

> retenu à un deuxième, qui en parlait au troisième, et ainsi de suite. À mesure de leur transmission le long de cette «chaîne de diffusion sociale», nous avons constaté que la quantité d'informations négatives augmentait (phénomène dit d'«amplification sociale du risque»).

Enfin, nous avons fourni à tous les participants les articles originaux. Mais cela n'a aucunement contribué à modifier les points de vue négatifs. Ainsi, à mesure qu'elle se transmet d'un individu à un autre, l'information devient non seulement de plus en plus négative, mais aussi de plus en plus prégnante, résistante à la remise en cause.

L'émotion joue un rôle majeur dans la diffusion de l'information. En 2015, en analysant des données sur Twitter concernant la «contagion émotionnelle», les chercheurs de l'OSoMe ont montré que les personnes surexposées à des contenus négatifs tendent à partager des posts négatifs, tandis que ceux surexposés à des contenus positifs partagent davantage les informations positives. Or, comme nous l'avons vu, les contenus négatifs se diffusent plus facilement que les contenus positifs. Dès lors, il est aisé de manipuler les émotions en lançant sur les réseaux sociaux des récits qui suscitent la peur ou l'anxiété.

C'est loin d'être une vue de l'esprit. Emilio Ferrara, de l'université de Californie du Sud, et ses collègues de la fondation Bruno-Kessler, en Italie, ont montré que, durant le référendum sur l'indépendance de la Catalogne en 2017, en Espagne, des robots sociaux ont été utilisés pour retweeter des récits violents et incendiaires, exacerbant ainsi le conflit social.

L'ASCENSION DES BOTS

Les bots sont simples à créer. Les médias sociaux fournissent des interfaces de programmation (API) qui permettent à quiconque d'installer assez facilement et de contrôler des milliers de bots. Or il suffit que quelques bots émettent des avis favorables lors de la publication d'un message pour que cela ait un impact considérable sur sa diffusion et sa popularité ultérieures.

À l'OSoMe, nous avons développé des algorithmes pour détecter ces bots. L'un d'eux, Botometer, est un outil public qui extrait 1 200 caractéristiques (profil, relations, langue, etc.) d'un compte Twitter. Il les compare à celles recueillies sur des dizaines de milliers de comptes précédemment identifiés comme bots, afin d'attribuer au compte Twitter une note indiquant sa probabilité d'être un faux profil robotique.

VULNÉRABILITÉ AUX FAUSSES INFORMATIONS

D'après une étude menée sur des utilisateurs de Twitter (résidant aux États-Unis) selon leurs penchants politiques, les libéraux autant que les conservateurs finissent par partager

des informations provenant de sites douteux (jugés ainsi par des vérificateurs indépendants). Toutefois, les utilisateurs à penchant conservateur sont un peu plus susceptibles de partager de fausses nouvelles.

Plus de 15 000 usagers de Twitter sont figurés sur cette matrice. La taille de chaque rond représente le nombre de comptes partageant le couple de coordonnées (biais politique, mésinformation), nombre allant de 1 à 429.

Pourcentage des tweets d'usagers partageant des liens de sources peu fiables

Risque de répandre de la mésinformation

Élevé

75

50

25

0

Faible

Très libéral

Centre

Très conservateur

Biais politique des usagers de Twitter

(inféré par les sources d'information partagées par l'utilisateur)

1

429

En pratique, nous avons estimé en 2017 que près de 15% des comptes Twitter actifs étaient des bots, et qu'ils avaient joué un rôle clé dans la diffusion de la désinformation durant la campagne présidentielle américaine de 2016. Dans les secondes qui suivaient la publication d'une fausse nouvelle – par exemple l'implication d'Hillary Clinton dans des rituels occultes –, cette «information» était retweetée par de nombreux bots, puis par des humains, séduits par l'apparente popularité du tweet.

Il suffit à un bot de suivre un compte, de cliquer sur des boutons «J'aime» ou de retweeter des posts pour infiltrer une communauté en ligne et influencer ses membres. Dans un modèle conçu à l'OSoMe, les agents sont des bots qui produisent des posts de qualité médiocre, de type «pièges à clics», et ne se retweetent que les uns les autres. Le modèle prend aussi en compte des agents authentiques, susceptibles de suivre ces bots.

Nous avons montré qu'en infiltrant seulement une petite fraction du réseau, ces bots peuvent réduire à néant la qualité de l'information du réseau entier. En suggérant aux agents authentiques de suivre d'autres comptes automatisés (technique dite de «follow trains»), ils accélèrent également la formation de chambres d'écho.

Certains manipulateurs agissent dans deux camps opposés, à des fins de polarisation politique ou pour gagner de l'argent. Nous avons ainsi démasqué un réseau de faux profils sur Twitter, tous coordonnés par une même entité. Certains affichaient leur soutien à la campagne en faveur de Donald Trump, tandis que d'autres se posaient en anti-Trump. Tous faisaient appel aux dons...

FREINER LA MANIPULATION EN LIGNE

En comprenant mieux nos biais cognitifs et la façon dont les bots les exploitent, nous pouvons nous protéger des manipulations. Notre laboratoire a mis au point des outils pour ce faire, par exemple Fakey, une application dont l'objectif est d'apprendre à repérer les fausses informations. Il s'agit d'un jeu qui simule le fil d'actualité d'un réseau social, affichant des articles réels issus de sources plus ou moins fiables. Les utilisateurs doivent décider lesquels vérifier, partager, etc. Notre analyse des données de Fakey confirme le comportement grégaire des utilisateurs: peu importe la crédibilité des informations, ils partagent plus facilement celles qu'ils croient largement diffusées.

Un autre outil, Hoaxy, également public, montre comment un post se répand sur Twitter. Les nœuds représentent des comptes Twitter réels et les arêtes qui les connectent indiquent la propagation du tweet de compte en compte. Chaque nœud a une couleur représentative du score qu'il a obtenu à Botometer, ce qui permet

BIBLIOGRAPHIE

G. Bronner, **Apocalypse cognitive**, PUF, 2021.

H. Y. Yan et al., **Asymmetrical perceptions of partisan political bots**, *New Media & Society*, publié en ligne le 16 juillet 2020.

N. Grinberg et al., **Fake news on Twitter during the 2016 U.S. presidential election**, *Science*, vol. 363, pp. 374-378, 2019.

D. Robert et al., **Bad news has wings : dread risk mediates social amplification in risk communication**, *Risk Analysis*, vol. 38, pp. 2193-2207, 2018.

X. Qiu et al., **Limited individual attention and online virality of low-quality information**, *Nature Human Behaviour*, vol. 1, article 0132, 2017.

B. Mønsted et al., **Evidence of complex contagion of information in social media : an experiment using Twitter bots**, *Plos One*, vol. 12(9), article e0184148, 2017.

aux utilisateurs de vérifier si des bots interviennent dans la diffusion. Des journalistes ont déjà utilisé ces outils pour rechercher les origines de campagnes de désinformation et repérer des opérations menées par des bots, comme celle visant à dissuader des électeurs de voter lors des élections de mi-mandat de 2018 aux États-Unis. Cependant, les manipulations sont de plus en plus difficiles à détecter, car les algorithmes deviennent très performants dans leur imitation du comportement humain.

Pour repérer des campagnes de désinformation lancées, par exemple, pour détourner l'attention du public, nous avons développé un outil appelé BotSlayer. Il suit les tweets, extrait des hashtags, évalue la probabilité des comptes d'être de faux profils, afin de signaler des thèmes qui se diffusent rapidement suspectés d'être amplifiés par des bots ou des comptes coordonnés. L'objectif est de permettre aux journalistes, aux associations et aux candidats politiques de repérer et de suivre en temps réel les campagnes d'influence insidieuse.

Ces outils sont utiles mais restent insuffisants pour limiter la prolifération de fausses nouvelles. L'éducation est certes un atout, mais elle ne peut aborder tous les sujets sur lesquels les gens sont susceptibles d'être bernés. Certains gouvernements et médias sociaux veulent agir pour réprimer la désinformation en ligne. Mais qui établit ce qui est faux, manipulateur et ce qui ne l'est pas? On pourrait accompagner certaines publications de messages d'avertissements, comme Facebook et Twitter ont commencé à le faire. Mais peut-on faire confiance aux personnes qui les émettent? En outre, le risque de nuire à la liberté d'expression existe. La portée mondiale de certains médias sociaux et leurs liens avec les gouvernements compliquent encore les choses.

Pour combattre la désinformation, une solution serait de rendre la création et le partage de l'information plus difficiles. Pour cela, il faudrait que l'accès à l'information ait un coût, par exemple en temps, en travail mental (faire un puzzle) ou en argent (en facturant des frais minimes). Certaines plateformes utilisent déjà les captchas ou la confirmation téléphonique comme sésames pour accéder aux comptes. Ces efforts pourraient être étendus afin de remplacer les incitations au partage à tout-va par des informations véritablement enrichissantes.

L'information prétendue gratuite ne l'est jamais réellement. En diminuant son coût, nous avons diminué sa valeur et ouvert la porte à sa falsification. Pour restaurer la santé de notre écosystème de l'information, nous devons restaurer une économie de l'information saine, qui nous protège des manipulations. ■

QUELQUES OUTILS

Modèle EchoDemo :
<https://osome.iuni.iu.edu/demos/echo>

Botometer :
<https://botometer.osome.iu.edu>

Fakey :
<https://fakey.iuni.iu.edu>

Hoaxy :
<https://hoaxy.osome.iu.edu>

BotSlayer :
<https://osome.iuni.iu.edu/tools/botslayer>