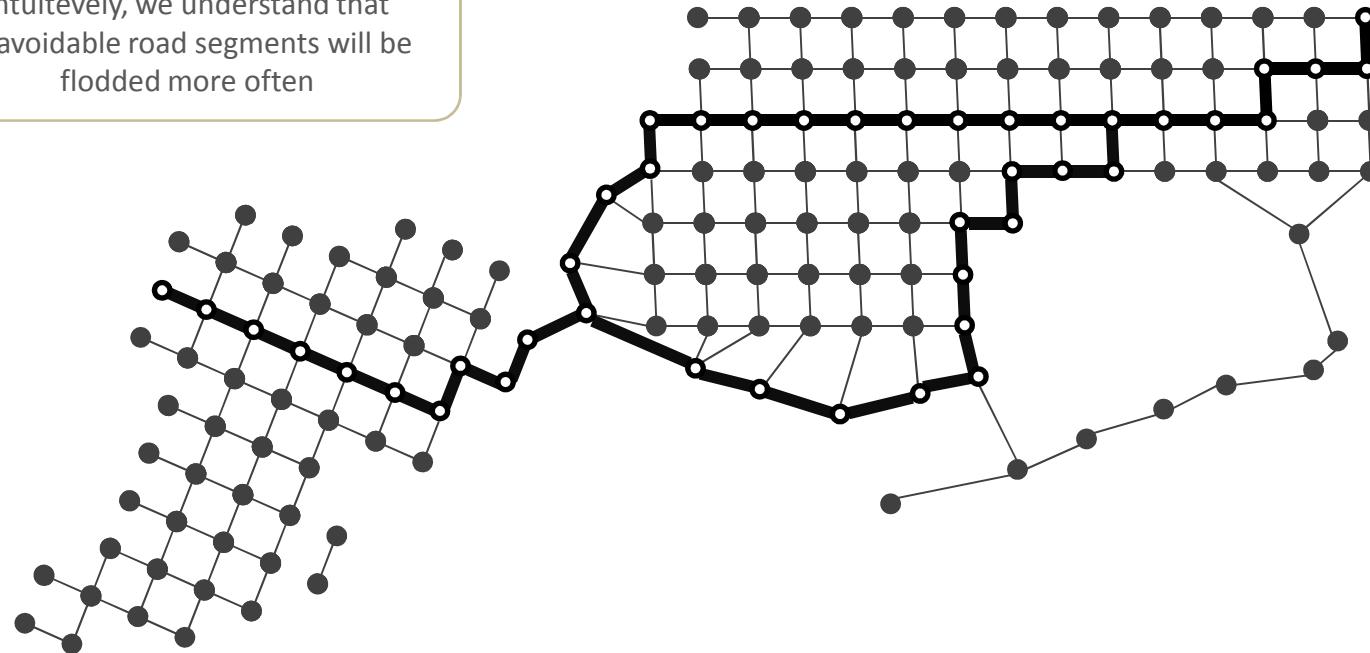
What a road network graph
can tell us about the traffic?



How to characterize an edge to predict its duration (Graph Features)

What a road network graph
can tell us about the traffic?

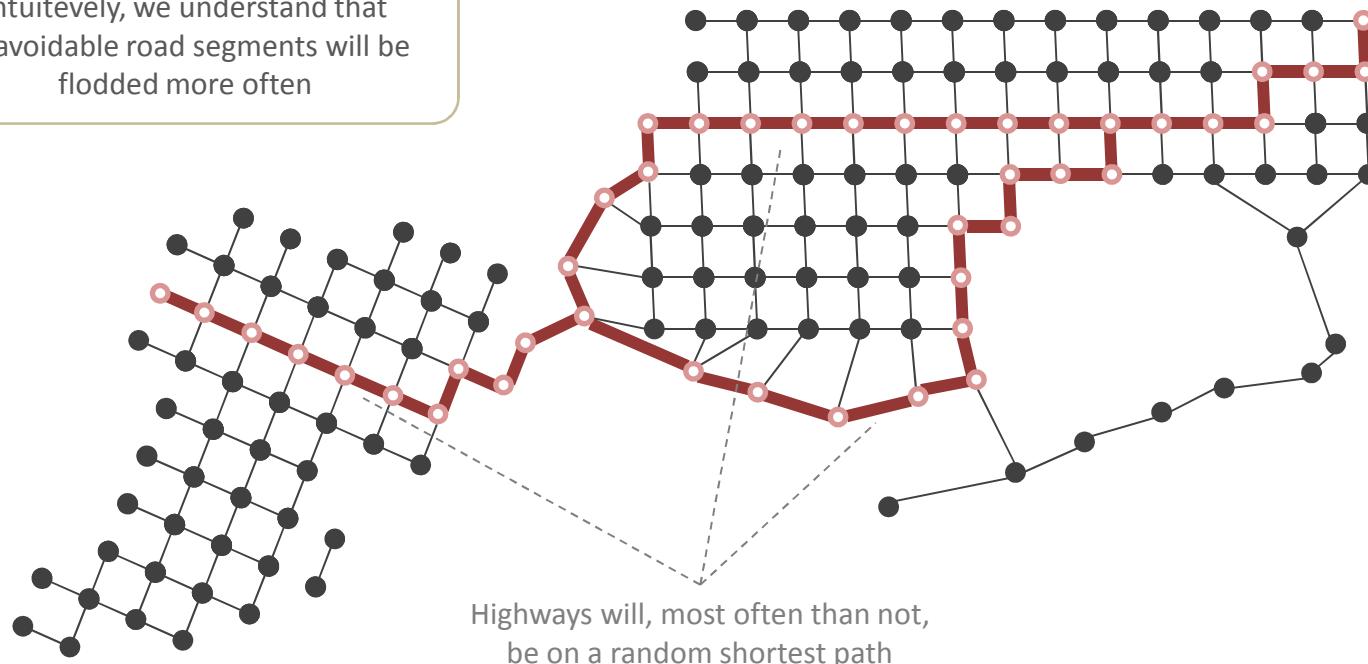
Intuitively, we understand that
unavoidable road segments will be
flooded more often



How to characterize an edge to predict its duration (Graph Features)

What a road network graph
can tell us about the traffic?

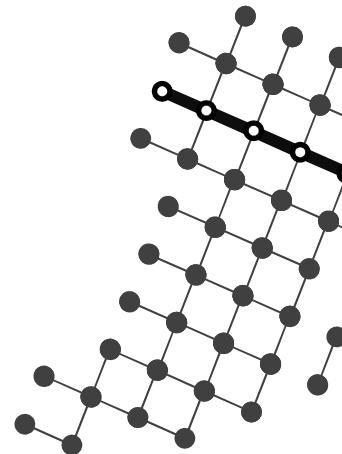
Intuitively, we understand that
unavoidable road segments will be
flooded more often



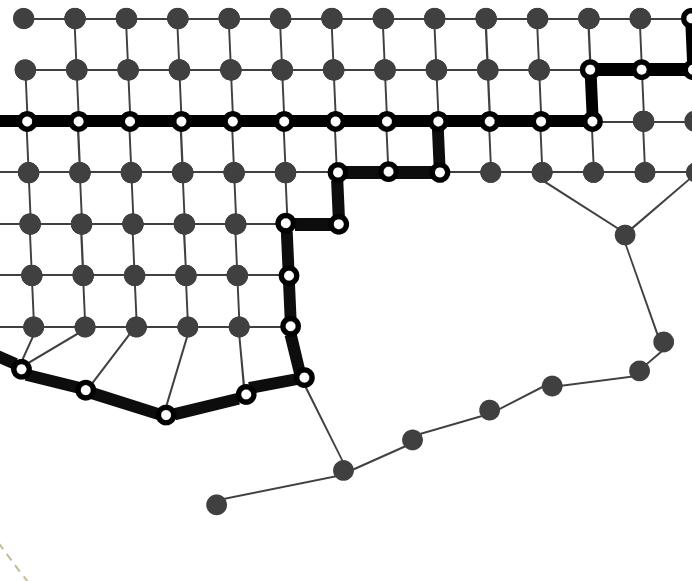
How to characterize an edge to predict its duration (Graph Features)

What a road network graph
can tell us about the traffic?

Intuitively, we understand that
unavoidable road segments will be
flooded more often

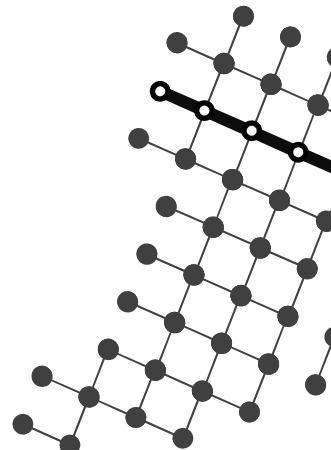


Bridge between two big components of a city is
an unavoidable road segment



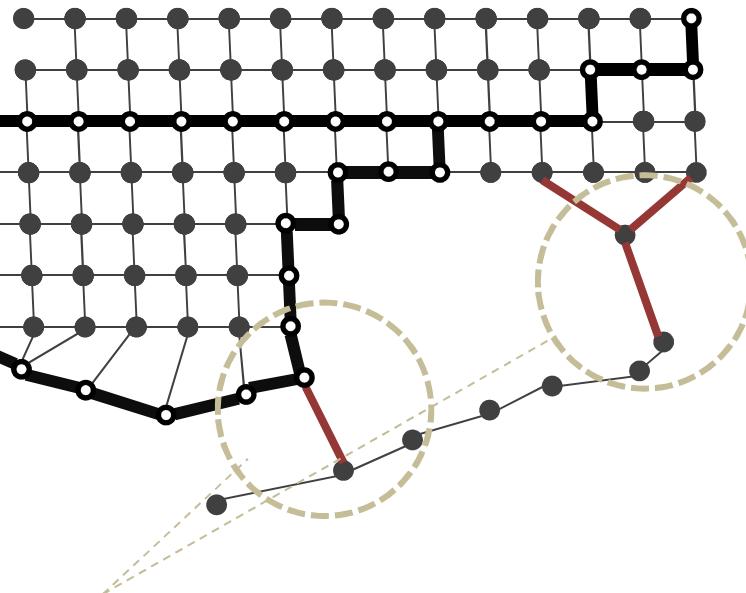
How to characterize an edge to predict its duration (Graph Features)

Intuitively, we understand that unavoidable road segments will be flooded more often



What a road network graph
can tell us about the traffic?

Those bridges are not unavoidable as it connects a small island to the rest of the city

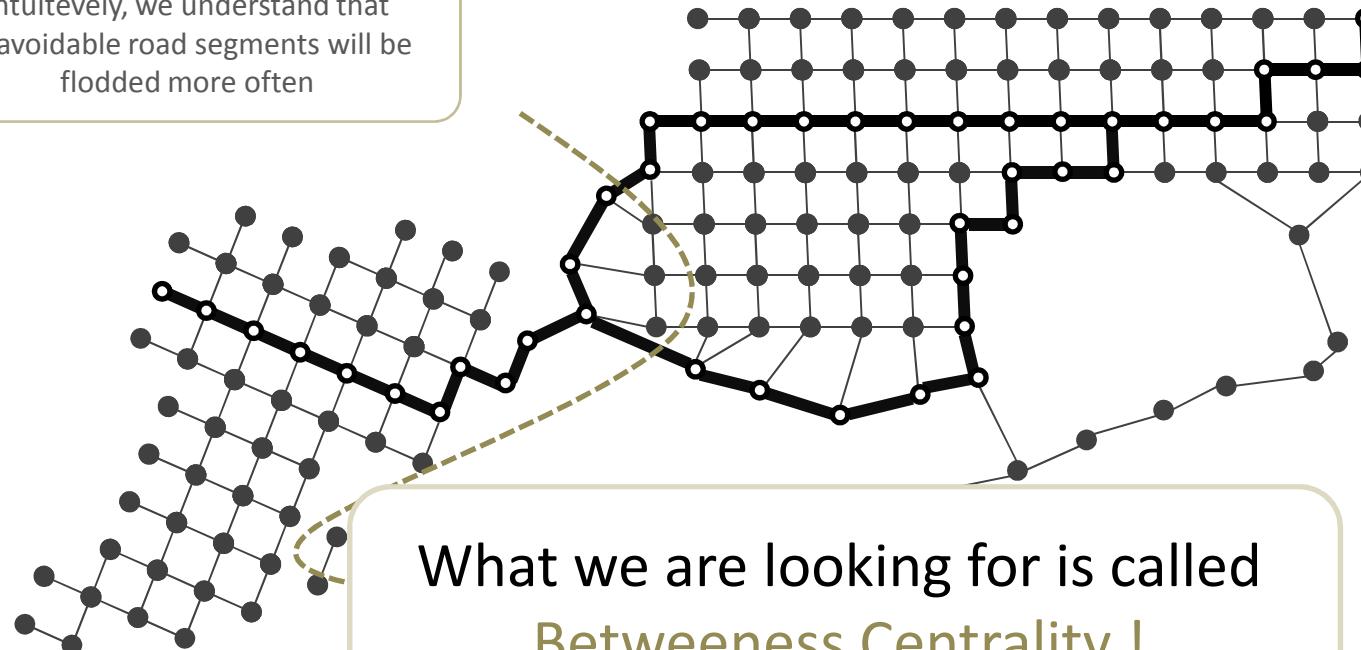


How to characterize an edge

to predict its duration (Graph Features)

What a road network graph
can tell us about the traffic?

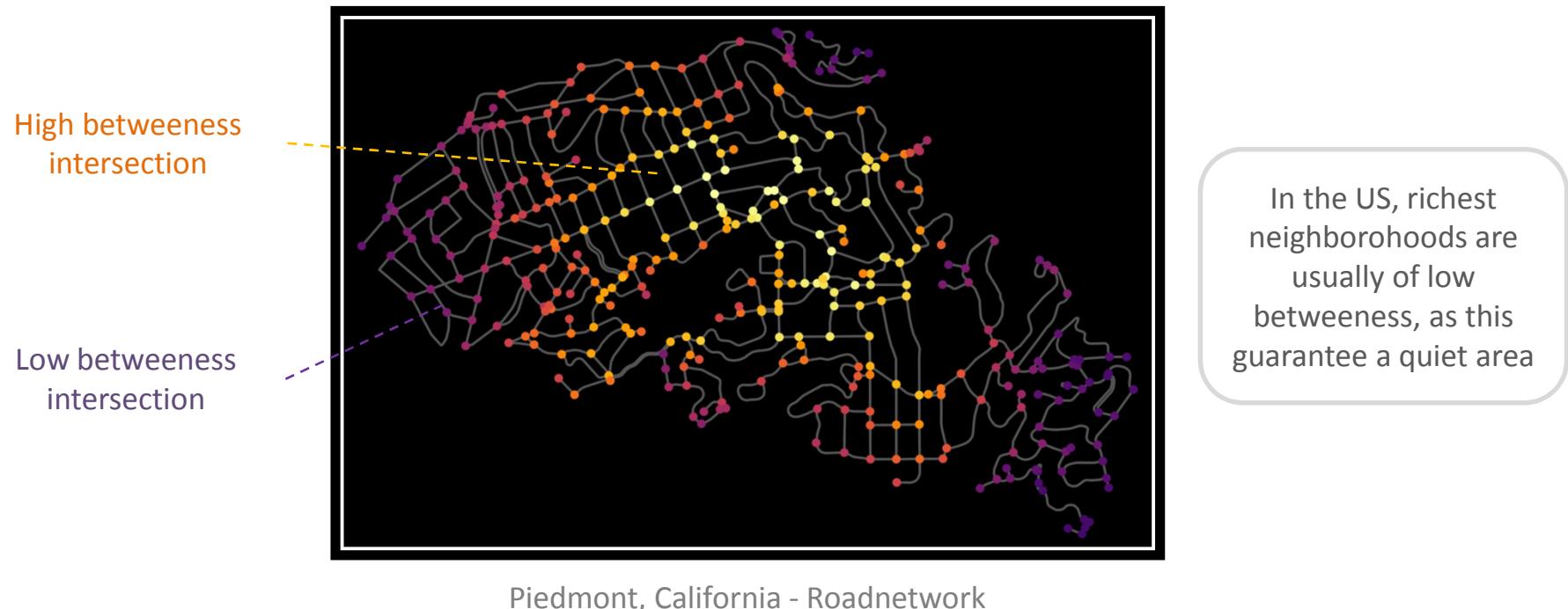
Intuitively, we understand that
unavoidable road segments will be
flooded more often



How to characterize an edge to predict its duration (Graph Features)

The **betweenness centrality** of a node* represent the frequency at which a random path goes through this specific node

*In this case, A node is an intersection between two segments roads



How to characterize an edge

to predict its duration (Graph Features)

New Feature to characterize an edge :

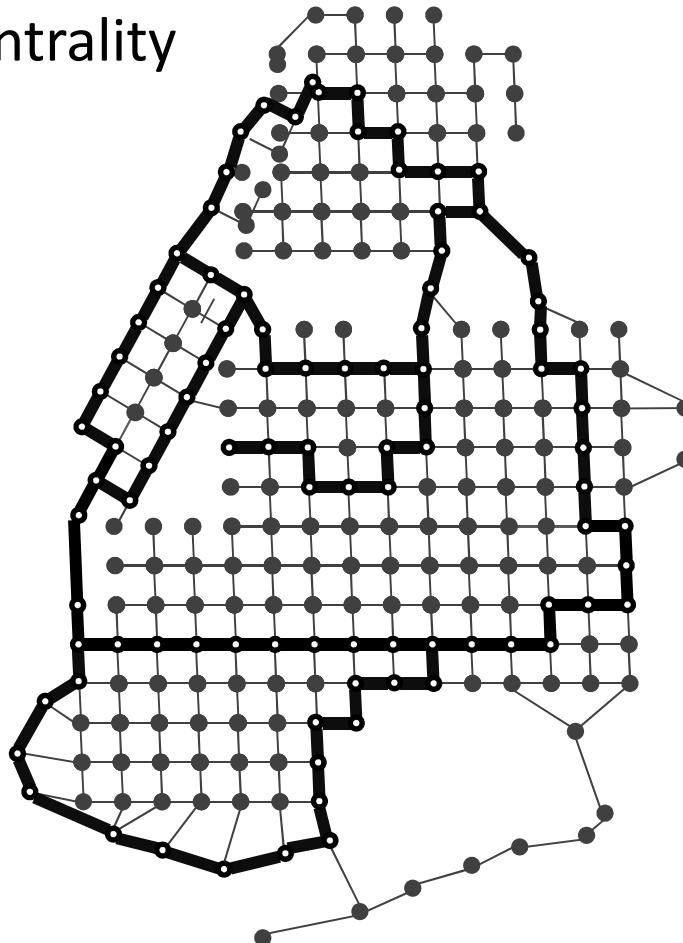
- Betweeness of
the graph's edge*

* Graph in which nodes are road segments,
edges are intersection between road

How to characterize an edge

to predict its duration (Graph Features)

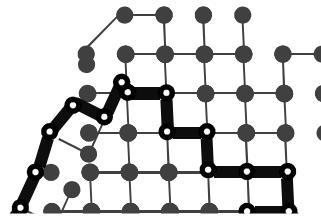
Is betweenness centrality
that good?



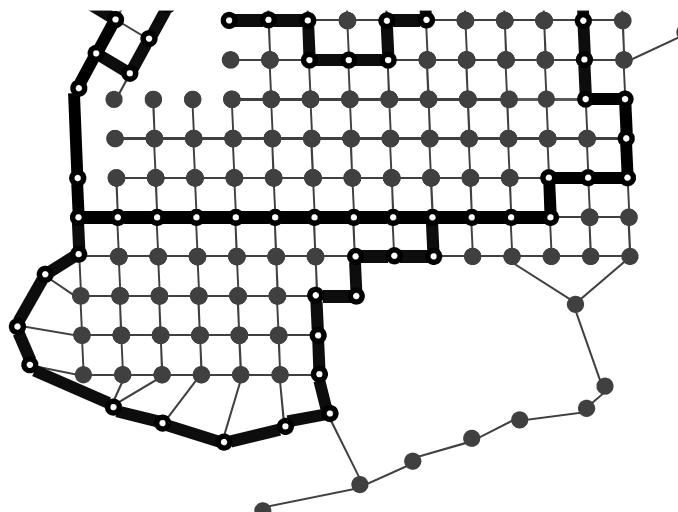
How to characterize an edge

to predict its duration (Graph Features)

Is betweenness centrality
that good?



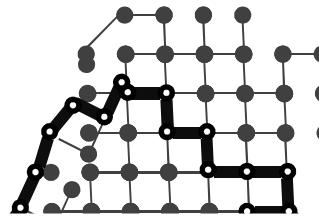
The betweenness centrality of a node* represent the frequency at which a random path goes through this specific node



How to characterize an edge

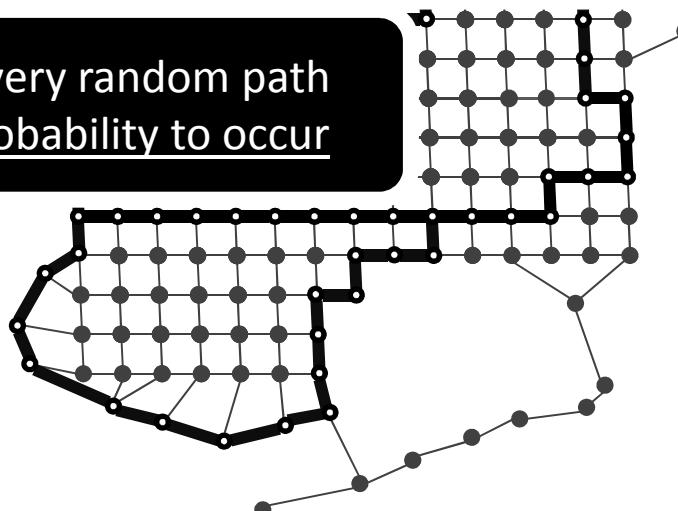
to predict its duration (Graph Features)

Is betweenness centrality
that good?

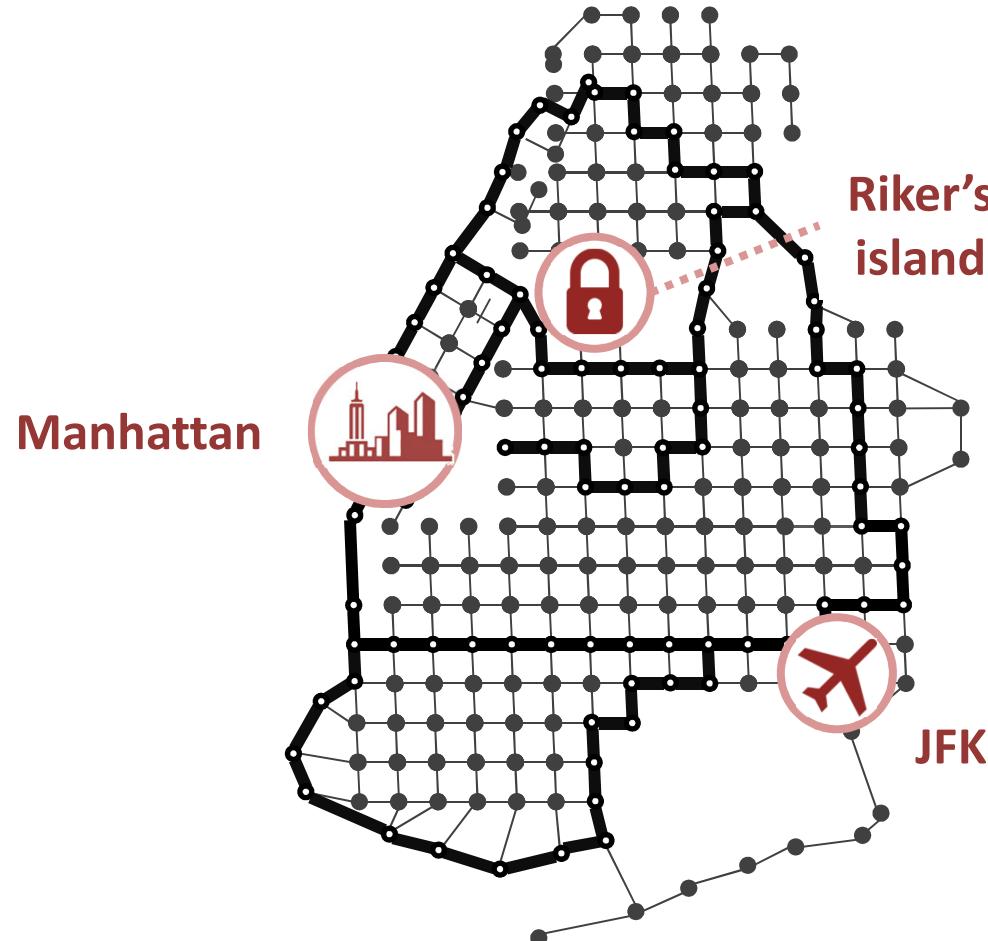


The betweenness centrality of a node* represent the frequency at which a random path goes through this specific node

It implies that every random path has the same probability to occur



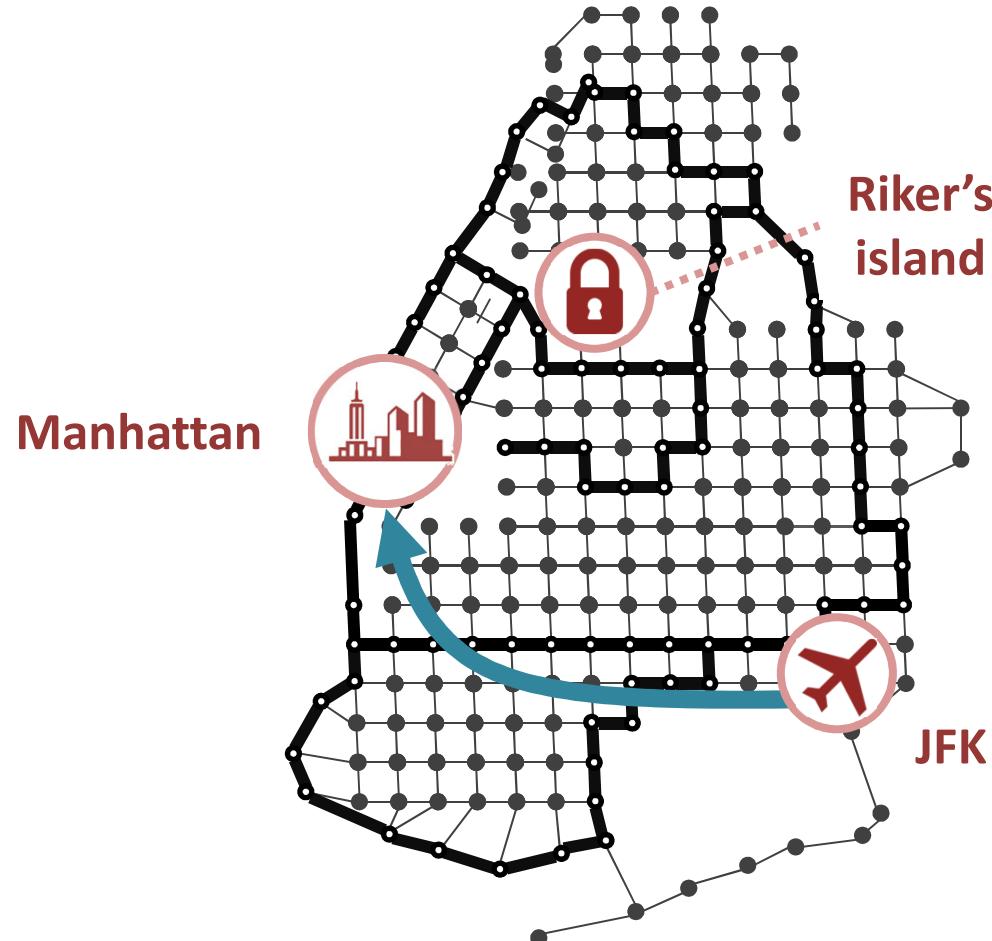
How to characterize an edge to predict its duration (Graph Features)



Exemple 1

Is betweeness centrality
that good?

How to characterize an edge to predict its duration (Graph Features)

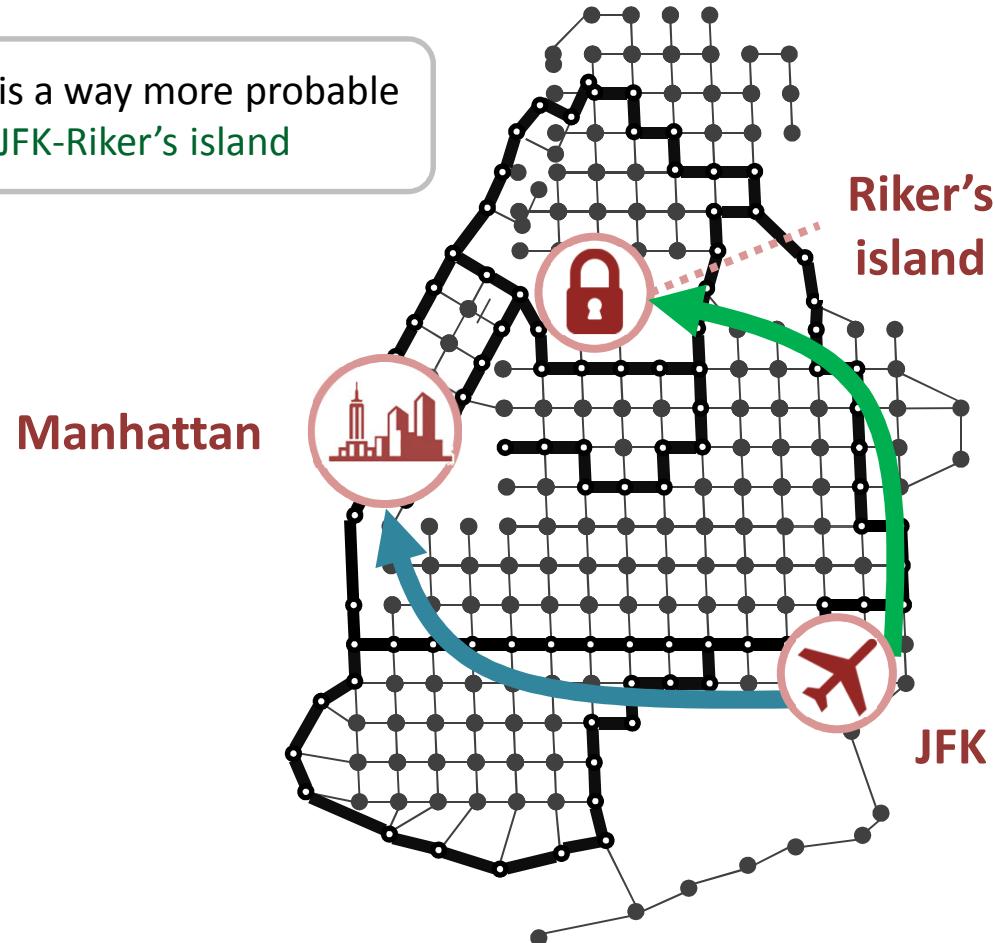


Exemple 1

Is betweenness centrality
that good?

How to characterize an edge to predict its duration (Graph Features)

JFK-Manhattan is a way more probable trip than JFK-Riker's island



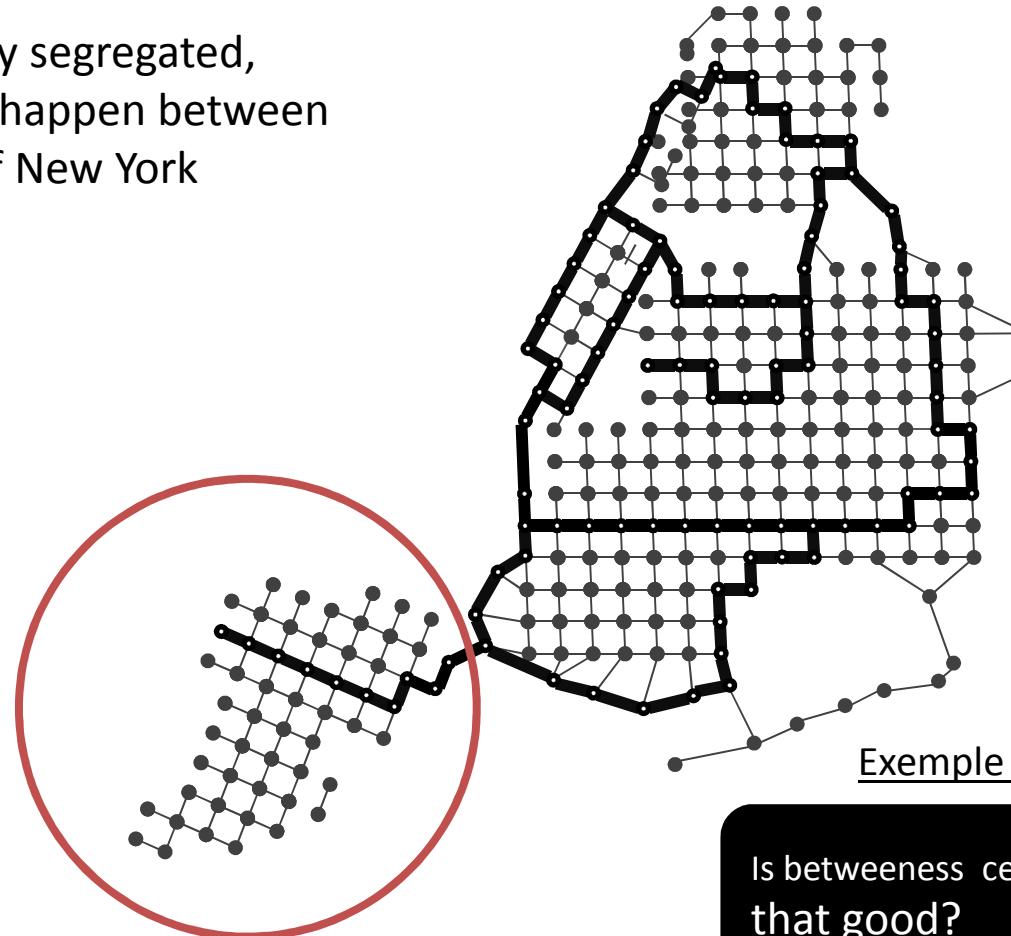
Exemple 1

Is betweenness centrality
that good?

How to characterize an edge

to predict its duration (Graph Features)

- ▲ Staten Island might be highly segregated, which means only few trips happen between Staten Island and the rest of New York

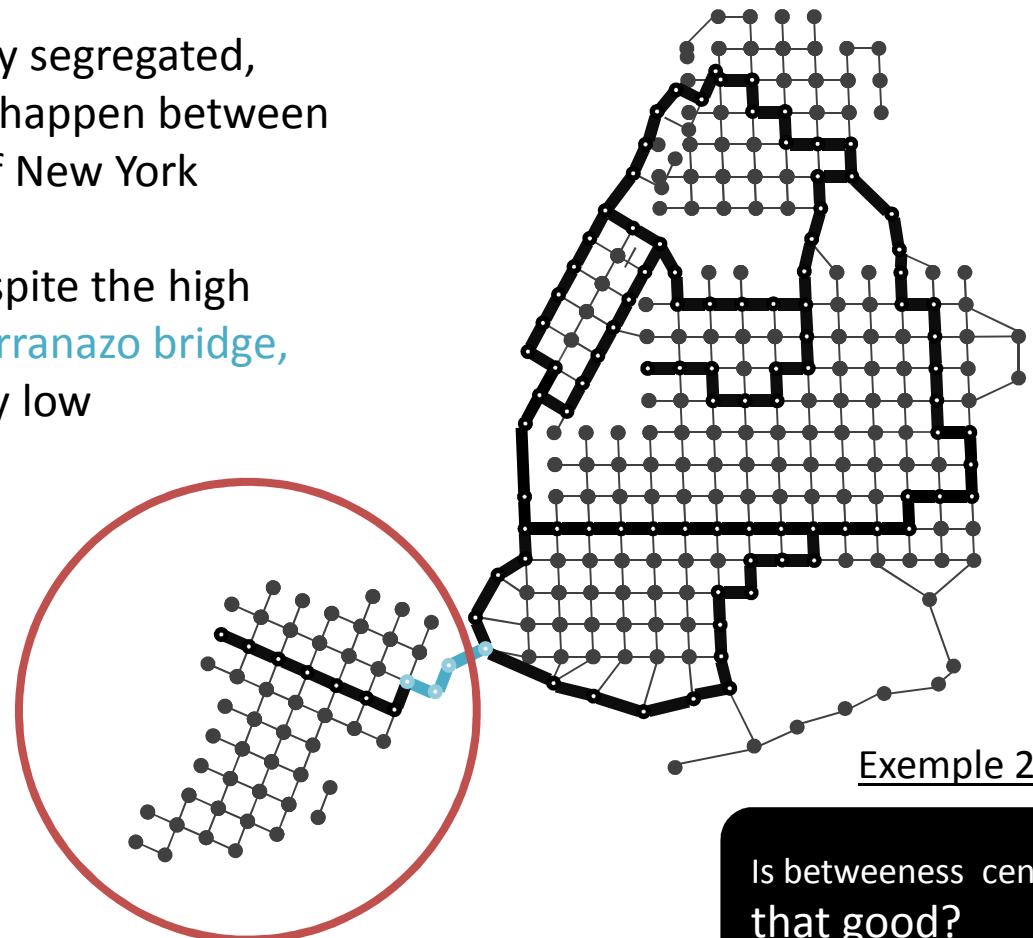


Is betweenness centrality
that good?

How to characterize an edge

to predict its duration (Graph Features)

- ▲ Staten Island might be highly segregated, which means only few trips happen between Staten Island and the rest of New York
- ▲ That would imply that despite the high betweenness centrality of Verrazano bridge, the traffic should be actually low on this road segment

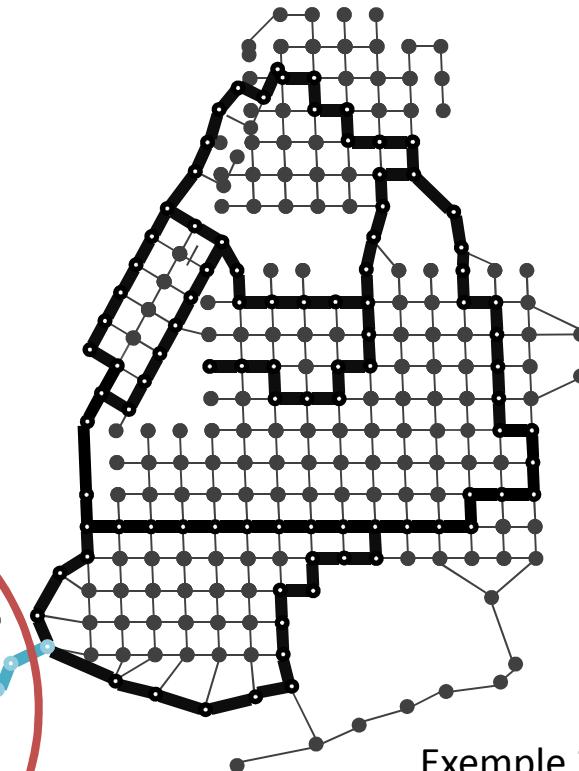
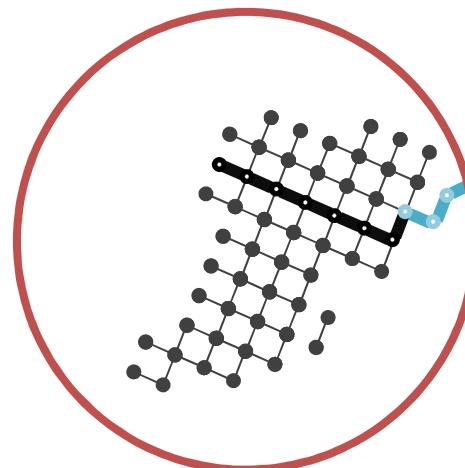


How to characterize an edge

to predict its duration (Graph Features)

- ▲ Staten Island might be highly segregated, which means only few trips happen between Staten Island and the rest of New York
- ▲ That would imply that despite the high betweenness centrality of Verrazano bridge, the traffic should be actually low on this road segment

How can we have this type of information?



Exemple 2

Is betweenness centrality that good?

How to characterize an edge

to predict its duration (Graph Features)

How can we get this information?

We do not know every trip's frequency! Do we?

We actually might have something somewhat correlated :

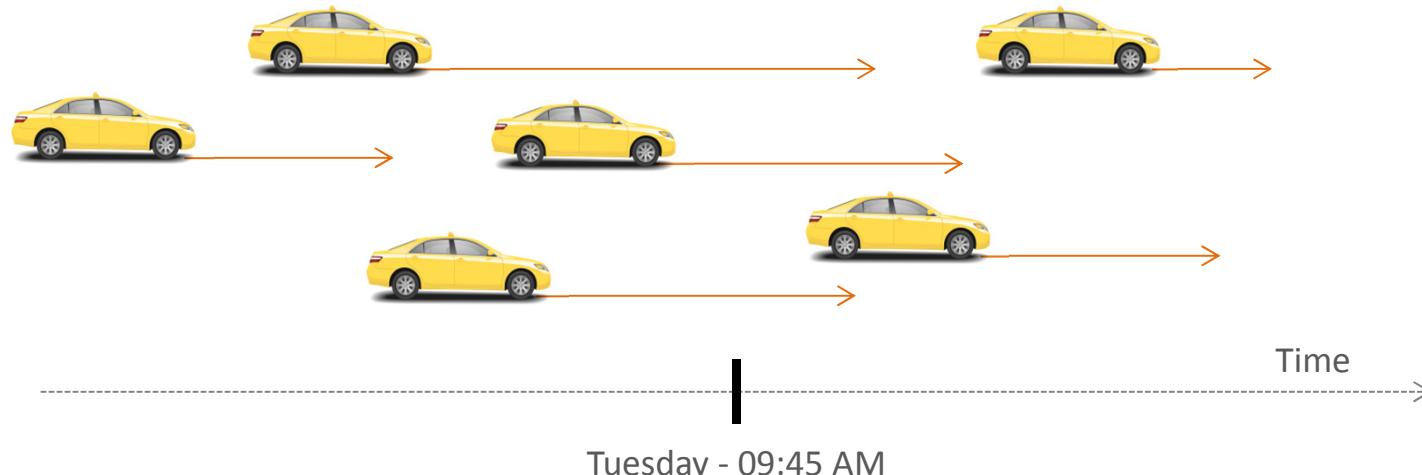
THE DATA

- 2,000,000 taxi trips
- Over a 6 months period
- ~10,000 trips by $[DayWeek_x_HourSlot]$

How to characterize an edge

to predict its duration (Graph Features)

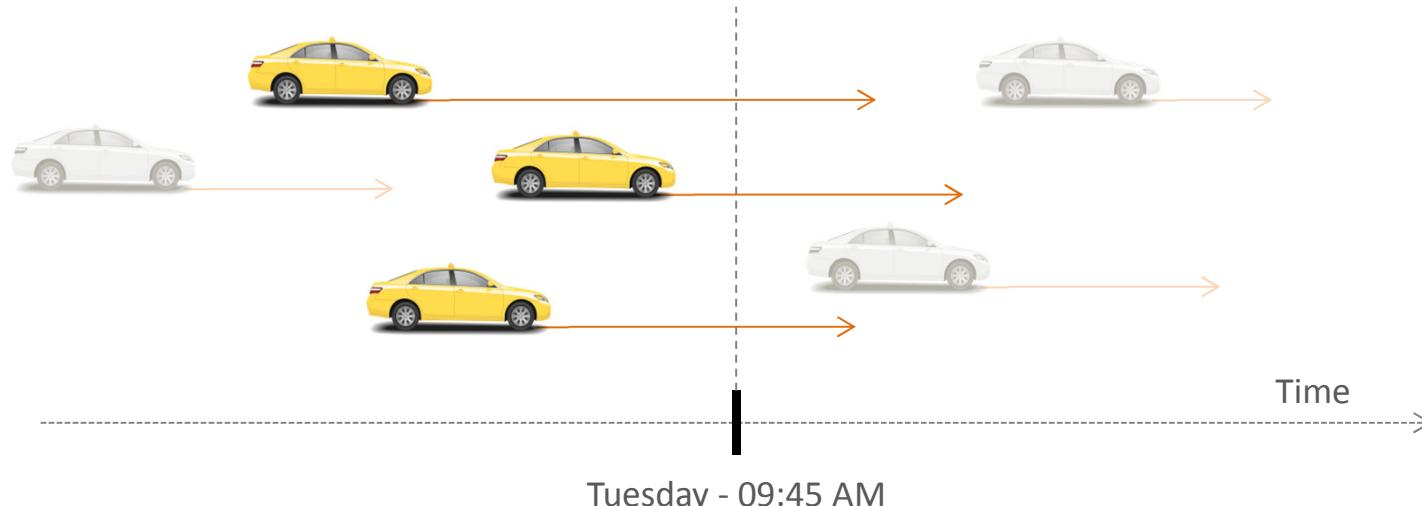
- Let's say, we want to capture the traffic at 09:45 AM on Tuesday
- We need to retrieve taxi trips that took place during
 - Tuesday at 09:45 AM
- That's 26 Tuesdays and about 10,000 selected trips



How to characterize an edge

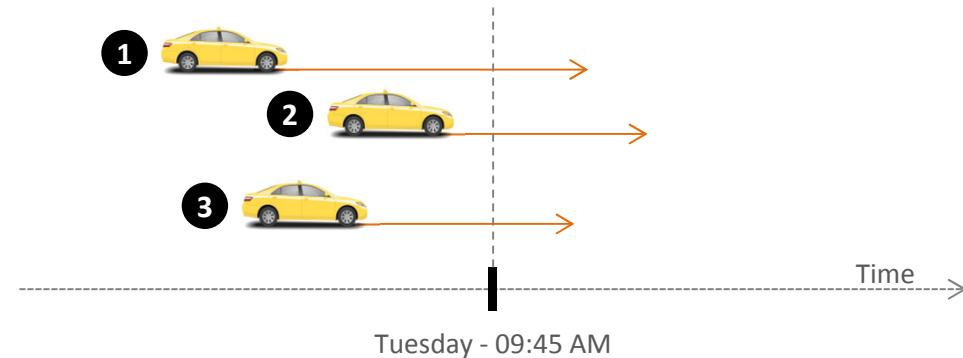
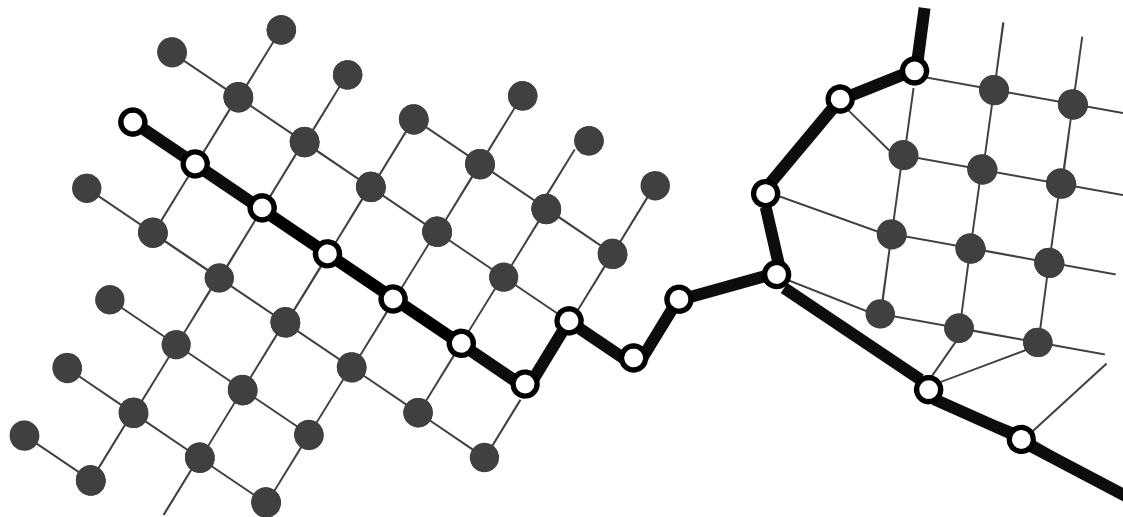
to predict its duration (Graph Features)

- Let's say, we want to capture the traffic at 09:45 AM on Tuesday
- We need to retrieve taxi trips that took place during
 - Tuesday at 09:45 AM
- That's 26 Tuesdays and about **10,000 selected trips**



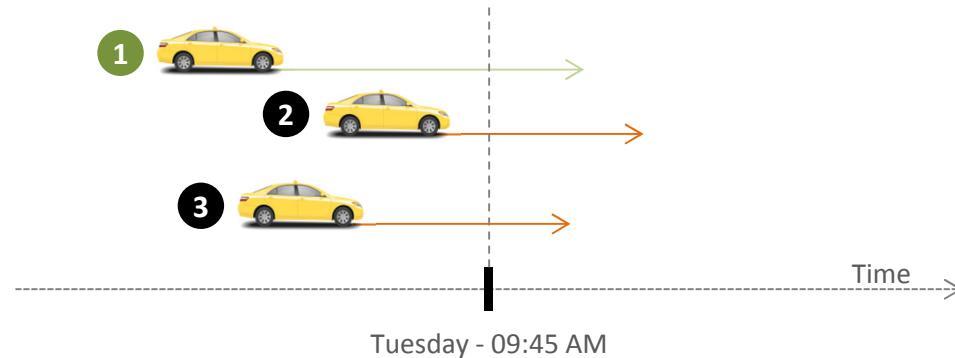
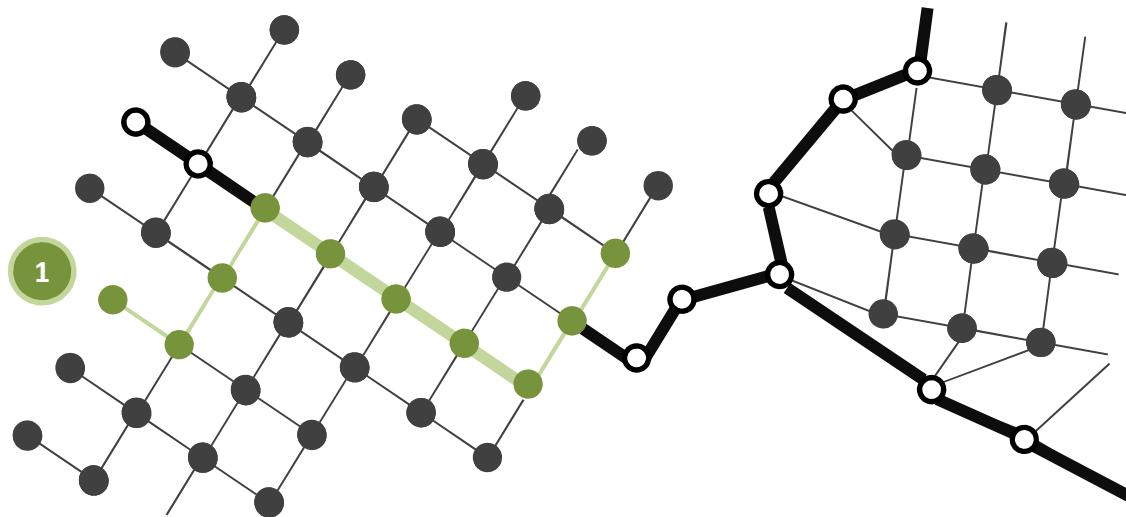
How to characterize an edge

to predict its duration (Graph Features)



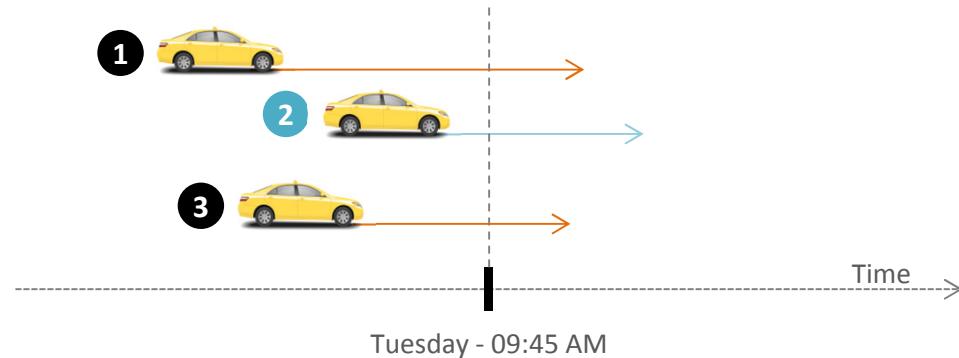
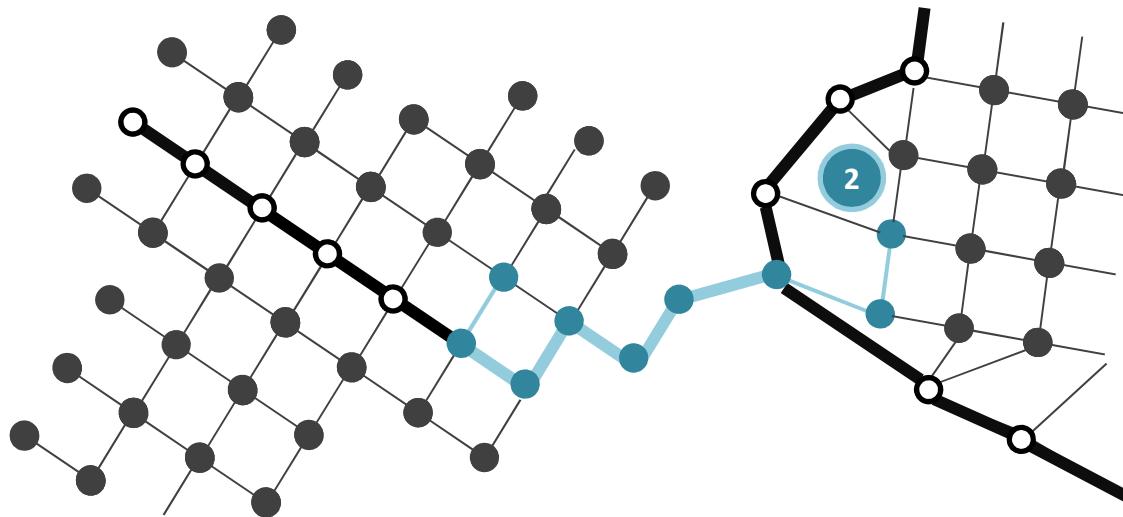
How to characterize an edge

to predict its duration (Graph Features)



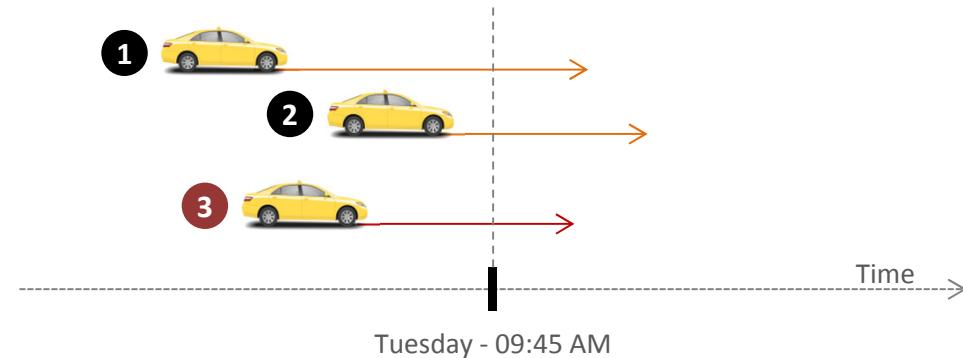
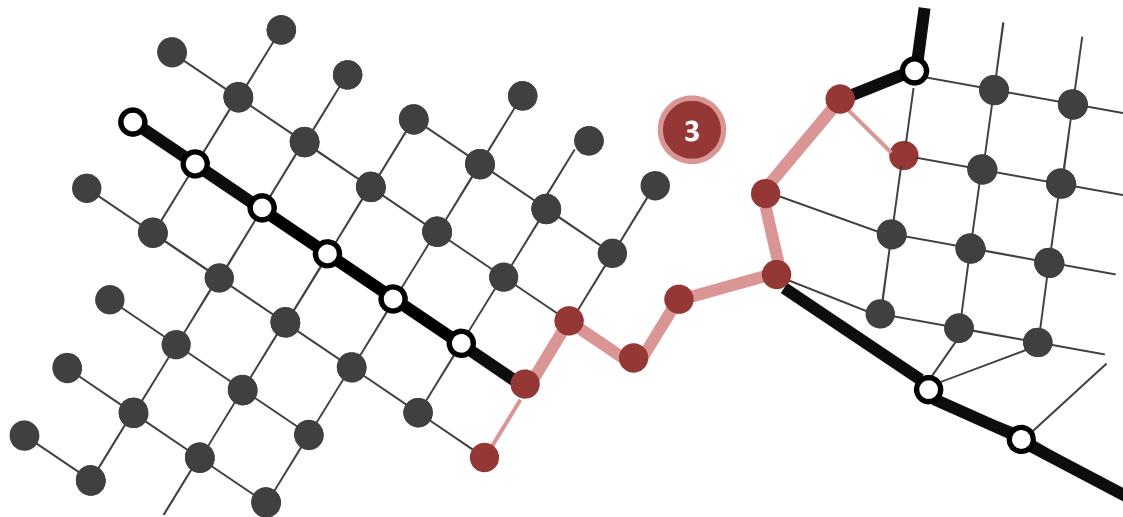
How to characterize an edge

to predict its duration (Graph Features)

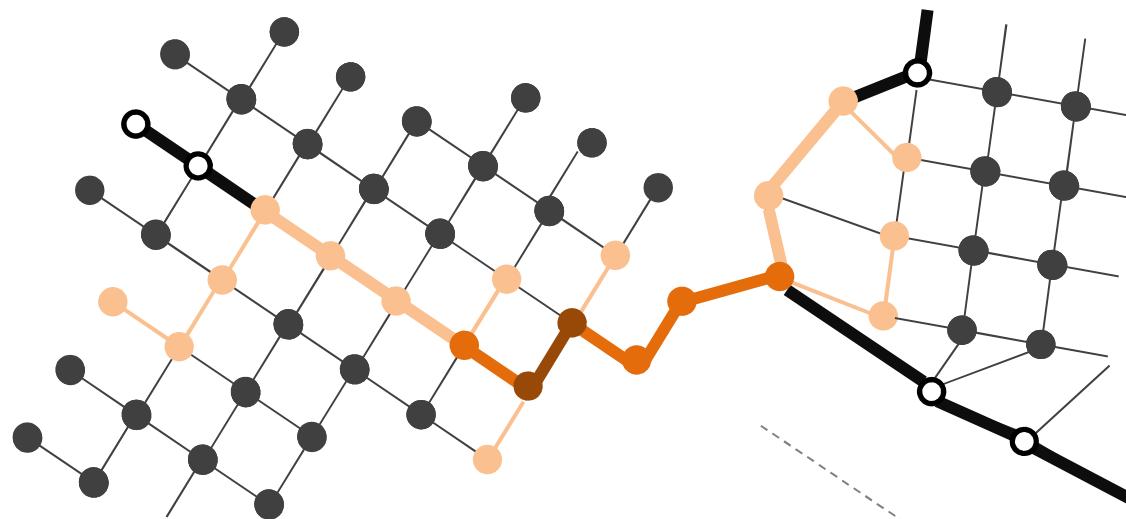


How to characterize an edge

to predict its duration (Graph Features)



How to characterize an edge to predict its duration (Graph Features)

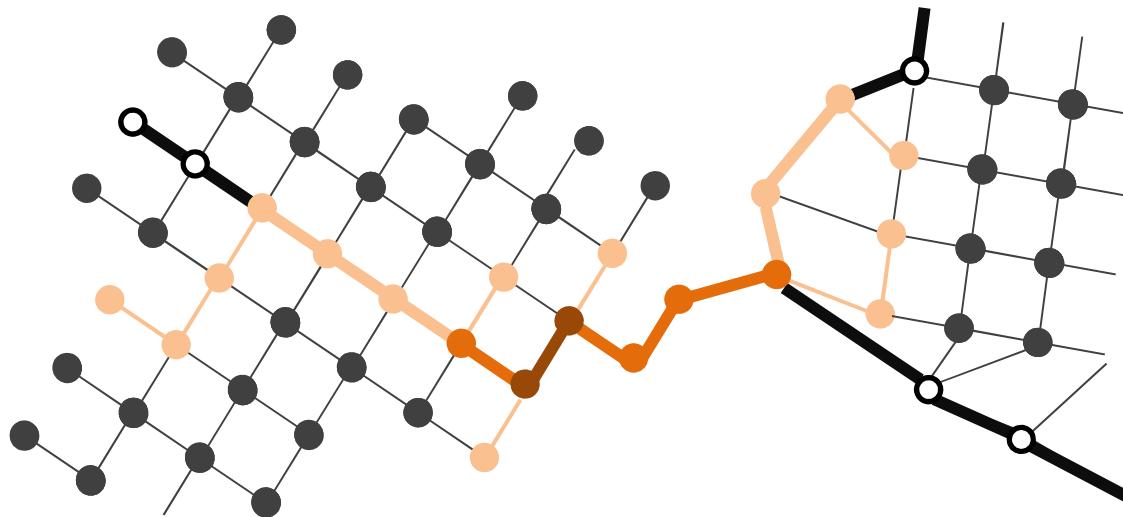


Tuesday - 09:45 Betweenness
(based on 3 paths)

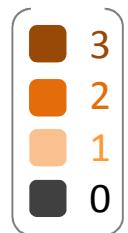
- 3/3 paths go through this segment
- 2/3 paths go through this segment
- 1/3 paths go through this segment
- 0/3 paths go through this segment

Good capture of
instant traffic!

How to characterize an edge to predict its duration (Graph Features)

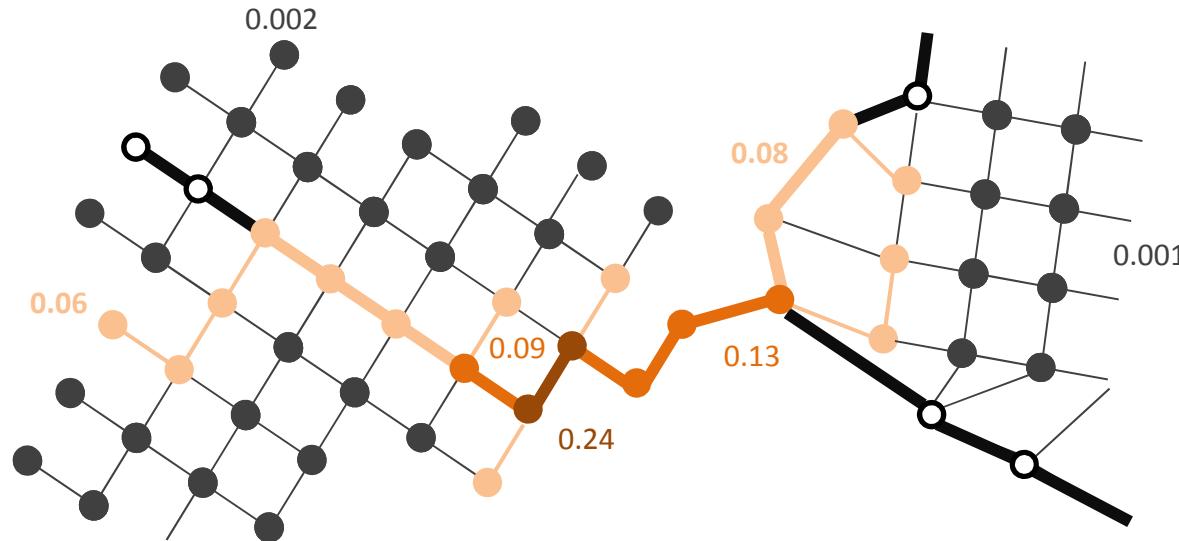


Tuesday - 09:45 Betweenness
(based on 3 paths)

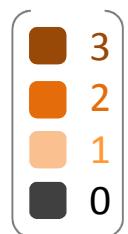


Draw from a
Dirichlet* distribution

How to characterize an edge to predict its duration (Graph Features)



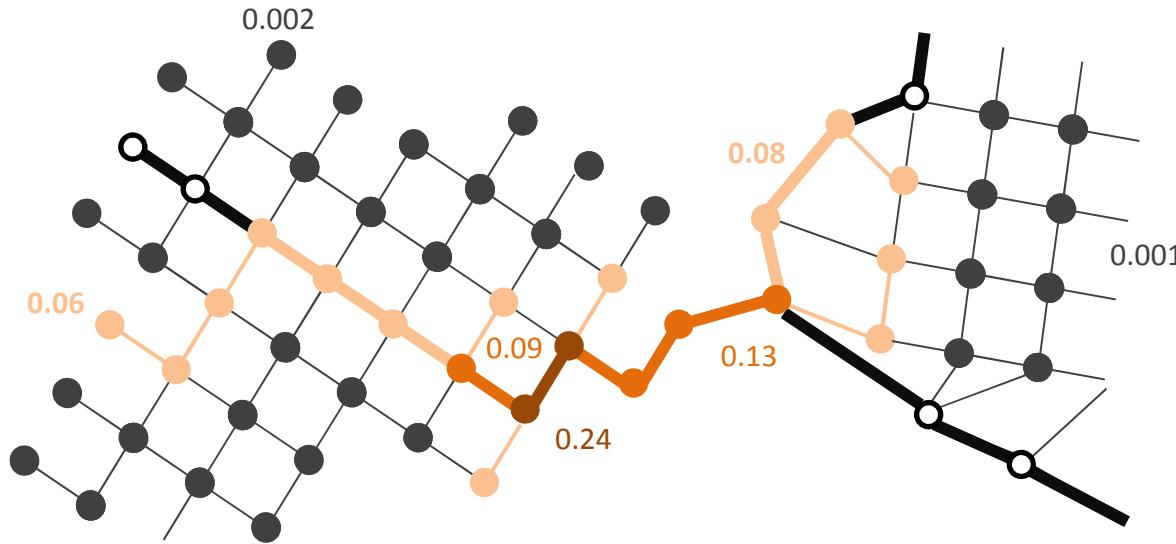
Tuesday - 09:45 Betweenness
(based on 3 paths)



Draw from a
Dirichlet* distribution

How to characterize an edge

to predict its duration (Graph Features)



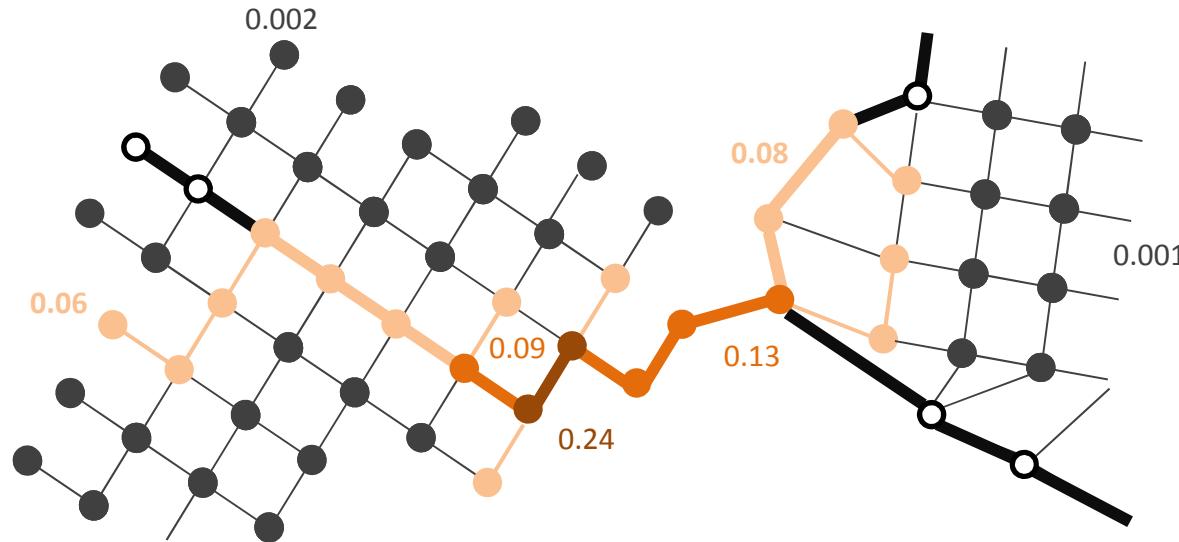
New Feature to characterize an edge



- Absolute betweenness (ex:**18**), i.e. number of path that goes through the edge
- Relative betweenness (ex:**0.09**), i.e. probability that a path goes through the edge

How to characterize an edge

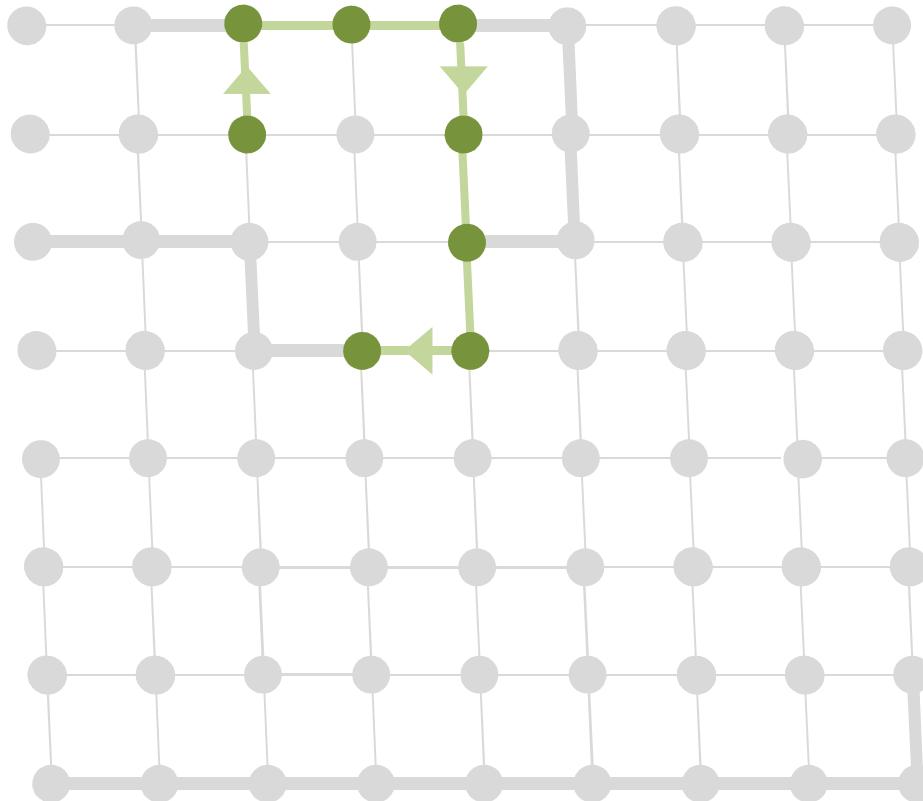
to predict its duration (Graph Features)



Be Careful! This betweenness is a representation of **taxi traffic**,
not *all vehicles traffic*. It might be somewhat correlated though

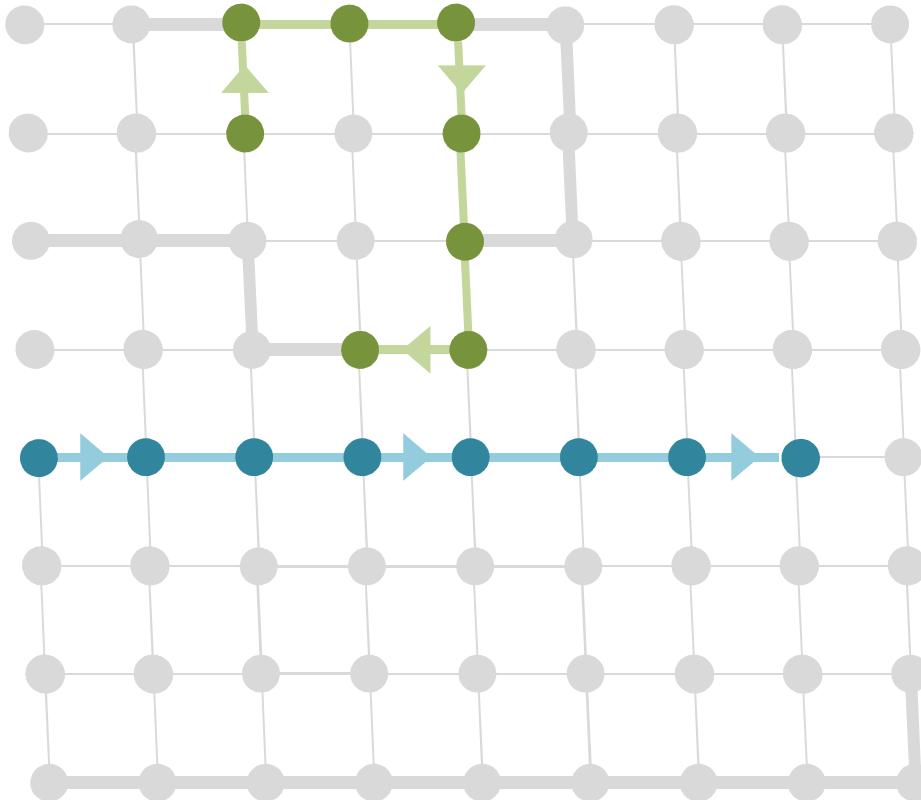
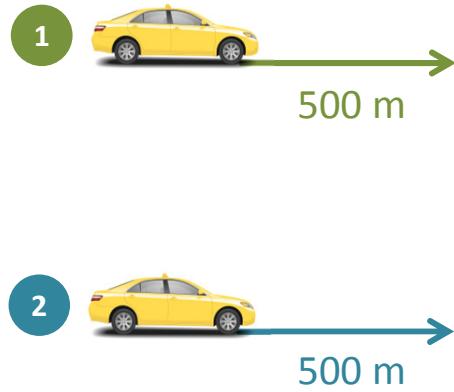
How to characterize an edge

to predict its duration (Graph Features)



How to characterize an edge

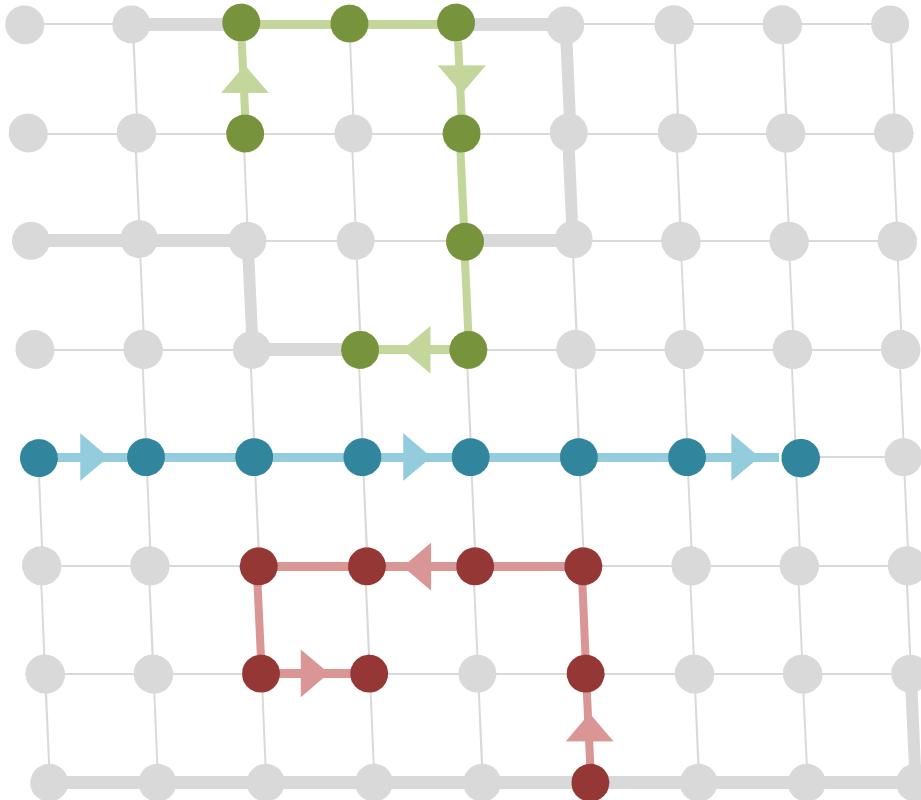
to predict its duration (Graph Features)



How to characterize an edge

to predict its duration (Graph Features)

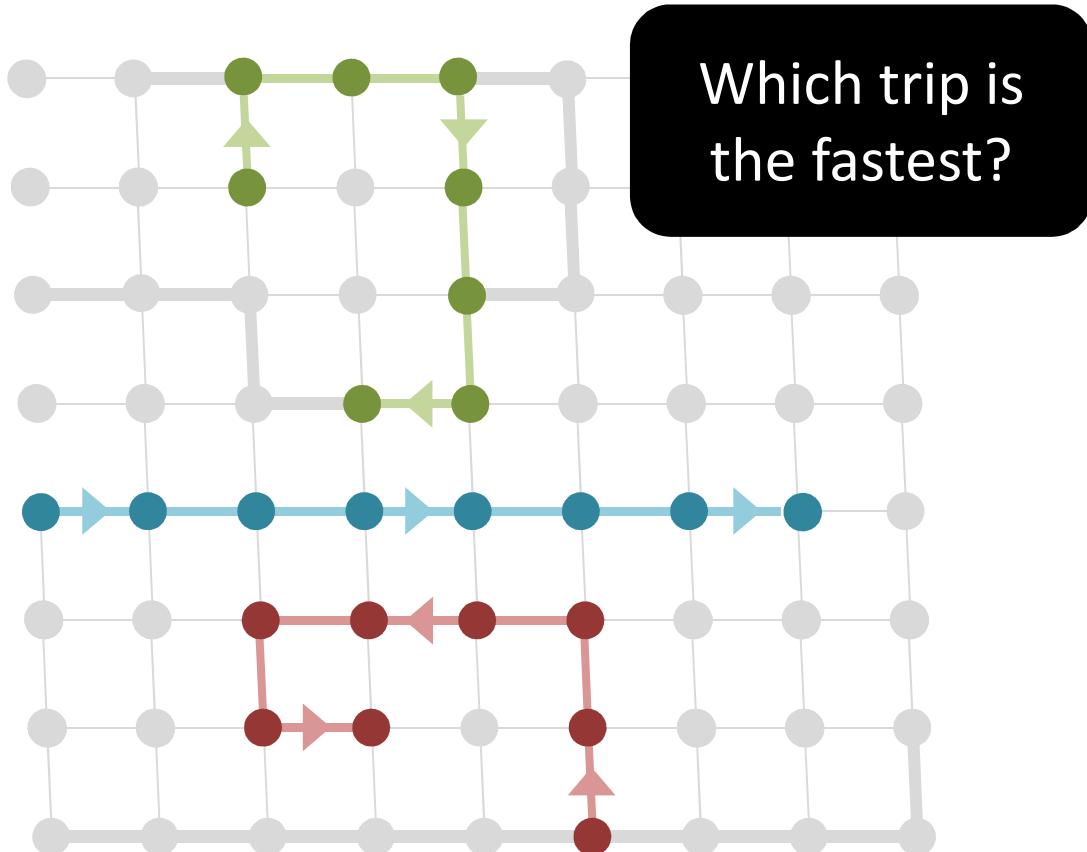
- 1   500 m
- 2   500 m
- 3   500 m



How to characterize an edge

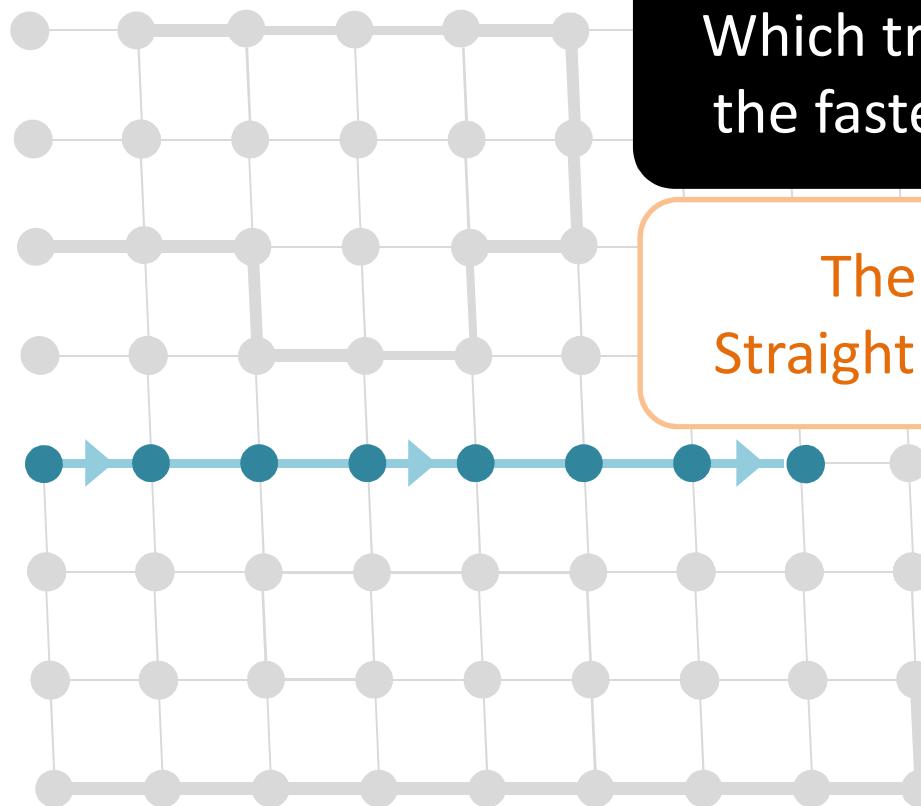
to predict its duration (Graph Features)

- 1   500 m
- 2   500 m
- 3   500 m



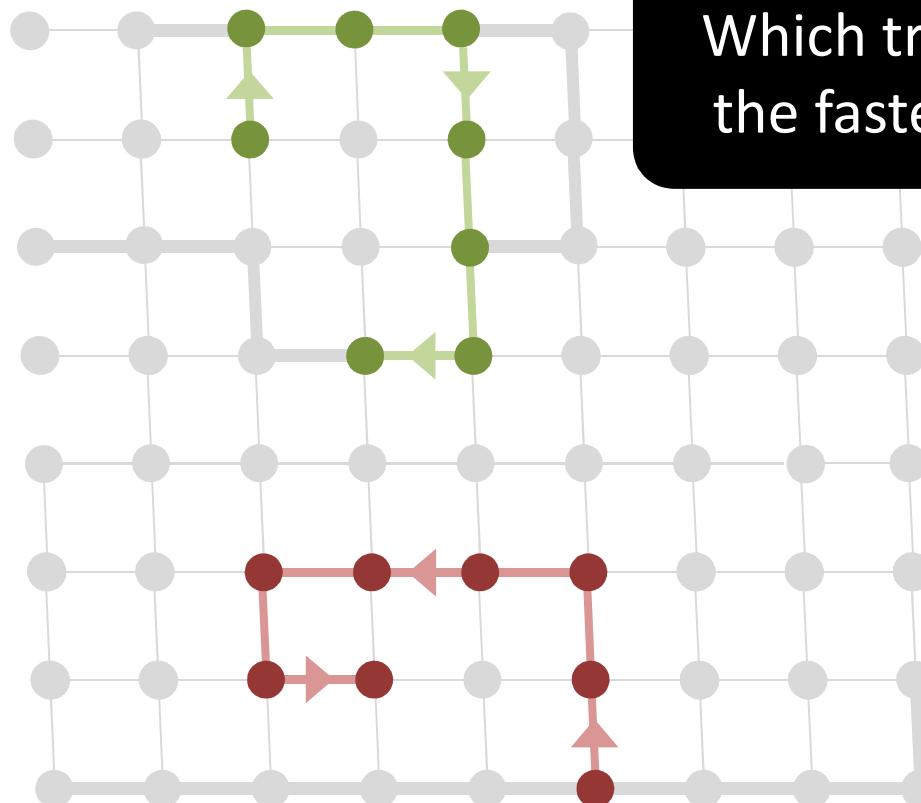
How to characterize an edge

to predict its duration (Graph Features)



How to characterize an edge

to predict its duration (Graph Features)



How to characterize an edge

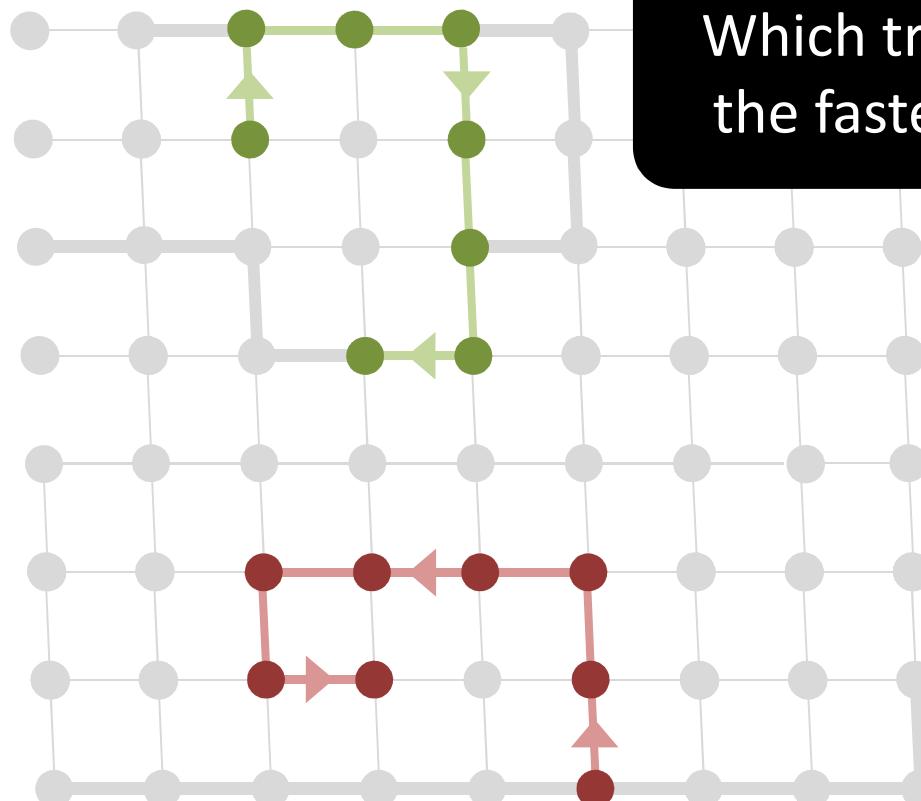
to predict its duration (Graph Features)



500 m
3 turns

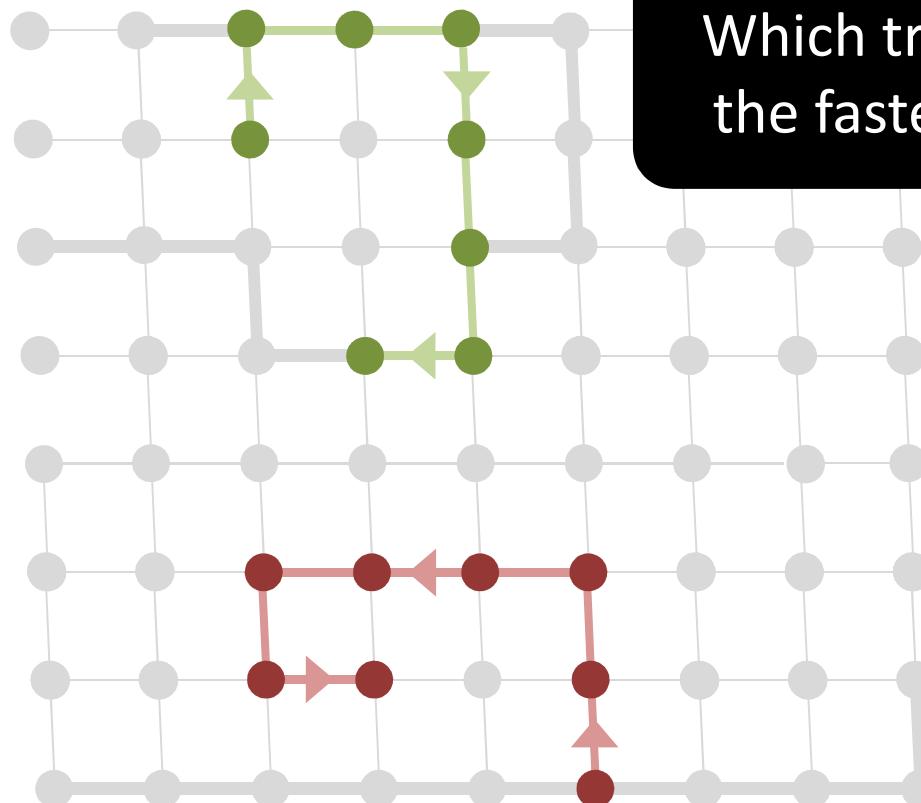
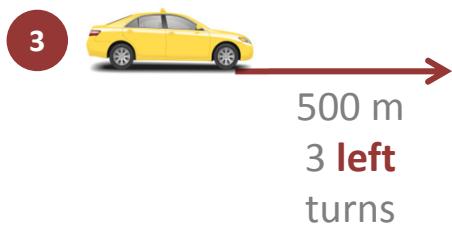
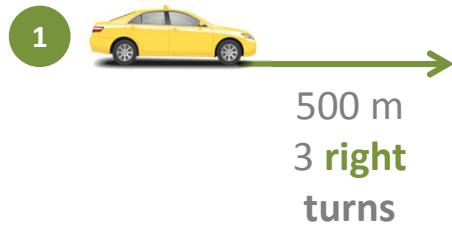


500 m
3 turns



How to characterize an edge

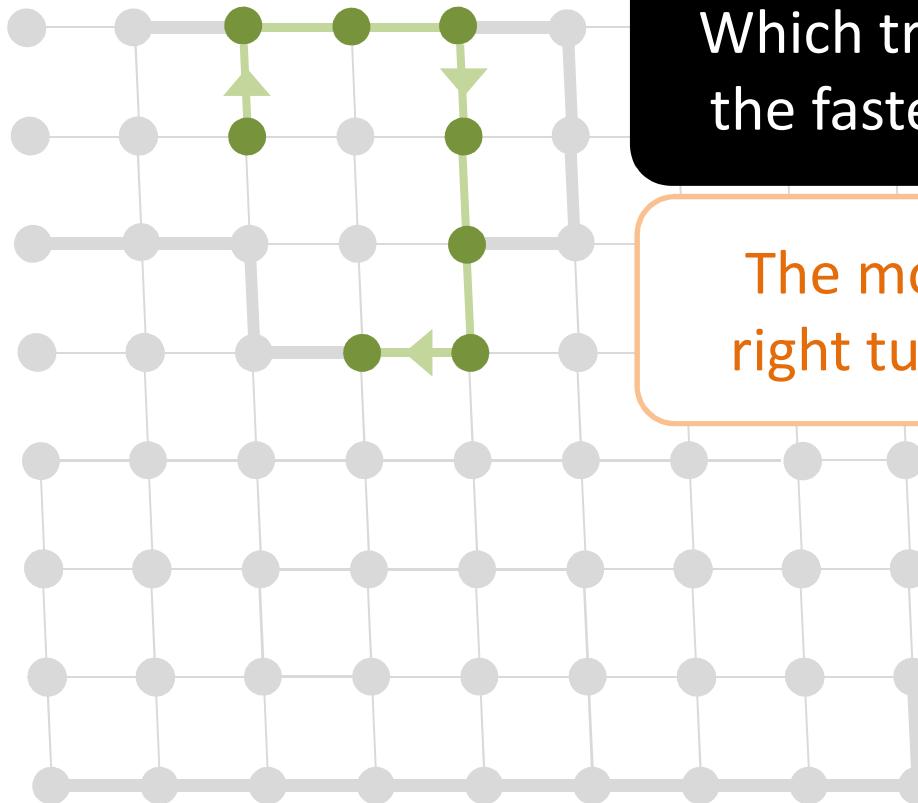
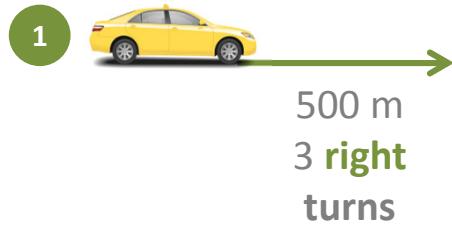
to predict its duration (Graph Features)



Which trip is
the fastest?

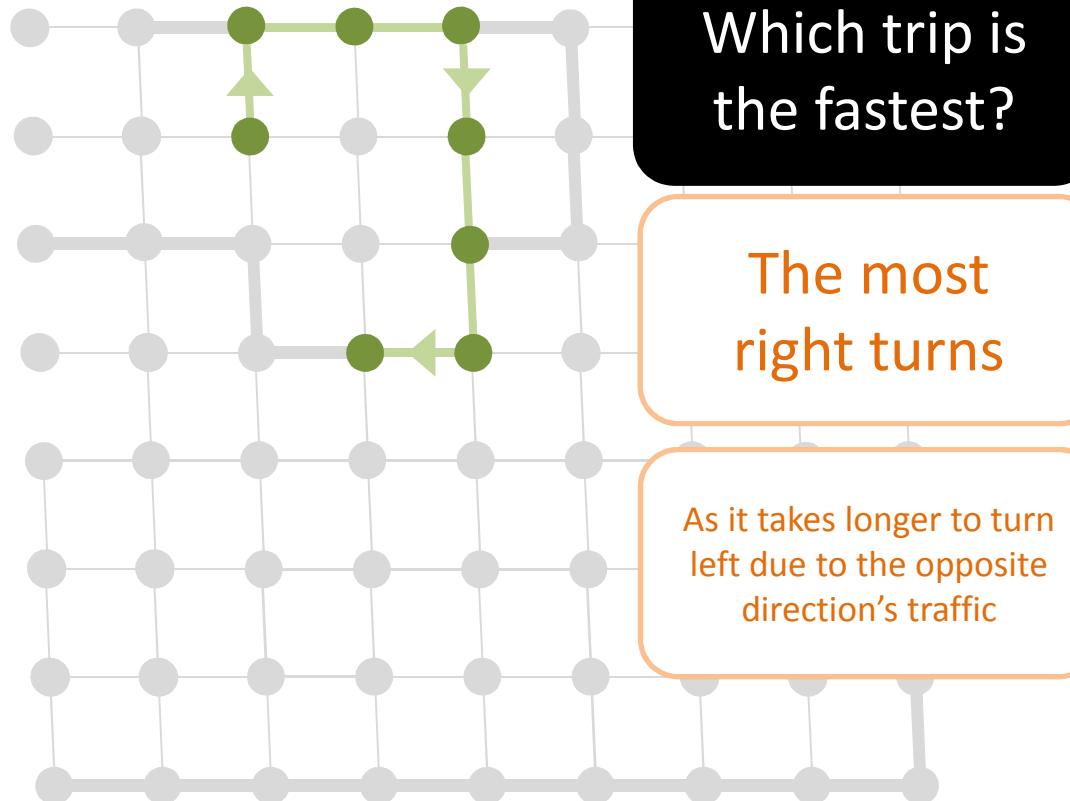
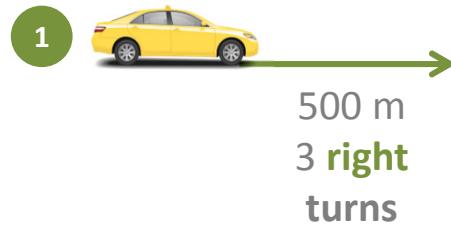
How to characterize an edge

to predict its duration (Graph Features)



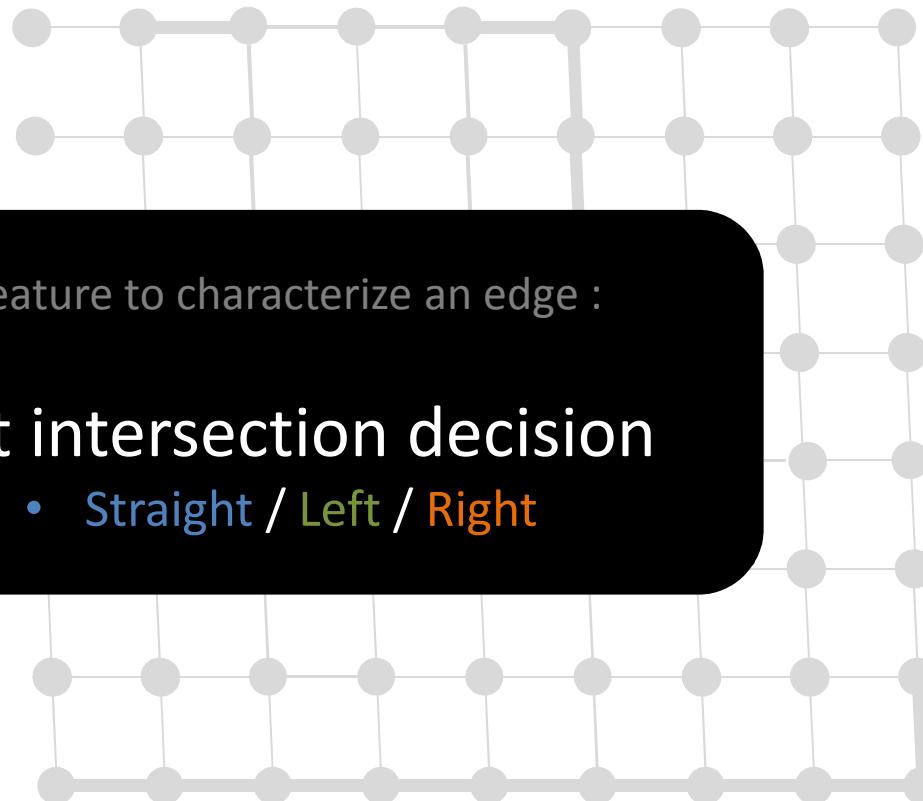
How to characterize an edge

to predict its duration (Graph Features)



How to characterize an edge

to predict its duration (Graph Features)



How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a road segment's duration?

How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a road segment's duration?



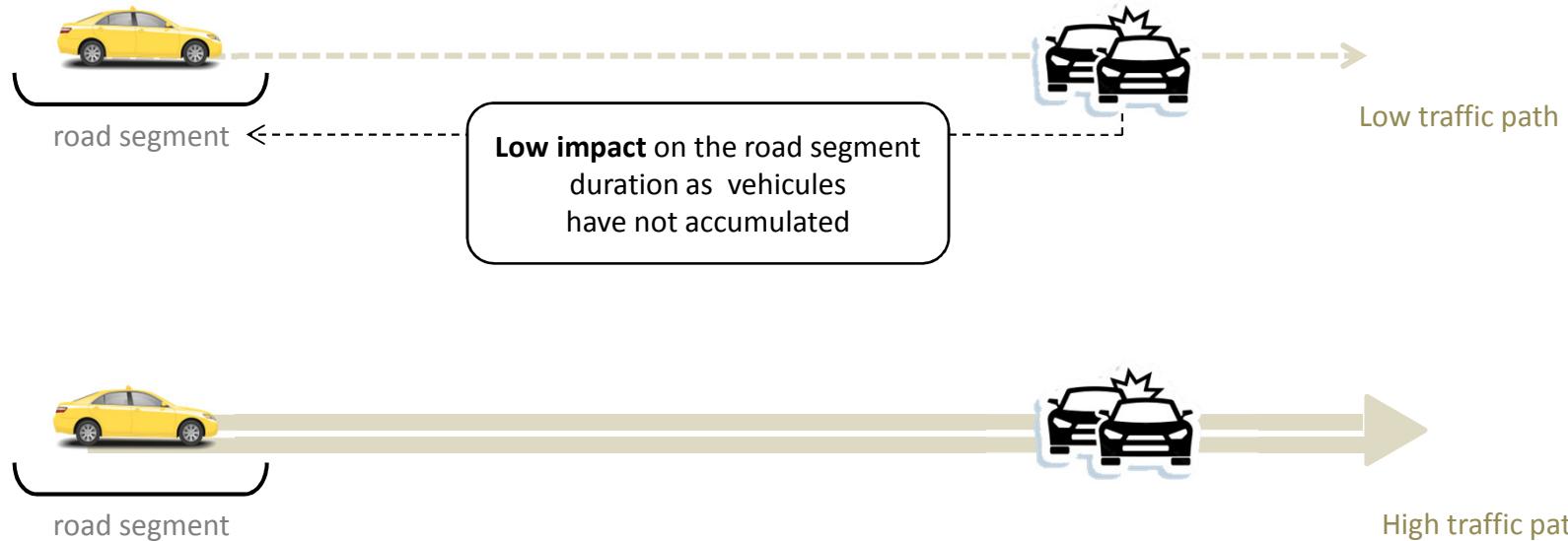
How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a road segment's duration?



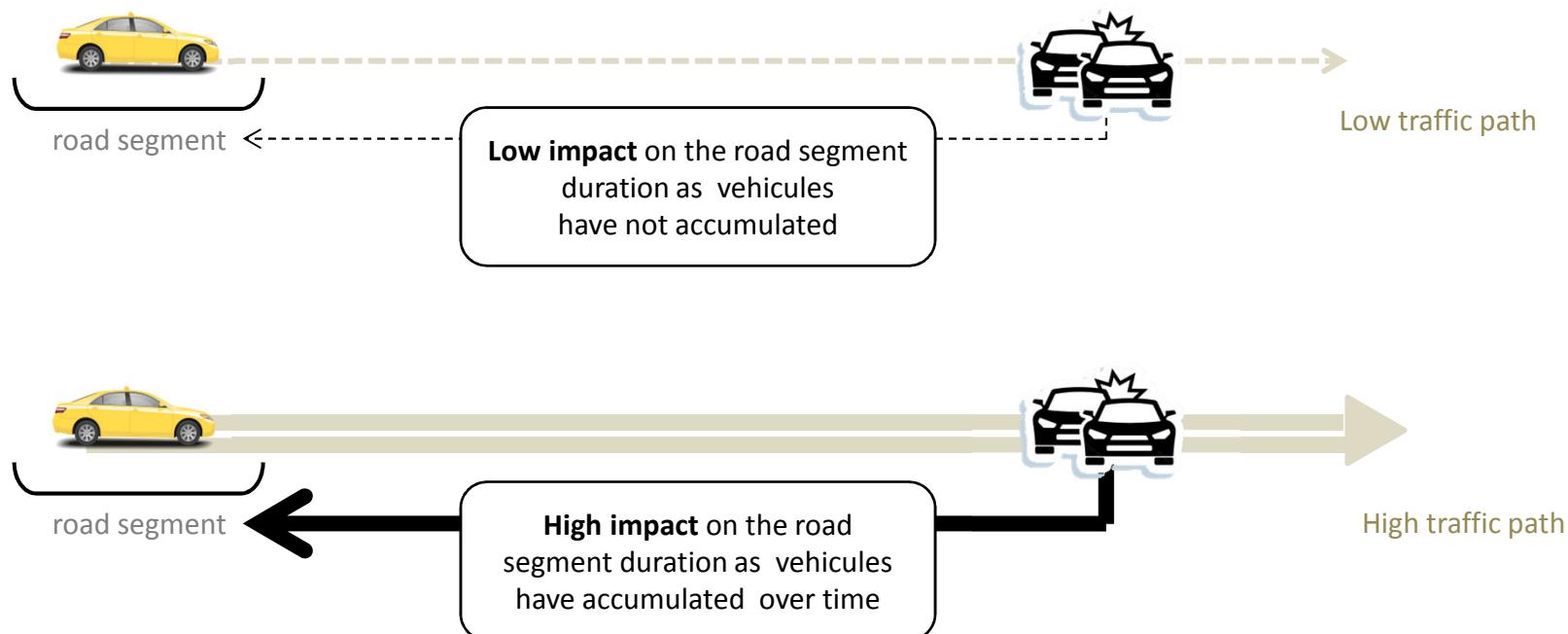
How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a road segment's duration?



How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a road segment's duration?

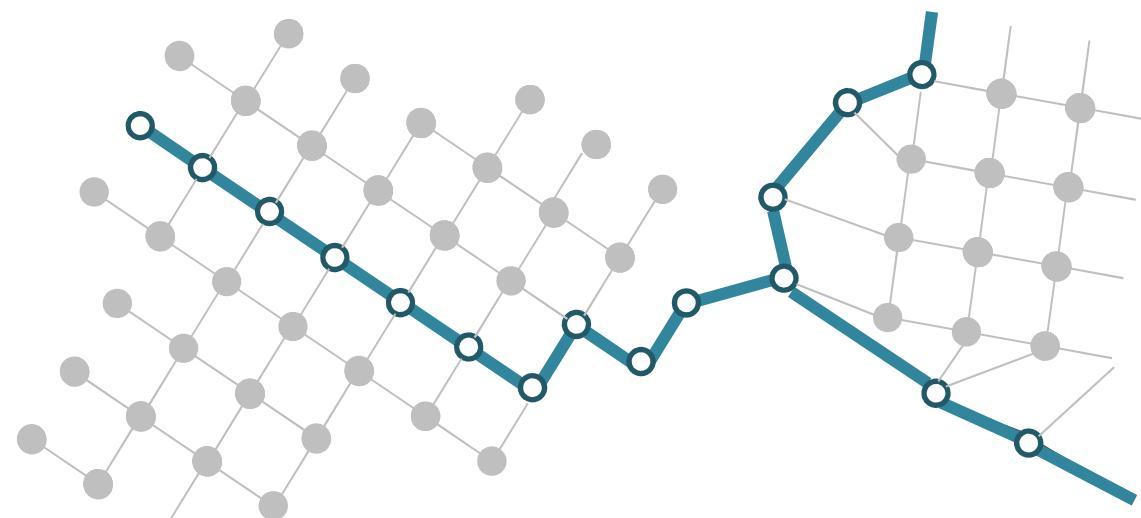


How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

Thus, for the sake of
simplicity



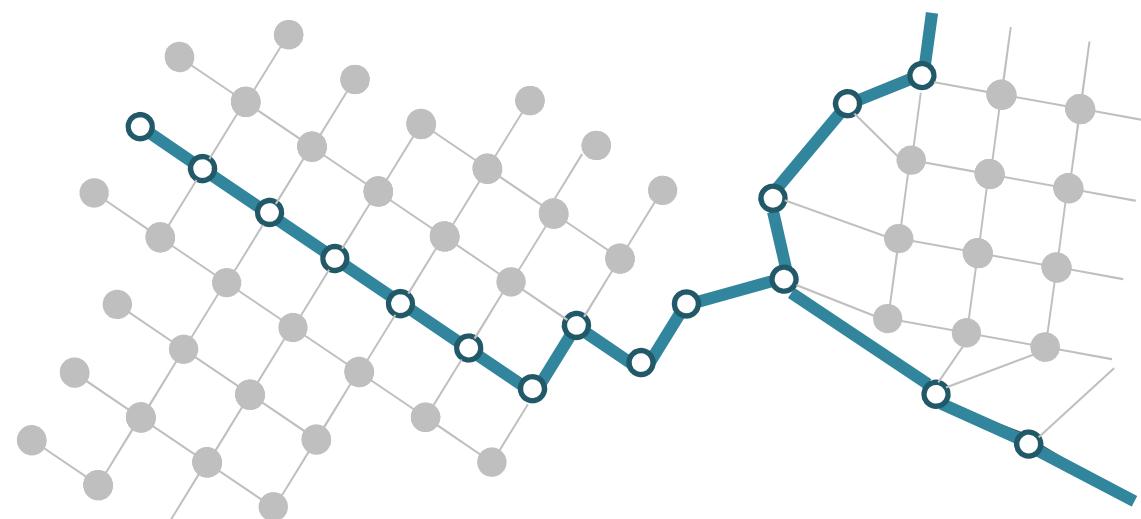
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

Thus, for the sake of
simplicity

- Only consider accidents **on**
high betweenness roads



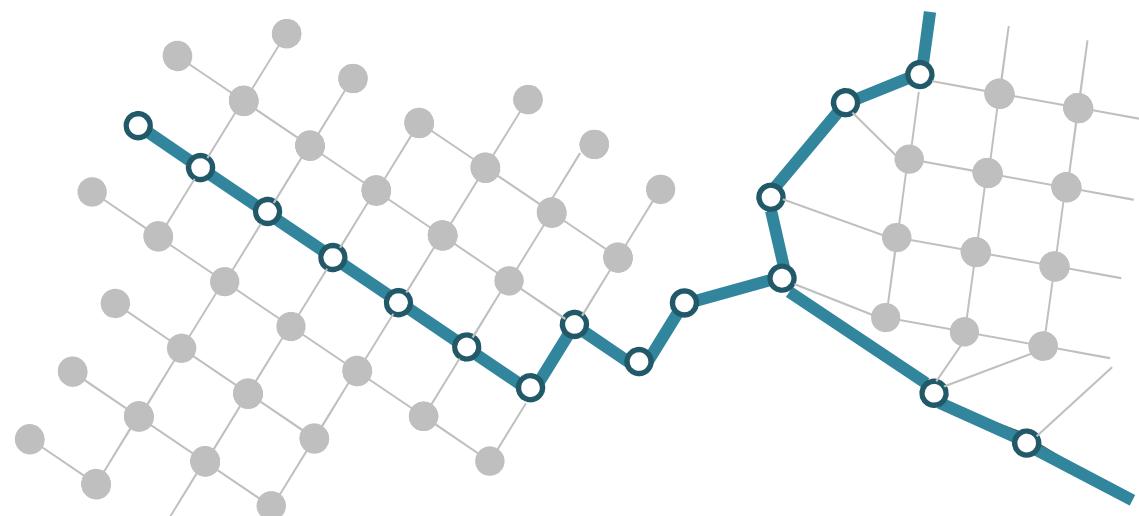
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

Thus, for the sake of simplicity

- Only consider accidents **on high betweenness roads**
- Only high **betweenness road segment** could be impacted by an accident



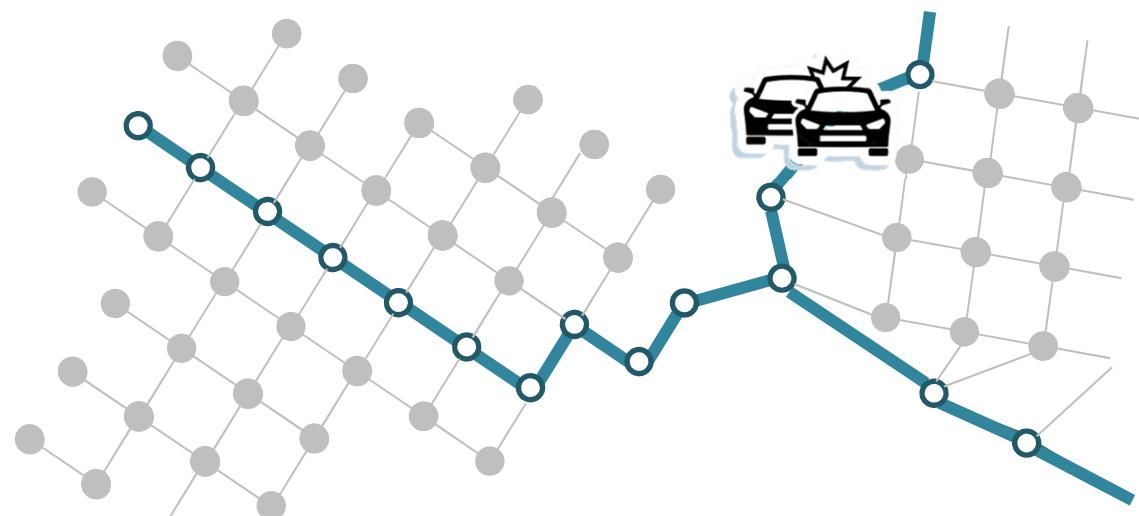
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

Thus, for the sake of simplicity

- Only consider accidents **on high betweenness roads**
- Only high **betweenness road segment** could be impacted by an accident



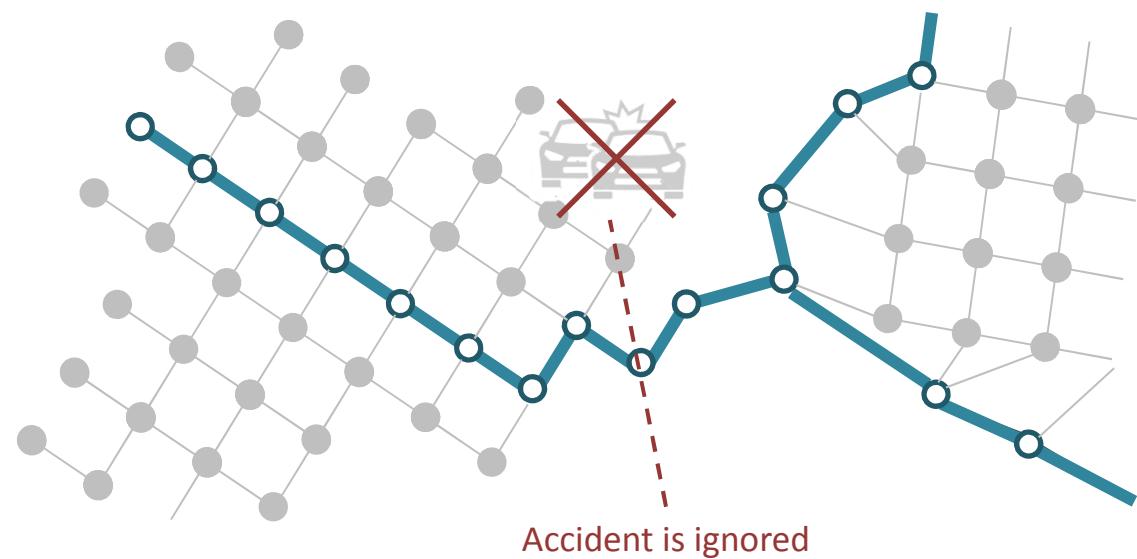
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a road segment's duration?

Thus, for the sake of simplicity

- Only consider accidents **on high betweenness roads**
- Only high **betweenness road segment** could be impacted by an accident

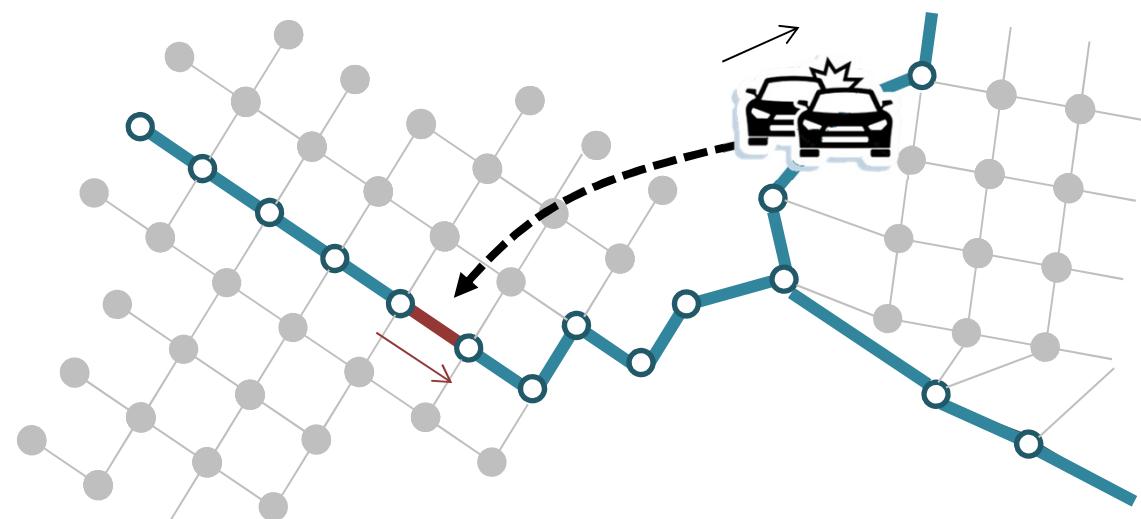


How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance :
 - Distance :
 - Shortest path :
 - Time elapsed :
 - Same direction :
 - Branches :
- Features to capture accident impact



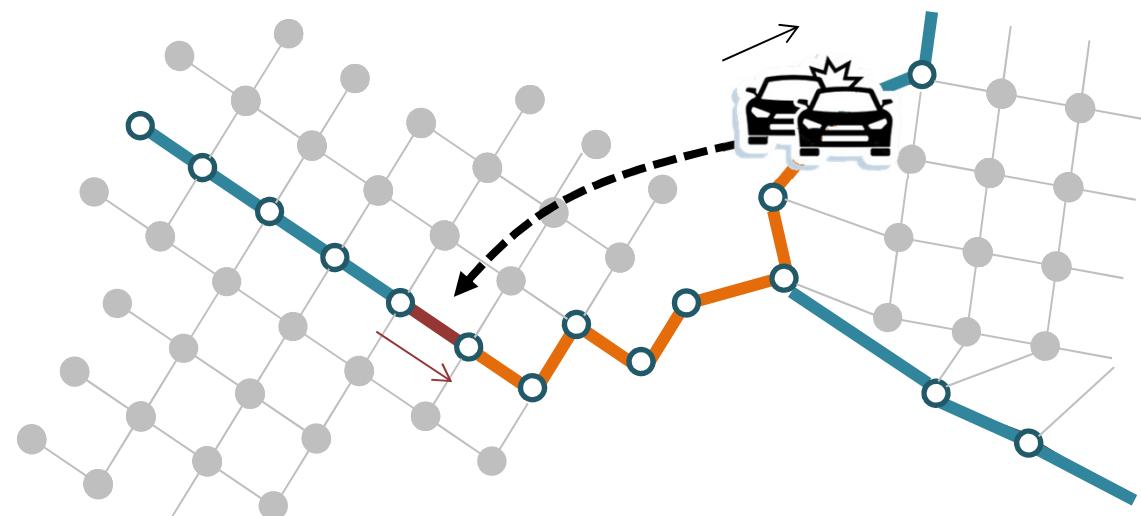
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance :
- Shortest path :
- Time elapsed :
- Same direction :
- Branches :

Features to capture
accident impact



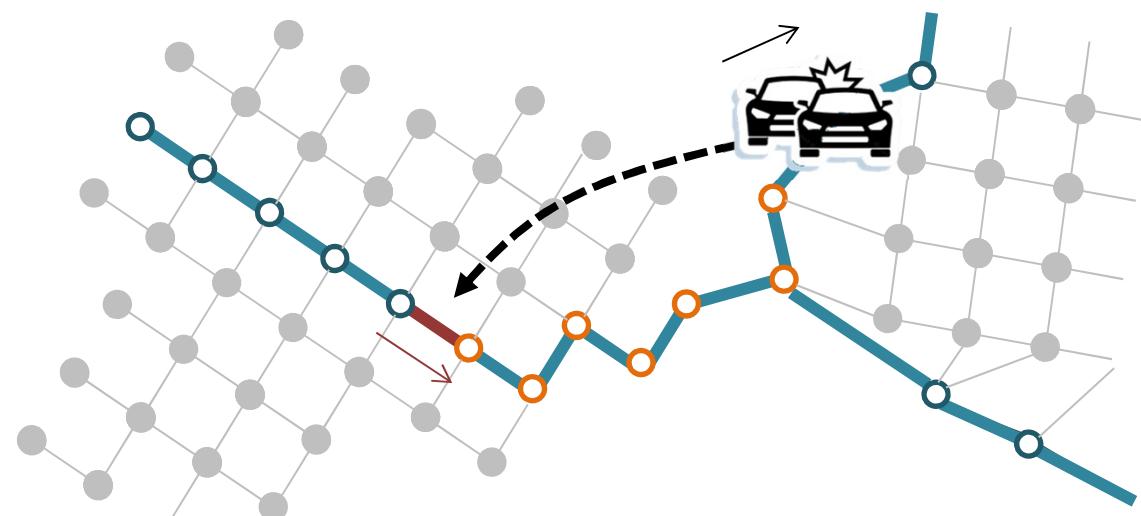
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path :
- Time elapsed :
- Same direction :
- Branches :

Features to capture
accident impact

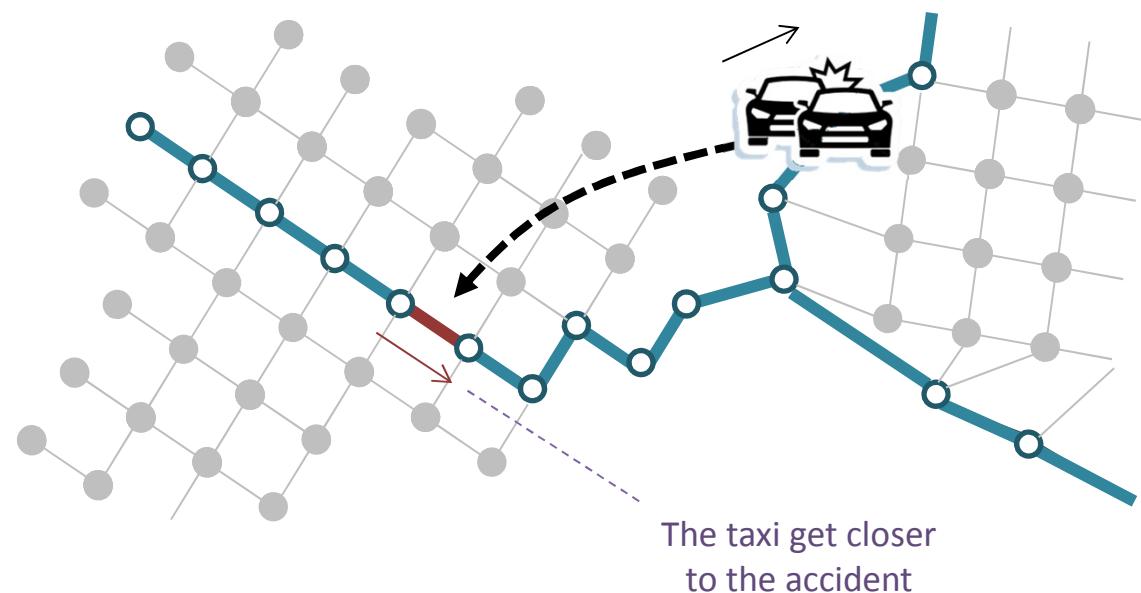


How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path : **True**
- Time elapsed :
- Same direction :
- Branches :

Features to capture
accident impact

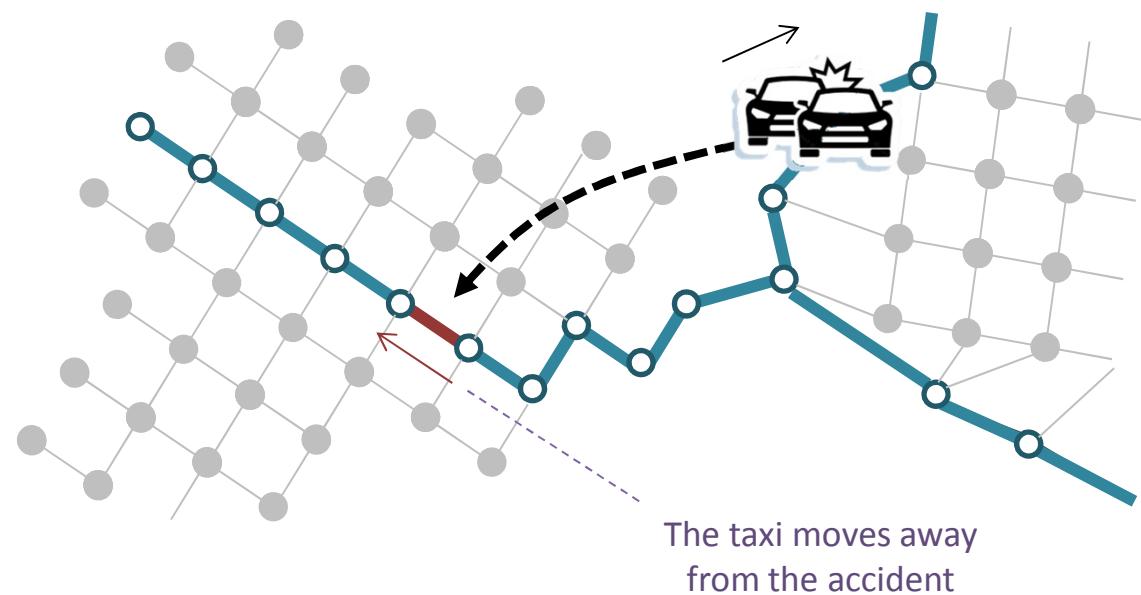


How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path : **False**
- Time elapsed :
- Same direction :
- Branches :

Features to capture
accident impact



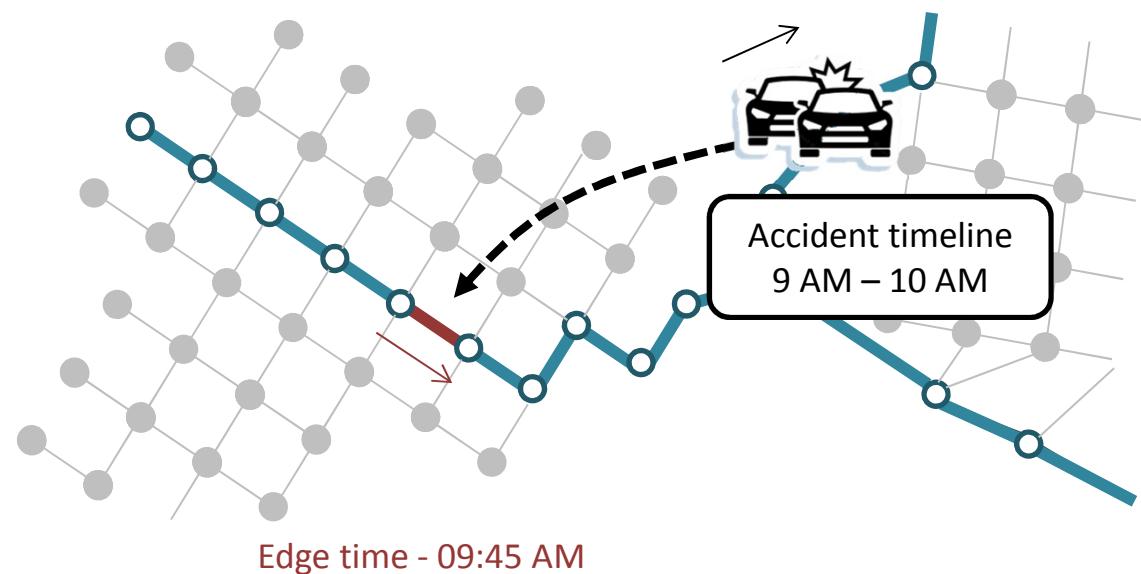
How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path : True
- Time elapsed : 45
- Same direction :
- Branches :

Features to capture
accident impact

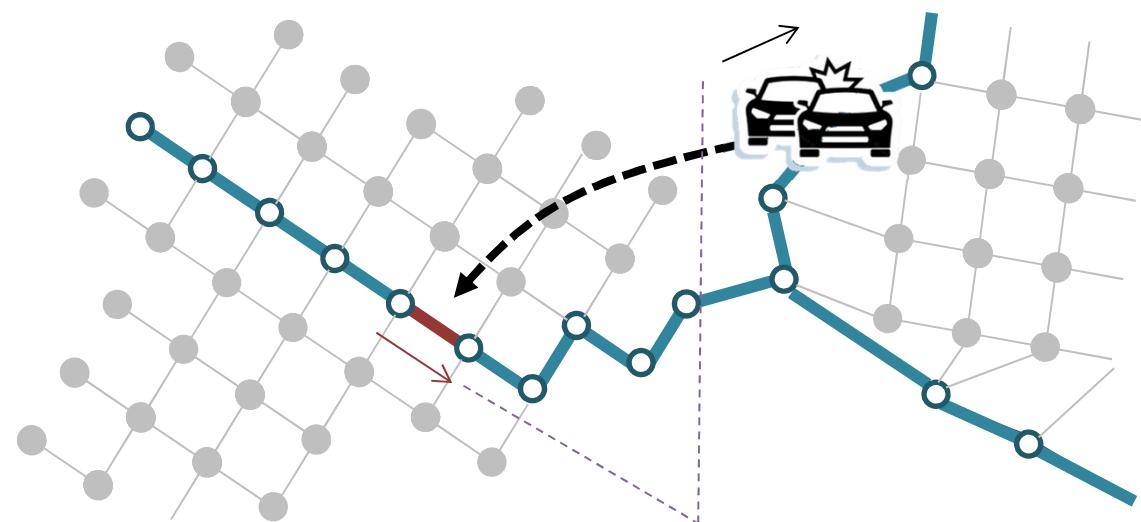


How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path : True
- Time elapsed : 45
- Same direction : **True**
- Branches :

Features to capture
accident impact



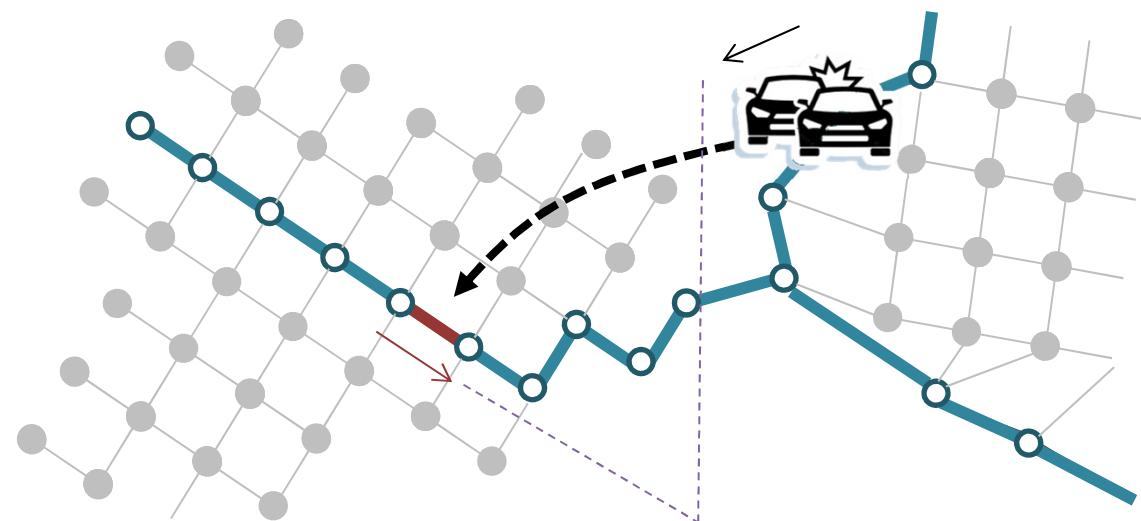
The accident took place on lanes going in the
same direction as the cab. Big traffic jam

How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path : True
- Time elapsed : 45
- Same direction : **False**
- Branches :

Features to capture
accident impact



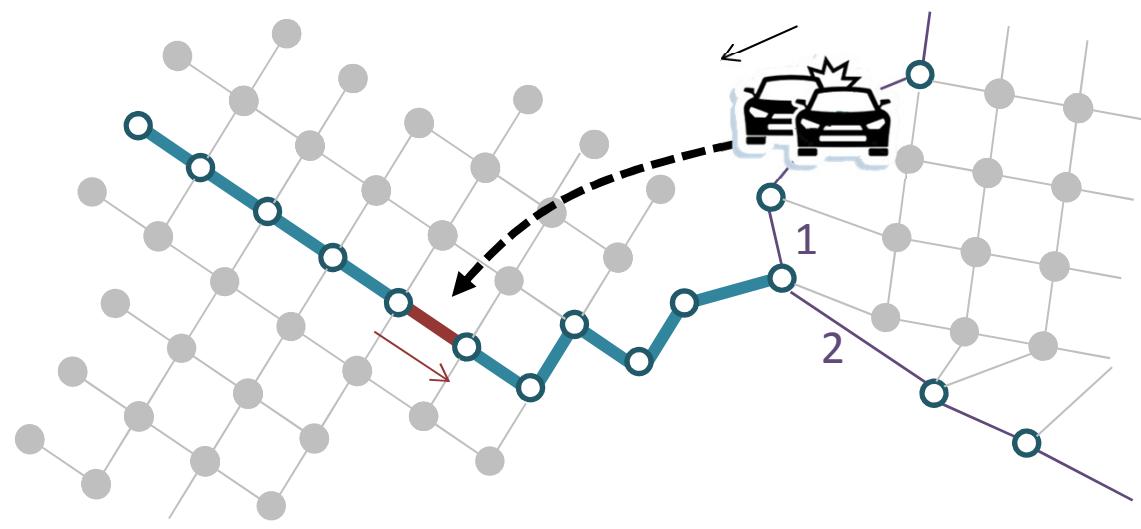
The accident took place on lanes going
in the opposite direction as the cab.
Possible slowdown as people brake to watch

How to characterize an edge to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

- Distance : 1580
- Distance : 7
- Shortest path : True
- Time elapsed : 45
- Same direction : False
- Branches : 2

Features to capture
accident impact



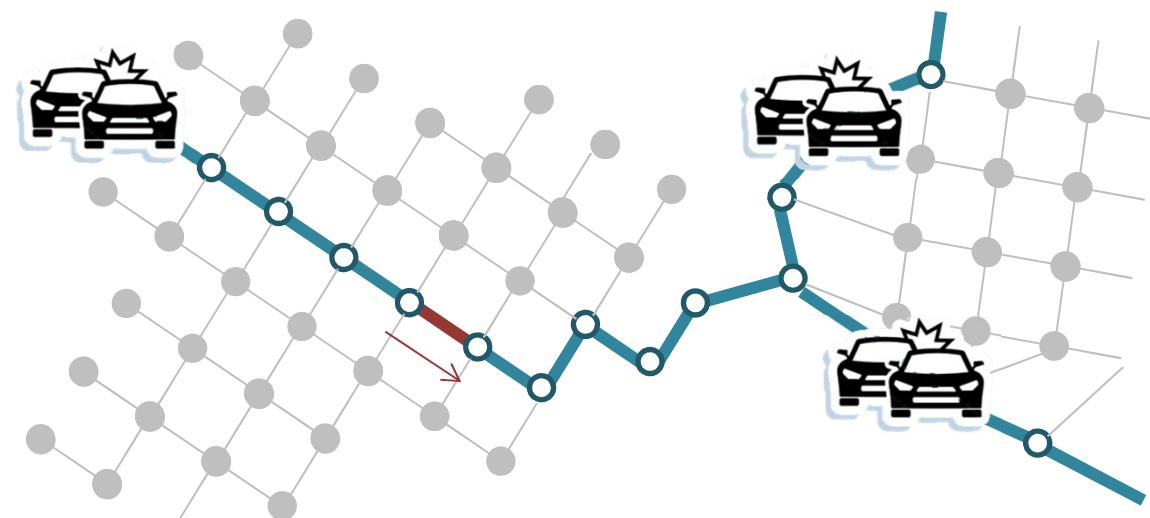
The more branches there are,
the less cars can accumulate

How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if more than
One accident?



How to characterize an edge

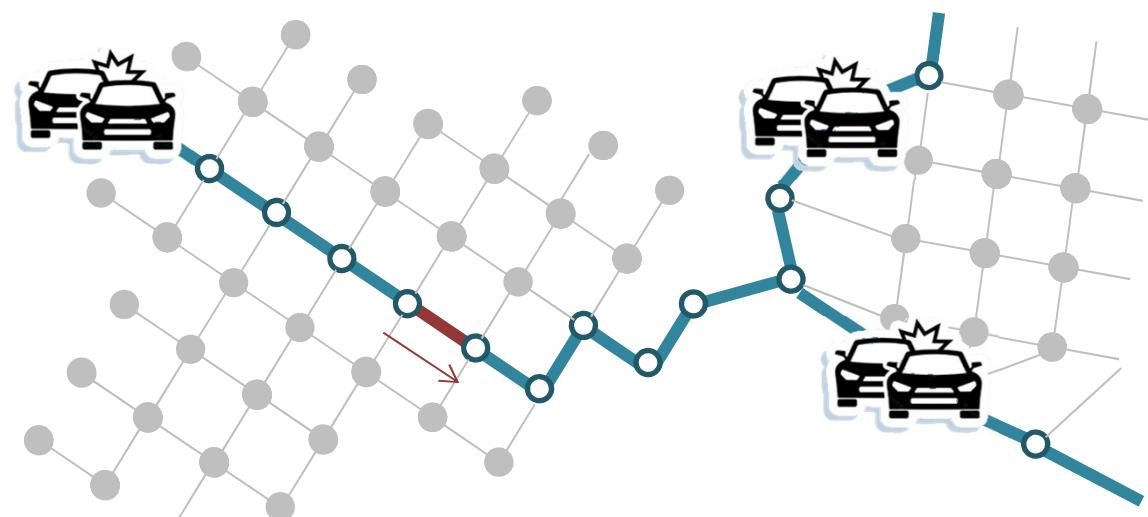
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if more than
One accident?

Keep only one based on :

- Shortest path
- Distance



How to characterize an edge

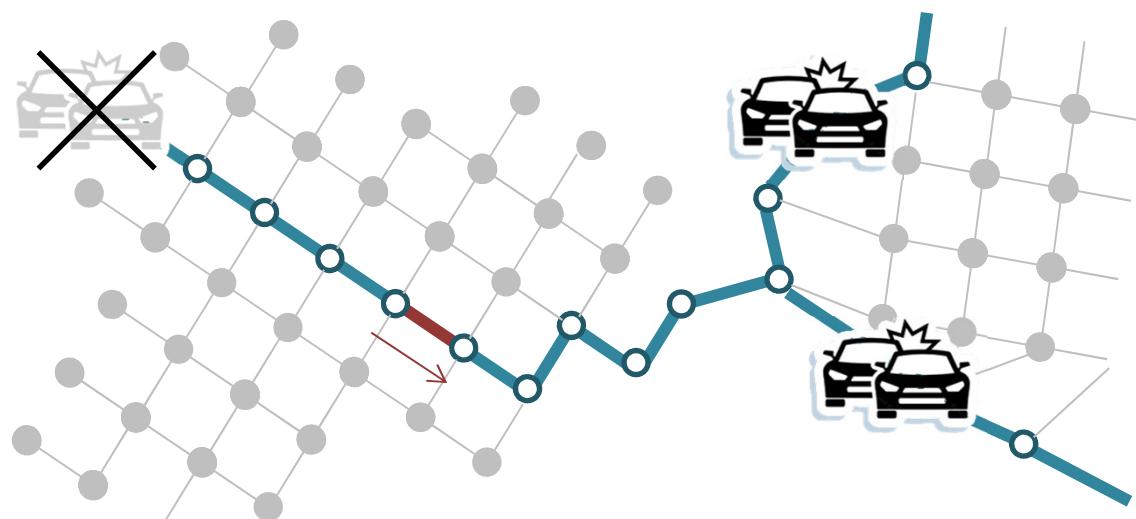
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if more than
One accident?

Keep only one based on :

- Shortest path
- Distance



How to characterize an edge

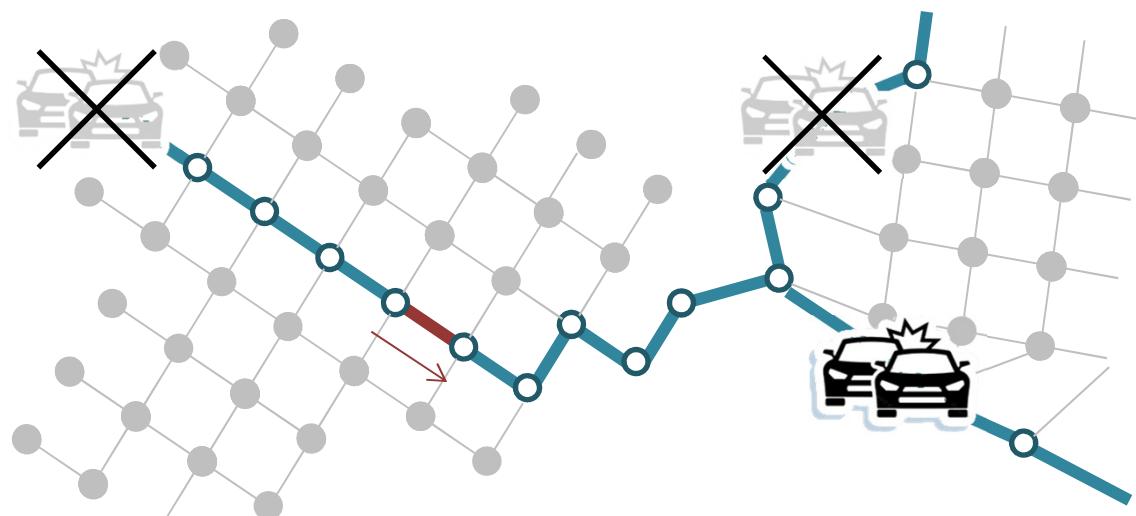
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if more than
one accident?

Keep only one based on :

- Shortest path
- Distance

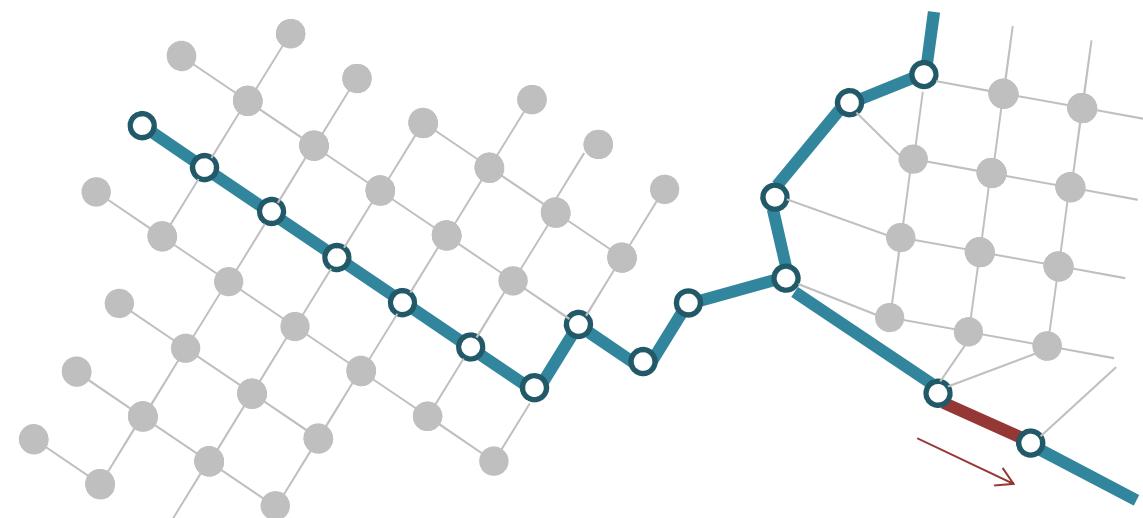


How to characterize an edge

to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?



How to characterize an edge

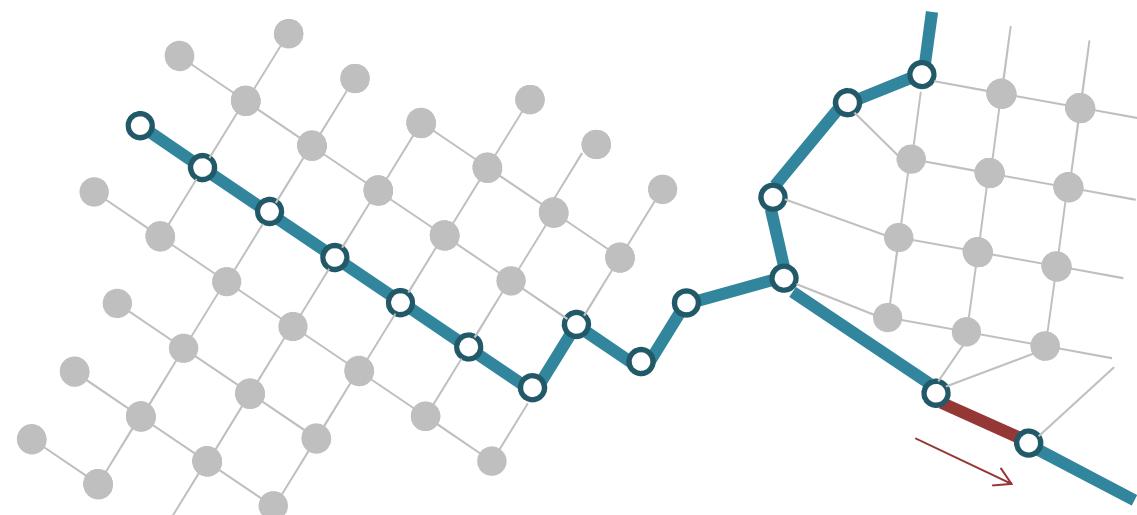
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?

Generate a
low impact accident :

- Far from the edge
- Not on the shortest path
- Very close/far in time



How to characterize an edge

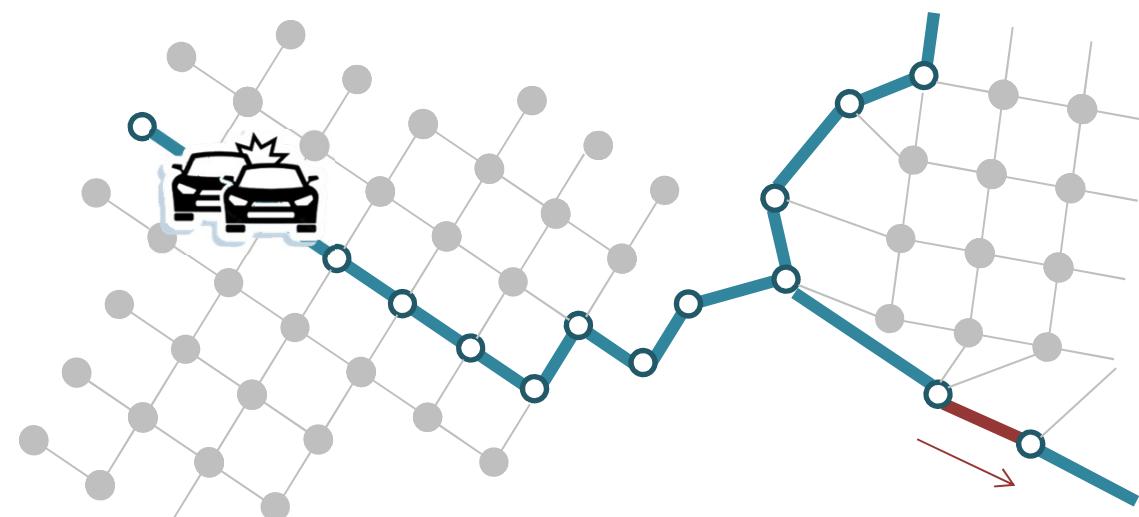
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?

Generate a
low impact accident :

- Far from the edge
- Not on the shortest path
- Very close/far in time



How to characterize an edge

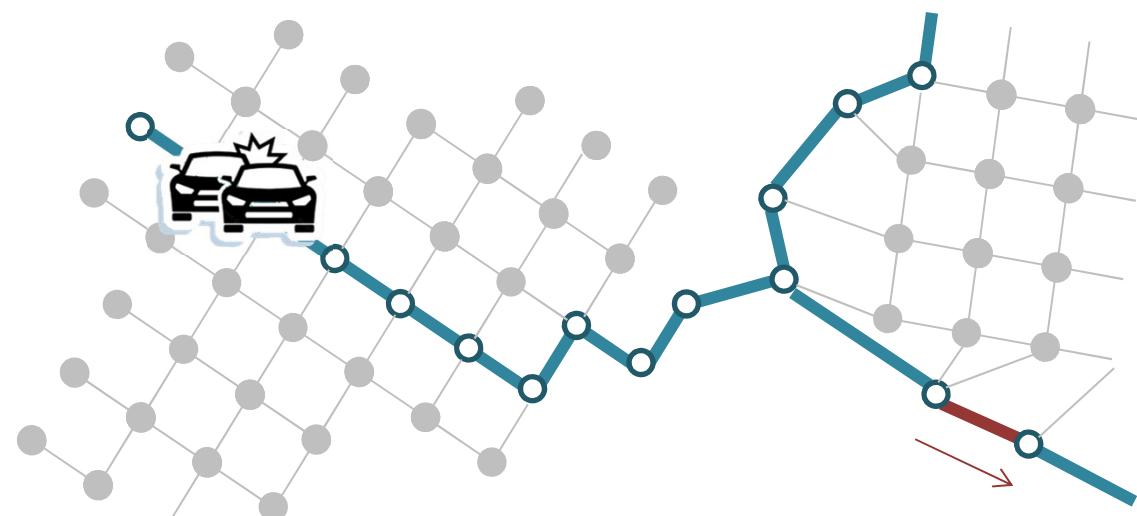
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?

Generate a
low impact accident :

- Far from the edge
- Not on the shortest path
- Very close/far in time



How to characterize an edge

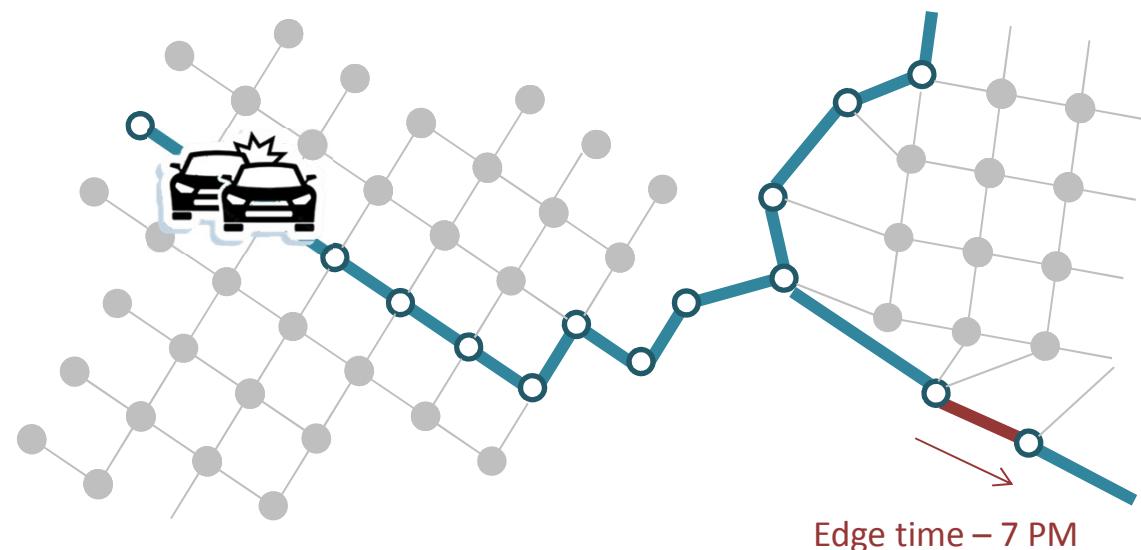
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?

Generate a
low impact accident :

- Far from the edge
- Not on the shortest path
- Very close/far in time



How to characterize an edge

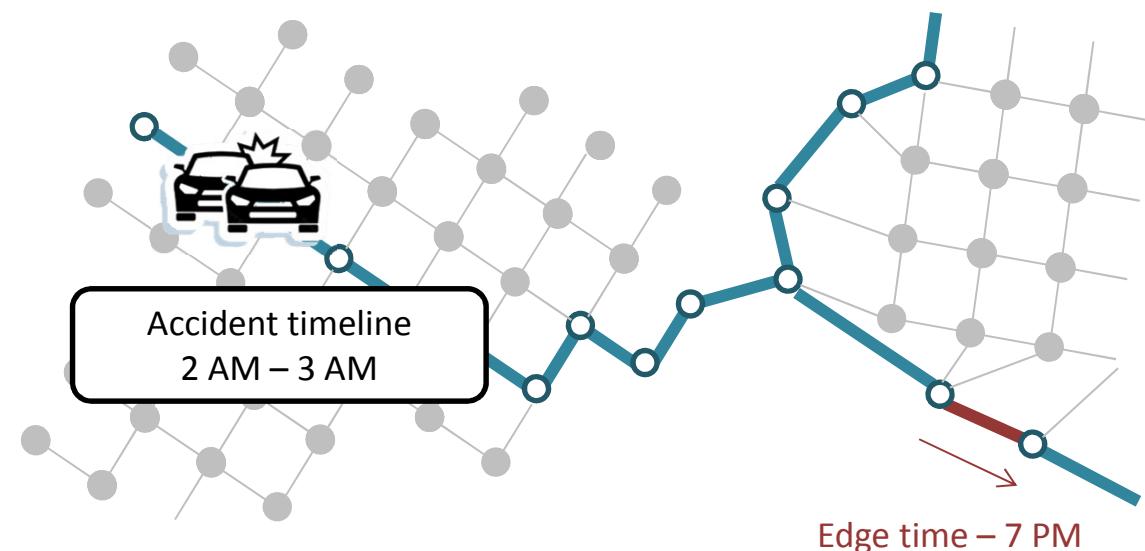
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?

Generate a
low impact accident :

- Far from the edge
- Not on the shortest path
- Very close/far in time



How to characterize an edge

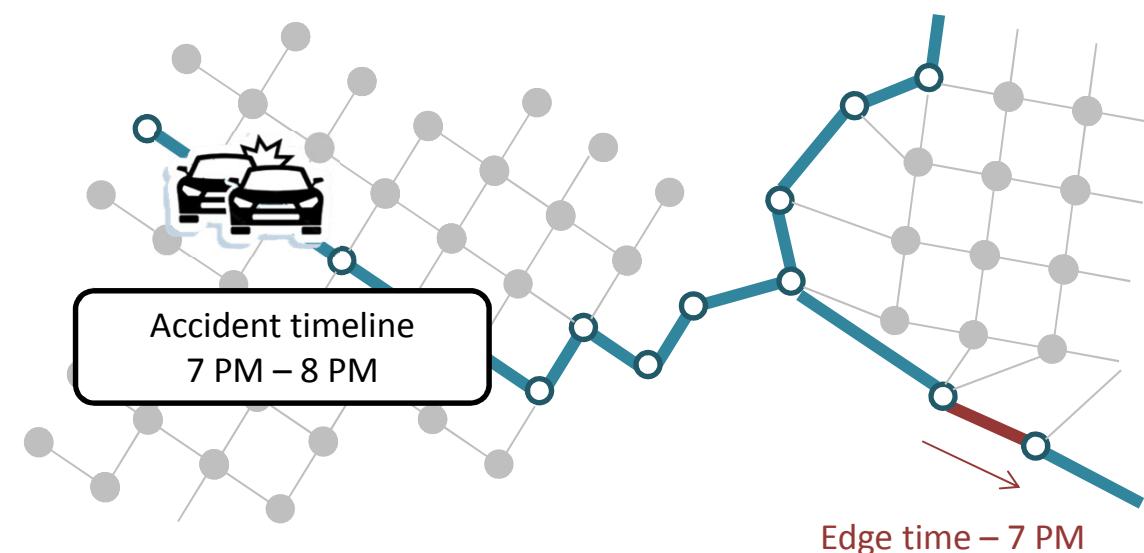
to predict its duration (Graph Features)

How an accident would impact
the prediction of a **road segment's** duration?

What if no
accident?

Generate a
low impact accident :

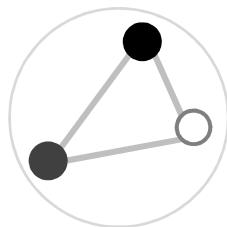
- Far from the edge
- Not on the shortest path
- Very close/far in time



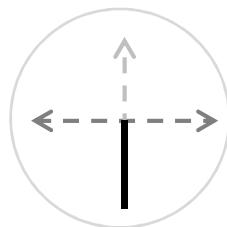
How to characterize an edge

to predict its duration (Graph Features)

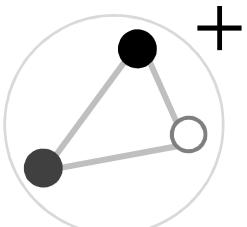
Let's recap our graph based features
that define a road segment



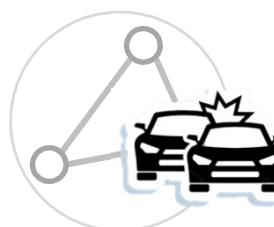
Regular
betweenness



Next intersection
decision

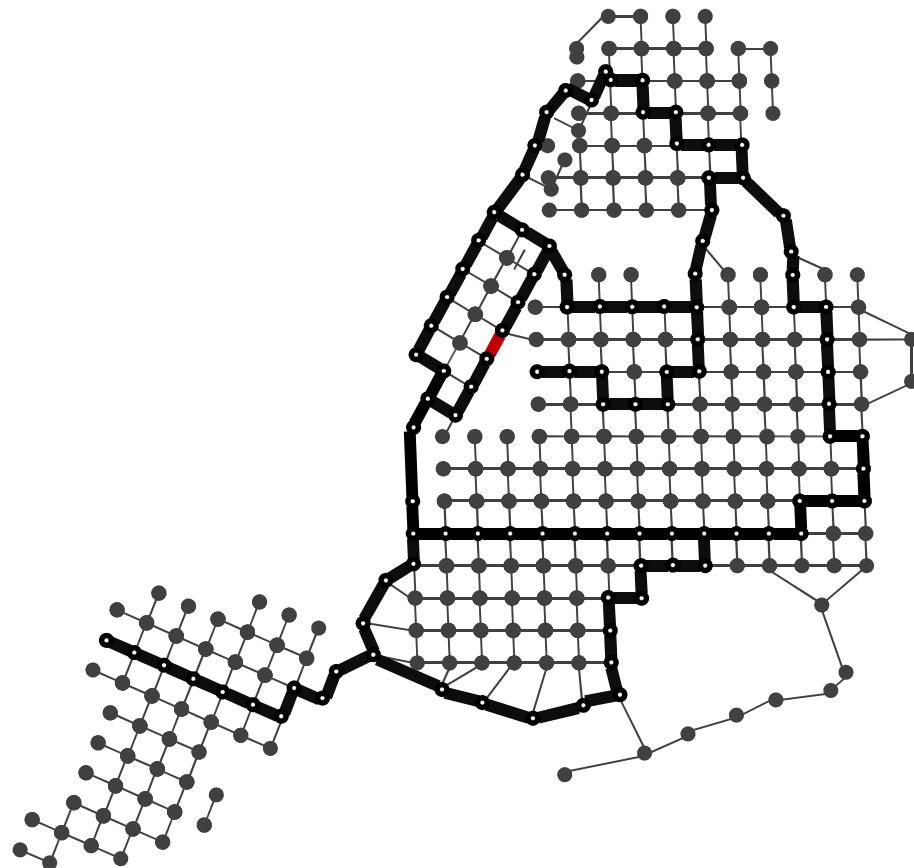


+
Trips based
betweenness

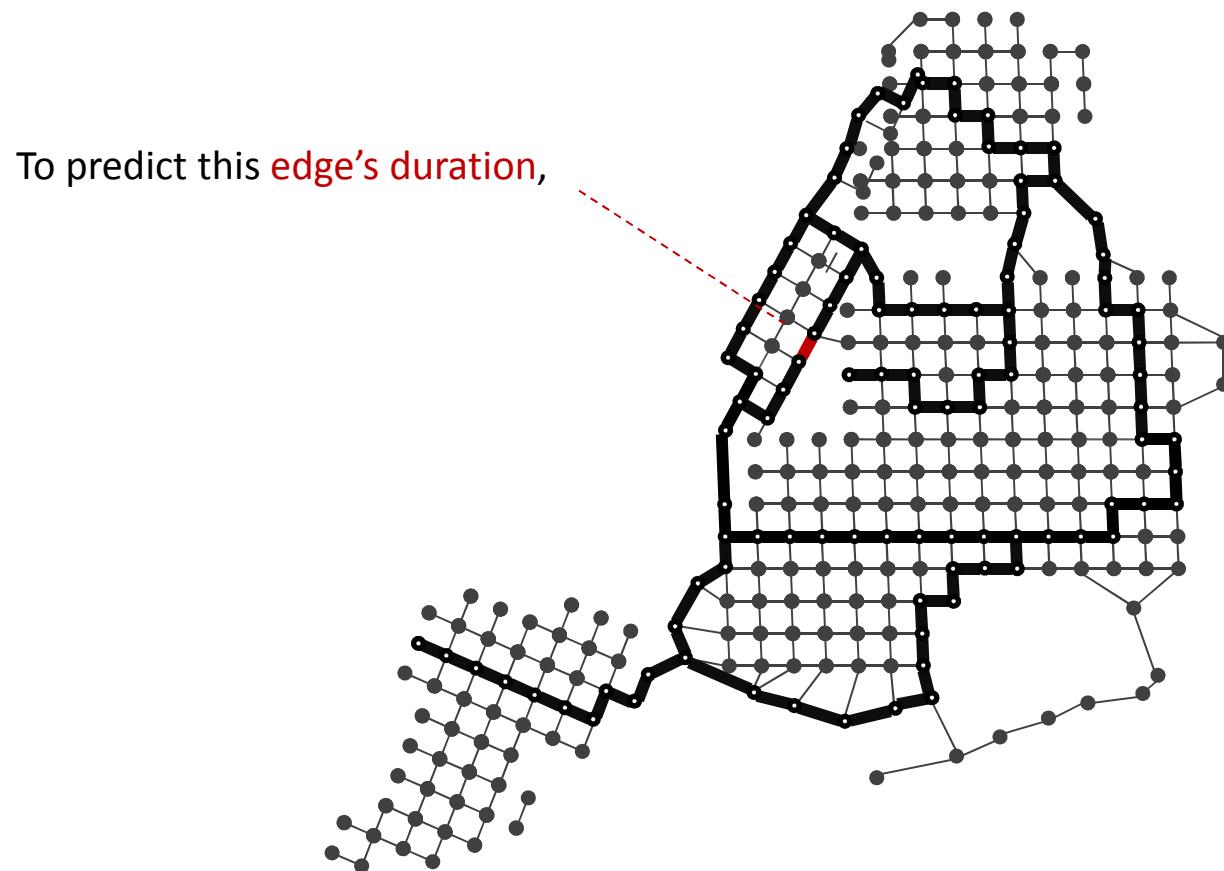


Closest accident
impact

Features used to predict road segment's duration

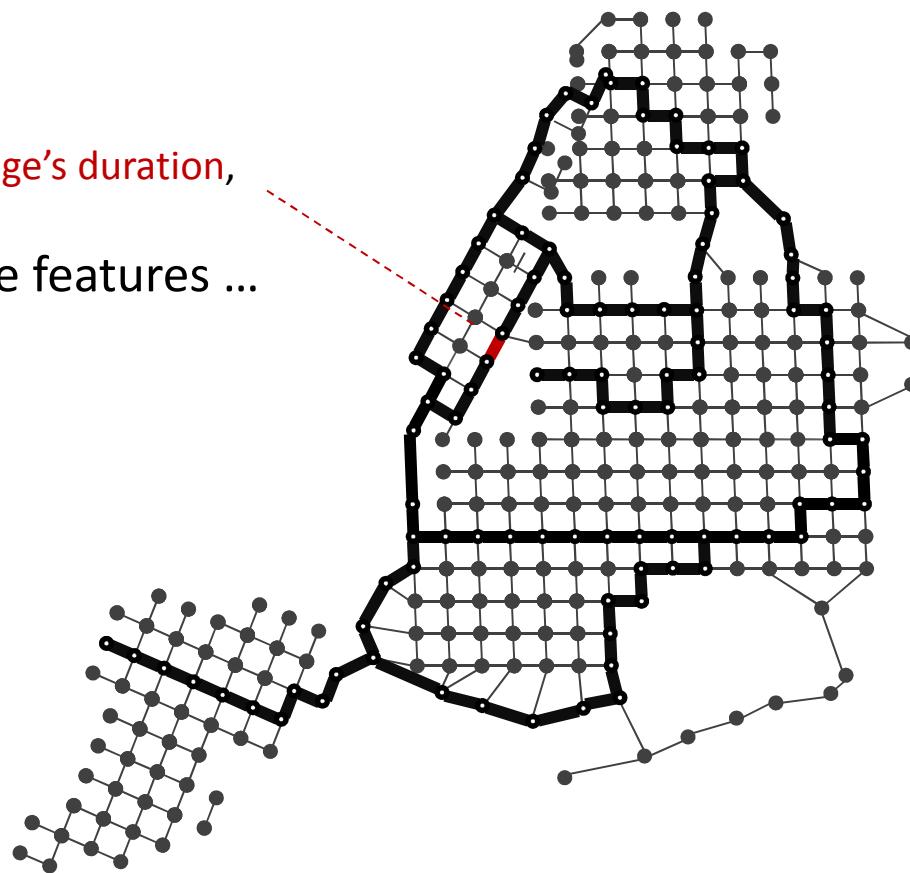


Features used to predict road segment's duration



Features used to predict road segment's duration

To predict this **edge's duration**,
we use those features ...



Features used to predict road segment's duration

Road

- ❖ Length
- ❖ Speed limit
- ❖ Number of lanes
- ❖ Two way edge
- ❖ Elevation gain
- ❖ Number of stops
- ❖ Number of traffic lights

Socio-economic

- ❖ Population density
- ❖ Land use
- ❖ Family composition
- ❖ Household median income

Temporal

- ❖ Starting time
- ❖ Day of the week
- ❖ Weather
- ❖ Luminosity
- ❖ Day of the week
- ❖ Work. d. / Wkd / Holid.
- ❖ N. days to switch
- ❖ N. days from switch

Graph

- ❖ Regular betweenness
- ❖ Trips based betweenness
- ❖ Next intersection decision
- ❖ Closest accident impact :
 - Distances
 - Shortest path
 - Time elapsed
 - Same direction
 - Branches

Features used to predict road segment's duration

Road

- ❖ Length
- ❖ Speed limit
- ❖ Number of lanes
- ❖ Two way edge
- ❖ Elevation gain
- ❖ Number of stops
- ❖ Number of traffic lights



Temporal

- ❖ Starting time
- ❖ Day of the week
- ❖ Weather
- ❖ Luminosity
- ❖ Day of the week
- ❖ Work. d. / Wkd / Holid.
- ❖ N. days to switch
- ❖ N. days from switch

Socio-economic

- ❖ Population density
- ❖ Land use
- ❖ Family composition
- ❖ Household median income

Original Features :

- Number of passengers
- Vendor ID

- ❖ Nearest intersection betweenness

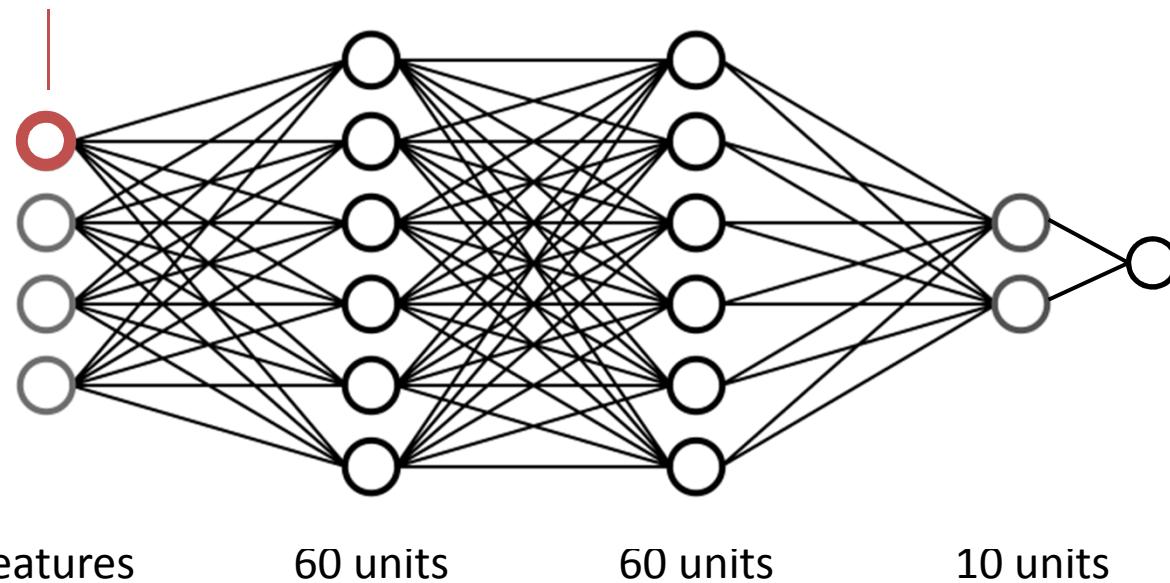
- ❖ Closest accident impact :

- Distances
- Shortest path
- Time elapsed
- Same direction
- Branches

The model

Neural Network
with 3 hidden layers

Predicted edge **starting time**,
Based on the previous edge of the sequence
except for the first one which is known



Batches of sequences of edges are fed to the model
The loss is based on the sum of all edge's duration prediction
(Similar to the RNN)

Data

Road network

- <https://data.cityofnewyork.us/City-Government/road/svwp-sbcd>
- <https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=932>
- <https://osmnx.readthedocs.io/en/stable/> (python package)

Land use

- <https://zola.planning.nyc.gov/about/>
- <https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-gis-zoning.page>

Median income

- <https://data.cccnewyork.org/data/table/66/median-incomes#66/107/62/a/a>

Density

- <https://www1.nyc.gov/site/planning/data-maps/open-data.page>

Accidents

- <https://www1.nyc.gov/>
- <https://crashmapper.org/#/>
- <https://data.cityofnewyork.us/>