



Traquer les maladies, un tweet à la fois

SANTÉ NUMÉRIQUE Les traces que nous laissons sur la Toile ne servent pas qu'à pister nos habitudes de consommation: elles sont également prisées des épidémiologistes, qui les examinent sous toutes les coutures pour étudier des maladies



L'analyse des tweets permet de récolter des informations épidémiologiques hors de tout contexte médical. (DADO RUVIC/REUTERS)



MARCEL SALATHÉ
 CHEF DU
 LABORATOIRE
 D'ÉPIDÉMIOLOGIE
 DIGITALE, EPFL

INTERVIEW

PROPOS RECUEILLIS PAR FABIEN GOUBET
 @fabiengoubet

Un épidémiologiste «un peu geek» venu mettre une touche de numérique dans les pratiques de sa communauté: voilà comment se présente Marcel Salathé, 40 ans. Fer de lance de l'épidémiologie digitale dans son laboratoire de l'EPFL, le chercheur coorganise aujourd'hui et demain à Genève un symposium international sur la santé numérique et personnalisée. Il y racontera comment il fouille, à l'aide d'ordinateurs, des milliards de tweets afin d'en extraire des informations pertinentes destinées à étudier, façon

Web 2.0, la propagation des maladies.

Vous travaillez dans un champ d'études émergent, l'épidémiologie numérique. En quoi consiste votre travail? Lorsque des confrères épidémiologistes me posent la question, je leur réponds souvent que je suis leur collègue un peu geek venu pour leur donner un coup de main! En tant qu'épidémiologistes numériques, nous les aidons à faire sens des gigantesques masses de données informatiques qui constituent pour eux des sources inédites d'informations. Alors que leur travail habituel repose sur des données de nature biologique ou médicale, son pendant numérique s'appuie sur les traces que nous laissons sur Internet, telles que nos publications sur les réseaux sociaux. Nous étudions ainsi les tweets postés sur Twitter ou les requêtes sur Google et essayons d'en déduire, par exemple, comment les épidémies de grippe se propagent.

Quel est l'intérêt de ce type de données? C'est leur contexte. En épidémiologie classique, les données proviennent d'exams médicaux, en tout cas d'un environnement purement médical. C'est un biais important: le fait d'être dans un cabinet, voire de remplir un questionnaire de santé, oriente fortement les réponses des participants. Un tel biais est inexistant sur Twitter: les gens parlent librement sans savoir que nous allons utiliser leurs

tweets a posteriori. De plus, chaque tweet est associé à une date de publication, voire à une géolocalisation, c'est donc une précision supplémentaire. Bien sûr, travailler avec des tweets est un défi dans le

sens où ces données sont dénuées de structure, il faut les «nettoyer» avant de pouvoir les utiliser. Mais c'est aussi une opportunité à saisir car ils renferment des informations qui nous échappent totalement en temps normal.



Le Temps / Sortir
1002 Lausanne
021 331 78 00
www.letemps.ch

Genre de média: Médias imprimés
Type de média: Presse journ./hebd.
Tirage: 36'802
Parution: 6x/semaine

N° de thème: 999.056
N° d'abonnement: 1086739
Page: 12
Surface: 67'989 mm²

Est-ce à dire que la manière dont on étudie les épidémies n'est plus satisfaisante? Non, pas du tout, et d'ailleurs nous ne sommes pas là pour

remplacer les épidémiologistes! Ils font très bien leur travail, mais il existe selon moi trois principaux défauts que l'on pourrait corriger avec un savoir-faire numérique. D'abord, dans les pays moins développés, l'épidémiologie classique atteint ses limites: très peu de gens peuvent consulter un médecin, alors qu'ils ont en général plus facilement accès à Internet. Deuxième point, les données épidémiologiques sont biaisées: elles reposent plus sur les maladies (puisque nous n'allons consulter que lorsque nous sommes malades) que sur la santé. Enfin, la majorité des données restent prisonnières des établissements et des systèmes informatiques fermés dans lesquels elles ont été collectées, ce qui limite leur utilisation.

Pourquoi la discipline émerge-t-elle si tard? Internet et les réseaux sociaux ne datent pourtant pas d'hier... En fait, les premières idées remontent aux débuts d'Internet. Mais démontrer qu'une nouvelle méthode a un intérêt prend du temps. C'est ainsi que fonctionne la science. Et puis il faut garder en tête que nous travaillons avec d'importantes quantités de données: il a donc fallu attendre que les réseaux sociaux soient massivement adoptés avant de démarrer. N'oublions pas non plus que nous ne collectons que des données publiquement accessibles (c'est le cas des tweets), pas des statuts Facebook privés ou des historiques individuels de recherche sur Google. Forcément, cela prend plus de temps.

Comment compareriez-vous la surveillance des épidémies d'il y a quinze ans avec celle que vous promouvez aujourd'hui? L'épidémiologie numérique est beaucoup plus rapide, plus

réactive. Elle permet de répondre à des questions insolubles avec les outils classiques. Prenons l'exemple du débat sur la vaccination, qui fleurit sur les réseaux sociaux. En analysant les commentaires, nous pouvons en extraire du sens et obtenir des informations que nous n'aurions jamais acquises en temps normal.

Comment analysez-vous les tweets? C'est un mélange de tâches informatiques et humaines. La plupart du travail est assuré par des machines capables d'analyser d'énormes quantités de données en très peu de temps. Elles réalisent, grâce à des algorithmes d'intelligence artificielle d'apprentissage automatique («machine learning»), un premier tri des tweets en éliminant les messages inutiles. Ensuite, des êtres humains contrôlent si les tweets conservés sont bien utilisables pour notre étude. A l'avenir, la partie humaine devrait cependant tendre à la diminution. Les algorithmes apprennent en effet de leurs erreurs et deviennent chaque jour de plus en plus performants. Bientôt, ils seront meilleurs que nous.

Qu'en pensent les grandes instances de santé publique telles que l'Organisation mondiale de la santé (OMS)? Il serait prématuré de dire qu'elles intègrent l'épidémiologie numérique dans leurs opérations. Disons qu'elles se montrent très enthousiastes et qu'elles suivent nos travaux, et ceux des autres équipes, avec intérêt. Que l'on parle de l'OMS ou des centres européen ou américain de prévention et de contrôle des maladies, il s'agit d'instances qui avancent avec prudence et c'est la moindre des choses lorsque l'on agit sur la santé publique.

La transformation numérique de la santé soulève des questions de confidentialité. Bien que vous travailliez sur des données publiques, vous sentez-vous concerné? Tout à fait. Ce qui est certain, c'est que de plus en plus

de données vont circuler du consommateur au service en court-circuitant les instances de santé. Il faudra donc réfléchir à la manière dont on veut les protéger, ce qui constitue un important défi. A ce titre, la Suisse représente un parfait laboratoire d'expérimentation à grande échelle. Certaines plateformes suisses innovantes telles que Healthbank.coop ou Midata.coop proposent ainsi de centraliser toutes nos données relatives à la santé en un seul et même endroit, un peu comme une banque stocke notre argent. Pour accéder à ces données, chaque service, chaque laboratoire doit demander une autorisation. C'est une idée intéressante à tester, car cela garantit à la fois la confidentialité tout en donnant un accès sécurisé aux données dont la science a besoin pour progresser. ■

Date: 22.09.2016

LE TEMPS



Le Temps / Sortir
1002 Lausanne
021 331 78 00
www.letemps.ch

Genre de média: Médias imprimés
Type de média: Presse journ./hebd.
Tirage: 36'802
Parution: 6x/semaine

N° de thème: 999.056
N° d'abonnement: 1086739
Page: 12
Surface: 67'989 mm²

EN VIDÉO

Neuf experts pour autant d'avis

Quels sont les enjeux sociétaux de la médecine dite de précision? Et ses impacts économiques? Quels sont les droits dont devrait disposer chaque patient face à l'utilisation des données concernant sa santé? Et cette «médecine personnalisée» trouve-t-elle des ramifications dans une approche plus globale de la santé? Pour nourrir le débat, «Le Temps», en partenariat avec la Fondation Leenaards, a approché neuf experts romands de tous ces domaines, tels Bertrand Kiefer, directeur de «Médecine & Hygiène», le généticien Denis Duboule, l'éthicienne Samia Hurst, le psychologue social Fabrizio Buttera, l'épidémiologiste digital Marcel Salathé ou le juriste Dominique Sprumont. A découvrir sur: www.letemps.ch/dossiers/medecinedeprecision. ■ LT