Digital Epidemiology: A new method to infer human behavior and pathogen transmission

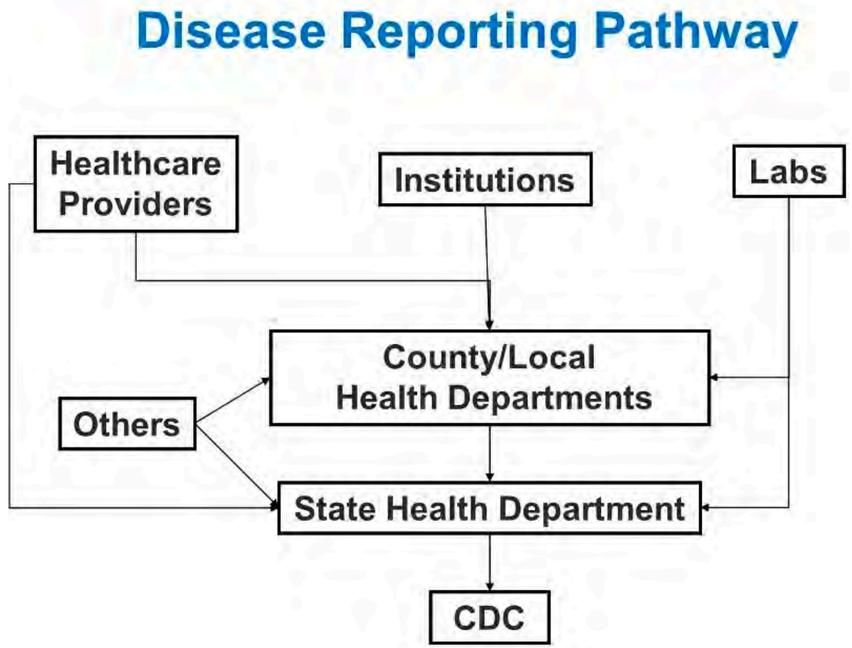
Benjamin Roche

This presentation is done in total independence from the event organizer. I have no conflict of interest to declare regarding the current presentation.

Digital Epidemiology

What is it exactly?

- Two recurrent problems in epidemiology
  - Epidemiological dynamics is always hard to quantify because a long process (local notification, transferring information, formatting, etc…)
  - Epidemiological dynamics rely on human behavior, which requires long study



Digital Epidemiology

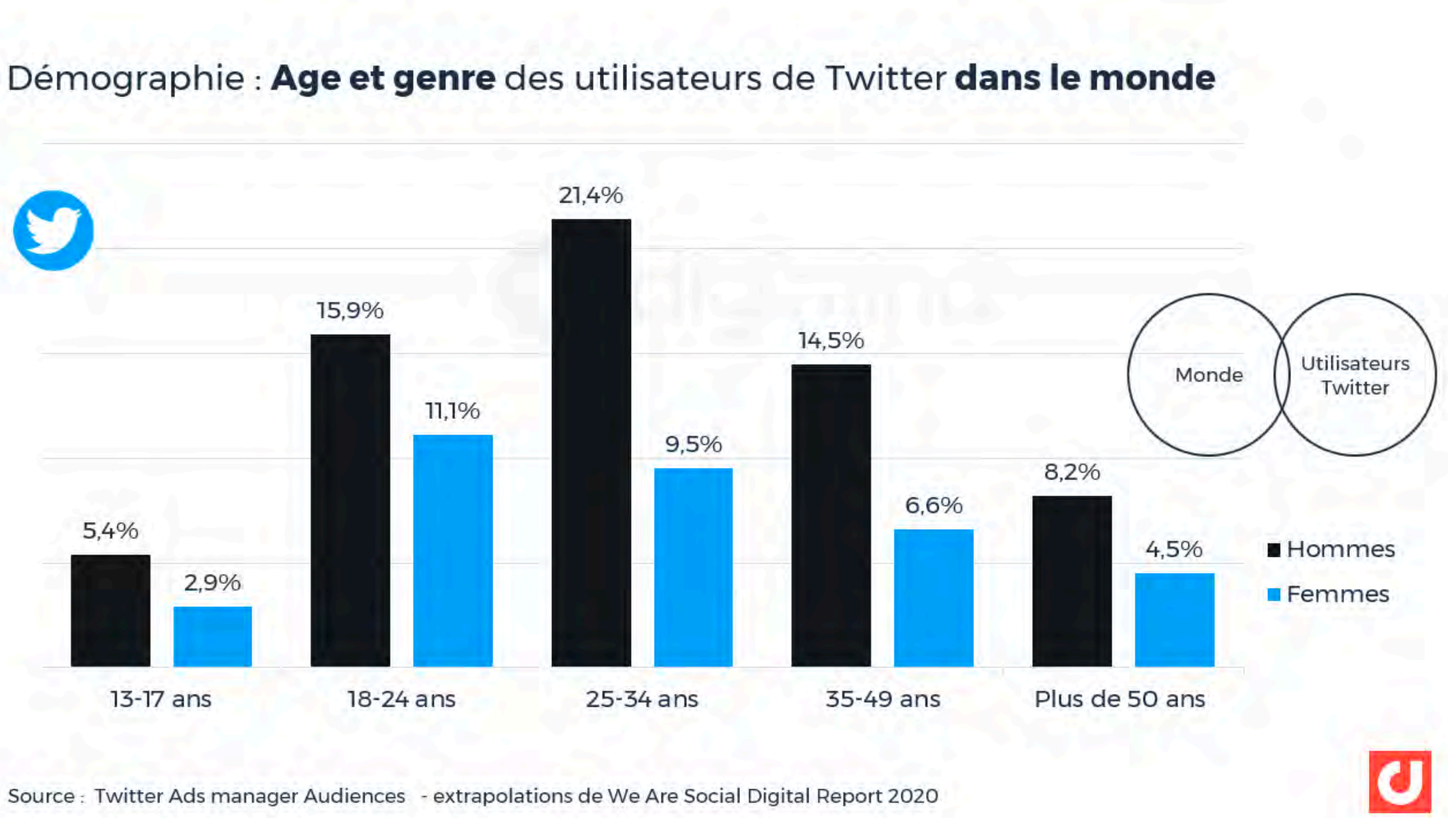
What is it exactly?

- Using the huge amount of data to infer:
  - Epidemiological dynamics
  - Human behavior regarding the emerging pathogens
- Classification of messages
- Incorporation into mathematical models to better understand epidemiological dynamics and/or improving forecasting

Relevance of digital epidemiology for epidemiological forecasting

Twitter is a biased proxy for human feelings

- Twitter is more used by young people
- Not representative of the whole population
- Impact ?



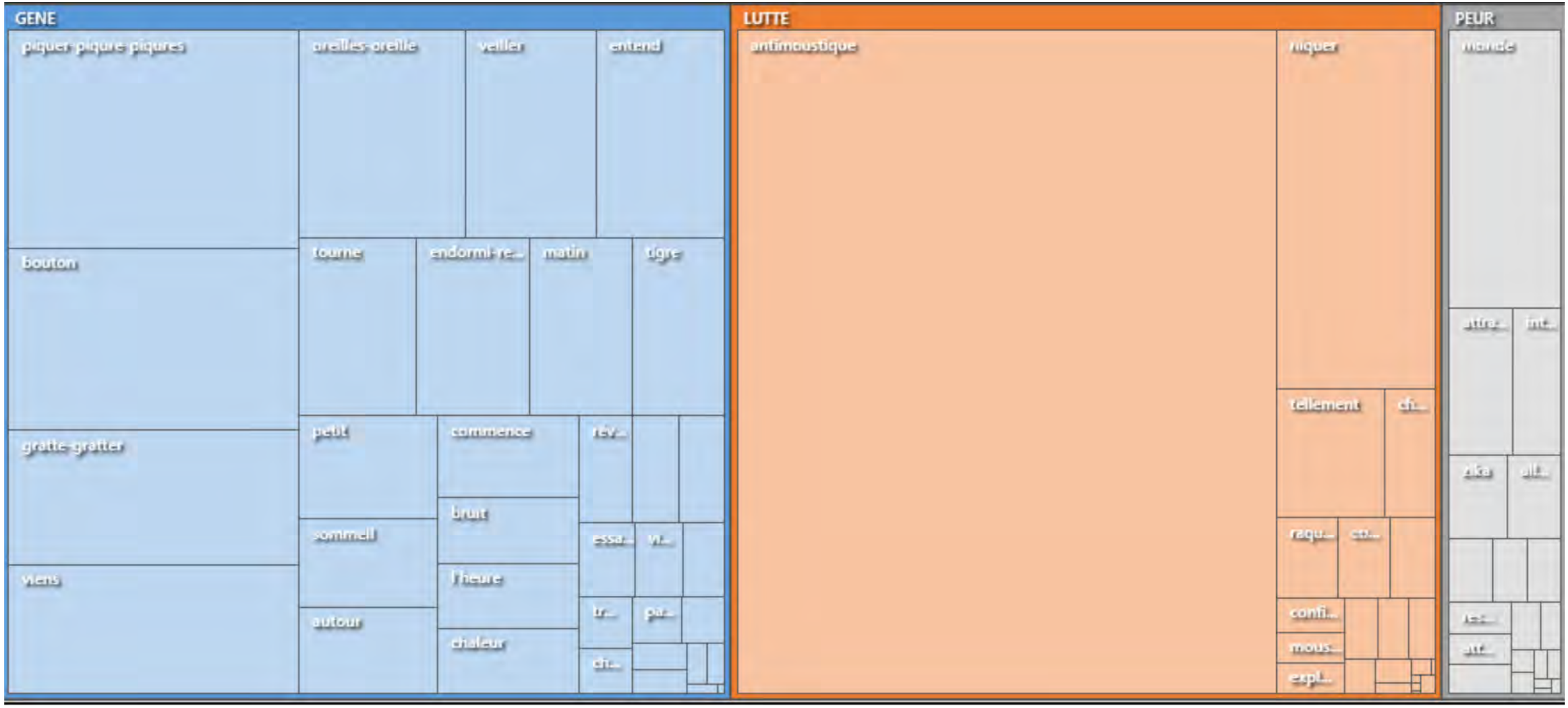
Collaboration with Public Health France

- Integration of specific questions regarding vector-borne diseases in large-scale polls
- Comparison of expected outcomes when considering data from twitter or large-scale poll?
- Is Twitter a reliable human behavior proxy for epidemiological forecasting?

Santé publique France

Health Barometer	Twitter
Poll, n=15,000 individuals	160,000 tweets with « mosquito"
7 questions on three feelings	Automatic classifications on three feelings

Lexical analysis



Mathematical modeling

- Age-structured model
- Transmission rate fluctuates through time with feelings fluctuations
- 1 age class for the model using Twitter data
- 8 age classes for the model using health barometer data

$$\frac{dS_i}{dt} = -\beta S_i \sum I$$

$$\frac{dI_i}{dt} = \beta S_i \sum I - \sigma I_i$$

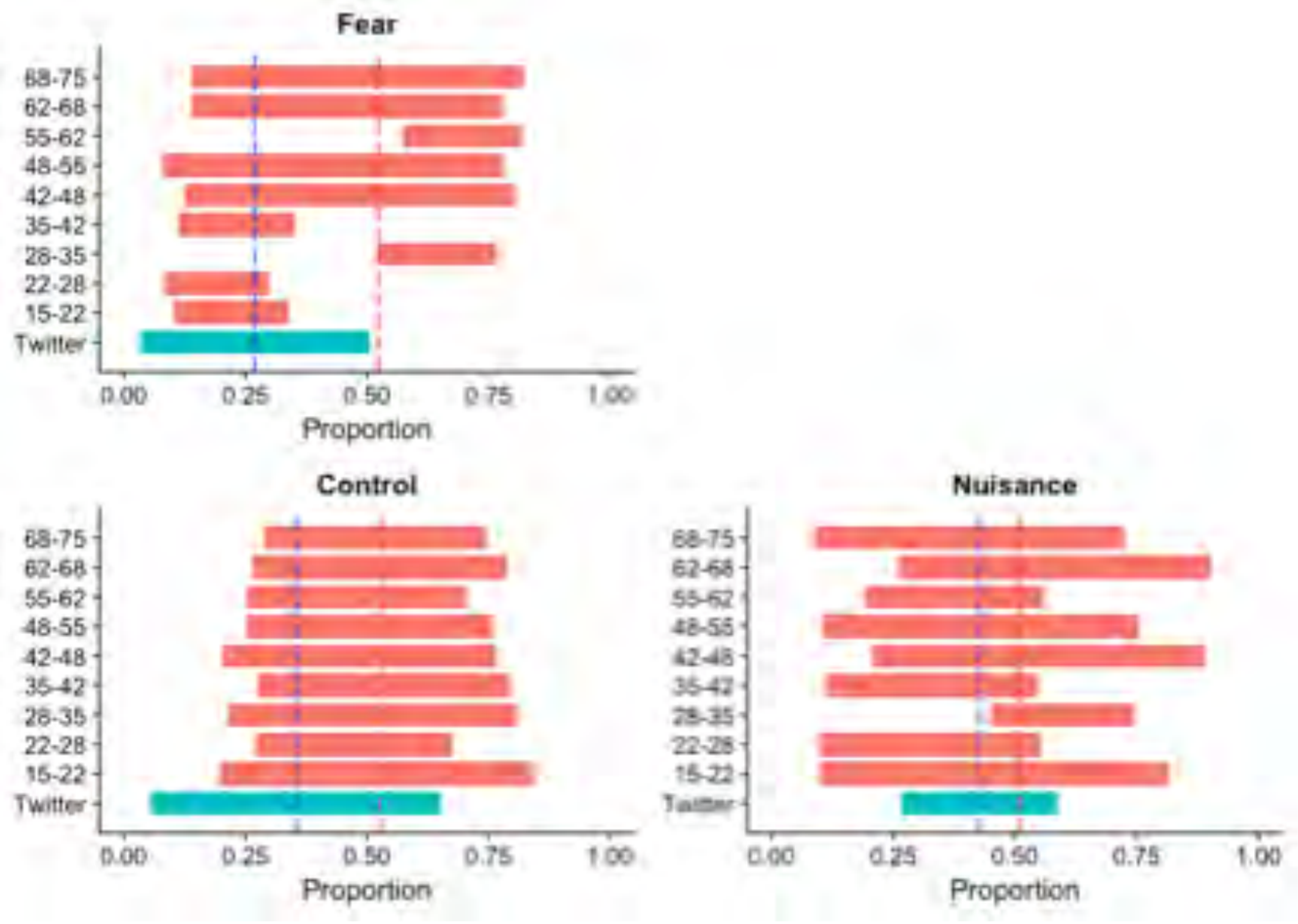
$$\frac{dR_i}{dt} = \sigma I_i$$

$$\beta_i = \beta_0 (1 + \sum \gamma_{ij} \delta_{ij})$$

Jourdain et al, submitted

Difference between data

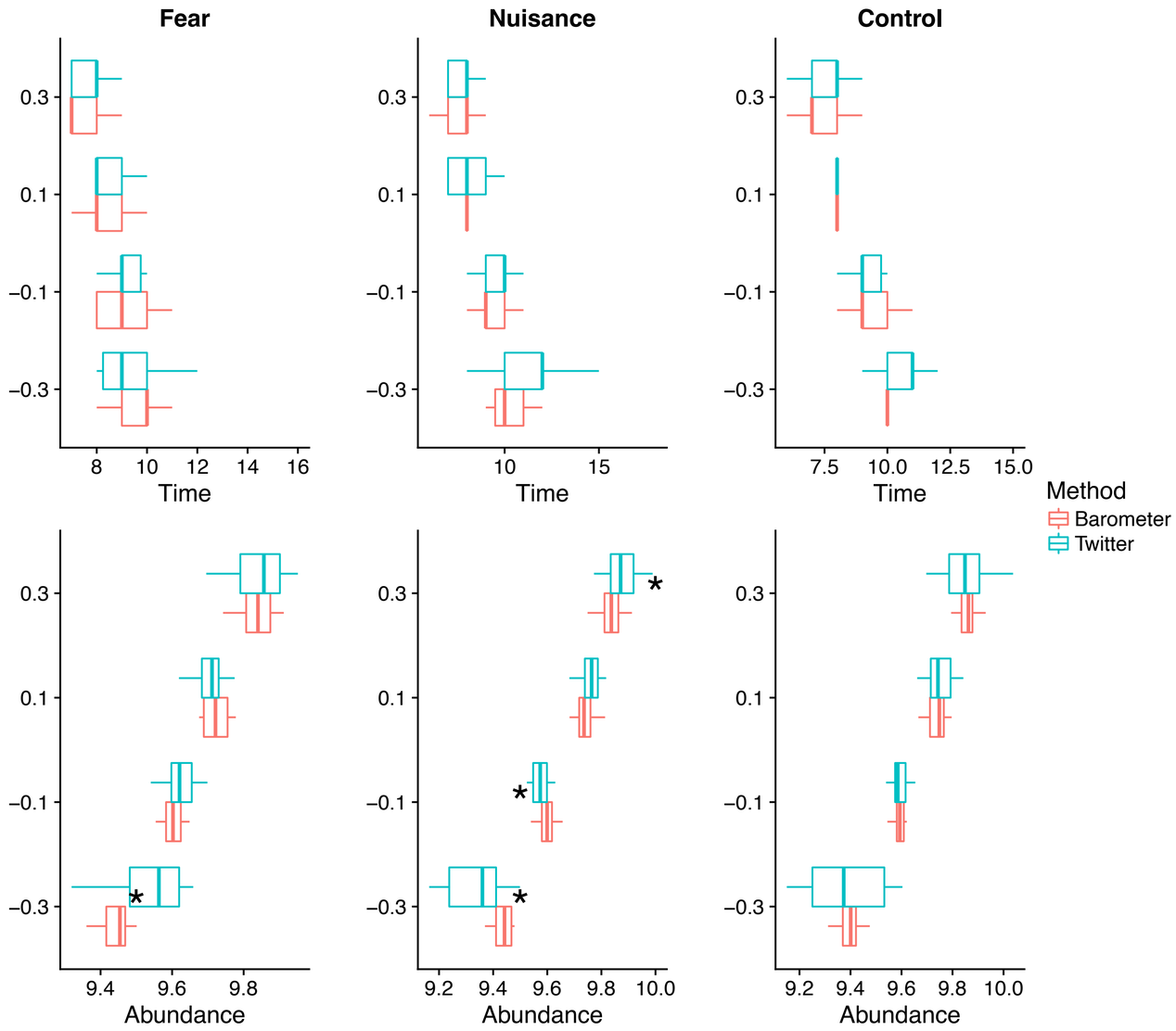
- Depending on the feeling considered, difference are more or less important
- Impact for forecasted dynamics?



Jourdain et al, submitted

Twitter is a good proxy for outbreak forecast

- Forecasting epidemiological dynamics is (almost) similar if using twitter or poll data
-



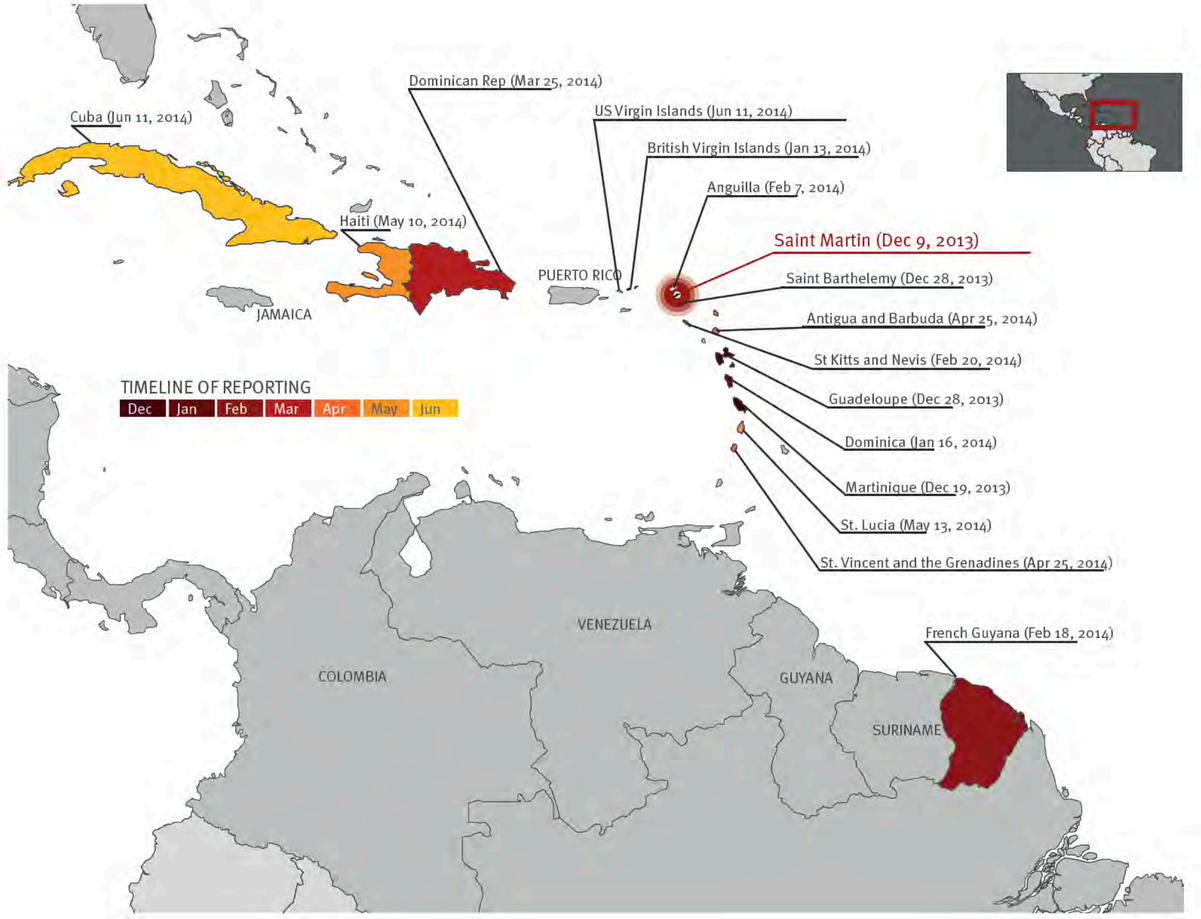
Jourdain et al, submitted

Summary & Discussion

- Twitter is a biased proxy, but reliable enough to be used for short-term forecasting
- Considered only simple mathematical models
  - Need to use more complicated transmission routes
- Considered only vector-borne diseases and 3 feelings
  - Has to be replicated across a larger number of diseases/transmission pathways
- Call for an urgent opening of social network data

The example of #Chikungunya in Carribeans

Chikungunya outbreak(s) in Caribbean islands



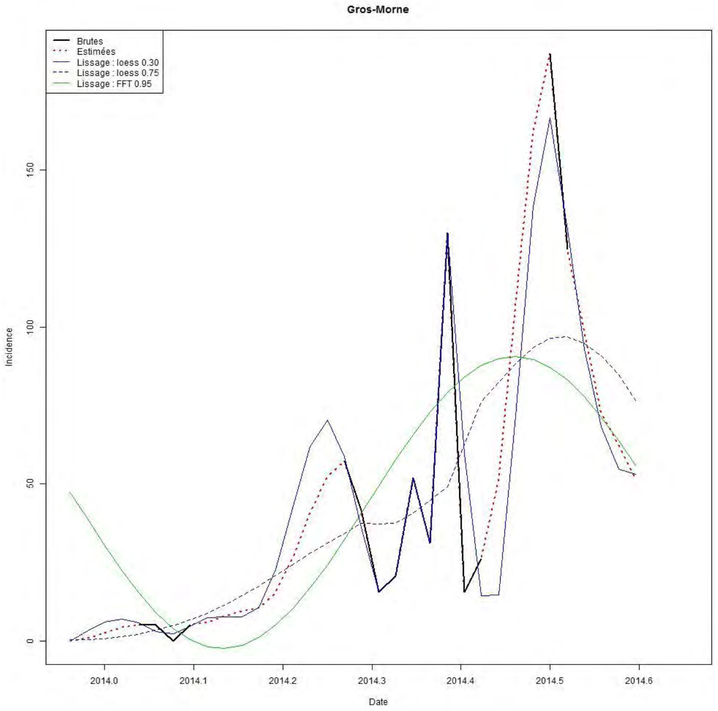
Cauchemez *et al*, Eurosurveillance, 2014

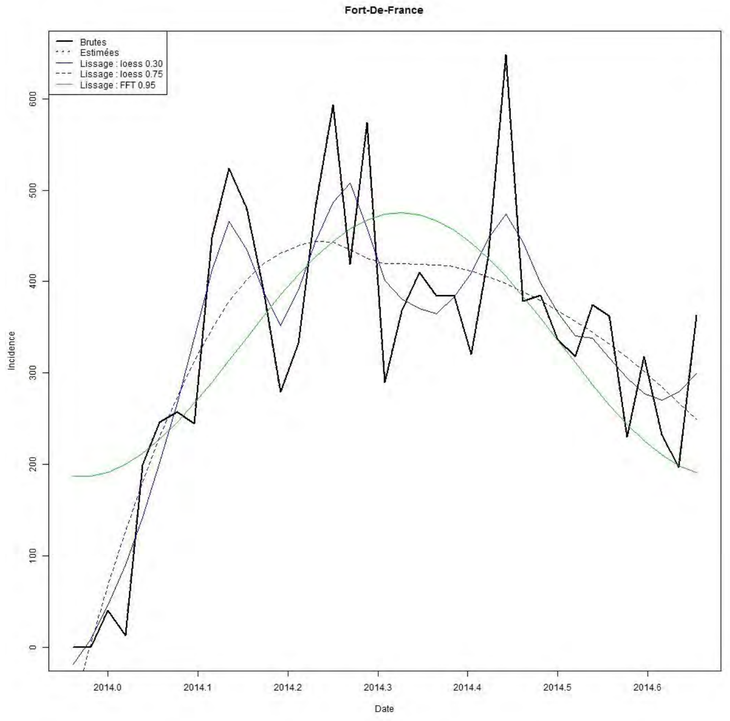
The epidemiological context in Martinique

- French oversea territory with around 400,000 inhabitants
- Sentinel network involving more than 20% of local MDs
- Recurrent epidemics of Dengue virus
- Insecticide resistance is (almost) fixed
-

Data available

- Epidemiological time series
  - Number of new suspected cases every week in each locality





Data available

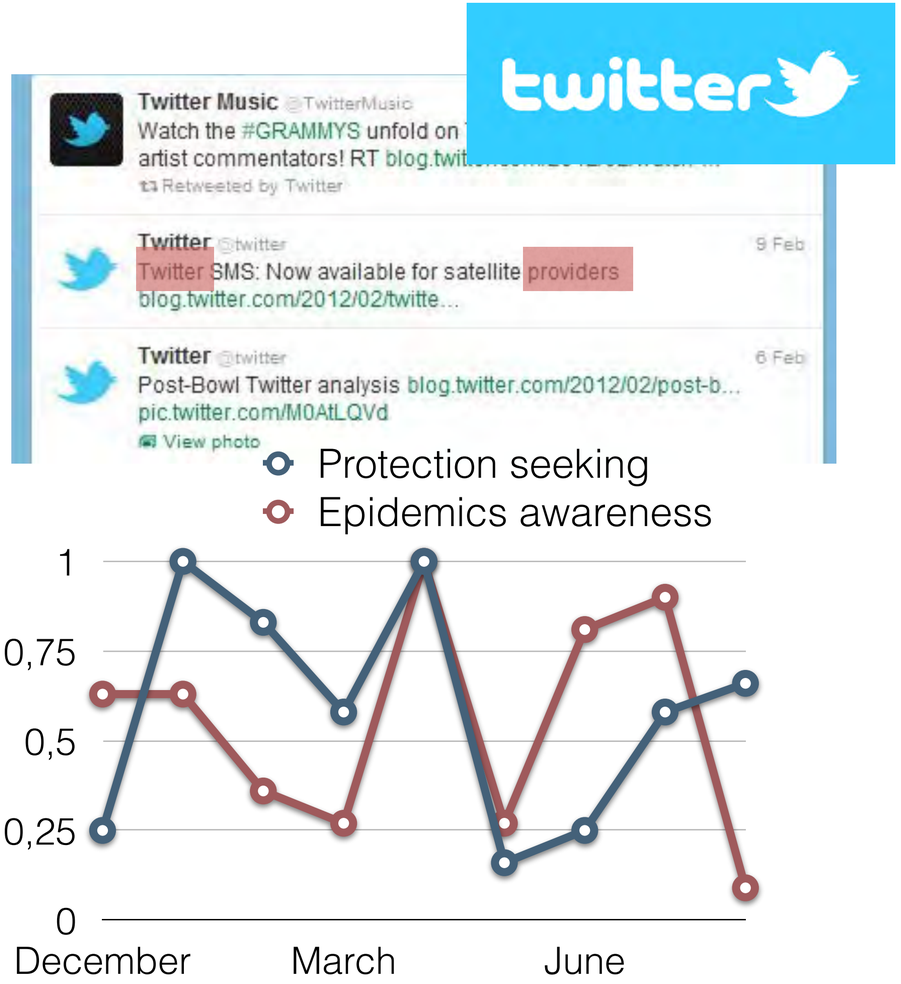
- Epidemiological time series
- Entomological data
  - Different kind of breeding areas monitored
  - Generalized linear model to extrapolate probability of mosquito presence through time during the outbreak according to climatic variables recorded

Roche et al, J Med Entomol, submitted

Epidemiological time series
Entomological data
Human behavior data?

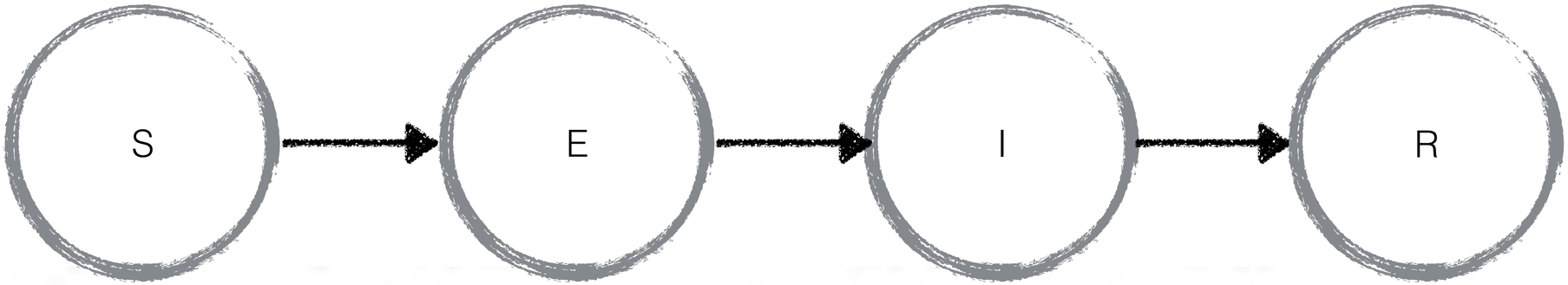
Using feelings expressed on Digital Social network to infer human behavior

- More than 200 tweets recorded during the first 8 months of the outbreak by local twitter accounts
- More than 2,000 messages have circulated when including re-tweets
- Quantitative estimates of Twitter activity of #Chikungunya to infer awareness of the epidemics
- Textual analysis of each message to identify when individuals where looking for protection against disease



Roche et al, Sci Rep (2016)

Temporal dynamics at the island scale



$$\beta(t) = x_0(x_1 a(t))(1 - x_2 p(t))(1 + x_3 q(t))$$

a(t): Probability of mosquito presence (Mosquito abundance)

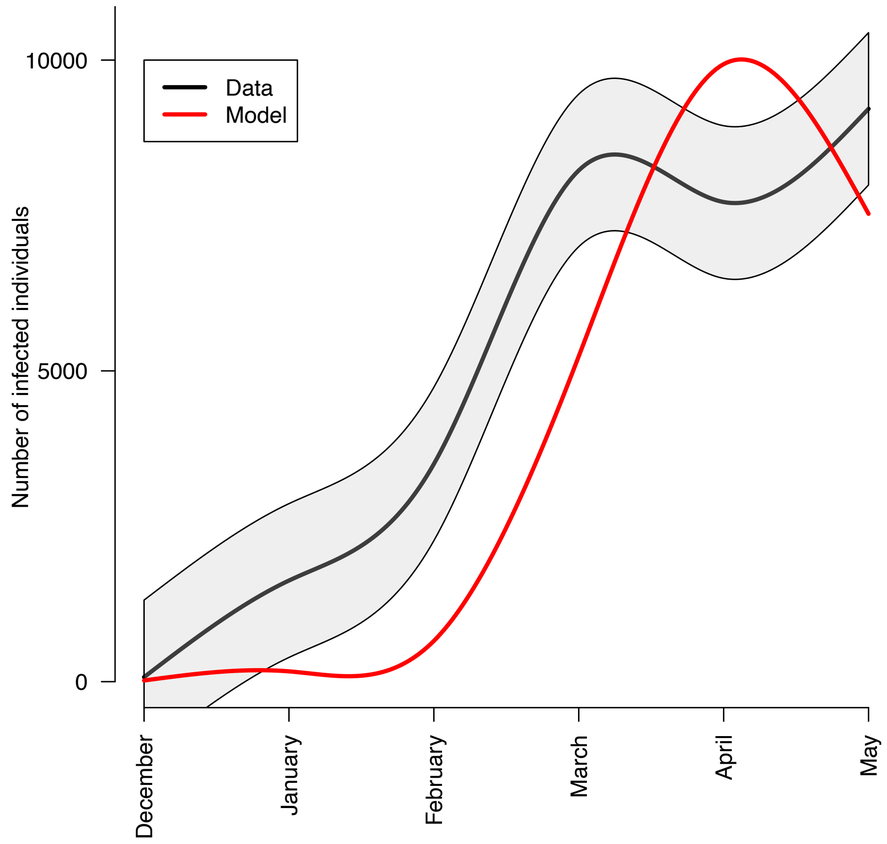
p(t): Proportion of tweets with protection seeking (Protection seeking)

q(t): Scaled number of tweets with #Chikungunya (Epidemics awareness)

Roche et al, Sci Rep (2016)

Temporal dynamics at the island scale

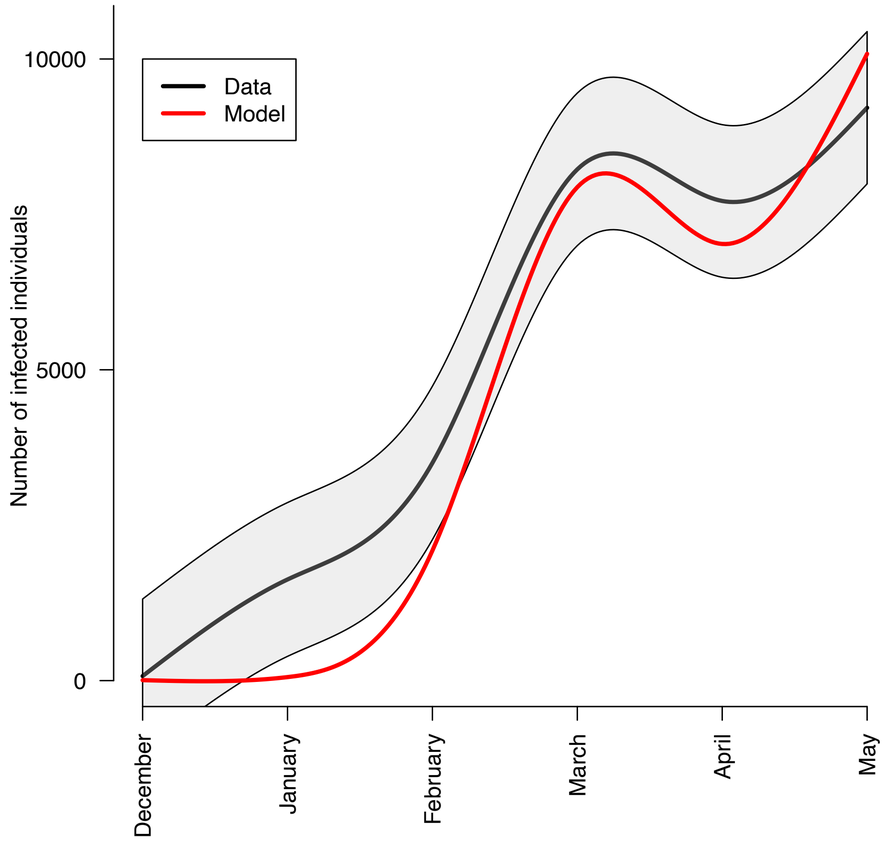
- Simple SEIR including only mosquito abundance miserably failed to mimic the observed pattern
-



Roche et al, Sci Rep (2016)

Temporal dynamics at the island scale

Parameter	MSE
None	10771
Mosquito abundance (MA)	8488
Protection seeking (PS)	8766
Epidemics awareness (EA)	10333
MA and PS	8041
MA and EA	7994
PS and EA	8634
MA, PS and EA	5715



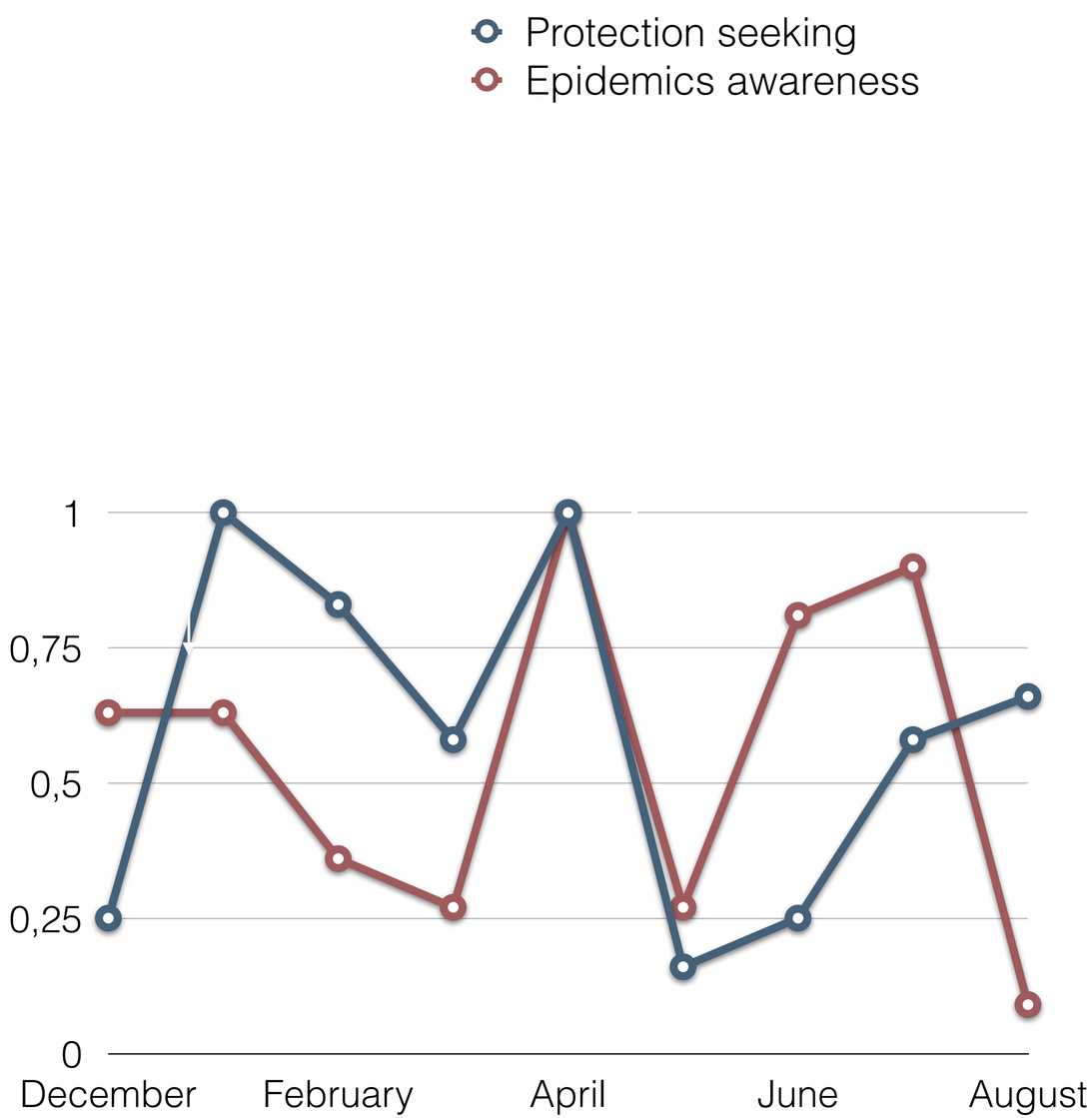
Roche et al, Sci Rep (2016)

Summary & Discussion

- Human behavior may have been an essential component in early stages of Chikungunya outbreak
- Twitter messages as a proxy for human behavior ?
  - Protection seeking dynamics quantified on Twitter matches with the number of requests made to vector control services…
  - Needs more comparative studies between textual analysis of Twitter messages and more « conventional » methods (*e.g.,* large-scale poll)

Understanding mistakes

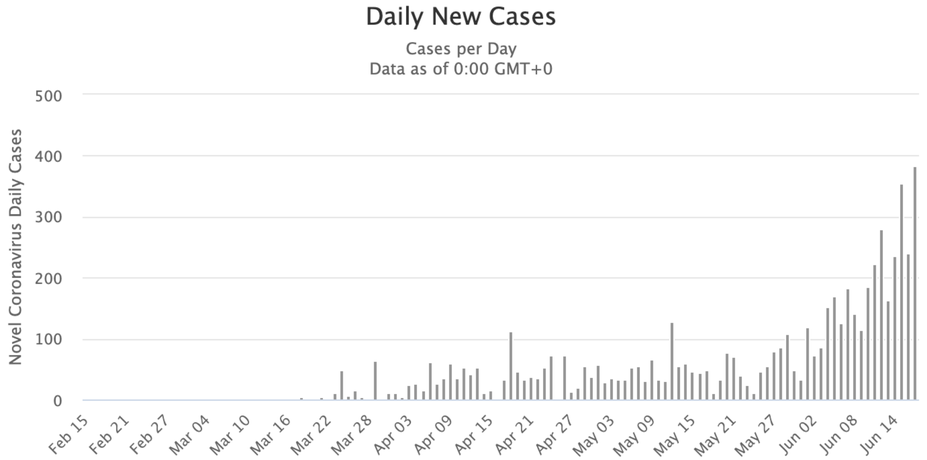
- Some communication campaigns may have been terrible
- Example: « Dengue/ Chikungunya, same fight »
-

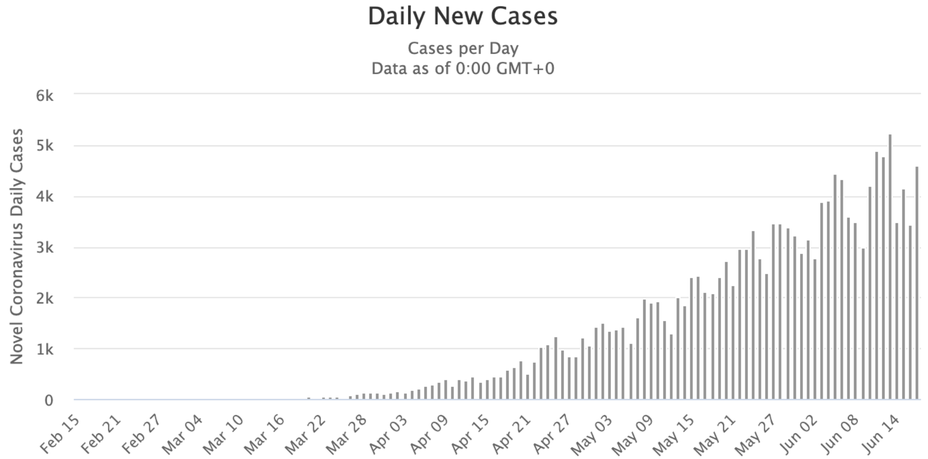


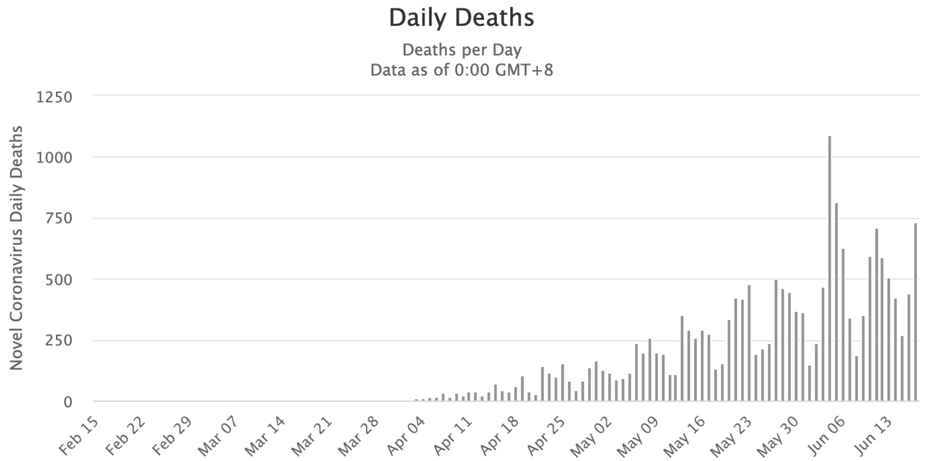
On-going project for COVID-19

Transmission fluctuations of COVID-19

- COVID-19 is highly fluctuating according to public health responses measures
- Number of confirmed cases are not reliable, ICU and deaths are, but they are delayed

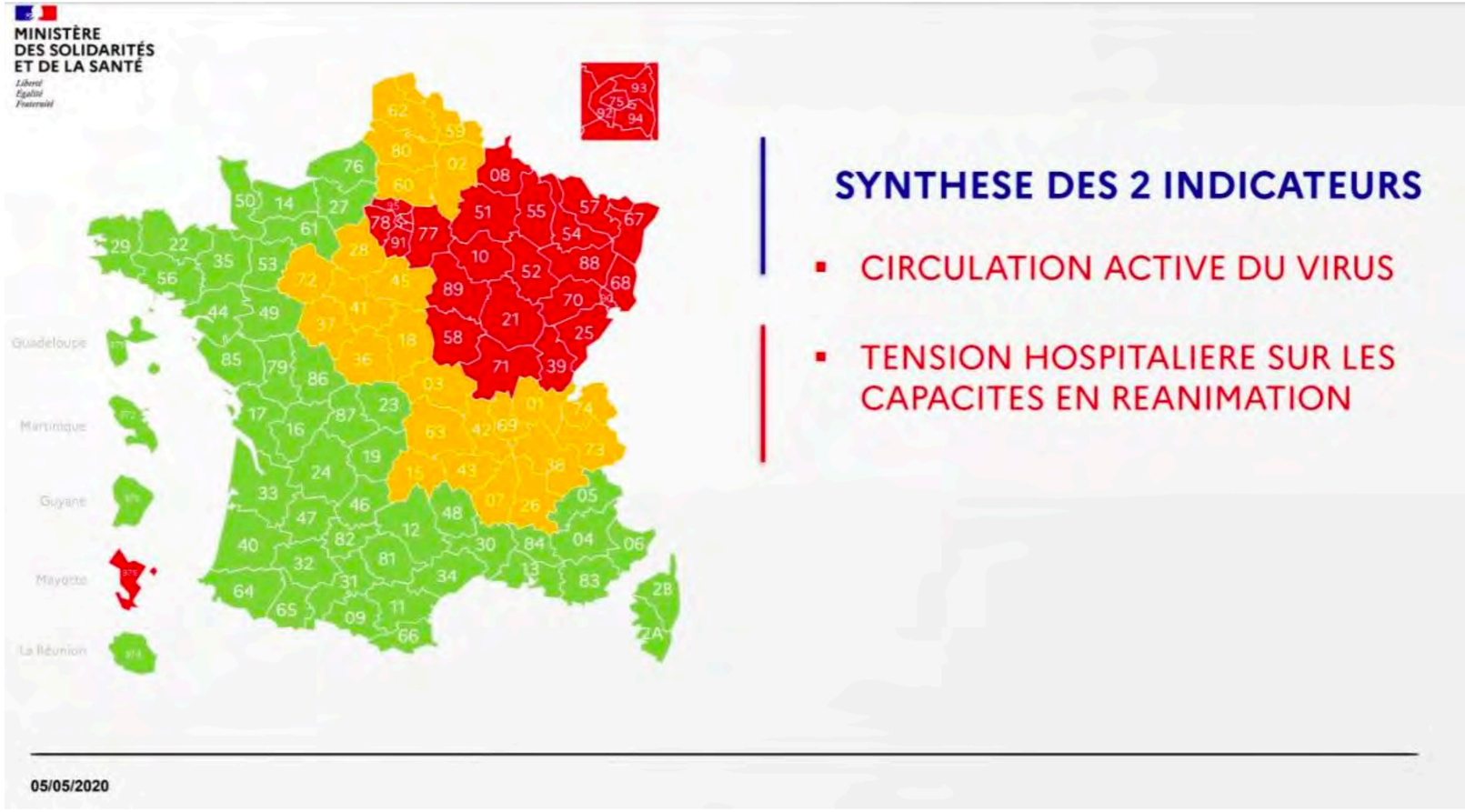






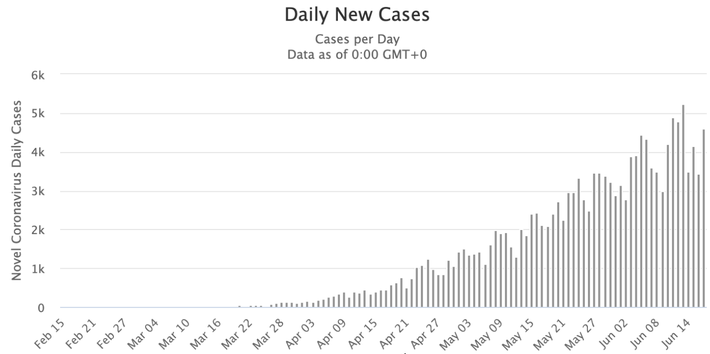
Anticipating these fluctuations

- Local fluctuations will drive governmental decisions
- Anticipating these fluctuations is challenging because:
  - Delay between « real » notifications and transmission increase
  - Rely on human behavior



DigEpi project

ANR COVID-19 Flash



twitter

Automatic classification

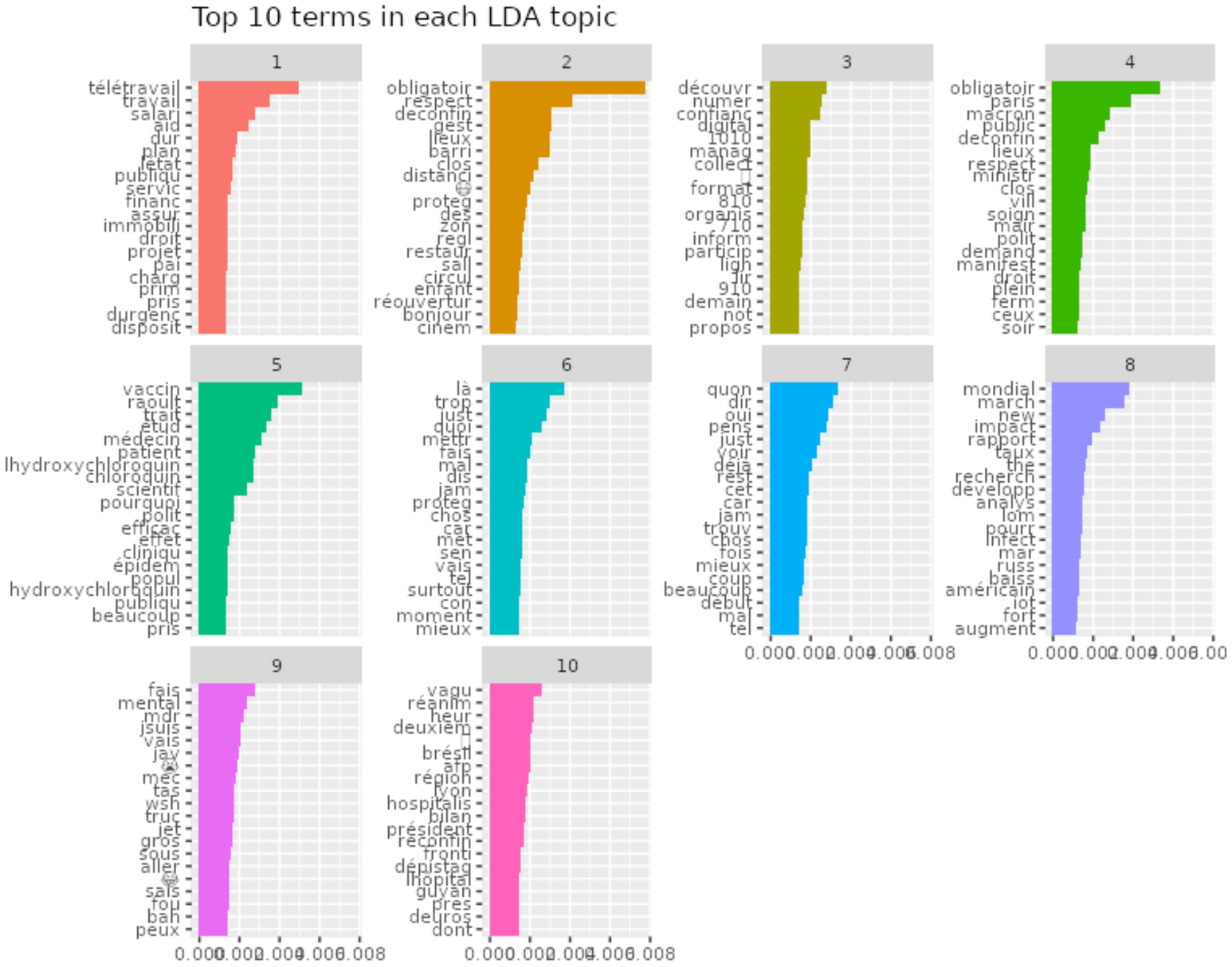
Parameter estimation through time

Synchrony estimation

Forecast at short-term

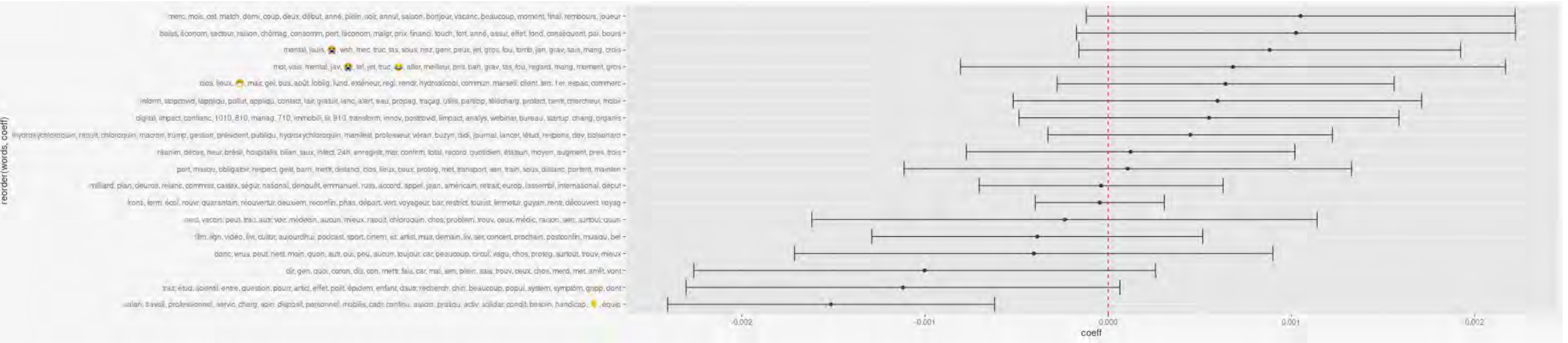
Classification results

- Different topics have been identified, which are well structured
- However, the words inside each topic are not intuitively connected together, hampering their use to understand human behavior



Association of each topic on transmission intensity

- Few topics are significantly associated with an increase or decrease of transmission intensity



Association of each topic on transmission intensity

- Few topics are significantly associated with an increase or decrease of transmission intensity

Digital Epidemiology for COVID-19

Target

Department

Marseille

Date range to train model:

2020-03-01 to 2020-07-15

Synchrony

Time shift between tweets and epidemics (days):

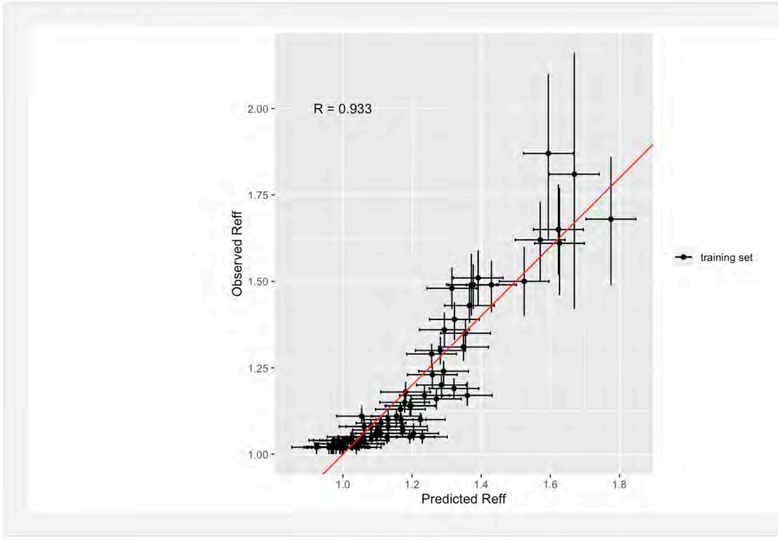
2 3 4 5 6 7

Time to Forecast

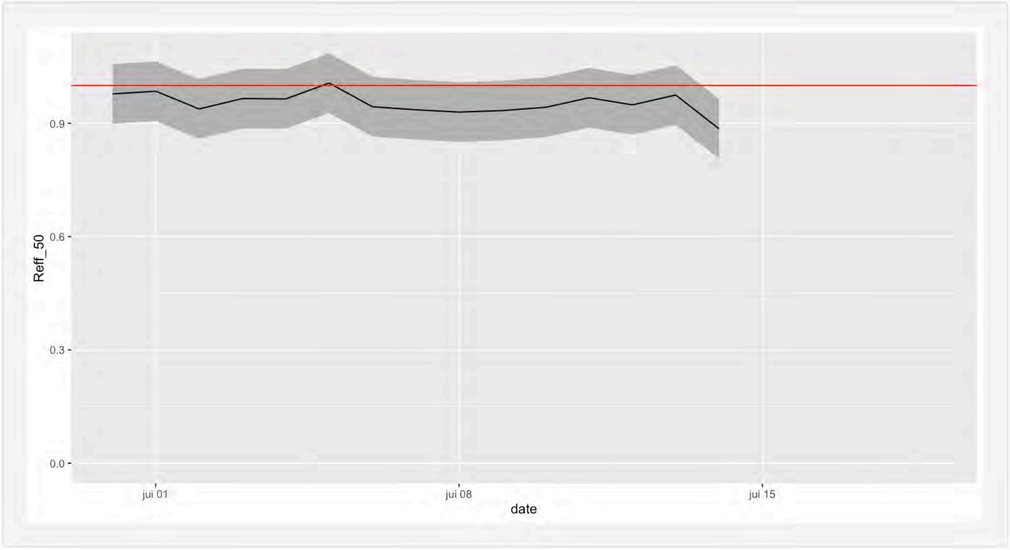
2 3 4 5 6 7

Forecast

Quality of fit



Forecast



Current steps

- Improving the accuracy of the classification procedure (participative approach)
- Complexifying the mathematical model involving twitter data

Digital Epidemiology: A new method to infer human behavior and pathogen transmission

Benjamin Roche