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 ?





Source : Twitter Ads manager Audiences - extrapolations de We Are Social Digital Report 2020

Démographie : Age et genre des utilisateurs de Twitter dans le monde





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?



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

GENE				LUTTE		PEUR
piquer pique pique piques	arelles arelle	veller	entenel	entimeustique	And the	Mada
Ecutori			tigne		tellament ti	atitica
gratts-gratter	i bleq i semineli		22224 146		raju- ett.	ينفذ
STERS		italeur			semit. mous expt.	JEL JEL



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

Jourdain et al, submitted

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

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

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

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





2014.0 2014.1 2014.2 2014.3 2014.4 2014.5 2014.6 Date



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











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



Temporal dynamics at the island scale



$\beta(t) = x_0(x_1a(t))(1 - x_2p(t))(1 + x_3q(t))$

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

Roche et al, Sci Rep (2016)

- p(t): Proportion of tweets with protection seeking (Protection seeking)
- q(t): Scaled number of tweets with #Chikungunya (Epidemics awareness)

Temporal dynamics at the island scale

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



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

500

400

300

200

100

Daily

Novel

- 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

Cases

Novel Cor



Daily New Cases

Cases per Day Data as of 0:00 GMT+0





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



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



0.000.000.004.006.008

Top 10 terms in each LDA topic

0.000.000.000.000.008



Association of each topic on transmission intensity

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

merc, mois, cet, match, demi, coup, deux, début, anné, plein, soir, annul, saison, bonjour, vacanc, beaucoup, moment, final, rembours, joueur baiss, économ, sedaur, reison, chômag, consomm, pert, léconom, malgr, prix, financi, touch, fort, anné, assur, effet, fond, consèquent, pai, bours mental, jsuis, 😭, wsh. mec. truc, tas, sous, nez, geni, peux, jet, gros, lou, tomb, jen, grav, sais, mang, ciciis mot vais mental lav 😭 tet jet trug 😆 aller melleur pris ban grav tas fou regard mang moment gros icios, lieux, 😷, mair, gel, bus, aoùt, lobig, lund, extérieur, regi, rendr, nydrosloobl, commun, marseil, client indy, postcovid, limbact analys, webinar, bureau, startup, chang, organiskont, ferm, écol, rouvr, guarantain, réouvertur, deuxiem, reconfin, phas, départ, vert, voyageur, bar, restrict, tourist, fermetur, guyan, rentr, découvert, voyag nest, vaccin, peut, trait, aut, whit medeoin, aucun, mieux, raouit, chioroguin, chos, problem, trouv, ceux, médic, raison, sen, sunour, guunfilm, idn. vidéo, liw, cultur, autourdhui, podcast, sport, cinem, sit, artist, mus, demain, liv, ser, concert, prochain, postconfin, muaiqu, bet donc, wrus, peut, nest, moin, quon, autr, our, peu, aucun, toujour, car, peaucoup, circui, vagu, chos, proteg, surfout, trouv, mieuxdir, gen, quor, coron, dia, con, mettr. faia, car, mal, sen, plain, saia, trouv, ceux, citos, merd, met, arrêt, vonttrait, étud, spienfil, entre, guestion, pour, artici, elfet, poirt, épidem, enfant, dautr, recherch, chim, beaucoup, popul, system, symptôm, gripp, dont--salari, travail, professionnel, servio, charg, soin, disposit, personnel, mobilis, cadr. continu. associ, pratiqui, ectivi solidar, pondit, bespin, handicap, 🎈 , équip -



Association of each topic on transmission intensity

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

merc; mois, get, match, demi, coup, deux, début, anné, plein, soir, annul, saison, bortjour, vacanc, beaucoup, moment, final, remibours, joueur baiss, économ, secteur, reison, chômag, consomm, pert, léconom, malgr, prix, financi, touch, fort, anné, assur, effet, fond, conséquent, pai, bours -

mot, yais, mental, jav, 😭, tel, jet, truo, 😂, aller, meilleur, pris. ban, grav, tas, fou, regard, mang, moment, gros -

iolos, lieux, 😷, mair, gel, bus, aoùt, loblig, lund, exténeur, regl, rendr, nydrosloopl, commun, marsell, client, len, 1 er, espac, commerc -

inform, stoprowd, lappliqu, pollut, appliqu, contact, lair, gratuit, lano, alert, eau, propag, traçag, utilis, particip, télécharg, protect, centr, chercheur, mobil -

oligital, impact, confianc, 1010, 810, manag, 710, immobili, liit 910, transform, innov, postcovid, limpact, analys, webinar, bureau, startup, chang, organis -

chloroquín, macron, trump, gestion, président, publiqu, hydroxychloroquín, manileát, professeur, véran, buzyn, didi, journal, lancet, lélud, respons, dev, bolsonard +

reanim, deces, neur, bresil, hospitalis, bilan, taux, intect, 24h, enregistr, mar, confirm, total, record, publicien, étatsun, moyen, augment, pres, trois -

port, masou, obligatoir, respect, geat, barri, mettr, distanci, clos, lieux, caux, proteg, met, transport, aan, train, sous, distanc, portent, mainten =

m urd, plan, deuros, relanc, commas, castax, ségur, national, denquêt, emmanuel, russ, accord, appel, jaan, américam, retrait, europ, lassembl, international, déput -

front, ferm, égol, rouvr, quarantain, réouvertur, deuxiem, reconfin, phas, départ, vert, voyageur, bar, restrict, tourist, fermetur, guyan, rentr, découvert, Voyag =

nest, vacon, peut, trait, autr, voir, médeoin, aucun, mieux, raouit, chloroguin, chos, problem, trouv, ceux, médic, raison, sen, surbui, guun -

film, ich, vidéo, livr, ou iur, aujourdhui, podcast, sport, cinem, sit, artist, mus, demain, liv, ser, concert, prochain, postconfin, musiqu, bel -

donc; wrus, peut, nest, main, quon, autr, oui, peu, sucur, toujour, car, peaucoup, circul, vagu, chos, proteg, surfout, trouv, mieux -

dir, gen, quor, coron, dia, con, mettr, faia, car, mal, sen, plain, saia, trouv, ceux, chos, merd, met, amét, vont-

trait, étud, scientil, entre, question, pouir, artici, effet, polit, épidem, enfant, dautr, recherch, chin, beaucoup, popul, system, symptóm, gripp, dont - 🛌

salari, travali, professionnel, servio, charg, spin, disposit, personnel, mobilis, cadr centinu, associ, pratiqu, activ, solidar, pondit, bespin, handicap, 🔫 , équip -

mental, jeuis, 😭, wsh, med, truc, tas, sous, nez, genir, peux, jet, gtos, lou, tomb, jen, grav, sais, mang, crois -



Digital Epidemiology for COVID-19

arget			Sync
Department			Time sh
Marseille		•	
Date range to train model:	_		12
2020-03-01	to	2020-07-15	2
			0





Quality of fit





chrony nift between tweets and epidemics (days): 5 6 Time to Forecast 7 1 11 3 5 6

Forecast

Current steps

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



Institut de Recherche pour le Développement FRANCE

Le projet DigEpi (pour Digital Epidemiology), financé par l'Agence Nationale pour la Recherche, a pour objectif d'apprécier l'impact des perceptions de la population française exprimés sur twitter vis à vis de la pandémie actuelle de COVID-19. Une fois ces perceptions identifiées et leur évolution au cours du temps quantifiée, nous souhaitons, au travers de modèles mathématiques et statistiques, comprendre comment ce type de données peut permettre d'alerter les autorités sanitaires de façon préventive sur une augmentation attendue de cas de coronavirus. Néanmoins, nous avons besoin d'apprendre à nos algorithmes comment classifier ces messages (c'est ce qu'on appelle l'apprentissage supervisé). C'est pour cela que nous avons besoin de vous.

Ainsi, vous pouvez contribuer activement à ce travail de recherche en identifiant pour chacun des messages affichés la (ou les) thématique(s) à laquelle (auxquelles) il se réfère et le sentiment qui y est associé ("rassuré", "adhésion", etc...). Vous pouvez également indiquer que le message ne convient à aucun des champs proposés, ou même si le message ne parle tout simplement pas de la pandémie de COVID-19. Les messages sont actualisés très fréquemment, ainsi que les thématiques proposées. Plus le nombre de participants est important, et plus nous pourrons avoir un algorithme fiable permettant d'aider les autorités de santé publique à anticiper les conséquences de cette pandémie. Seules vos classifications sont enregistrées, aucune donnée personnelle n'est stockée.

Le code de l'application est disponible sur GitHub. Pour toute question, vous pouvez envoyer un email à benjamin.roche@ird.fr.

IRD 🛅 Banques 🔰 Twitter 存 Facebook 🚦 Franceinfo 🖪 YouTube 😵 Sports.fr 📊 LinkedIn 💦 ResearchGate 🎧 GitHub S Verts-Luisants 🚱 Blocks 🚱 Shiva 🚱 ..:: ADUM - Espac...





Epidémie de COVID-19 : soyez acteur de la recherche

Tweet à classer:

Si le vaccin soignait aussi la calvitie le problème il serait résolu

Ce tweet ne parle pas du COVID-19

Connaissances et informations liées à la maladie/au virus

Confiance Adhésion	Neutre	Nuancé	Méfiance Opposition	
Propagation de	la maladie			
Rassuré	Neutre	Doute	Inquiet	
Mesures pour la	distanciation phys	sique (confinement, masqu	ues, isolement, femeture	2
Adhésion à la mesure	Neutre	Incompréhension de la mesure	Rejet de la mesure	
Vaccination				
Adhésion à la mesure	Neutre	Incompréhension de la mesure	Rejet de la mesure	
Action des pouv dépistage)	voirs publics (politio	que sanitaire, mise en œuv	re des mesures, stratég	i
Confiance Adhésion	Neutre	Nuancé	Méfiance Opposition	
Impact psycholo	ogique personnel (stress, anxiété, déprime, s	olitude)	
Positif	Neut	re	Négatif	
Impact économ	ique (perte de reve	nus impossibilité de trava	iller)	



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

Benjamin Roche