

APPLICATION DE TRACKING STOPCOVID : IDENTIFICATION DES THÉMATIQUES DE DISCUSSIONS ET DES OPINIONS PAR L'ANALYSE DES RÉSEAUX SOCIAUX

Paméla Voillot¹, Pierre Arwidson², Pierre Foulquié¹, Anne-Juliette Serry², Adel Mebarki¹, Nathalie Texier¹, Stéphane Schück¹

¹Kap Code, 28 rue d'Enghien 75010 Paris, France
²Santé publique France, 12, rue du Val d'Osne, 94 415 Saint-Maurice, France

INTRODUCTION

L'application StopCovid, mise en ligne le 2 juin 2020, permet de prévenir les individus qui ont été proches d'une personne testée positive au Covid-19 pendant au moins 15 minutes et à moins d'un mètre [1]. Celle-ci est téléchargeable sur la base du volontariat. Le fonctionnement de l'application a été approuvée par la CNIL et le code source est public. Malgré ces précautions, cette application divise la population. L'application a été téléchargée plus de 2,3 millions de fois à la mi-août. À ce jour, 1 169 usagers se sont déclarés positifs sur l'application permettant de repérer 72 contacts à risque.

Les réseaux sociaux sont un espace privilégié pour plus de 32 millions d'internautes. Ces plateformes occupent une place grandissante comme lieu d'expression où les internautes échangent sur leurs santé, leurs préoccupations et où la parole est libérée [2].

L'identification des opinions et des thématiques de discussions des internautes autour de l'application durant le confinement pourrait aider à mieux appréhender le comportement de la population envers cet outil numérique en post-confinement.

MATÉRIEL ET MÉTHODES

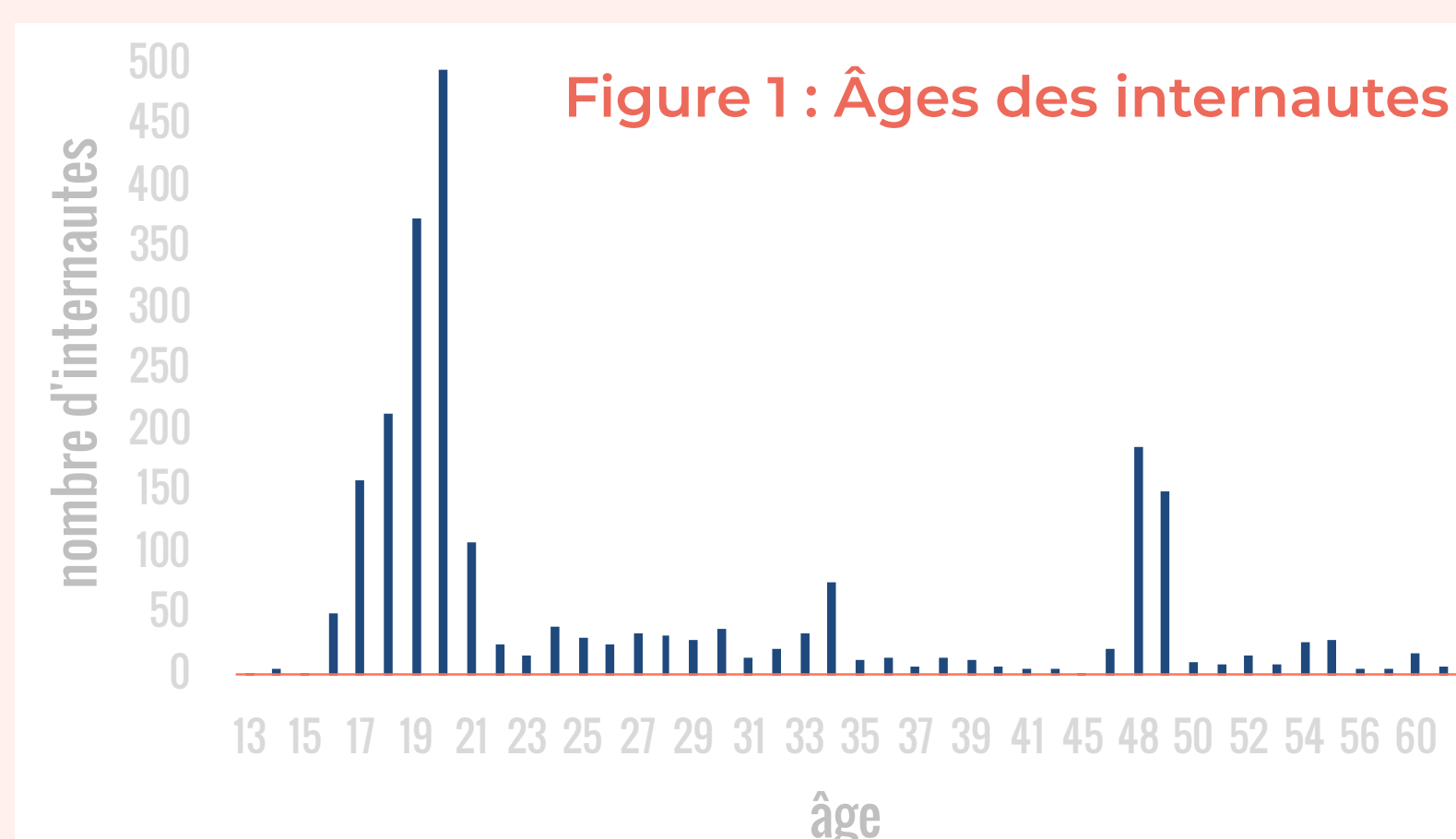
L'étude a été menée via l'outil Detec't®. Les messages associés à l'application StopCovid (dataset StopCovid) ont été extraits de Twitter entre le 13/04/2020 et le 06/05/2020.

La méthode d'analyse de sentiment de Microsoft Azure Cognitive Services a été utilisée pour identifier la perception générale des messages vis-à-vis de StopCovid (neutre, négative ou positive).

Par la suite et afin de caractériser le contenu de ces messages Twitter, l'algorithme Biterm Topic Model (BTM) a été appliqué sur l'ensemble du corpus afin de permettre une identification automatique des différentes thématiques de discussions abordées [3].

RÉSULTATS

Le dataset StopCovid contenait 7 110 tweets rédigés par 4 248 internautes différents avec une majorité de jeunes postants (moyenne = 27 ans) (figure 1).



Nous avons pu observer un accroissement du nombre de messages rédigés à ce sujet fin avril, concomitant avec l'avis de la commission supérieure du numérique du 24 avril 2020 sur les conditions de mise en œuvre de l'application StopCovid (figure 2).

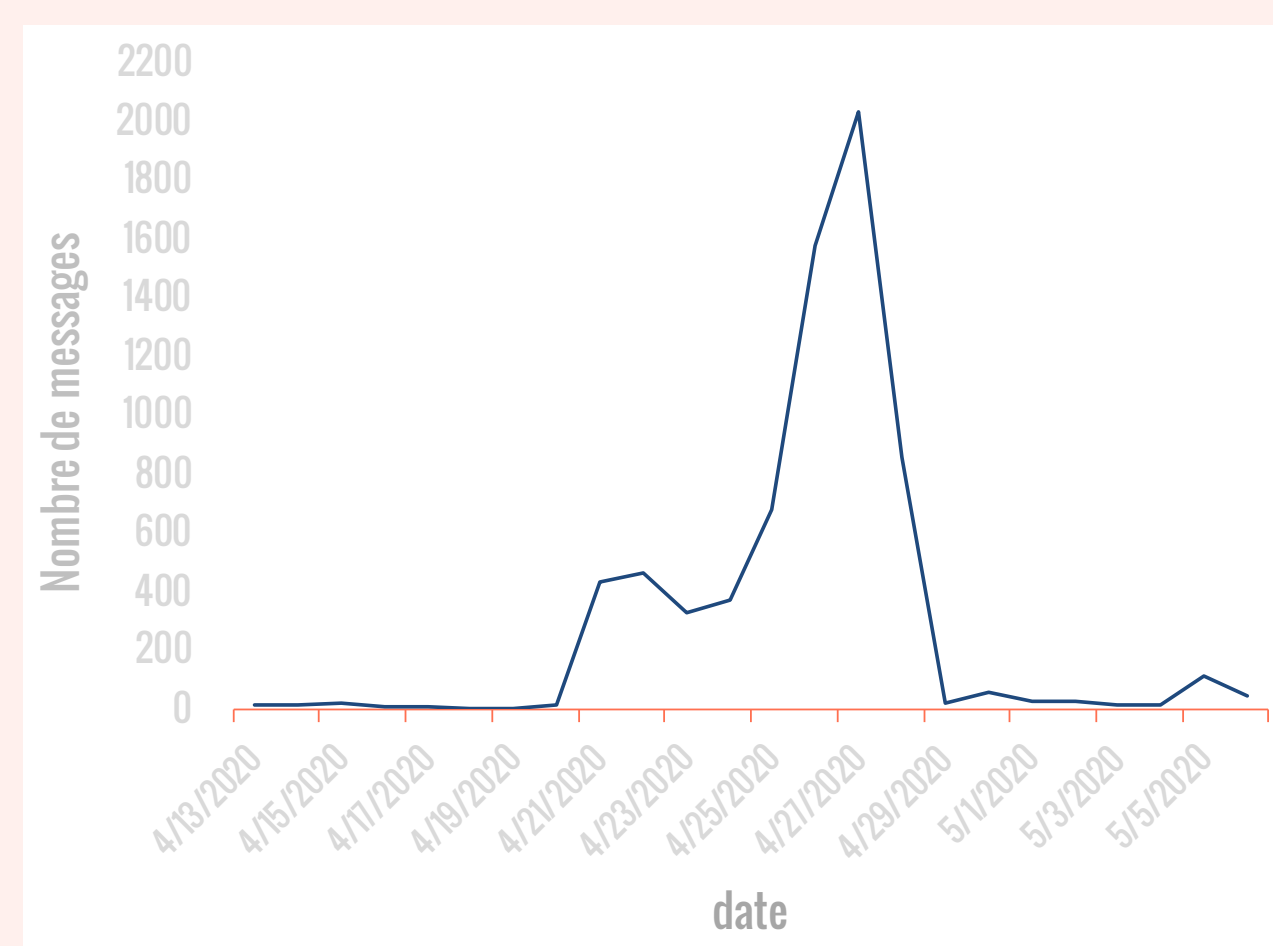


Figure 2 : Dates de publication des tweets concernant l'application StopCovid

L'analyse de la perception générale des messages a révélé une majorité de neutralité (53,29% des messages) (figure 3).

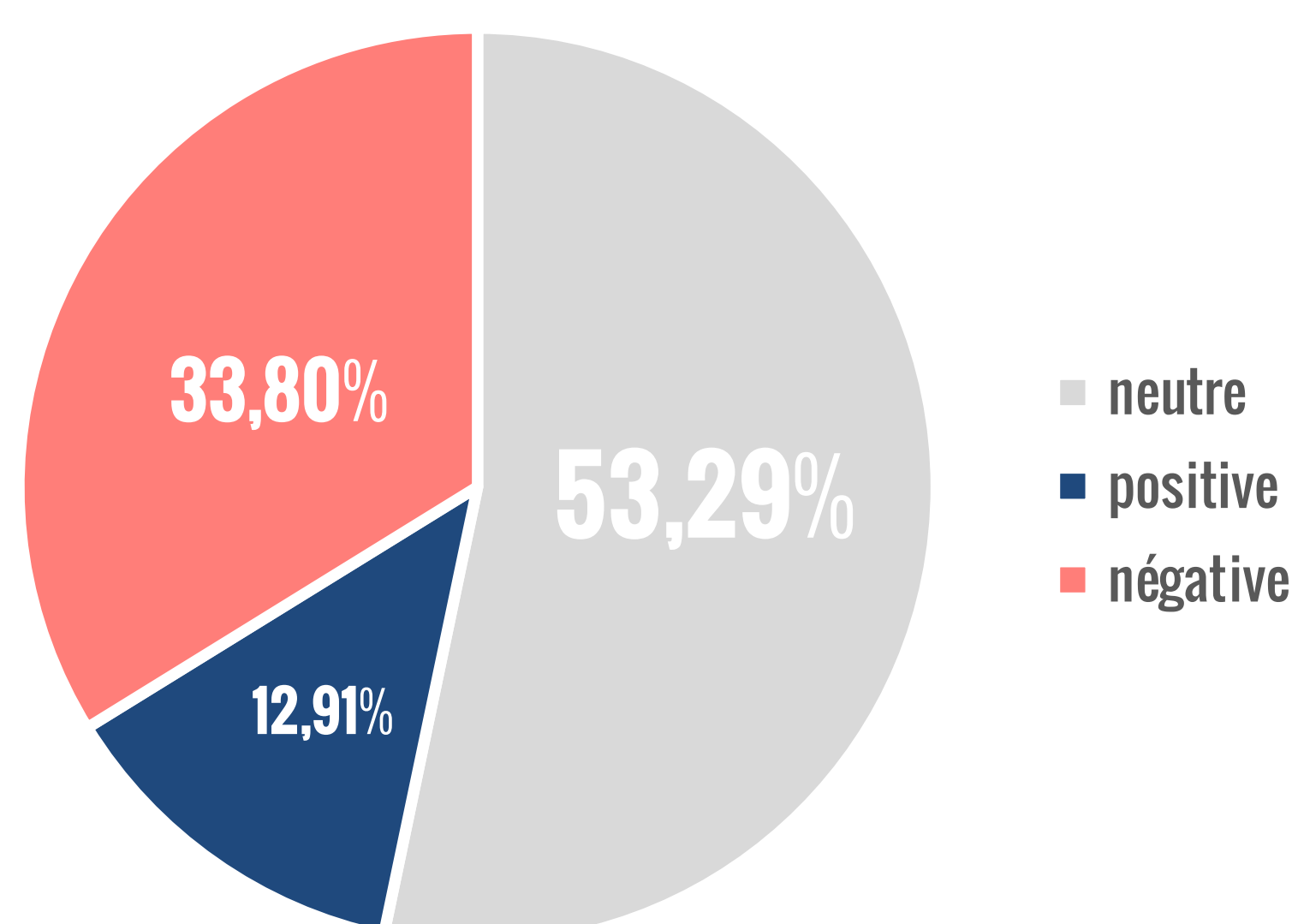


Figure 3 : Perception des messages des internautes concernant l'application StopCovid (%)

Ils se rapportent à des partages d'informations générales et diverses telles que les participants au projet ou encore l'objectif d'une telle application. Cette neutralité est suivie d'un grand nombre de messages avec une perception négative (33,80%) remettant principalement en question l'intérêt même du dispositif. Toutefois, 12,9% des messages avaient une perception positive, mentionnant l'utilité sanitaire du projet.

Par la suite, l'application de l'algorithme BTM a permis de mettre en avant 9 grandes thématiques de discussions dont 8 d'intérêt (figure 4).

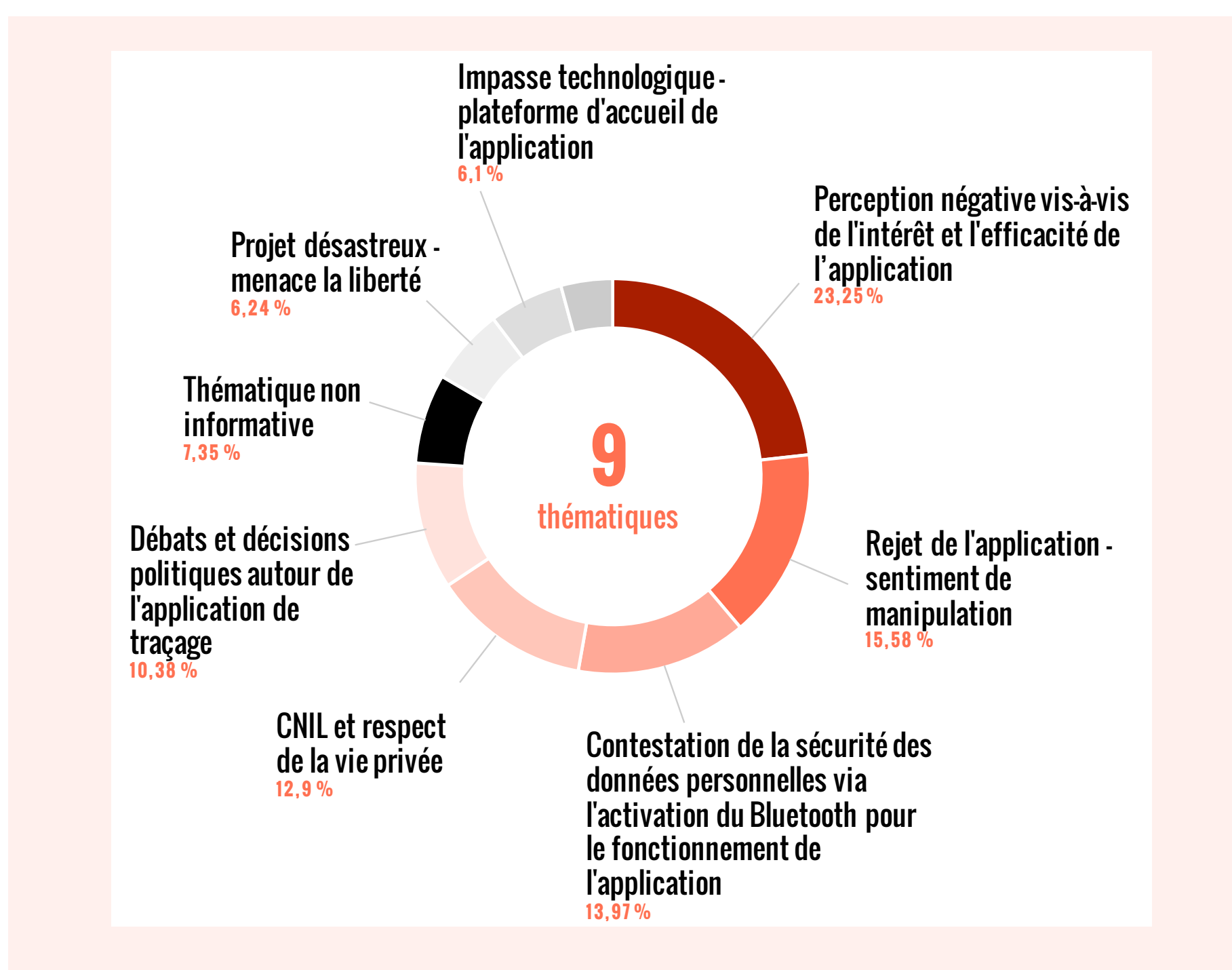


Figure 4 : Thématiques de discussions principales

La principale thématique se réfère aux opinions négatives des internautes vis-à-vis de l'intérêt et de l'efficacité de l'application (23,25% des messages) (figure 5).

« Cessez ce caprice de StopCovid ! Vraiment ce gadget n'aura aucune efficacité... Comptez vous procurer un smartphone à chaque citoyen ?! Arrêtez les frais ! Dépensez utilement et intelligemment > Masques Tests... »

Figure 5 : Ex. de tweet concernant la thématique « Perception négative vis-à-vis de l'intérêt et l'efficacité de l'application »

La seconde thématique la plus représentée dans le corpus se réfère au sentiment de manipulation par culpabilisation pour imposer l'utilisation de StopCovid (15,58%) (figure 6).

« Je refuse l'injonction culpabilisante que nos politiques, technos et autre relais médiatiques ne vont pas tarder à nous imposer sur StopCovid. Ce sera sans moi. »

Figure 6 : Ex. de tweet concernant la thématique « Rejet de l'application - sentiment de manipulation »

La troisième thématique met en avant l'insécurité des données personnelles via l'utilisation permanente du Bluetooth pour le bon fonctionnement de l'application (13,97%) (figure 7).

«Tiens, une nouvelle faille de sécurité dans une implémentation Bluetooth ! Nous maintenons : imposer Bluetooth risque de mettre en danger la vie numérique les utilisateurs de StopCovid»

Figure 7 : Ex. de tweet concernant la thématique « Contestation de la sécurité des données personnelles via l'activation du Bluetooth pour le fonctionnement de l'application »

Les thématiques suivantes s'en remettent au respect de la vie privée (12,9%), aux débats politiques (10,38%), aux menaces de liberté (6,24%), puis à l'impasse technologique concernant la plateforme d'accueil de l'application (6,1%) et à la mobilisation générale pour son développement (4,23%).

CONCLUSION

Cette étude a permis d'observer l'existence d'une communauté d'internautes en débat et en échange d'information concernant l'application de tracking StopCovid.

L'analyse du contenu des messages sur Twitter a principalement permis de mettre en avant des perceptions et opinions négatives vis-à-vis du dispositif. Cependant, certaines limites inhérentes à toutes les études sur les médias sociaux demeurent. Aucun média ne s'est avéré représentatif de la population générale [4]; par exemple, seuls 34 % des internautes français utilisent activement Twitter.

Malgré cela, nous pouvons observer que l'efficacité, l'insécurité des données personnelles et l'atteinte à la vie privée sont au cœur des échanges sur les réseaux sociaux à ce sujet.

Ce type d'étude infodémiologique pourrait servir aux stratégies de Santé publique afin de comprendre le ressenti des utilisateurs français et d'améliorer leur information.

RÉFÉRENCES

- [1] Info Coronavirus COVID-19 - STOPCOVID | Gouvernement, <https://www.gouvernement.fr/info-coronavirus/stopcovid>, 31 août 2020.
- [2] Fox S, Duggan M. PewInternet. 2013 Jan 15. Health Online 2013 URL: <http://www.pewinternet.org/2013/01/15/health-online-2013/> [accessed 2018-02-27] [WebCite Cache ID 6xXiCgyLK].
- [3] Blei DM, Lafferty JD, 2009. "Topic models," in Text Mining: Classification, Clustering, and 670 Applications, Vol. 10, eds A. N. Srivastava and M. Sahami (Boca Raton, FL: Chapman and 671 Hall/CRC), 34.
- [4] Hootsuite & We Are Social. Digital 2020: Global Digital Overview. DataReportal – Global Digital Insights <https://datareportal.com/reports/digital-2020-global-digital-overview>