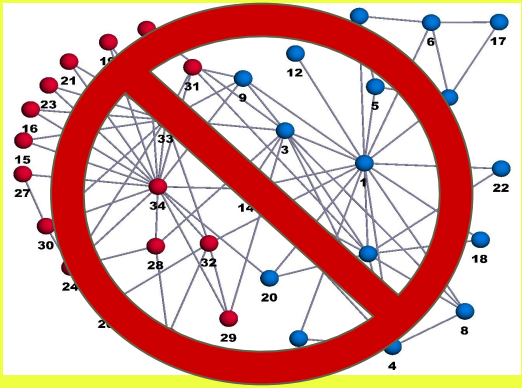
Night life and road safety: a comparison of 7 Italian cities

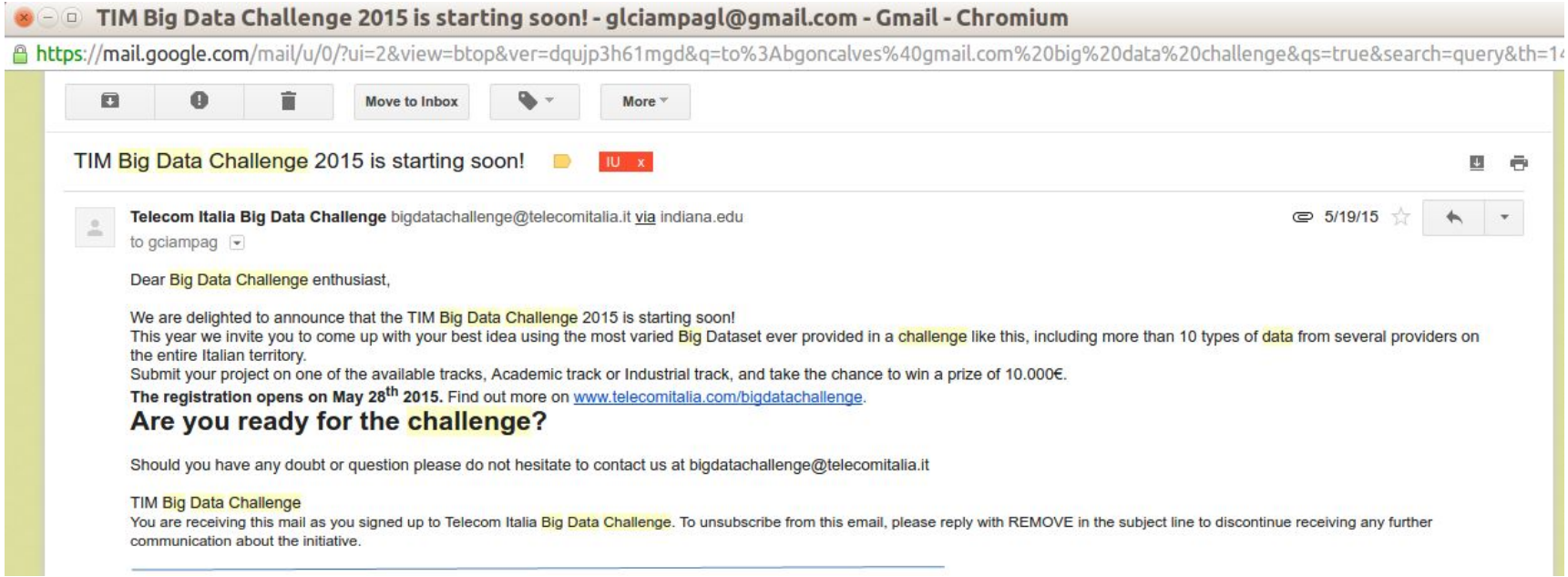
Giovanni Luca Ciampaglia

INDIANA UNIVERSITY
NETWORK SCIENCE INSTITUTE

WARNING: NO networks in this talk!



Genesis



TIM Big Data Challenge 2015 is starting soon! - glciampagl@gmail.com - Gmail - Chromium

https://mail.google.com/mail/u/0/?ui=2&view=btop&ver=dqujp3h61mgd&q=to%3Abgoncalves%40gmail.com%20big%20data%20challenge&qs=true&search=query&th=14

Move to Inbox | More

TIM Big Data Challenge 2015 is starting soon! IU x

Telecom Italia Big Data Challenge bigdatachallenge@telecomitalia.it via indiana.edu 5/19/15

to gciampag

Dear Big Data Challenge enthusiast,

We are delighted to announce that the TIM Big Data Challenge 2015 is starting soon!
This year we invite you to come up with your best idea using the most varied Big Dataset ever provided in a challenge like this, including more than 10 types of data from several providers on the entire Italian territory.
Submit your project on one of the available tracks, Academic track or Industrial track, and take the chance to win a prize of 10.000€.
The registration opens on May 28th 2015. Find out more on www.telecomitalia.com/bigdatachallenge.

Are you ready for the challenge?

Should you have any doubt or question please do not hesitate to contact us at bigdatachallenge@telecomitalia.it

TIM Big Data Challenge
You are receiving this mail as you signed up to Telecom Italia Big Data Challenge. To unsubscribe from this email, please reply with REMOVE in the subject line to discontinue receiving any further communication about the initiative.

Giovanni Luca Ciampaglia <gciampag@indiana.edu> 5/19/15

to NaN

Not sure if anybody is interested in participating to this?

G

Giovanni Luca Ciampaglia

✎ 919 E 10th · Bloomington 47408 IN · USA
☞ http://www.glciampaglia.com/
✆ +1 812 855-7261
✉ gciampag@indiana.edu

Bruno Goncalves <bgoncalves@gmail.com> 5/20/15

to Giovanni

Hi,

This looks interesting, but I don't seem to be able to find a description of the dataset?
Best,

B

****
Bruno Miguel Tavares Gonçalves, PhD
Homepage: www.bgoncalves.com
Email: bgoncalves@gmail.com
****

nge - Chromium

My Drive - Googl | The Contest | Big | Open Big Data | [| Mobile Traffic Sic | Untitled present

n/bigdatachallenge/contest.html

vetti

BIGDATACHALLENGE2015

TIM
Il futuro firmato Telecom Italia

Big Data for Big Ideas. Innovation gets a free hand!

TIM **Big Data Challenge** is a contest created to stimulate innovation concerning Big Data. We are looking for professionals, researchers and enthusiasts from **all over the world**, ready to challenge each other and develop the Big Data projects of the future.

This year's theme is **Big Data for Competitiveness Boost**: how can data help a country grow and become more competitive?

The last edition proved to be incredibly popular, with more than **1,100 people** taking part from more than **20 different countries**, submitting **over a 100 design ideas**.

What about you? Are you ready for the challenge?

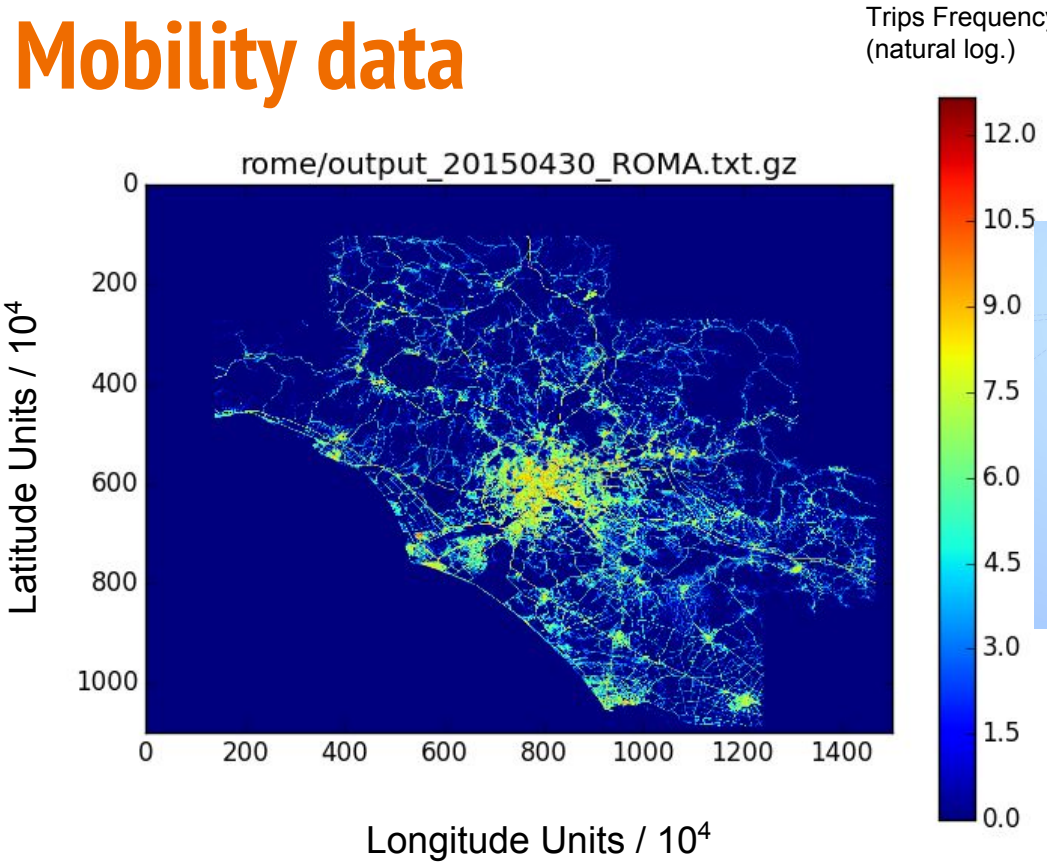
ABOUT US | LIVING INNOVATION | CHOOSING A BETTER FUTURE | SHARING CULTURE | INVESTORS | SUSTAINABILITY | PRESS | CAREER | OUR OFFERING | SOCIAL HUB

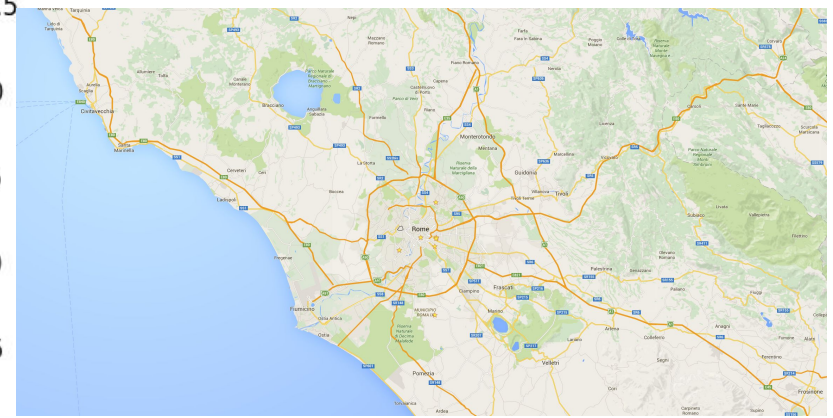
Mobile | Sitemap | Privacy | Legal notes

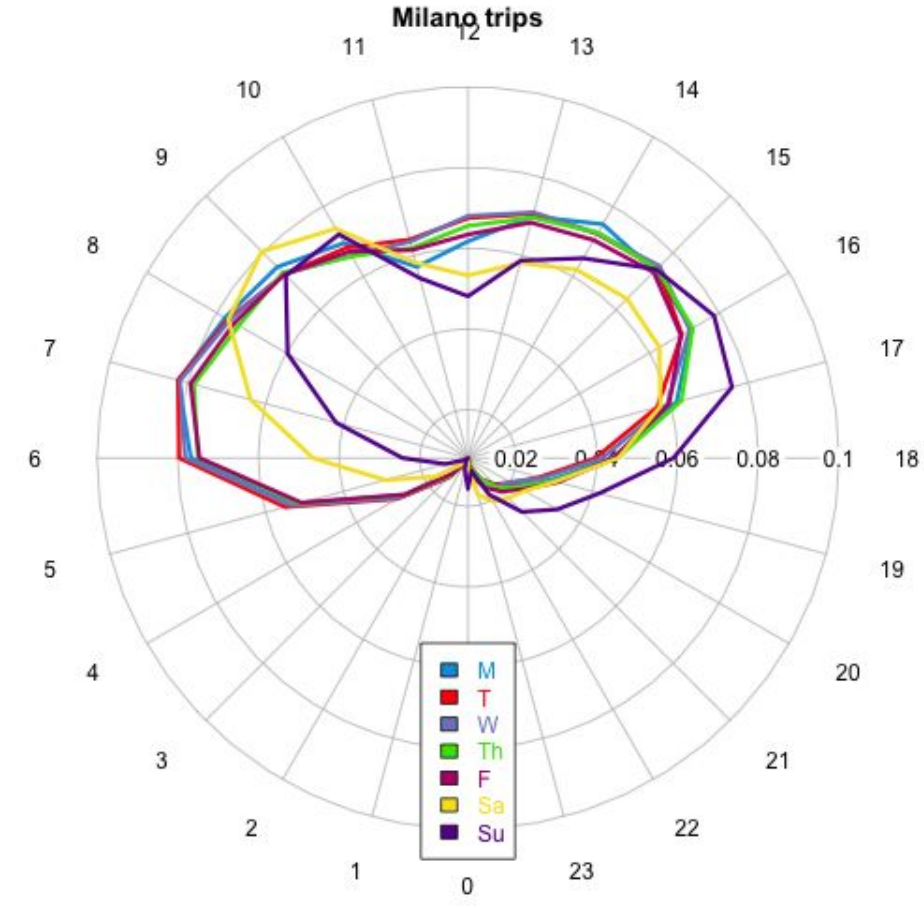
TBDC 15: The data

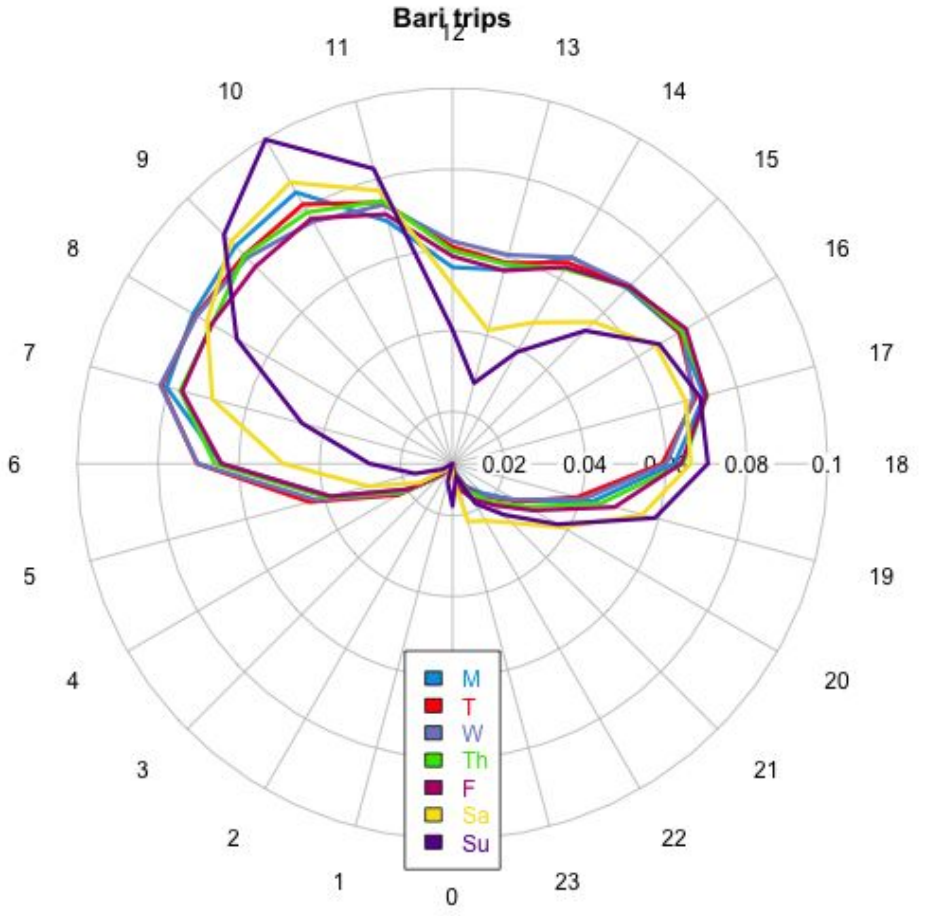
- 7 major Italian cities: (Rome, Naples, Milan, Turin, Venice, Bari, Palermo)
  - North to south
  - Includes greater metropolitan areas in most cases
- Diverse dataset
  - 2x **Mobility** (**Infoblu**, Viasat) datasets
  - Calls + SMS + Internet (TIM)
  - Presence (computed from mobile users data)
  - Demographics (gender, age-range and living area of callers)
  - Economics (List of companies, headquarters, branches of firms from Cerved DB)
  - Social (geolocalized data via API; didn't get those...)
  - **Car accidents** (geolocalized claims from Unipol insurer)
  - Census data (ISTAT) + various shapefiles

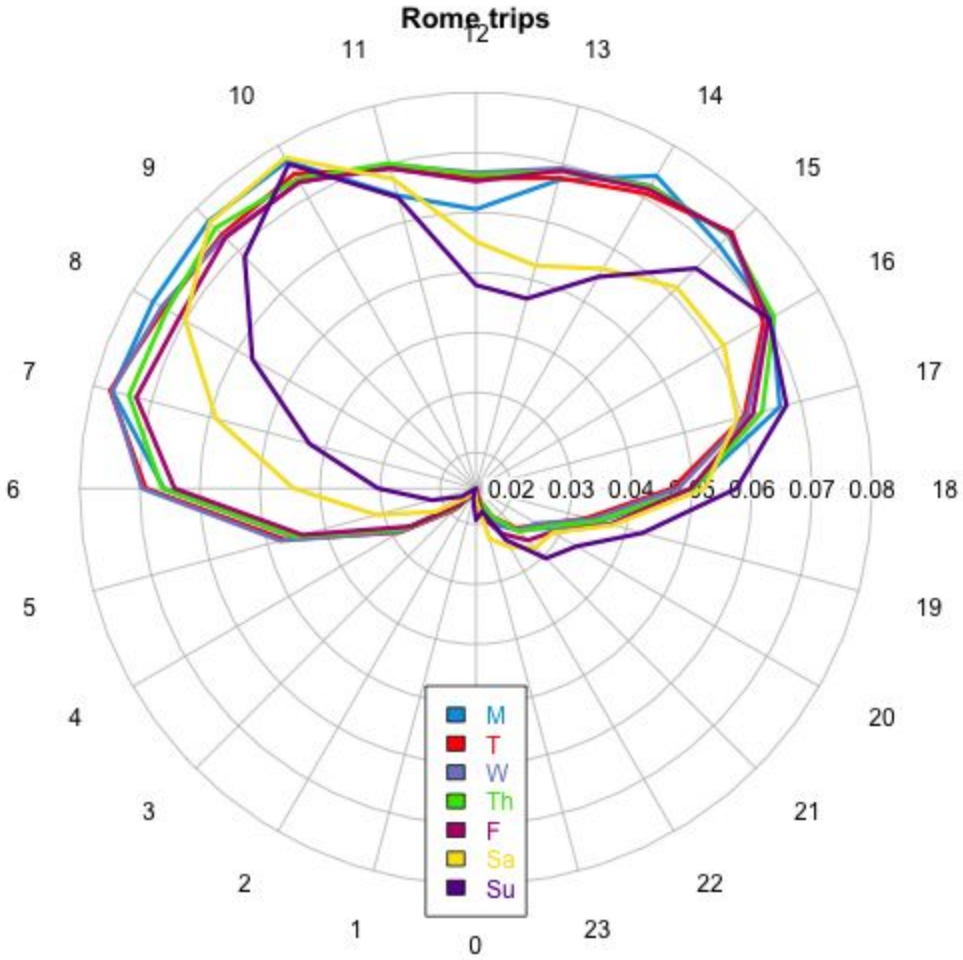
Mobility data



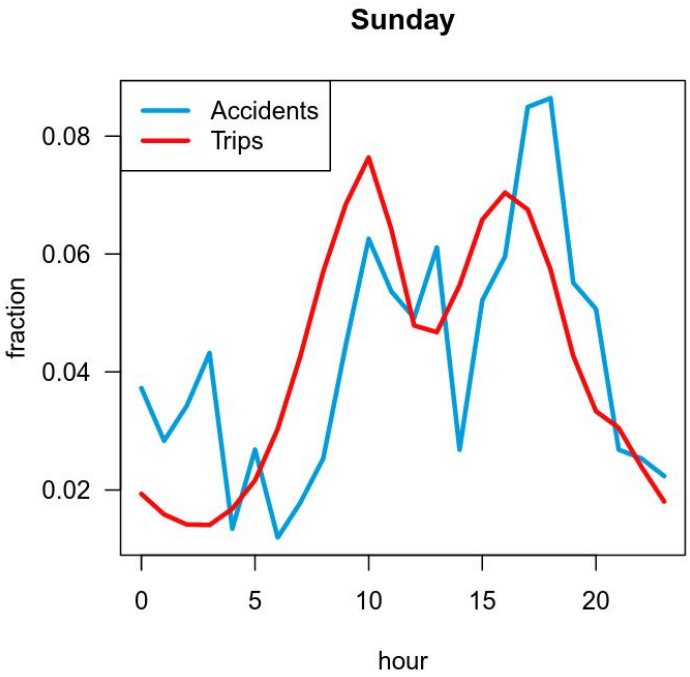


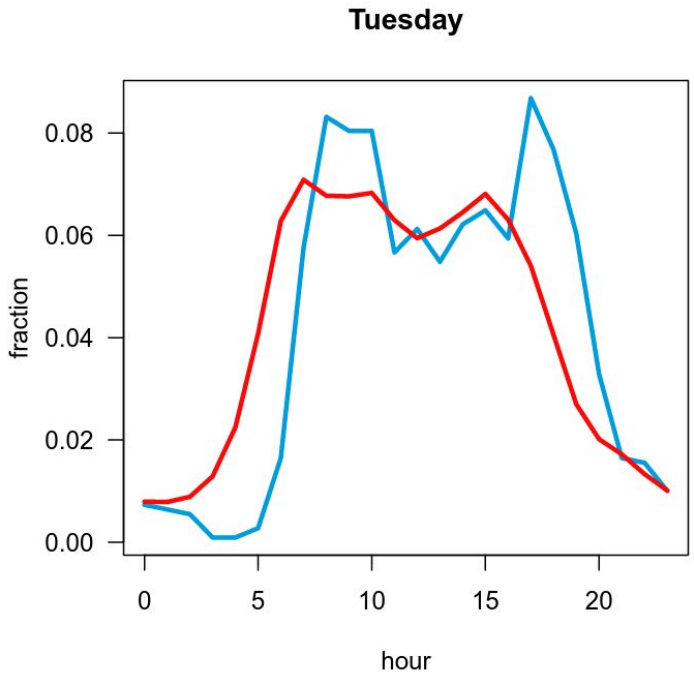

Milano trips
12
13
14
15
16
17
18
19
20
21
22
23
0
1
2
3
4
5
6
7
8
9
10
11
0.02
0.06
0.08
0.1
M
T
W
Th
F
Sa
Su

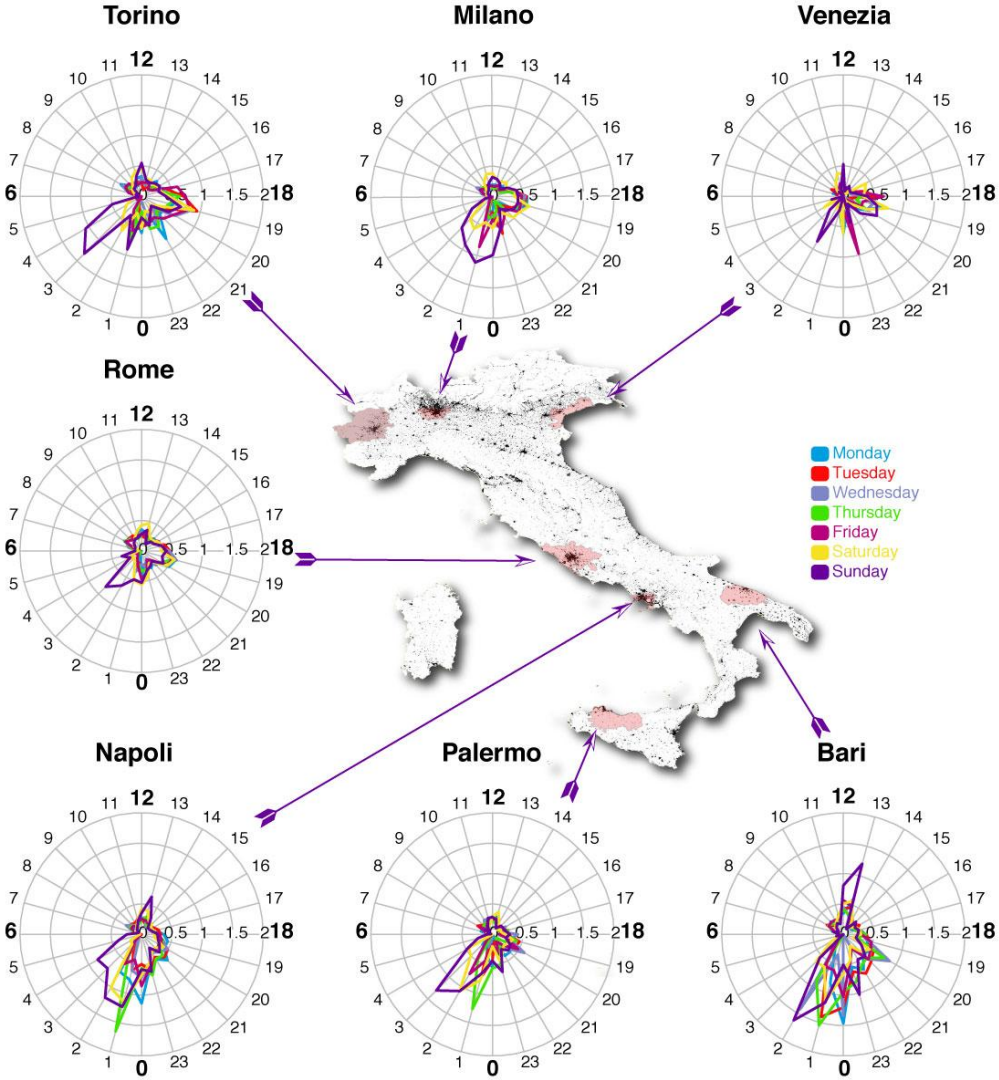

Bari trips
12
13
14
15
16
17
18
19
20
21
22
23
0
1
2
3
4
5
6
7
8
9
10
11
0.02
0.04
0.08
0.1
M
T
W
Th
F
Sa
Su


Rome trips
12
13
14
15
16
17
18
19
20
21
22
23
0
1
2
3
4
5
6
7
8
9
10
11
0.02
0.03
0.04
0.05
0.06
0.07
0.08
M
T
W
Th
F
Sa
Su


ROMATODAY


Sunday
Accidents
Trips
0.08
0.06
0.04
0.02
fraction
0
5
10
15
20
hour

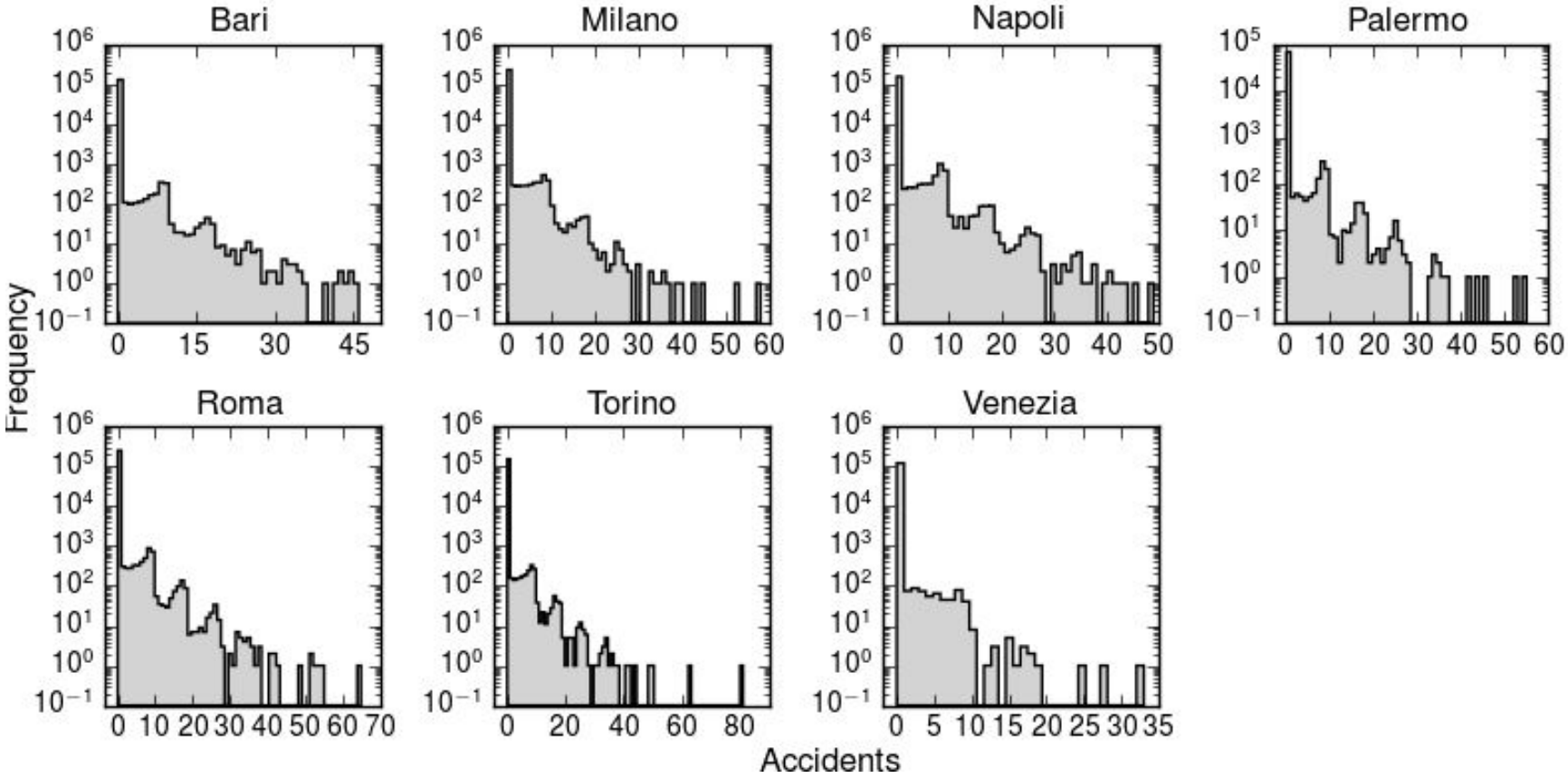

Tuesday
0.08
0.06
0.04
0.02
0.00
fraction
0
5
10
15
20
hour


Torino
Milano
Venezia
Rome
Napoli
Palermo
Bari
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday

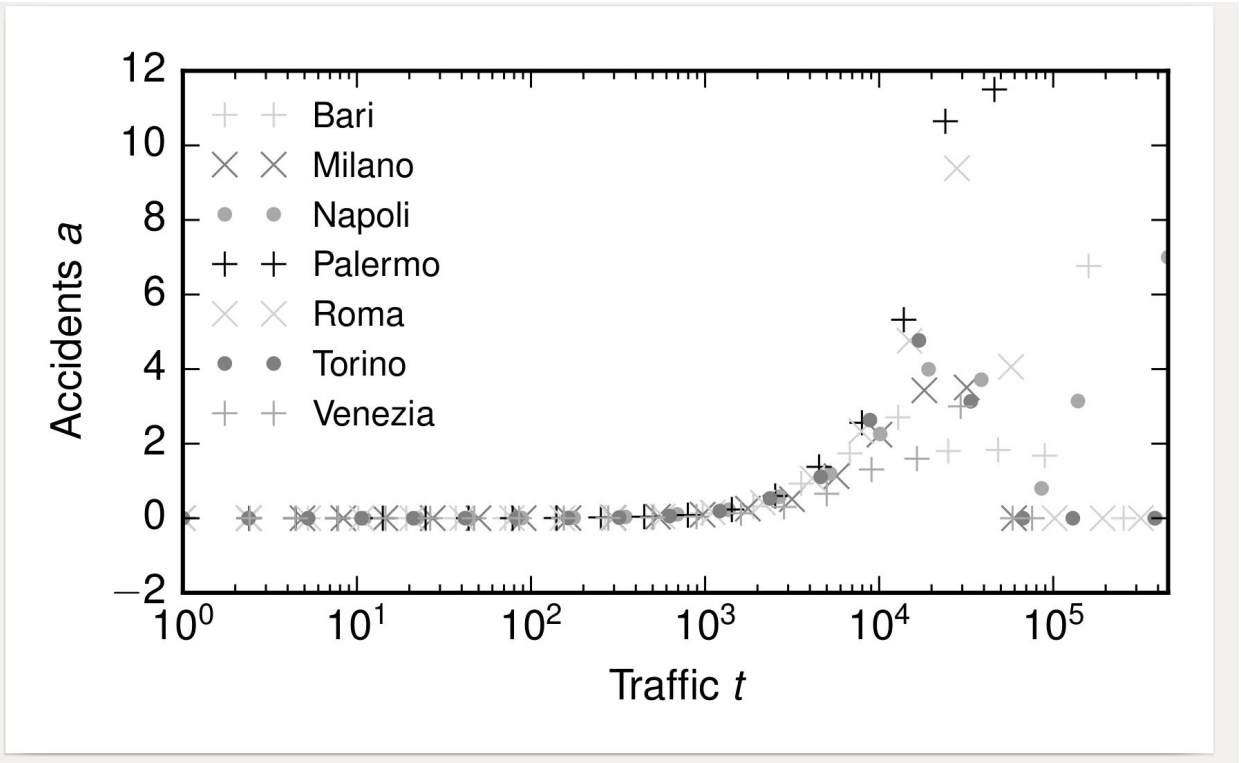
Research Questions

1. Is there a relation between traffic, speed, and accidents?
2. Can we predict what are the most risky areas for accidents?
3. Can we glean more if adding social data?
    a. Text (tweet) while driving
    b. Guessing DUI driving

Accidents data



Traffic vs accidents



Zero-inflated models

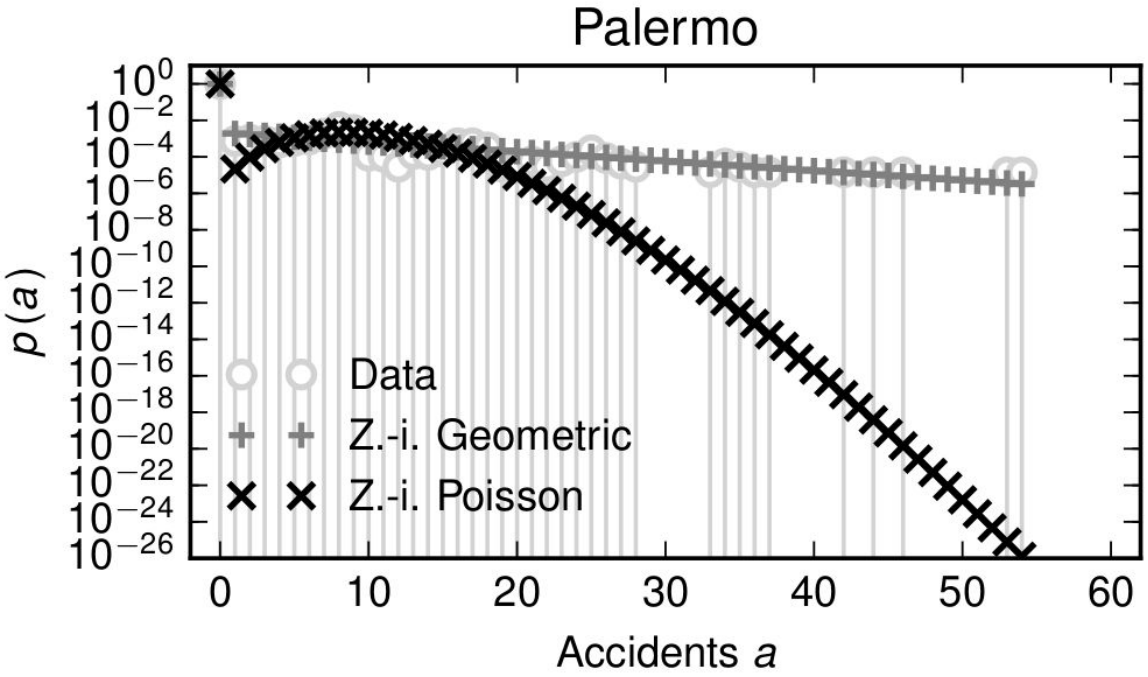
The geometric distribution with support starting at 0 is the distribution of number of failures k before the first success in a series of Bernoulli trials each with success probability p:

$$\Pr(k\,;p) = p\,(1-p)^k$$

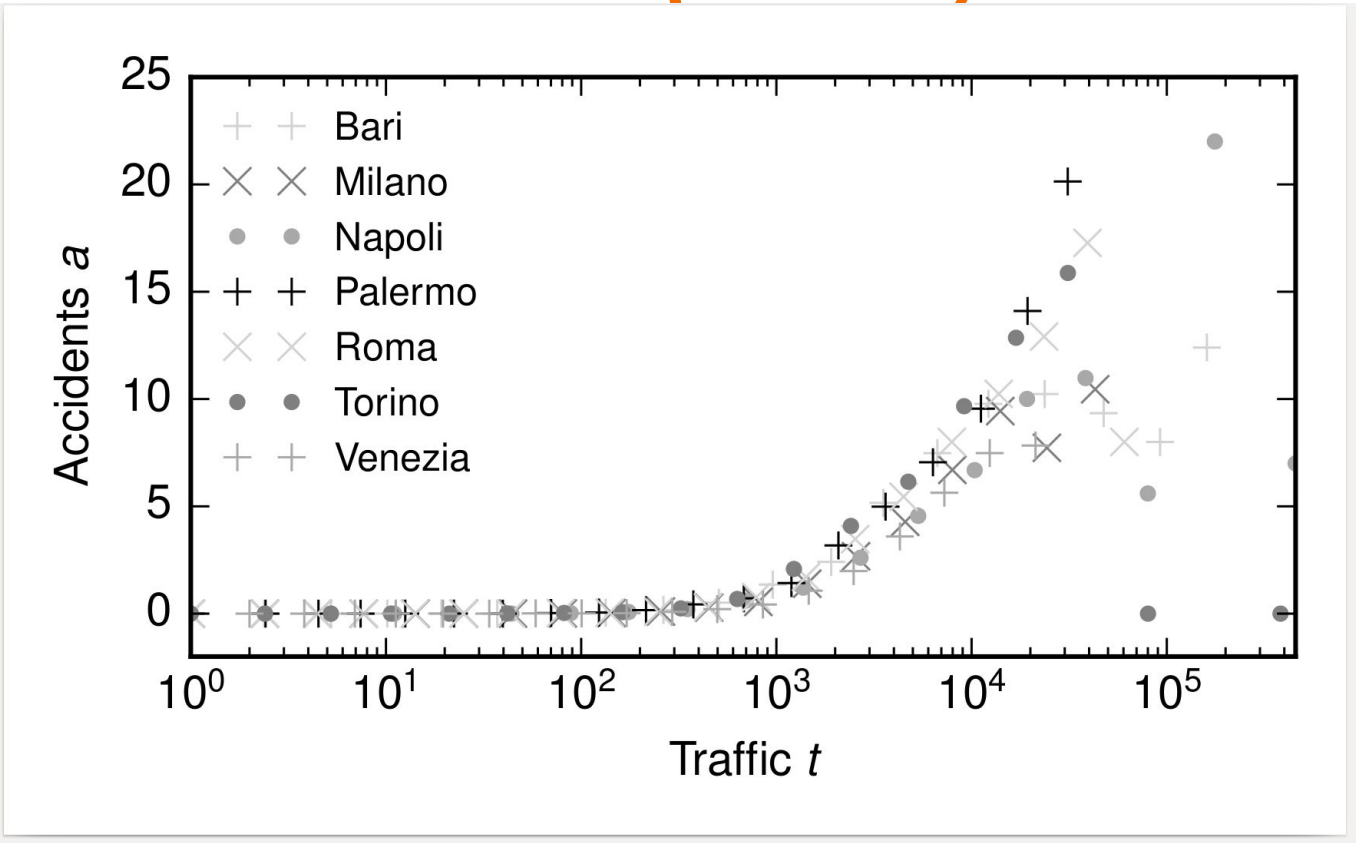
The zero-inflated model is a mixture of a process generating zeros with fixed probability π, and a normal geometric process. That is, for each observation first we toss a coin and with probability π produce a zero, otherwise we draw from the Geometric distribution:

$$\Pr(k\,;\pi,p) = \begin{cases} \pi + (1-\pi)\,p & \text{if } k = 0 \\ (1-\pi)\,p\,(1-p)^k & \text{if } k > 0 \end{cases}$$

Zero-inflated Fit Results

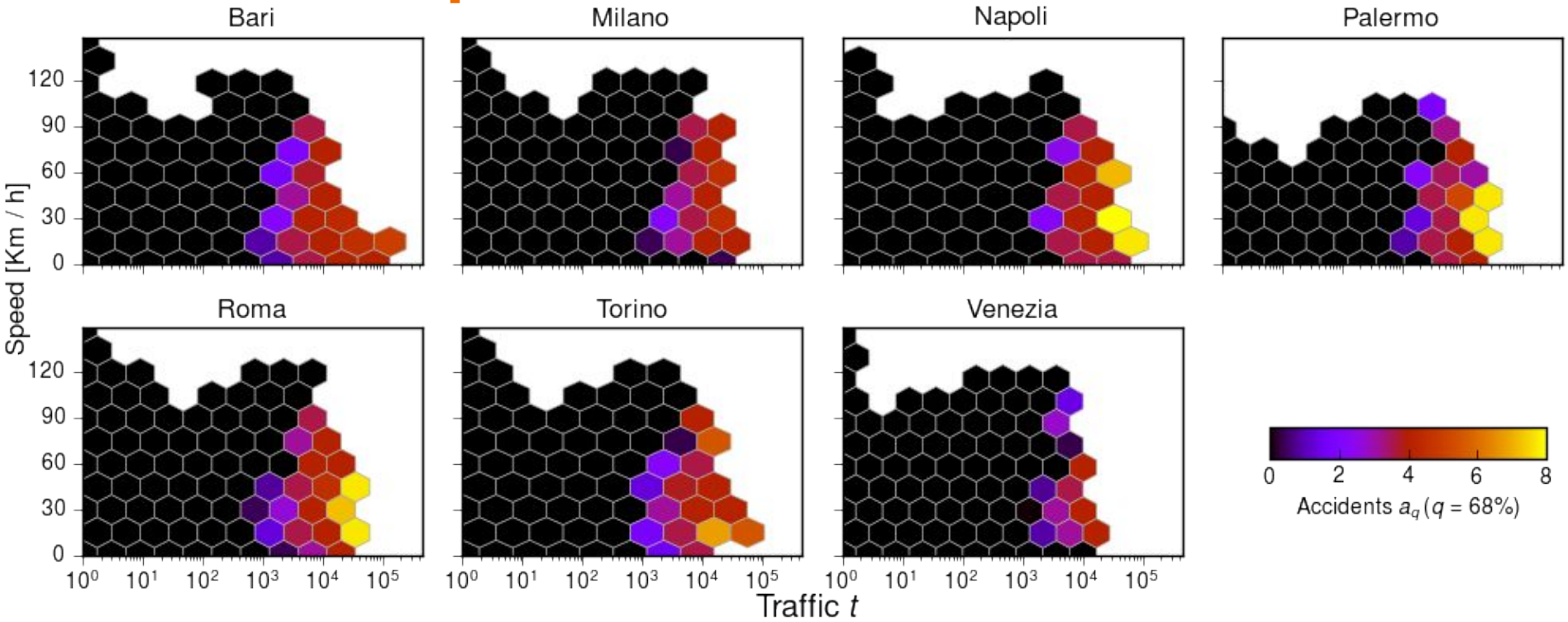


Traffic vs accidents *(cleaned)*

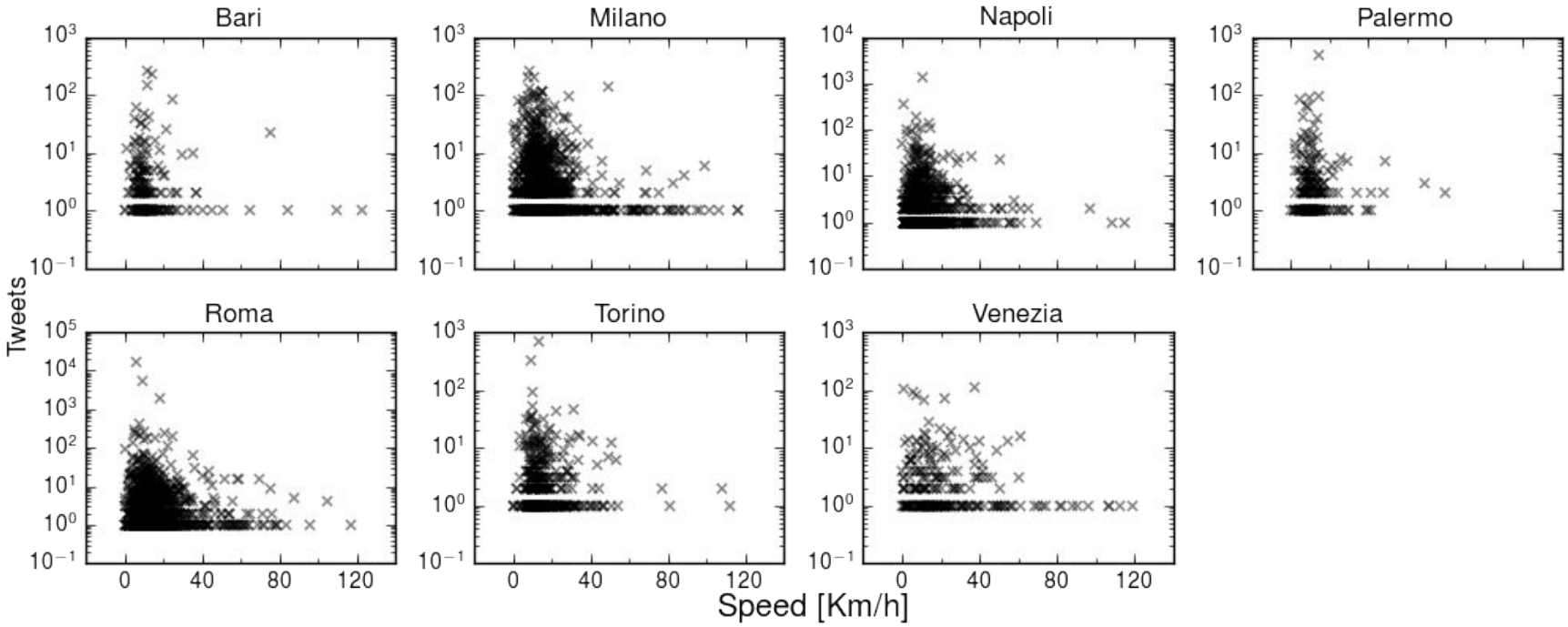


Bari	7.45%
Milano	11.15%
Napoli	17.45%
Palermo	8.26%
Roma	10.88%
Torino	7.90%
Venezia	9.24%

Accidents vs speed



Tweets vs speed



Where are we now?

- Prediction task
  - Target: accidents in a cell
  - Predictors: speed, traffic, tweets
  - Actually adding tweets does NOT improve error
- Looking at routes
  - Where are the trips that results in accidents originate from and are directed to?

Thanks!

José Ramasco - IFISC, Spain

Bruno Gonçalves - NYU, USA

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

INDIANA UNIVERSITY
BLOOMINGTON

BIGDATACHALLENGE2015