# Enriching Traffic Information with a Spatiotemporal Model based on Social Media

Bruno P. Santos
Paulo H. L. Rettore
Heitor S. Ramos
Luiz F. M. Vieira
Antonio A. F. Loureiro

DCC
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO
UFMG
UFAL

# Agenda

- Contextualization
- Motivation

- T-MAPS
  - A case Study

Introduction → Data Acquisition → T-Maps Modeling → Conclusão

- Data sources
- Spatiotemporal coverage
- Twitter as a traffic sensor

- Related work
- Future directions

# Introduction

- The **quality of life** in a city is, in part, a **reflection** of the **mobility** which the city offers

- It implies in the **constant** need for **planning** and **management** of the transportation system.

# Introduction

- Who are interested in?
  - Governments
  - Researchers
  - Industries
- How are they studying and planning the Intelligent Transportation Systems (ITS)?
  - Using raw data sources
    - Inductive loops (velocity, density, and flow)
    - Traffic cameras
    - Origin-destination matrix

# Introduction

- Who are interested in?
  - Governments
  - Researchers
  - Industries
- How are they studying and planning the ITS?
  - Raw Data sources
    - Inductive loops (velocity, density, and flow)
    - Traffic camera
    - Origin-destination matrix

✘ Unfortunately, the **access** to these data sources is, in general, **limited** to those who are connected to *governmental entities or large corporations.*

# Introduction

- The Location-Based Social Media (LBSM)
  - Ex: Twitter, Instagram, and Foursquare

- LBSM as an alternative
  - Low cost
  - Users sharing their
    - Thoughts
    - Viewpoints
    - Their feelings
    - **<u>Traffic conditions</u>**

# Introduction

- In this work, we **investigated** the **traffic scenario** from the **lens** of **LBSM**

- We conducted a study to understand better the **relationship** between the **real traffic** scenario and the **data** provided by **Twitter**

# Introduction

- Goals
  - **LBSM data collection and its characterization** as a data source to describe the traffic scenario
  - **Twitter MAPS (T-MAPS):** we propose a low-cost spatiotemporal model to improve the description of traffic conditions based on tweets.

# Data collection

- Ordinary users X Specialist users

# Data collection

- Ordinary users X **Specialist users**

| Acc name | # Tweets |
|---|---|
| @511NYC | 126925 |
| @TotalTrafficNYC | 20267 |
| @WazeTrafficNYC | 7850 |
| ... | ... |
| @NYC DOT | 3680 |
| **Total of 21 accs:** | **655K** |

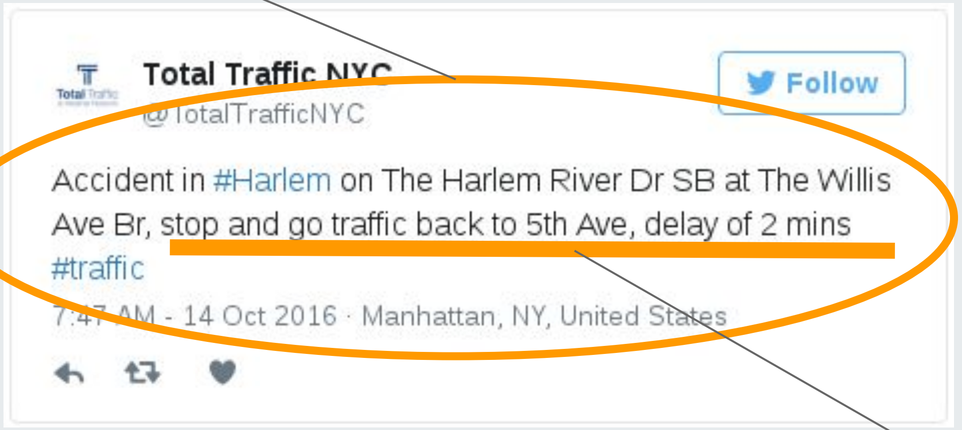
**Total Traffic NYC**
@TotalTrafficNYC

Follow

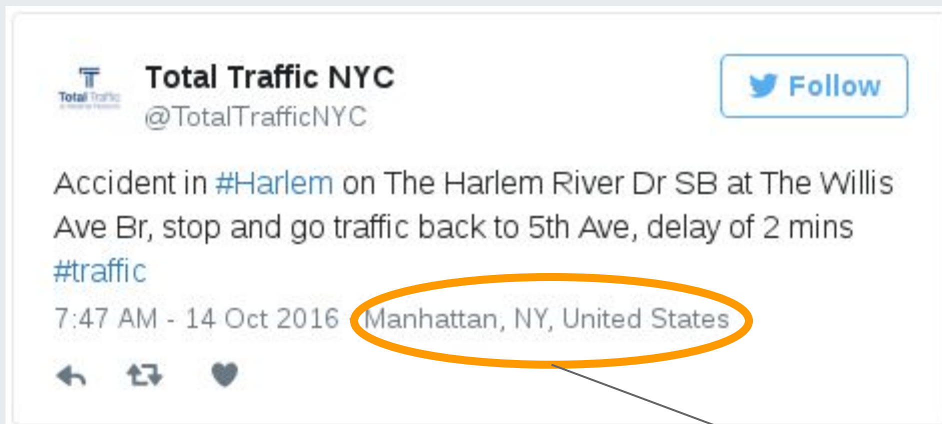
Accident in #Harlem on The Harlem River Dr SB at The Willis Ave Br, stop and go traffic back to 5th Ave, delay of 2 mins #traffic

7:47 AM - 14 Oct 2016 · Manhattan, NY, United States

**Traffic event description**




**Traffic condition**

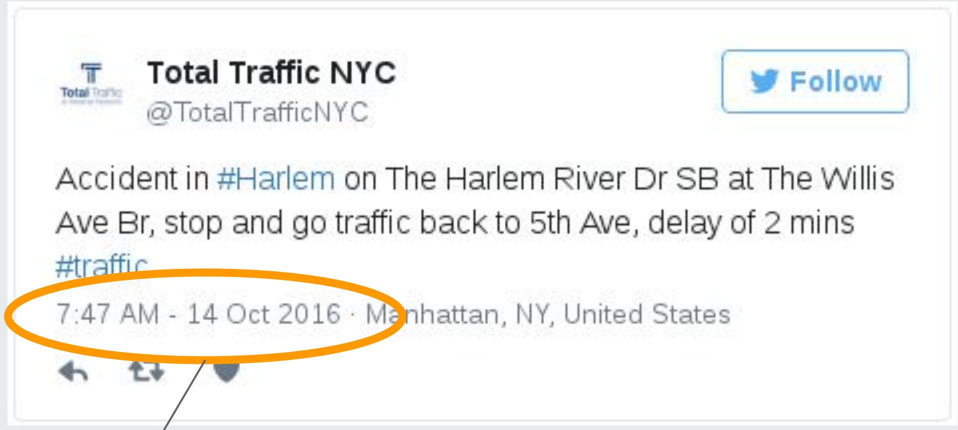


Total Traffic NYC
@TotalTrafficNYC
Follow

Accident in #Harlem on The Harlem River Dr SB at The Willis Ave Br, stop and go traffic back to 5th Ave, delay of 2 mins #traffic

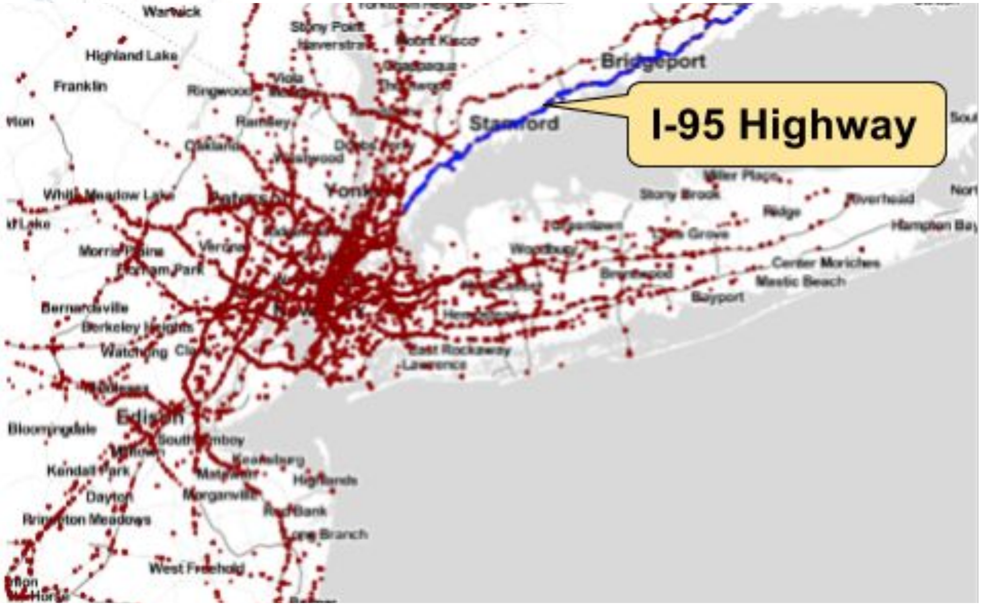
7:47 AM - 14 Oct 2016 Manhattan, NY, United States

**Location description. Some Tweets have geotag.**

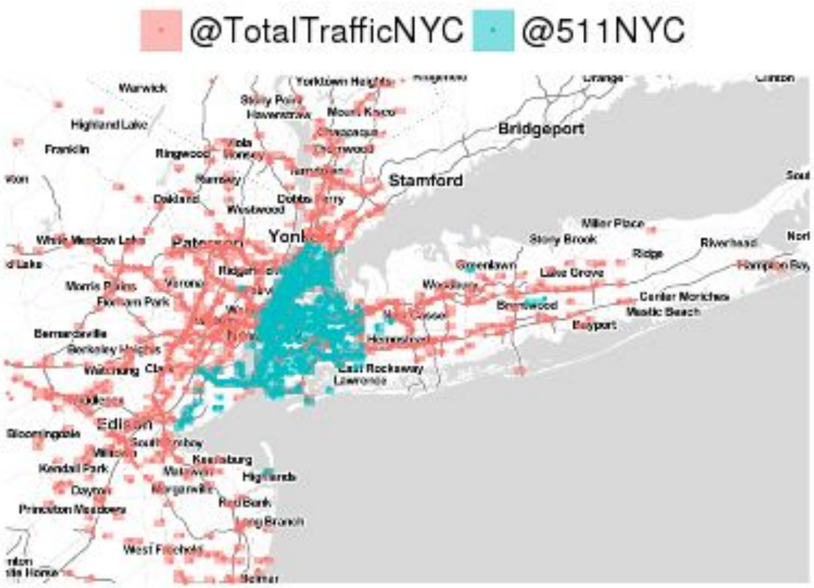



**Tweet input hour**

## Data collection
# Spatial coverage



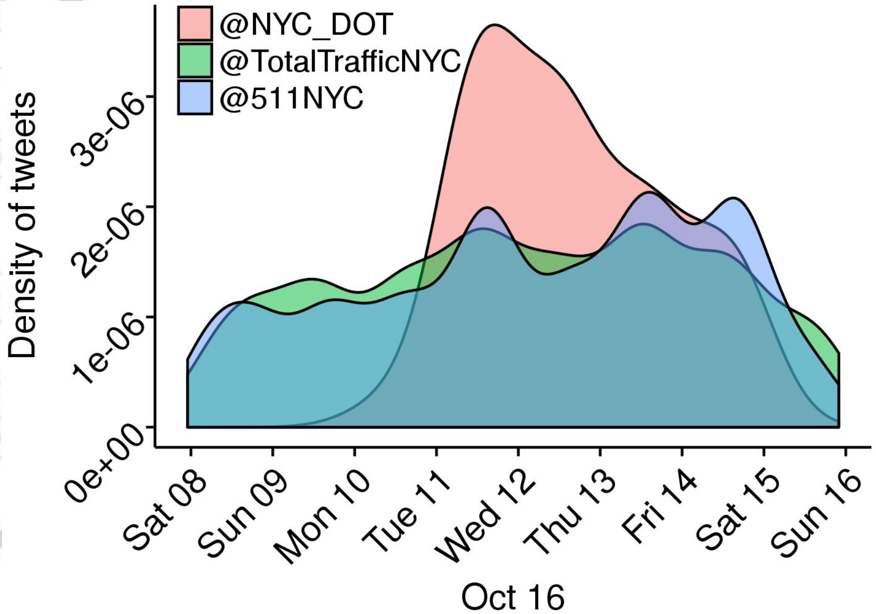

Tweets in NY

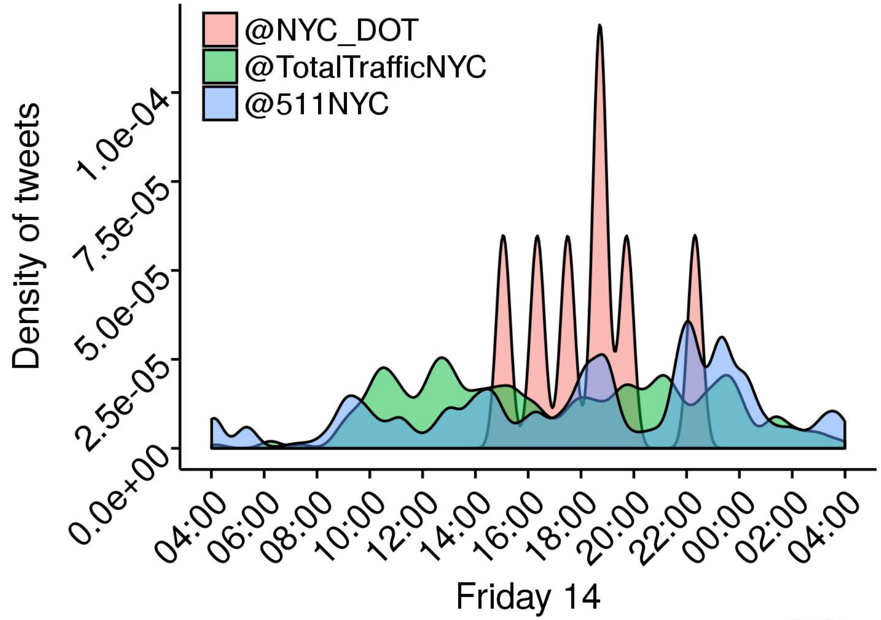



Spatial coverage in NY of two specialist accs

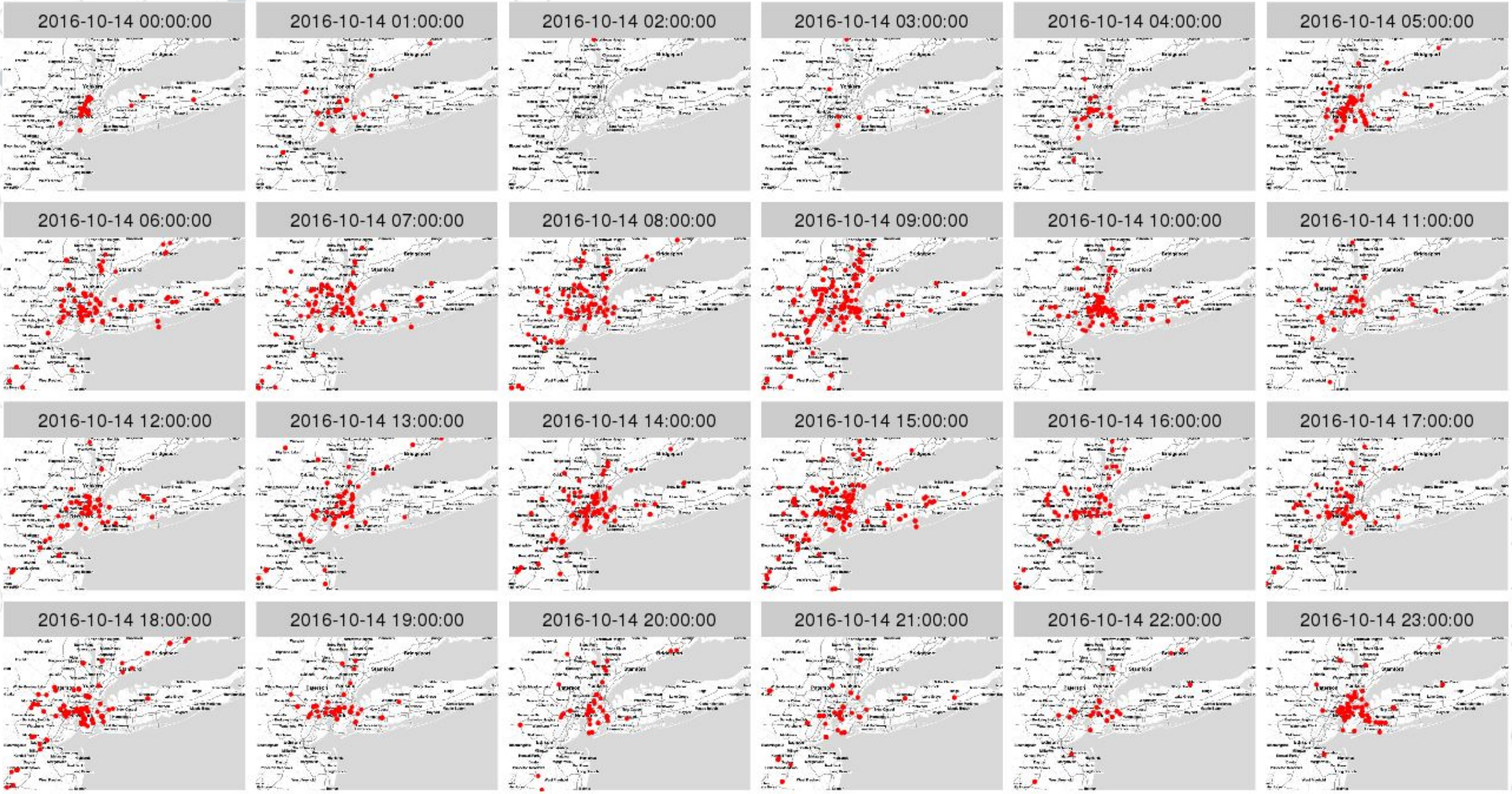
# Data collection

# Temporal coverage



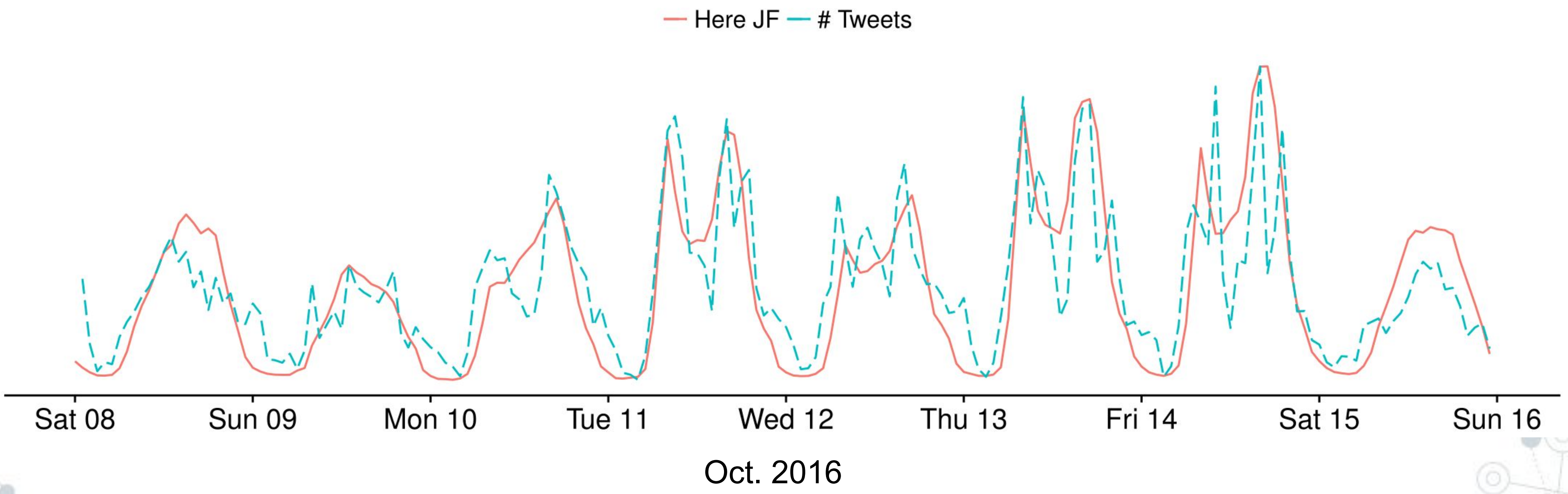

**Temporal coverage of three accs**


2016-10-14 00:00:00
2016-10-14 01:00:00
2016-10-14 02:00:00
2016-10-14 03:00:00
2016-10-14 04:00:00
2016-10-14 05:00:00
2016-10-14 06:00:00
2016-10-14 07:00:00
2016-10-14 08:00:00
2016-10-14 09:00:00
2016-10-14 10:00:00
2016-10-14 11:00:00
2016-10-14 12:00:00
2016-10-14 13:00:00
2016-10-14 14:00:00
2016-10-14 15:00:00
2016-10-14 16:00:00
2016-10-14 17:00:00
2016-10-14 18:00:00
2016-10-14 19:00:00
2016-10-14 20:00:00
2016-10-14 21:00:00
2016-10-14 22:00:00
2016-10-14 23:00:00

# Twitter as a traffic sensor

- How related are the tweets to the traditional traffic sensor?

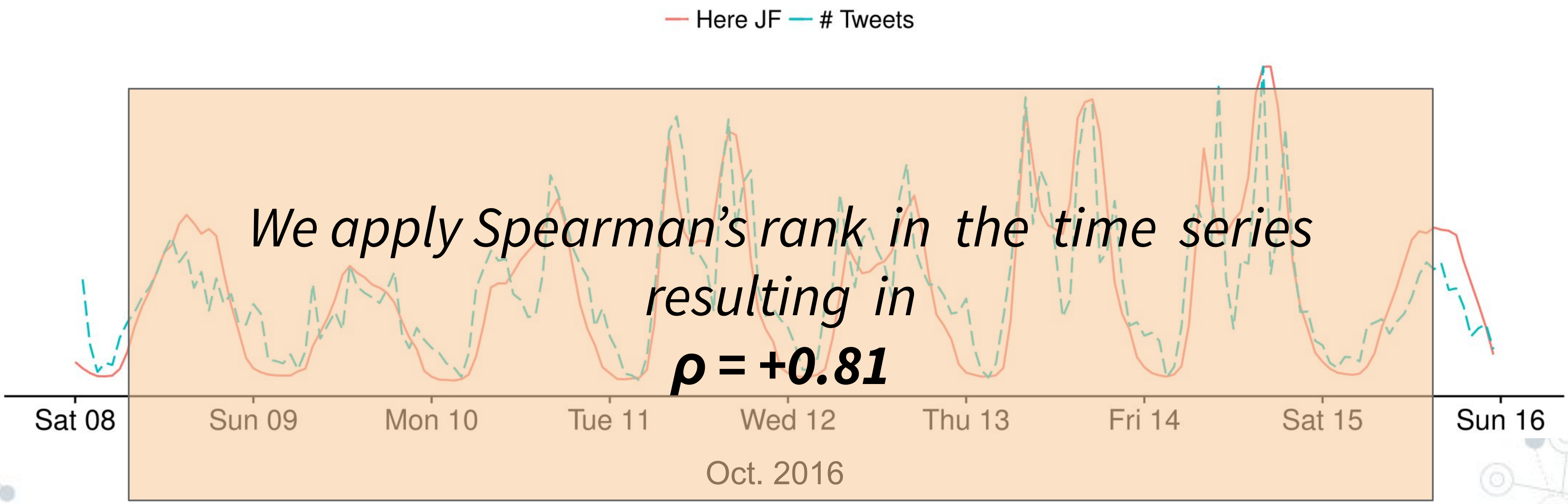
# Twitter as a traffic sensor




**Data:**

- **Jam Factor (Here developer API)**
- ***# tweets***

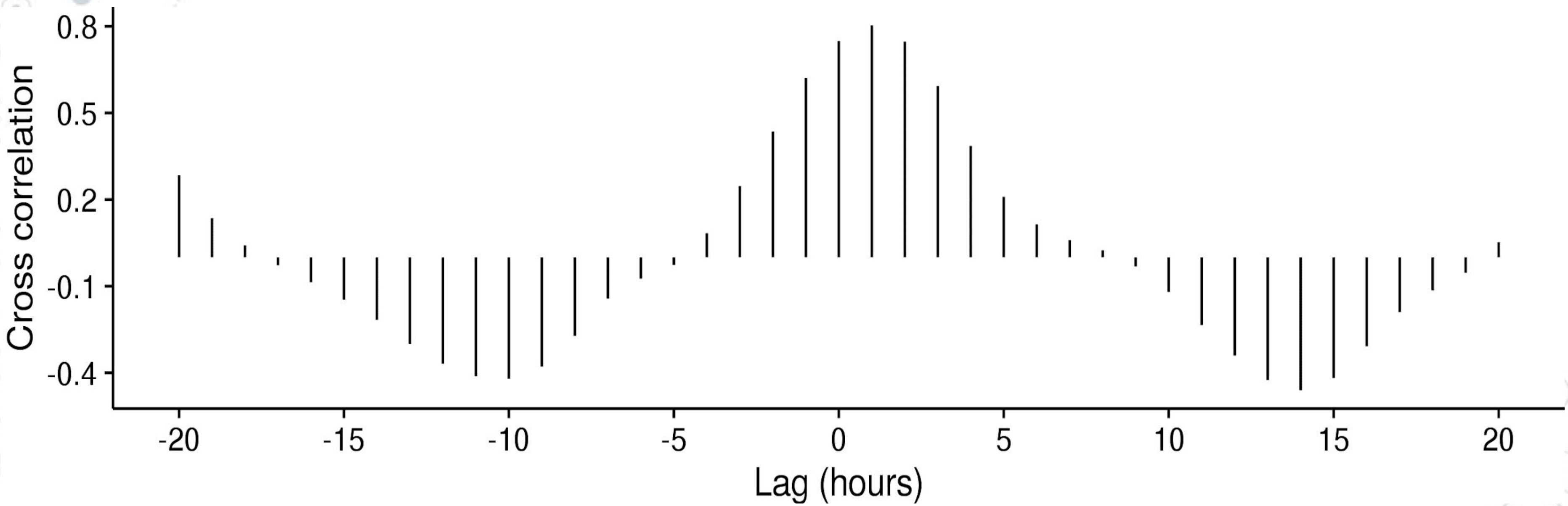
# Twitter as a traffic sensor



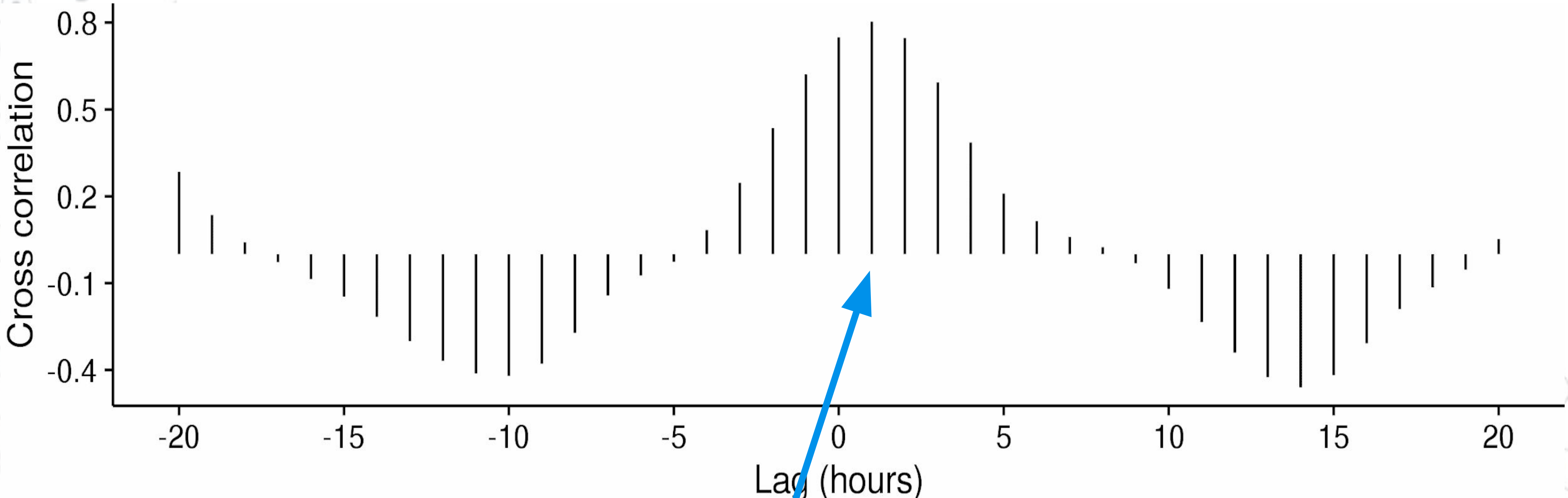

*We apply Spearman's rank in the time series resulting in*
$\rho = +0.81$

**Data:**
- **Jam Factor (Here developer API)**
- ***# tweets***

# Twitter as a traffic sensor



# Twitter as a traffic sensor




The highest correlation (0.8) appears when the lag is +1 meaning
**#tweets curve is 1 hour ahead of JF**

# Twitter MAPS (T-MAPS)

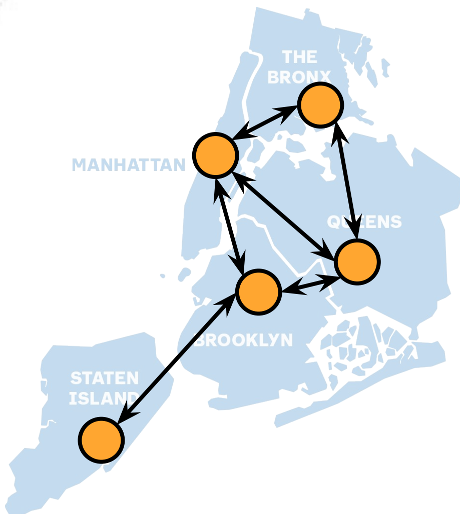
- Modeling Process
  1. Data acquisition
  2. Filtering and data fusion process
  3. Metrics

# T-MAPS
# Step 1 - Data Acquisition


New York City boundaries

Data from LBSM

# T-MAPS
# Step 2 - Filtering and Data fusion



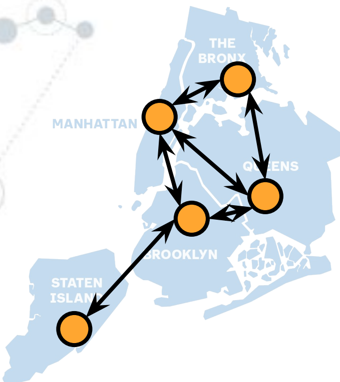

G(V,E)

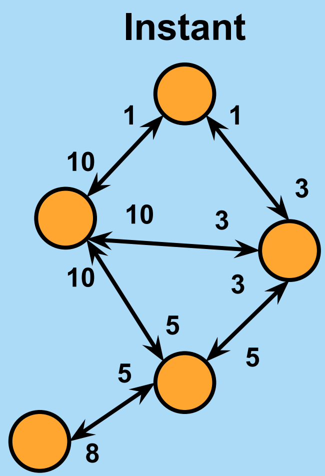
- V(G) = map sub-regions
- E(G) =adjacent regions

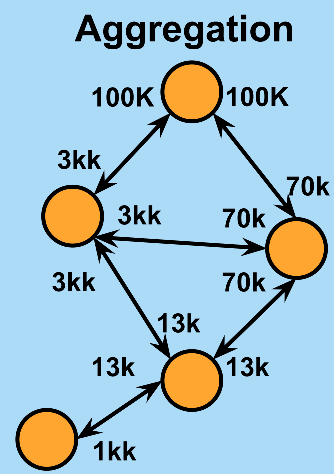
- Spatiotemporal assignment
- Filtering and bind data to regions
- Remove data inconsistent

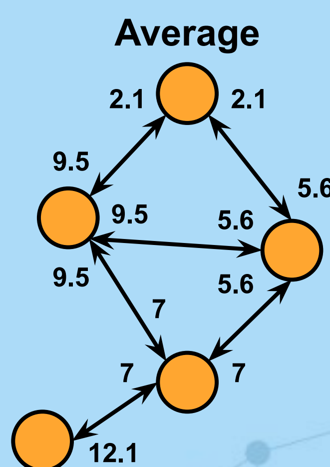
## T-MAPS

# Step 3 - Data Fusion and Metrics



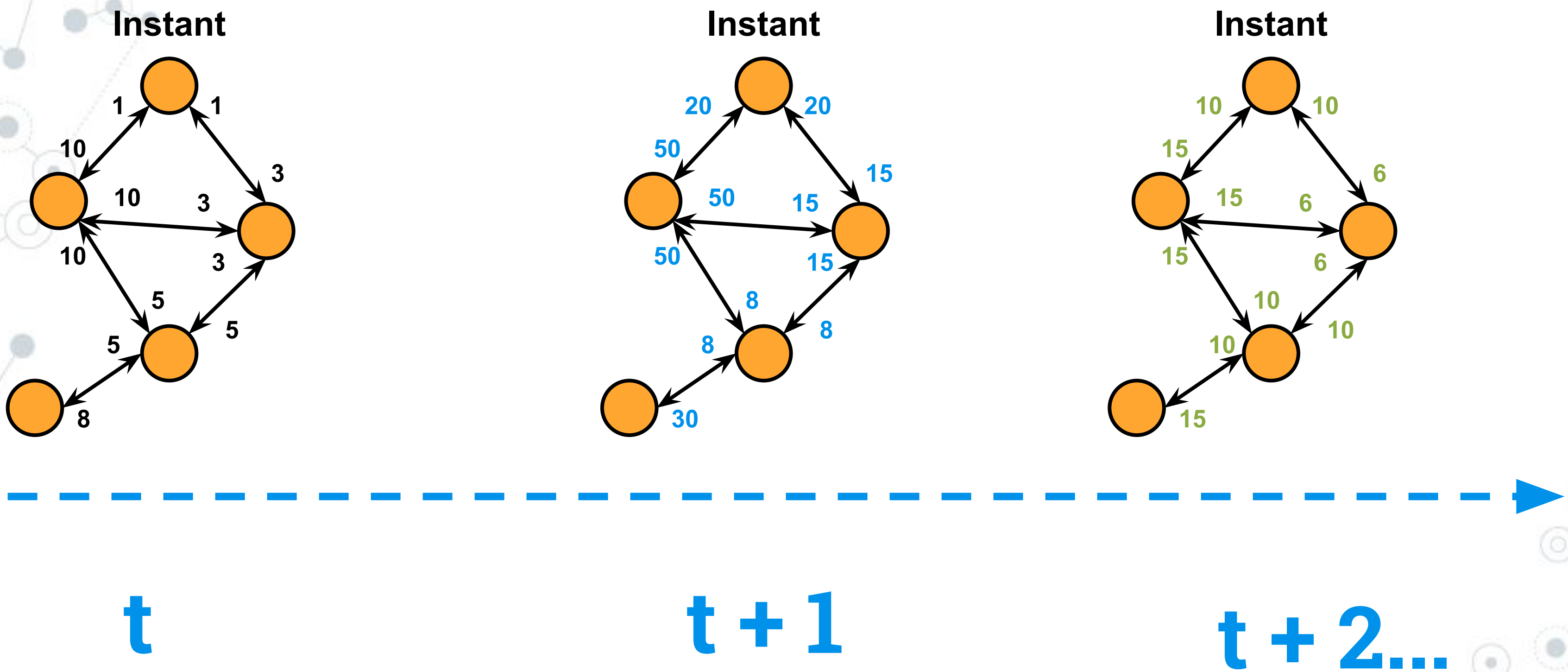



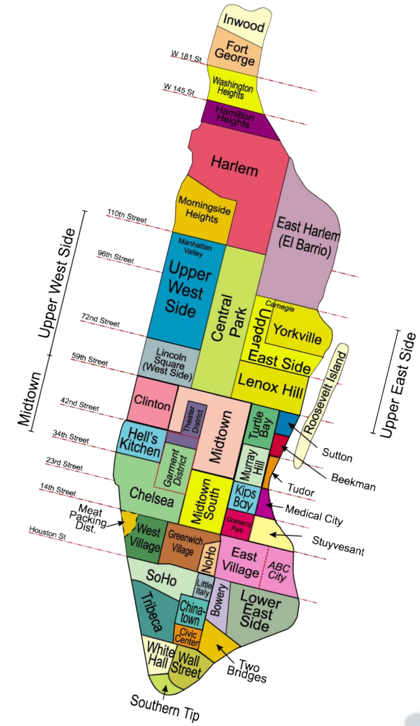
## T-MAPS

# Time Discretization - metric - instant



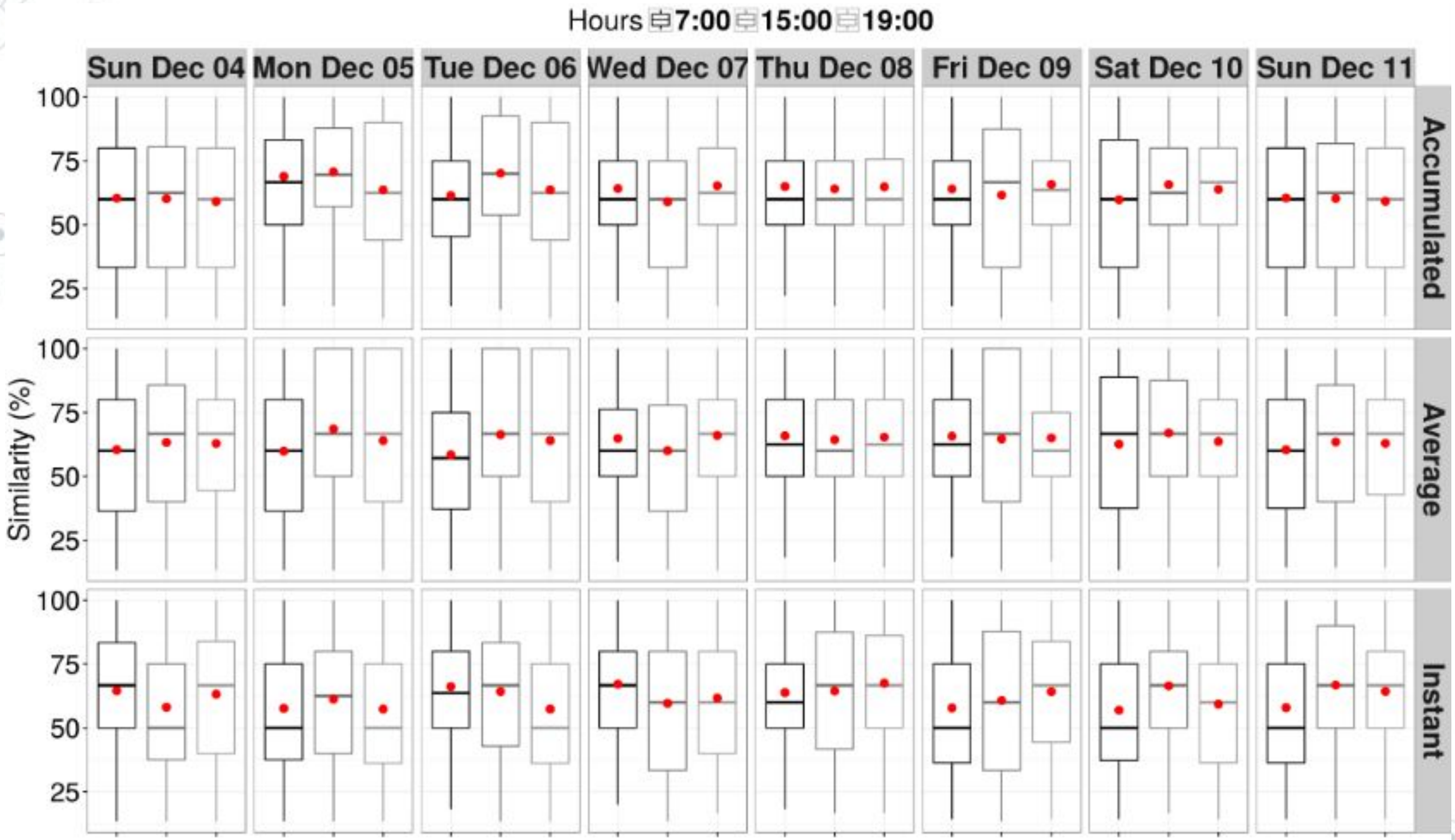
# A Case Study - T-MAPS

- Manhattan
  - 29 official borough
  - 21 specialist accounts from Twitter.
    - ~ 280 K geotagged *tweets*
      - Oct - Dec 2016
- T-MAPS
  - Shortest Path - Dijkstra Algorithm
- Google Directions



## A Case Study

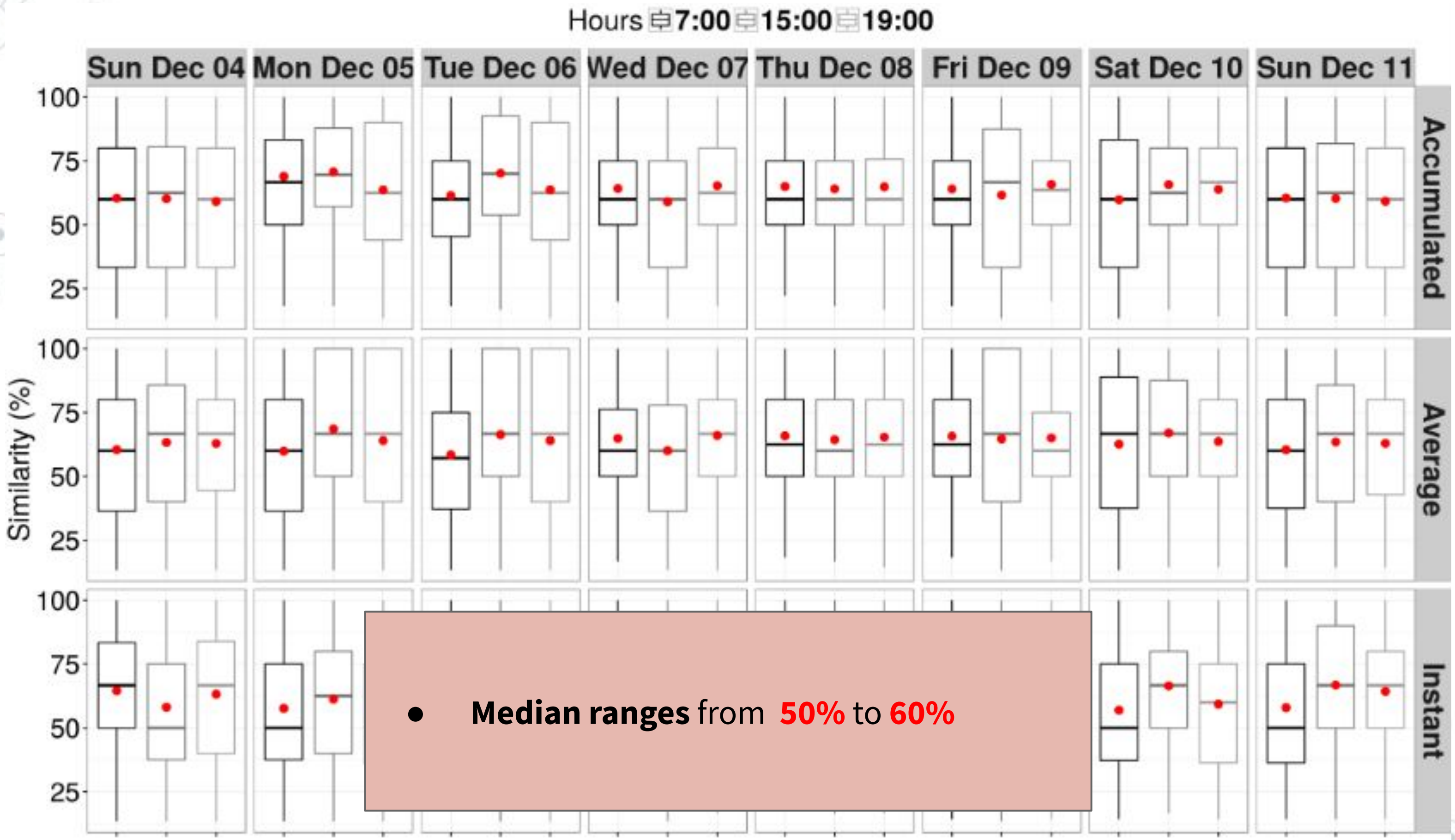
# T-MAPS Applicability




Route similarity between T-MAPS and Google Directions (dots represent the mean)

## A Case Study

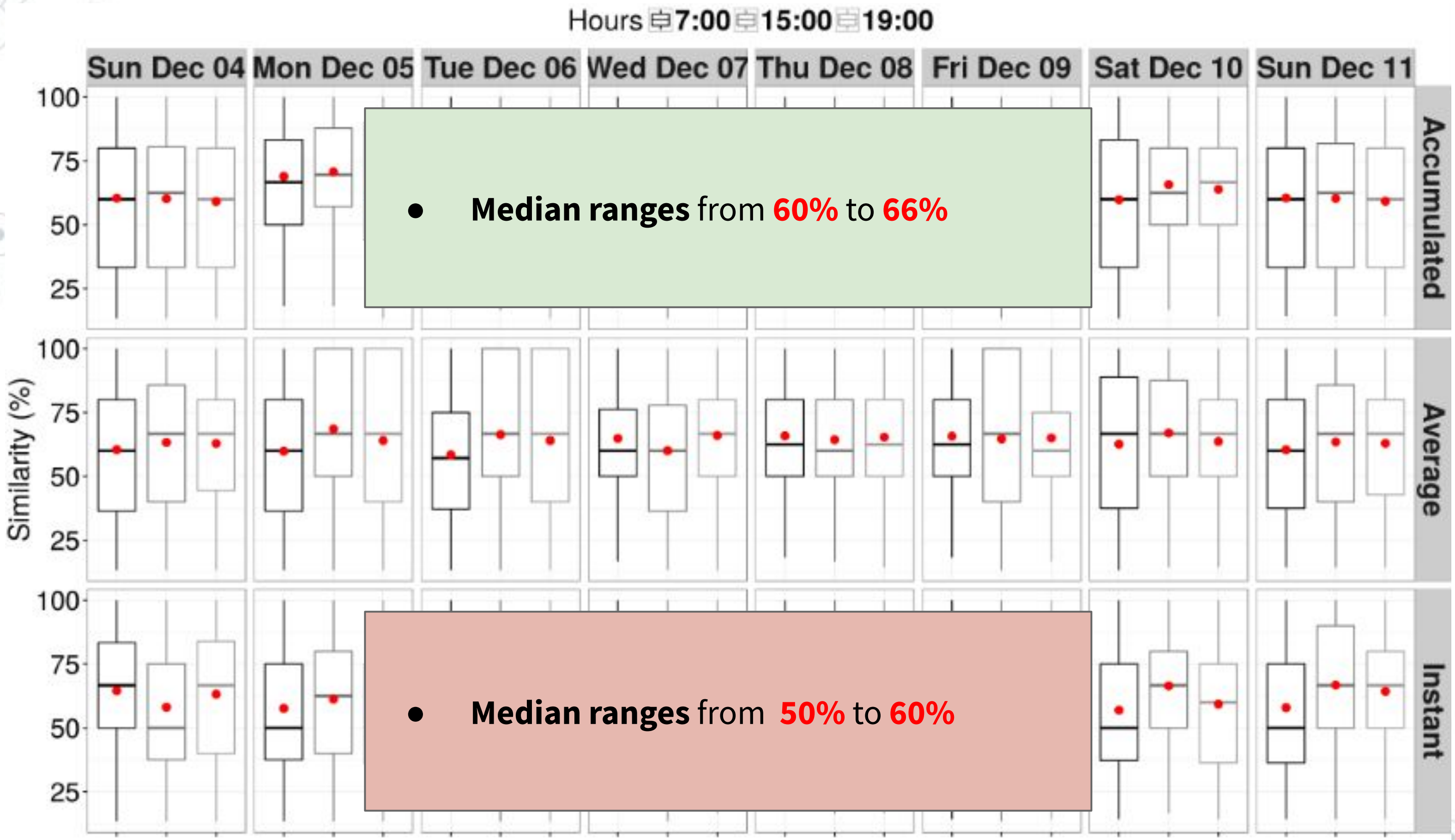
# T-MAPS Applicability




Route similarity between T-MAPS and Google Directions (dots represent the mean)

## A Case Study

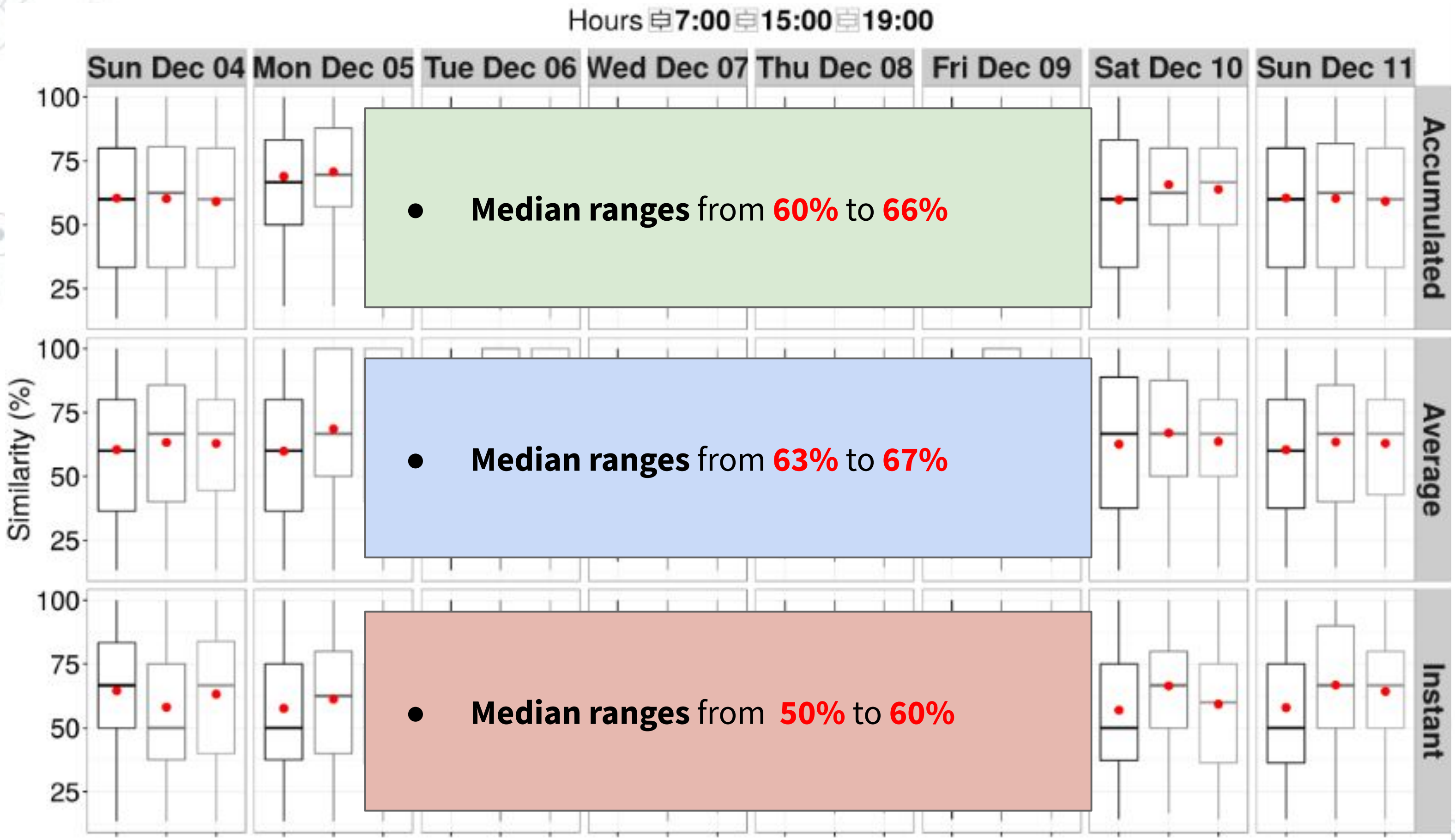
# T-MAPS Applicability




Route similarity between T-MAPS and Google Directions (dots represent the mean)

## A Case Study

# T-MAPS Applicability




Route similarity between T-MAPS and Google Directions (dots represent the mean)

# T-MAPS
# ROUTE DESCRIPTION SERVICES

**LBSM data collection**

- RM punctuation
- RM stop words

✘ '/,.;*"?!/...

✘ the, is, at, which…

R library:
[syuzhet, tm, stringr, wordcloud]

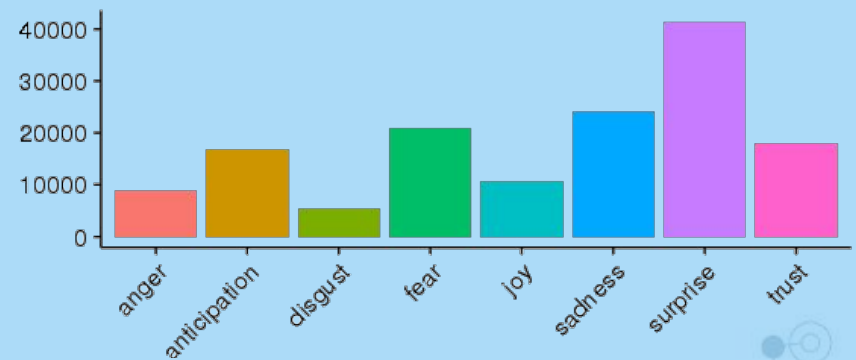
# T-MAPS
# ROUTE DESCRIPTION SERVICES

- **LBSM data collection**
  - RM punctuation
  - RM stop words
- **Stemming**

Jamming, jammed → **Jam**
Ave, Av→ **Avenue**
St →**Street**

# T-MAPS
# ROUTE DESCRIPTION SERVICES

- **LBSM data collection**
  - RM punctuation
  - RM stop words
- **Stemming**
  - **Score**



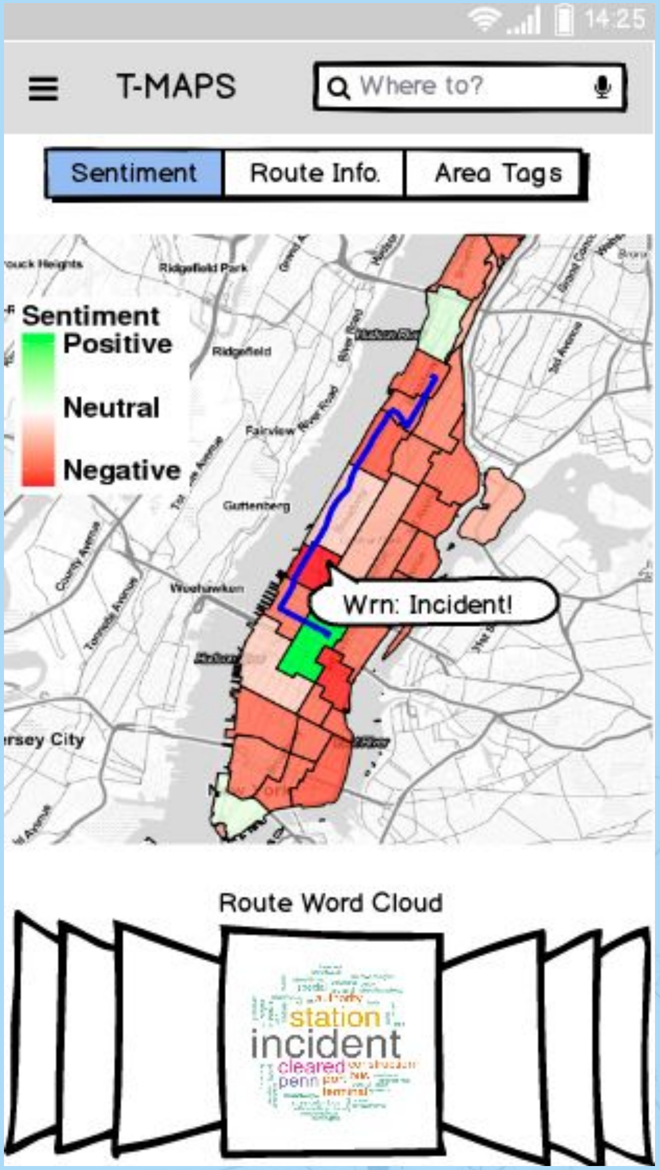
# T-MAPS
# ROUTE DESCRIPTION SERVICES

- **LBSM data collection**
  - RM punctuation
  - RM stop words
- **Stemming**
- **Score**
- **Pos ou Neg?**

| ## | Ang. | Anticip. | Disg. | Fear | Joy | Sad. | Surpri. | Trust | Neg. | Pos. |
|---|---|---|---|---|---|---|---|---|---|---|
| ## 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| ## 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ## 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ## 4 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 2 | 1 | 2 |
| ## 5 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 2 |
| ## 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# T-MAPS
# ROUTE DESCRIPTION SERVICES

- **LBSM data collection**
  - RM punctuation
  - RM stop words
- **Stemming**
  - **Score**
  - **Pos ou Neg?**
  - **Routes and Map**



# T-MAPS
# ROUTE DESCRIPTION SERVICES

**LBSM data collection**

RM punctuation

RM stop words

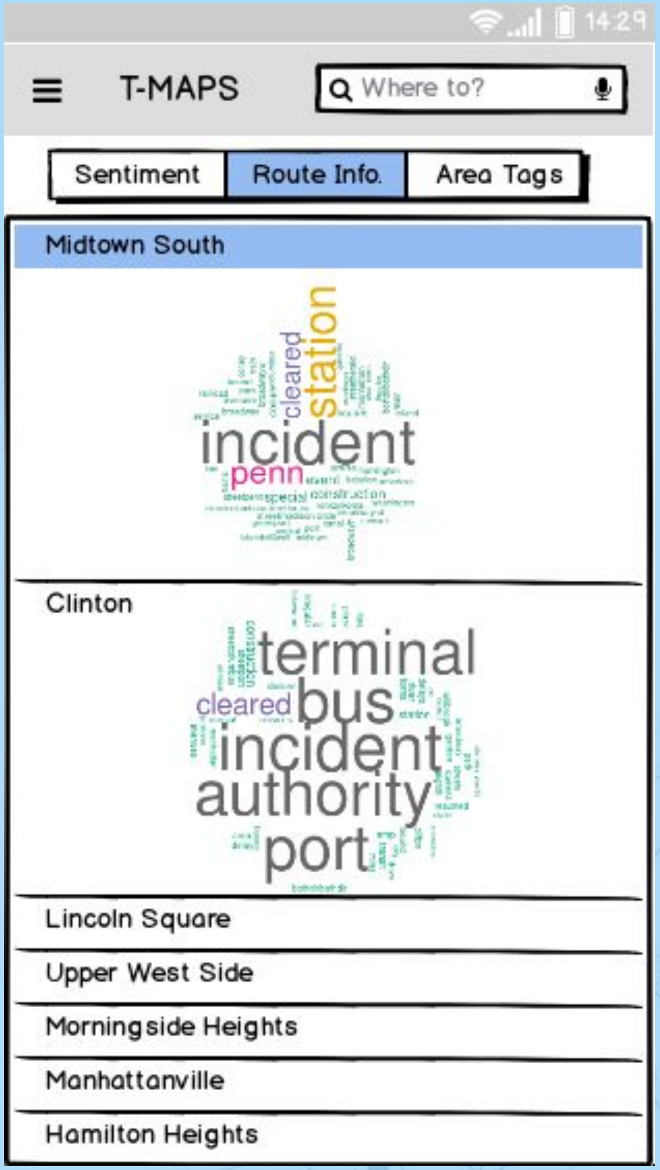
**Stemming**

**Word cloud**

**Score**

**Pos ou Neg?**

**Routes and Map**



# Conclusions

- We presented the T-Maps, a low-cost spatiotemporal model to enhance traffic and transit navigation context, using tweets
  - Three route description services, Route Sentiment, Route Information, and Area' Tags.
  - The similarity reached **62%**, and for a quarter of the evaluated trajectories, the similarity achieved up to **100%.** Compared with Google Direction route recommendation

# Conclusions

- As future work
  - Extend the T-Maps applying strategies to process the data and offer more valuable information
  - Employ regular users accounts from LBSM and uses reputation models to handle conflicting information
  - Extend T-Maps to larger regions, taking into account the computational problem

# Thanks!

## Questions?

{bruno.ps, rettore, lfvieira, loureiro}@dcc.ufmg.br
heitor@ic.ufal.br