Big Data Analysis for Multi-Modal Performance Analysis

Aude Marzuoli, Emmanuel Boidot, Eric Feron, Ashok Srivastava and
Philippe Monmousseau

NEXTOR Symposium
Georgia Institute of Technology

Friday, October 26, 2018

Georgia
Tech

Georgia
Tech

Content

Georgia
Tech

Current situation

Let's all agree:

Airports are the main bottlenecks of the Air Transportation System

Capacity constraints and inefficiencies at airports entry roads, parking, security, immigration, customs, gates, ramps, runways

$$\Rightarrow \text{congestion and delays}$$

Many different stakeholders that do not necessarily rely on each other to take decisions.

Georgia
Tech

Current situation

So far, information is dispersed:

- Airlines only have access to information about their passengers
- Airports have access to:
  - customs and security records
  - queue lengths
  - shuttle traffic, parking occupancy...
- Third-parties can collect traces from WiFi or Bluetooth beacons

No one has an overall view of the situation, only partial prism.
Currently Bureau of Transportation Statistics (BTS) provides most used metrics for the US Air Transportation System.

Georgia
Tech

Fact

Many studies have shown that flight delays do not represent total delays of door-to-door journey of passengers.

Georgia
Tech

Fact

Many studies have shown that flight delays do not represent total delays of door-to-door journey of passengers.

New prism is necessary

Passengers are the *raison d'etre* of the ATS and should be at the center of the evaluation of its performance

New impulse for a shift from flight-centric metrics to passenger-centric metrics: NextGen in the USA and SESAR in Europe.

Georgia Tech

Content

Georgia
Tech

Description of the data set

Data set

> March 27th, 2017 - April 9th, 2017 (2 weeks) :
> 5 billion records/ping per day

Each ping contains:

- Anonymized user id
- Timestamp
- Approximate location (longitude + latitude)

No data from non-consentant users, and only anonymous use of data.

Georgia
Tech

How to find passengers?
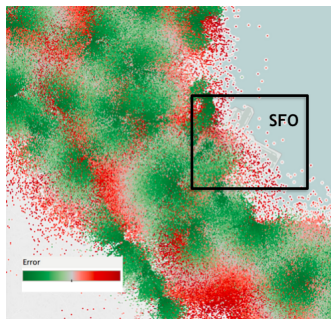
Figure: Zoom on mobile phone location data for one day near San Francisco International Airport.

Approach:

- Top 35 airports delimited as boxes
- Choose users appearing in 2 different airports on the same day
- Remove those within two airports in same metropolitan area
- Keep only first and last time stamp within each airport boundary

$\approx 595,000$ records

Georgia Tech

Is it consistent?

Rule of thumbs

$\left.\begin{array}{l} 875,000 \text{ passengers/day} \\ 78.1\% \text{ mobile phone penetration} \\ 30\% \text{ market share} \end{array}\right\} \Rightarrow \approx 228,000 \text{ "visible" passengers}$
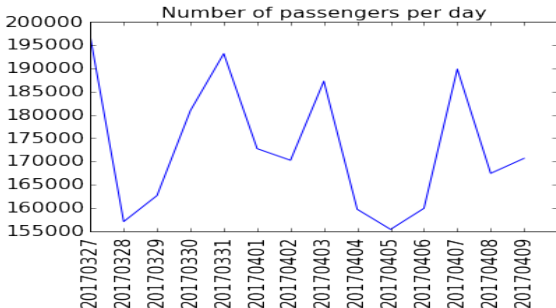
Figure: Number of passengers per day extracted from data

Content

Georgia
Tech

Passenger volume per airport

First Validation Step

Contrast passenger volumes with flight volumes from BTS

Georgia
Tech

Passenger volume per airport
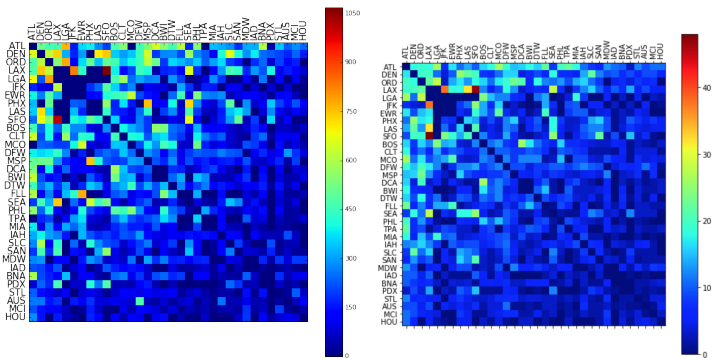
(a) Number of passengers per origin-destination airport pair, estimated from mobile phone location data.

(b) Number of flights flown per origin-destination airport pair, reported by BTS

Strikingly similar patterns! (Data from 03-27-17)

Passenger volume per airport

First Validation Step

Contrast passenger volumes with flight volumes from BTS

- Hub-and-spoke structure of the National Airspace
- Same origin-destination pair with max traffic: LAX-SFO
- "Anomalies" from mobile data set actually validated by BTS (e.g. SAN-SFO, SLC-DEN, AUS-DFW, BWI-BOS)

Georgia
Tech

Time spent in airports

What can be said about the time each passenger spends at the airport?

- Airlines know when/if passenger checked-in and when boards plane
- Airports can know parking in/out ticket
- Customs know when accepted in/out
- Security knows when security check is conducted
- WiFi can help with tracking some flow
- etc.

Overall, no one knows exactly how long a passenger stays at an airport!

Georgia
Tech

Time spent in airports

Distribution of time spent in any airport by passengers, from March 27th to April 9th, 2017.
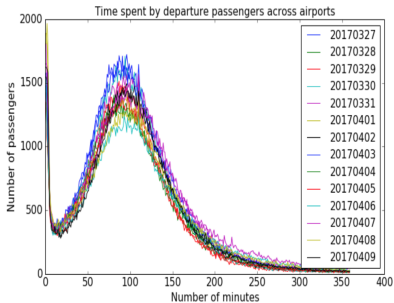
(c) Departing passengers.
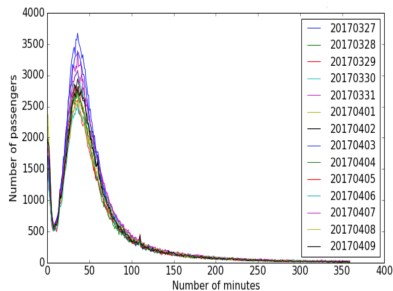
(d) Arrival passengers

Time spent in airports
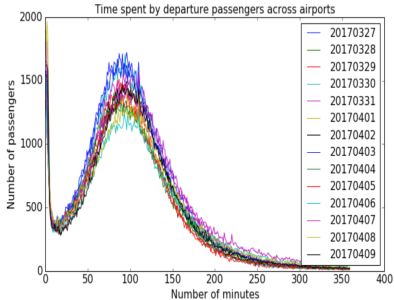
(e) Departing passengers.
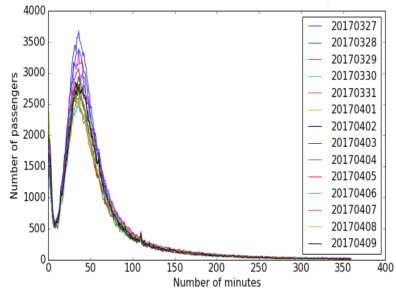
(f) Arrival passengers

Consistent with common sense: $\approx$ 90 min average waiting time for domestic departure.

Georgia
Tech

Monitoring degraded operation conditions

Assessing performance

Can mobile phone data tell us something about delays and cancellations?

Georgia
Tech

Monitoring degraded operation conditions
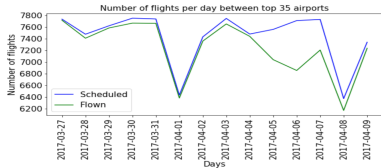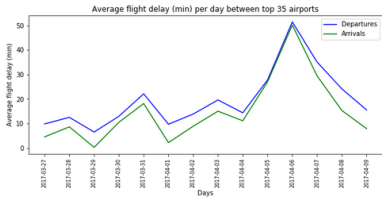
(h) Number of flights scheduled vs flown per day across top 35 airports.

(g) Flight delay per day from the Bureau of Transportation Statistics.
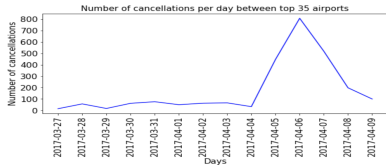
(i) Number of cancellations per day across top 35 airports.

Data from BTS

Monitoring degraded operation conditions
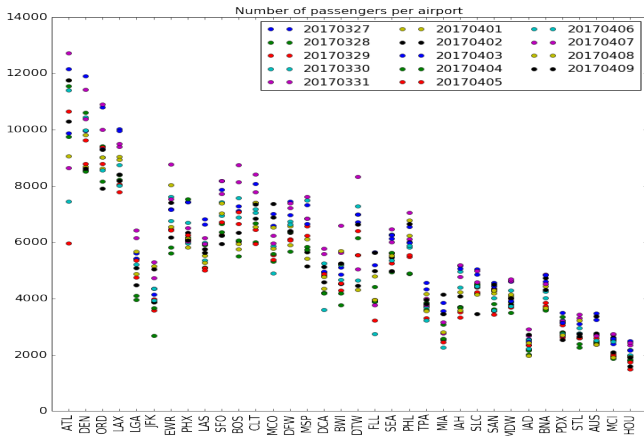
Figure: Number of passengers per day per airport from mobile data set.

Monitoring degraded operation conditions
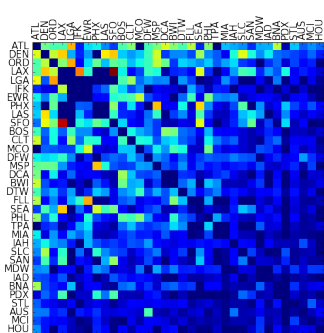
(a) Mobile phone data from 03-27-17   (b) Mobile phone data from 04-05-17
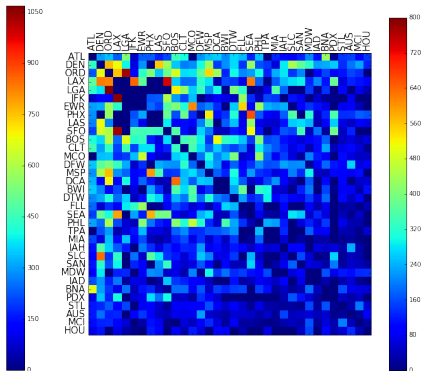
Atlanta lost its hub status!

Georgia
Tech

Monitoring degraded operation conditions

Assessing performance

Can mobile phone data tell us something about delays and cancellations?

We were able to see that there was a major disturbance within ATL and JFK traffic, let's dig into it!

Georgia
Tech

ATL degraded conditions monitoring
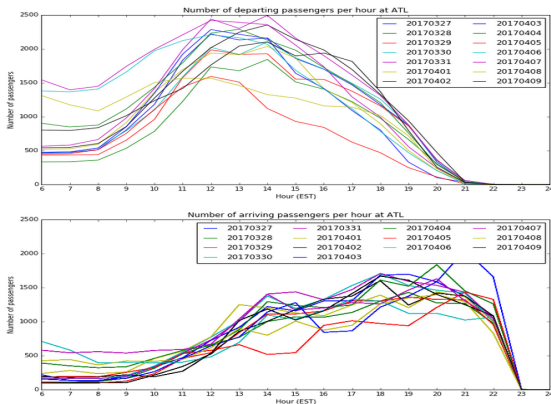
Number of passengers at departure and at arrival per hour at ATL airport over two weeks: 04-05-17 stands out!

ATL degraded conditions monitoring
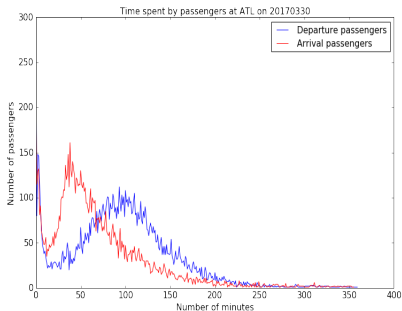
(c) Distribution of wait time at ATL airport on 2017-03-30
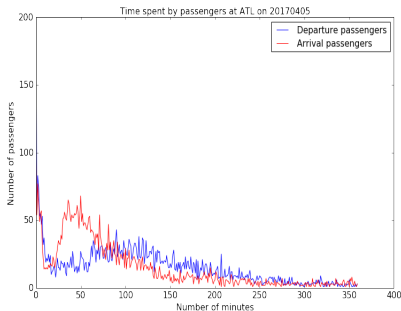
(d) Distribution of wait time at ATL airport on 2017-04-05

Impact of degraded operations on the time spent by passengers at ATL airport.

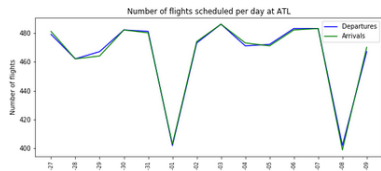# ATL degraded conditions monitoring
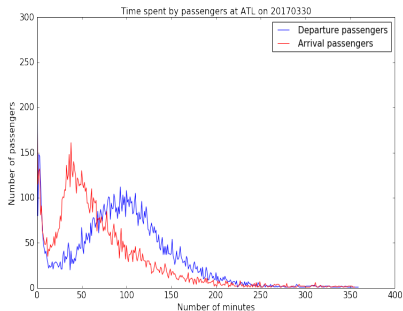
(e) Distribution of wait time at ATL airport on 2017-03-30
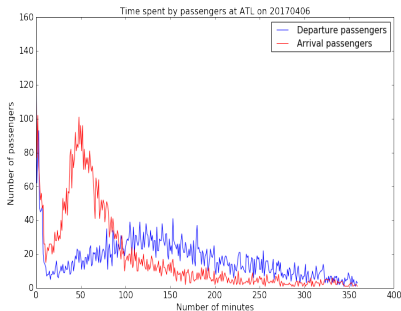
(f) Distribution of wait time at ATL airport on 2017-04-06

Impact of degraded operations on the time spent by passengers at ATL airport.

ATL degraded conditions monitoring
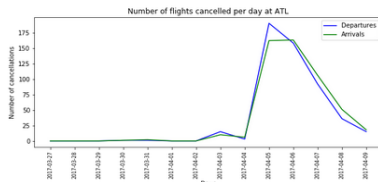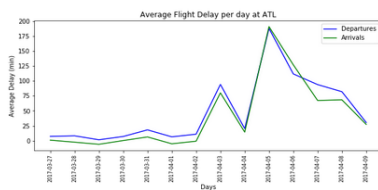
Flight-centric view of ATL airport from BTS

Content

Georgia
Tech

So far, we have shown that the presented data set extraction process was valid and could provide the same insights as conventional flight-centric data.

Let's see how this data outperforms traditional metrics!

Georgia
Tech

Improved analysis of time spent at airport

(i) At departure airport (j) At arrival airport
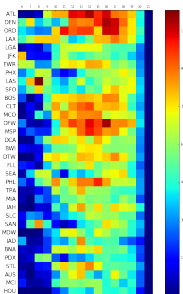
Time spent by passengers per hour at airport on March 27, 2017.

Noticeable airports for departure:

- LAS (Las Vegas):
  - Vegas strip congested
  - Slot machines in airport
- LAX (Los Angeles):
  - Major restructuration of airport in spring 2017

And more time spent in the evening at arrival.

Georgia Tech

Connecting passengers
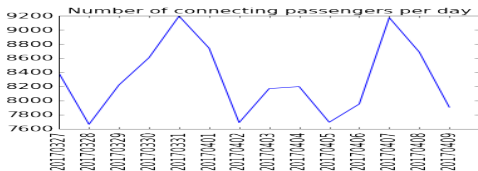
(k) Number of connecting passengers per day from mobile phone data.
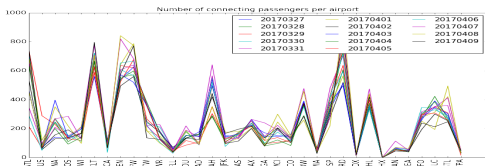
(l) Number of connecting passengers per airport per day.

Connecting passengers figures extracted from mobile phone data by

filtering passengers visiting more than 2 airports in a day.

- Available for all airlines simultaneously
- Better passenger and flight management?
- Total number low: no info on international flights
- Consistent with known hubs
- Volume consistent over days (except ATL 5 Apr.)

Georgia
Tech

Door-to-door journey analysis

Note: Airport only one portion of the journey

Can mobile phone data help have a global analysis of the full door-to-door journey?

Georgia
Tech

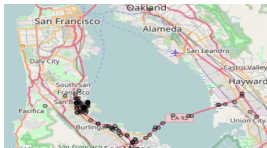## Door-to-door journey analysis

Note: Airport only one portion of the journey

Can mobile phone data help have a global analysis of the full door-to-door journey?

- Once passenger set extracted, possible to filter their full trajectory per day
- Variable from one user to another depending on "addiction" to mobile phone

Georgia
Tech

Door-to-door journey analysis

Overview of a trajectory over one day for a passenger traveling from SFO to DCA via LAX.
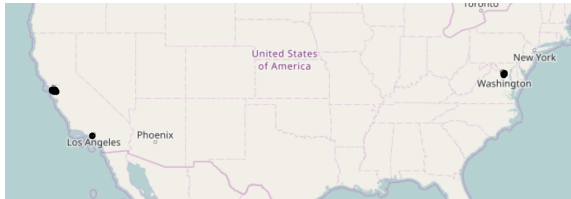
Zooming in on a trajectory over one day for a passenger traveling from SFO to DCA via LAX.

Georgia Tech
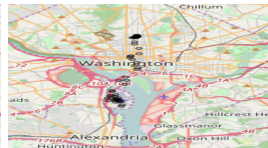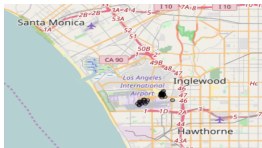
Door-to-door journey analysis

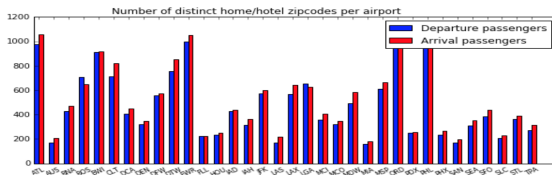Analysis of the number of zipcodes visited by passengers per departure or arrival airport.
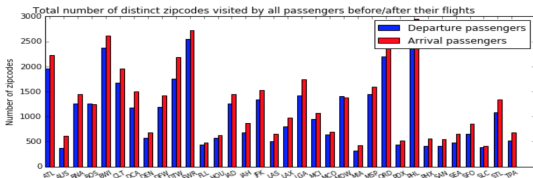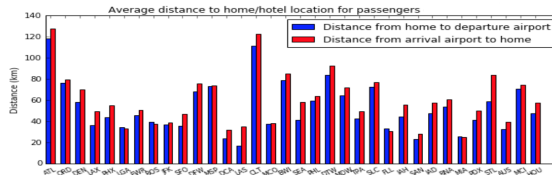

(p) Total number of zipcodes
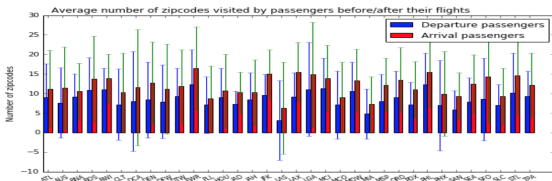
(q) Average number of zipcodes

Application to Urban planning

Domestic passenger behavior analysis: Distance to airport using the number of zipcodes visited to or from the airport as a new metric


(r) **Number of distinct zipcodes corresponding to passengers' home or hotel location per departure or arrival airport**

(s) **Average distance to passengers' home or hotel location per departure or arrival airport.**
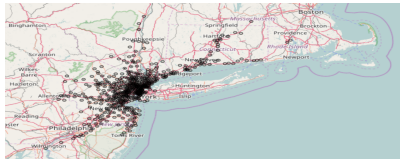
Application to Urban planning

Domestic passenger behavior for New York City airports: zipcodes visited before going to or after arriving at JFK, LGA and EWR.
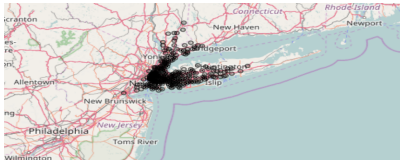
(t) John Fitzgerald Kennedy airport (JFK).

(u) La Guardia airport (LGA).

(v) Newark airport (EWR).
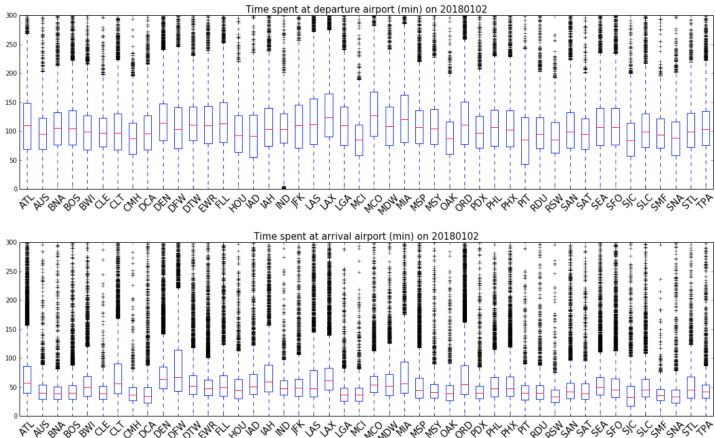
Recent work: New Benchmark

Figure: Time spent at airports by passengers on January 2nd
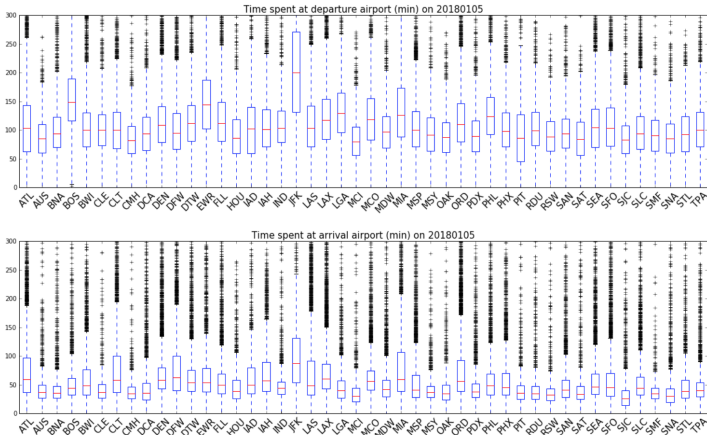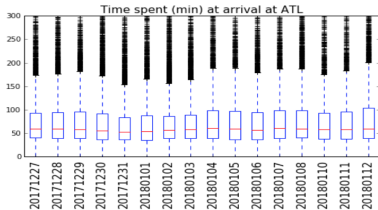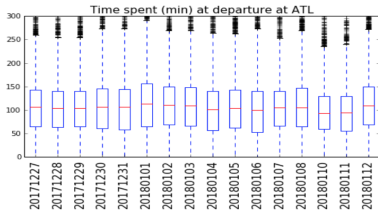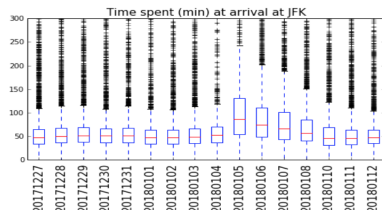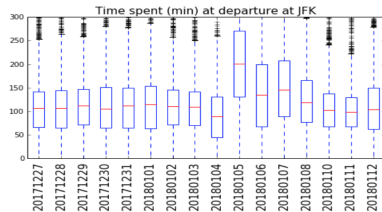
Recent work: New Benchmark

Figure: Time spent at airports by passengers on January 5th
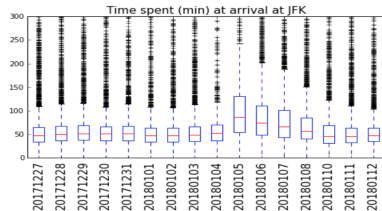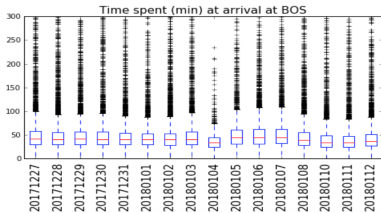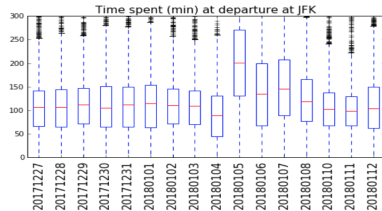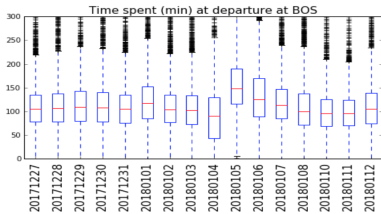
Recent work: New Benchmark

(a) Time spent at ATL by passengers   (b) Time spent at JFK by passengers
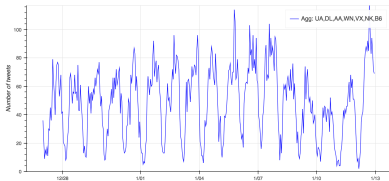
Georgia Tech

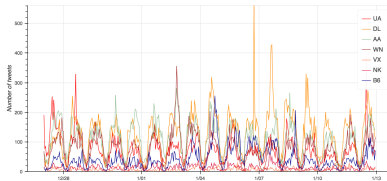# Recent work: New Benchmark

(c) Time spent at BOS by passengers   (d) Time spent at JFK by passengers

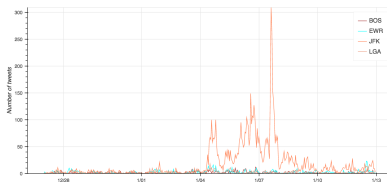# Recent work: Tweet Analysis

(e) Hourly volume of tweets filtered with delay keywords

(f) Hourly volume of tweets by airlines

(g) Hourly volume of tweets filtered with cancellation keywords

(h) Hourly volume of tweets by airports

Georgia Tech

Content

Georgia
Tech

Conclusion

- Mobile phones are valid sensors for air traffic passengers

Georgia
Tech

Conclusion

- Mobile phones are valid sensors for air traffic passengers
- New reliable system-wide passenger-centric metrics

Georgia
Tech

Conclusion

- Mobile phones are valid sensors for air traffic passengers
- New reliable system-wide passenger-centric metrics
- Objective measurement of impact of severe perturbations

Georgia
Tech

Conclusion

- Mobile phones are valid sensors for air traffic passengers
- New reliable system-wide passenger-centric metrics
- Objective measurement of impact of severe perturbations
- Useful informations for urban planning around airports

Georgia
Tech

Discussion

Any Questions?