



Une application du modèle ELM (*Elaboration Likelihood Model*) au partage d'information sur Twitter : étude du rôle de la forme du message et du profil de l'émetteur

Sophie Balech

DANS **INNOVATIONS 2022/3 n° 69** , PAGES 129 À 161

ÉDITIONS **DE BOECK SUPÉRIEUR**

ISSN 1267-4982

ISBN 9782807398146

DOI 10.3917/inno.pr2.0135

Date de mise en ligne : 07/10/2022

Article disponible en ligne à l'adresse

<https://shs.cairn.info/revue-innovations-2022-3-page-129?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour De Boeck Supérieur.

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

Une application du modèle ELM (*Elaboration Likelihood Model*) au partage d'information sur Twitter : étude du rôle de la forme du message et du profil de l'émetteur

Sophie BALECH

Institut d'Administration des Entreprises Amiens
Université Picardie Jules Verne - Laboratoire CRIISEA
sophie.balech@u-picardie.fr

RÉSUMÉ

Cet article s'intéresse aux comportements de partage de l'information des utilisateurs au sein d'une plateforme de micro-blogging, Twitter. Nous proposons un modèle explicatif de la performance d'un message en prenant en compte les signaux extérieurs (source et forme du message) au-delà du contenu et du sens du texte, et nous le testons empiriquement, sur un corpus de près de 800 000 tweets originaux émis par environ 235 000 utilisateurs sur une période de 7 mois concernant l'épidémie de Covid-19 en France. Nous montrons ainsi l'importance de la crédibilité de la source du message et de sa stratégie sur la plateforme, mais aussi celle de la forme du message, sa composition et son degré d'élaboration. Ces éléments sont nuancés en fonction du niveau d'engagement de la source dans le sujet de conversation sur lequel elle intervient et par le contexte dans lequel ces messages sont émis et reçus.

MOTS-CLÉS : Bouche-à-oreille électronique, Plateforme de micro-blogging, C2C, Twitter, NLP, Partage d'information, Modèle ELM

CODE JEL : M13

ABSTRACT

An Application of the ELM Model to Information Sharing on Twitter: Studying the Role of Message Form and Sender Profile

This paper focuses on the information sharing behaviour of users within a micro-blogging platform, Twitter. We propose an explanatory model of the performance of a message by taking into account the external cues (source and form of the message) beyond the content and meaning of the text, and we test it empirically, on a corpus of nearly 800,000 original tweets sent by about 235,000 users over a period of 7 months concerning the Covid-19 epidemic in France. We thus show the importance of the source's credibility and its strategy on the platform, but also of the form of the post, its composition and its degree of elaboration. These elements are nuanced by the level of engagement of the source in the topic of conversation on which it intervenes and by the context in which these messages are sent and received.

KEYWORDS: eWOM, Micro-blogging Platform, C2C, Twitter, NLP, Information Sharing, ELM Model

JEL CODE: M31

Twitter est le 5^e réseau social en France avec 12 millions d'utilisateurs. Cette plateforme est prisée pour le partage d'information en temps réel sur des sujets d'actualité (Asselin, 2021) et représente une forme d'économie de la contribution (Béraud, Cormerais, 2011) visant à transmettre des idées nouvelles et à créer du débat social. L'efficacité d'un message sur cette plateforme se mesure principalement par le nombre de partages que reçoit ledit message, les retweets (l'efficacité peut aussi se mesurer par le nombre de fois qu'un message a été sauvegardé par d'autres utilisateurs, ce qu'on appelle les favoris). En effet, un message retweeté par un utilisateur a été reconnu comme digne d'intérêt et a suscité suffisamment d'engagement pour que ce dernier souhaite le partager à sa communauté (ses *followers*, les membres de son audience) (Boyd *et al.*, 2010). Plus un message est viral, plus il montre ses capacités à susciter de l'engagement et plus son audience s'accroît, ce qui fait partie des objectifs des campagnes de marketing viral sur les réseaux sociaux (Sohn *et al.*, 2013). Cet article cherche à déterminer les éléments qui, au-delà du contenu même du message, favorisent sa diffusion. En s'appuyant sur le modèle ELM (Elaboration Likelihood Model) (Petty, Cacioppo, 1986 ; Petty *et al.*, 1983), nous cherchons à mettre en

évidence les signaux qui entraînent un comportement de partage au sein d'une plateforme de micro-blogging.

De nombreux travaux, notamment en informatique, communication ou journalisme, se sont intéressés aux processus de diffusion de l'information sur Twitter, mettant en avant notamment les cascades informationnelles, le rôle des membres dans un réseau, ou l'identification des influenceurs (voir par exemple Araujo *et al.*, 2017 ; Li, Liu, 2017 ; Yoo *et al.*, 2019). D'autres recherches se sont portées sur les impacts du bouche-à-oreille électronique sur les performances commerciales (Li *et al.*, 2019 ; Rui *et al.*, 2010) ou les caractéristiques clés des messages qui entraînent une forte diffusion dans différents contextes (voir par exemple Berman *et al.*, 2019 ; Nesi *et al.*, 2018 ; Pressgrove, McKeever, 2018 ; Van de Velde *et al.*, 2015).

Cependant, la théorie de la communication persuasive n'a que peu été utilisée pour expliquer les comportements de retweets. Nous proposons un modèle explicatif de la performance d'un message en prenant en compte des éléments relevant de la route périphérique et nous le testons empiriquement, sur un corpus de près de 800 000 tweets originaux émis par environ 235 000 utilisateurs sur une période de 7 mois concernant l'épidémie de Covid-19. À travers l'utilisation d'une méthodologie innovante, mêlant traitement du langage naturel (NLP) et analyse économétrique, le matériau brut des tweets est transformé en différentes variables permettant de tester notre modèle explicatif de la performance d'un message. Nous montrons ainsi l'importance de la source du message et de sa stratégie sur la plateforme, mais aussi celle de la forme du message, sa composition et son degré de sophistication. Ces éléments sont nuancés en fonction du niveau d'engagement de la source dans le sujet de conversation sur lequel elle intervient et par le contexte dans lequel ces messages sont émis et reçus.

L'article présente, dans un premier temps, le cadre théorique de l'ELM qui a permis de construire le modèle testé, en insistant sur le rôle de la source et de la forme du message dans la diffusion de l'information ainsi que l'effet du contexte et de l'engagement de l'émetteur. Le jeu de données et la méthodologie retenue sont présentés dans un deuxième temps, puis les résultats du modèle sont présentés, en commençant par le test du modèle global puis en abordant le rôle de l'engagement de l'émetteur et du contexte d'émission. Ces résultats sont discutés dans une dernière partie.

Cadre théorique

Le cadre de l'Elaboration Likelihood Model (ELM) (De Barnier, 2006 ; Petty, Cacioppo, 1986 ; Petty *et al.*, 1983) vise à comprendre comment se forment les attitudes qui conditionnent le comportement dans le cadre de la persuasion publicitaire. Le modèle met en évidence deux routes de traitement de l'information, une route centrale reposant sur le traitement cognitif des arguments clés, et une route périphérique reposant sur des signaux divers émis par la source ou les éléments de forme du message. Ce cadre s'est depuis vu adapté à la communication en ligne, notamment avec des études concernant les avis en ligne (pour une revue, voir Soulard, 2015) ou portant sur les médias sociaux (Teng *et al.*, 2014), en se concentrant sur le rôle des messages publicitaires ou des avis en ligne sur les choix de consommation. Nous proposons ici d'appliquer ce modèle aux contenus générés par les utilisateurs sur une plateforme de micro-blogging, Twitter, et nous intéressons donc à la persuasion des messages dans un environnement C2C. Plus précisément, considérant qu'un internaute naviguant sur un réseau social est exposé à trop d'informations pour toutes les traiter via la route centrale, nous nous intéressons aux éléments qui vont intervenir dans le traitement périphérique de l'information, et nous cherchons à voir le rôle que jouent certains signaux dans la persuasion entre utilisateurs, conduisant à produire du bouche-à-oreille électronique (eWOM). Conformément aux travaux sur le cadre ELM, ces signaux de la route périphérique peuvent être distingués en deux catégories : ceux émis par la source et les éléments de forme du message. Nous nous intéressons également à l'existence d'effets modérateurs sur la persuasion du contexte de réception du message et de l'engagement de l'émetteur dans le sujet de conversation étudié.

La source et la forme comme déterminants de la diffusion de l'information

La source du message

Les nombreuses études qui se sont intéressées à la diffusion d'information sur Twitter ont mis en évidence le rôle clé de certaines caractéristiques de la source dans l'efficacité de son message, notamment le rôle du nombre de personnes qui suivent le compte ou l'expérience de l'utilisateur de la plateforme (mesurée en temps depuis la création du compte ou en nombre d'actions effectuées) (Berman *et al.*, 2019 ; Nesi *et al.*, 2018 ; Van

de Velde *et al.*, 2015). Si ces variables semblent fondamentales pour expliquer le nombre de retweets (un utilisateur expérimenté se sera approprié les normes de la plateforme et en aura une bonne maîtrise, lui permettant de poster du contenu en adéquation avec les attentes des autres utilisateurs ; de même, une forte audience permet d'avoir une grande visibilité sur la plateforme, assurant ainsi une plus grande probabilité d'être partagé), nous pouvons aussi nous intéresser à la stratégie des utilisateurs postant le message vis-à-vis de la plateforme et de sa communauté. En effet, certains utilisateurs font preuve d'une grande proactivité pour étendre leur communauté, et ainsi leur sphère d'influence, marquant à la fois des formes de réciprocité importante ou des niveaux d'activité forts. À l'inverse, certains utilisateurs ont un comportement beaucoup plus passif, que l'on peut qualifier comme de recherche de contenus. Les utilisateurs faisant preuve d'une stratégie proactive envers la plateforme et sa communauté sont ainsi plus à même d'avoir une large audience pour ses messages. En suivant les travaux de Munzel et Kunz (2014), nous pouvons distinguer deux stratégies proactives : une stratégie visant à publier de très nombreux posts afin d'être visible grâce à la multitude des actions effectuées et des messages diffusés (les *creators* de Munzel et Kunz, 2014) ; une stratégie cherchant à gagner des membres dans sa communauté par des actions de réciprocité (*like*, suivi de l'activité des autres membres / *following*, partage des messages émis par d'autres membres) afin d'avoir une forte implication de la part de son audience (les *multipliers* de Munzel et Kunz (2014)). Une stratégie réactive sera quant à elle caractérisée par un comportement de consommation de contenus, de recherche d'informations, sans interaction forte avec la plateforme ou d'autres membres (les *lurkers* de Munzel et Kunz, 2014). Nous pouvons donc nous attendre à ce que les membres ayant une stratégie proactive aient plus de succès pour voir leur message se diffuser sur la plateforme que les utilisateurs adoptant une stratégie réactive. Nous formulons donc l'hypothèse suivante :

H1 : les stratégies proactives (vs réactives) de l'émetteur vis-à-vis de la plateforme favorisent la diffusion du message.

La crédibilité de la source est un signal important pour la persuasion s'appuyant sur la route périphérique. La crédibilité de la source fait référence à deux dimensions essentielles : sa fiabilité, jugée à travers son expertise et la confiance qu'on peut lui accorder ; son attrait, déduit de sa familiarité (effort de dévoilement des informations personnelles), de son caractère sympathique (physique et comportementale) et de sa similarité (la ressemblance supposée entre la source et le récepteur du message)

(McCracken, 1989). Dans une revue de littérature sur l'application du modèle ELM aux avis en ligne, Soulard (2015) montre que les recherches se sont intéressées à deux points particuliers : la réputation de l'auteur de l'avis en ligne (que l'on peut relier à l'expertise de McCracken (1989)) et la divulgation des informations personnelles (la familiarité de McCracken, 1989). Ces deux dimensions se retrouvent sur Twitter, où un utilisateur peut connaître l'identité du compte émetteur d'un message via son nom, mais aussi avoir accès aux informations que ce dernier choisit de divulguer à travers sa photo et sa description de profil, et peut savoir si ce dernier est un utilisateur que Twitter certifie (à travers la procédure de vérification de compte). Les récepteurs d'un message sont donc à même de porter un jugement quant à la crédibilité de son auteur, à travers les éléments de réputation, d'audience de l'émetteur (la taille de sa communauté) et des informations personnelles dévoilées. En accord avec les résultats de Li *et al.* (2019), selon lesquels la crédibilité de la source joue un rôle positif dans le succès commercial de produits via les mécanismes de bouche-à-oreille électronique, nous nous attendons à voir un effet de la crédibilité de la source dans l'efficacité que rencontre un message sur une plateforme de réseau social :

H2 : plus l'émetteur est jugé crédible (vs non-crédible), plus son message est diffusé.

La forme du message

Les études portant sur la viralité des tweets se sont intéressées, en plus des variables d'audience et d'expérience de la source, aux caractéristiques de la forme du message. Au-delà des sujets de conversation (*topics*) traités dans les messages, certaines études se sont intéressées aux éléments de composition (inserts d'image, de lien, de mot-clé renvoyant à un sujet précis, d'interpellation d'autres utilisateurs ou encore d'émojis). Les éléments de composition favorisent les interactions (interpellation d'utilisateurs, url), la transmission d'émotion (emoji, image) et d'information (hashtag, image). Jenders *et al.* (2013) et Van de Velde *et al.* (2015) mettent en évidence le rôle de la présence d'url, de hashtag ou de mentions dans la probabilité d'être retweeté, tandis que Quesenberry et Coolson (2019) ont ajouté la présence d'émojis dans leur modèle. Ces études ont montré le rôle de ces éléments de composition supplémentaires par rapport à du simple texte dans la diffusion des messages sur Twitter. Aussi, nous formulons l'hypothèse suivante :

H3 : plus le message contient d'éléments de composition supplémentaires (vs uniquement du texte), plus il est diffusé.

À ces variables de composition s'ajoutent aussi des variables relevant plutôt de la forme stylistique du message, ce que nous appelons le degré de sophistication. Berger et Milkman (2012) ont mis en évidence le rôle des sentiments exprimés dans les contenus pour la diffusion d'informations en ligne, élément que l'on retrouve sur Twitter. La longueur des messages a également été prise en compte dans les modèles explicatifs de la viralité des tweets, comme c'est le cas dans l'étude de Lahuerta-Otero *et al.* (2018). Pour compléter ces variables traditionnellement mobilisées pour expliquer la diffusion des tweets, nous ajoutons une mesure de lisibilité qui permet de rendre compte de la facilité de compréhension d'un texte (DuBay, 2004), ainsi qu'une mesure de diversité lexicale qui rend compte de la richesse du vocabulaire utilisé. Ainsi, un tweet sera considéré comme sophistiqué lorsque sa rédaction est travaillée, ce qui passe par un vocabulaire riche, une compréhension aisée et l'expression de sentiment. Formellement, nous chercherons à tester l'hypothèse suivante :

H4 : plus le degré de sophistication du message est important (vs faible), plus il est diffusé.

L'engagement de l'émetteur dans le sujet de conversation et l'effet du contexte

Selon les situations, l'efficacité d'un message sur les réseaux sociaux ne sera pas la même. En particulier, plus l'émetteur est engagé dans un sujet de conversation, plus il possède d'expérience, de connaissance sur les codes associés au sujet de conversation et d'informations, et plus son message devrait être efficace, c'est-à-dire partagé et diffusé. Ainsi, un utilisateur engagé, c'est-à-dire qui multiplie les interactions sur le sujet de conversation, verra ses messages plus adaptés au sujet et à son audience, et les éléments de forme seront utilisés de manière adéquate, afin de favoriser la diffusion de ses messages. De plus, un utilisateur très engagé sera jugé d'autant plus crédible (degré d'expertise sur le sujet) ce qui devrait aussi favoriser la diffusion de ses messages (Luo *et al.*, 2013) et il aura d'autant plus d'impact sur son audience qu'il développe des stratégies proactives. Aussi, selon le niveau d'engagement de l'émetteur dans le sujet de conversation étudié, le rôle des différentes variables précédentes devrait varier, ce qui nous conduit à formuler l'hypothèse suivante :

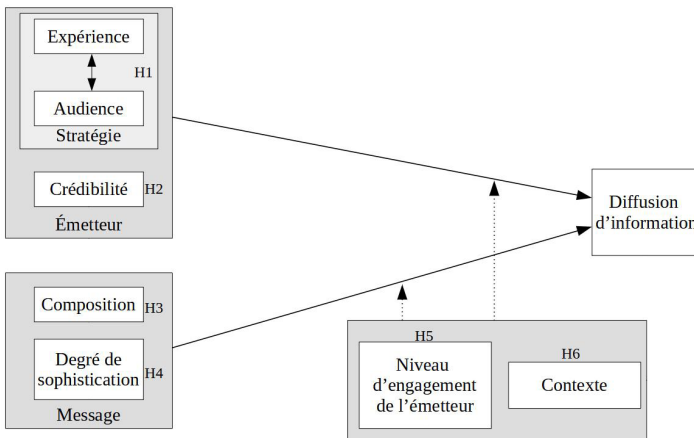
H5 : il existe un effet modérateur du niveau d'engagement de l'émetteur dans le sujet de conversation sur les variables de source et de forme dans la performance du message sur une plateforme de micro-blogging.

De plus, l'ELM met en évidence des effets modérateurs agissant sur la probabilité d'élaboration du récepteur de la communication, notamment le rôle de l'implication de ce dernier vis-à-vis du message ou encore le contexte dans lequel il reçoit ce message. Aussi, selon le contexte dans lequel le message est émis et reçu, les effets des variables liées à la source et à la forme du message auront une importance différente. En particulier, lors de l'apparition d'un nouveau sujet de conversation, l'effet de nouveauté peut permettre un niveau de partage d'informations plus important que lorsque ce sujet est déjà bien connu, où peut apparaître un effet de lassitude ou d'apprentissage. De même, un contexte marqué par une forte anxiété peut favoriser la diffusion d'information aux caractéristiques particulières au contraire d'un contexte marqué par la sérénité. Nous formulons donc l'hypothèse suivante :

H6 : il existe un effet modérateur du contexte sur les variables de source et de forme dans la diffusion d'un message sur une plateforme de micro-blogging.

Ces différents éléments sont formalisés dans le modèle proposé dans la figure 1, reprenant les différentes variables de la route périphérique retenues dans cette recherche et leur impact sur la persuasion et la diffusion de l'information.

Figure 1 – Modèle testé : variables de la route périphérique agissant sur la diffusion d'information et rôle modérateur du niveau d'engagement de l'émetteur et du contexte



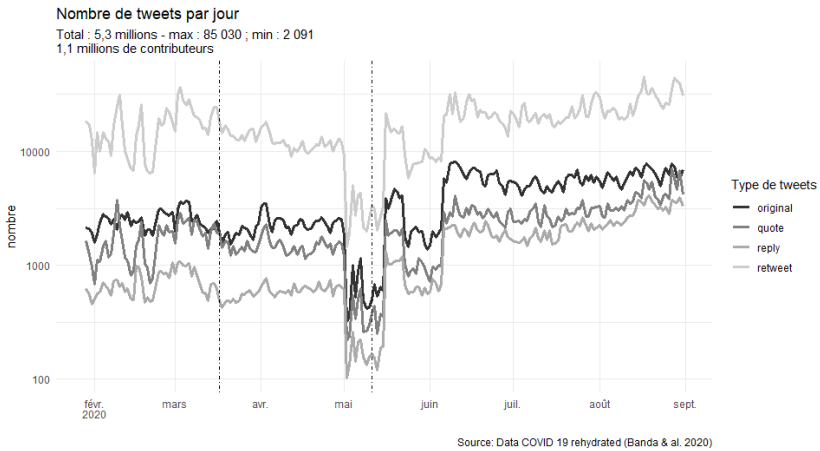
Données et méthodes

Jeu de données

La crise liée à l'épidémie de Covid-19 représente une situation d'incertitude majeure, dans laquelle le recours aux réseaux sociaux pour s'informer, se rassurer ou tout simplement partager son expérience est important. Nous étudions les comportements de partage de l'information sur une plateforme privilégiée par les internautes pour réagir à ces situations d'incertitude, Twitter (Hernandez-Suarez *et al.*, 2019). Une telle situation d'incertitude représente un contexte pertinent pour tester notre modèle, puisque les messages sont nombreux, les informations diverses (*fake news*, chiffres morbides, recettes de pain, etc.), et le climat émotionnel anxio-gène (Erku *et al.*, 2021 ; Cheng *et al.*, 2020 ; Li *et al.*, 2020). Nous nous concentrons particulièrement sur la période d'apparition de l'épidémie, l'expérience inédite de confinement et la phase de relâchement et de quasi-quiétude qui a suivi, caractérisée par la mise en place de nouvelles normes sociales que sont les gestes barrières.

Le jeu de données utilisé est extrait du corpus de tweets constitué par Banda *et al.* (2020). Les mots-clés utilisés pour collecter les tweets en temps réel sont différentes variations autour des hashtags Coronavirus et Covid. Dû aux conditions de service de Twitter, le corpus contient uniquement les identifiants de plus de 194 millions de tweets collectés entre janvier et décembre 2020. Nous avons réhydraté cet ensemble pour obtenir les contenus des tweets et de nombreuses variables disponibles via l'API de Twitter. Ces variables concernent les caractéristiques des tweets, celles liées à leur auteur et à leur performance sur la plateforme. Puis, nous avons extrait un sous-corpus en langue française. Le jeu de données utilisé comprend un total de 5,3 millions de tweets émis par 1,1 million de contributeurs, entre le 31 janvier et le 1^{er} septembre 2020. La répartition dans le temps et entre les différentes formes de post (tweet original, retweet, *reply* et *quote*) est présentée sur la figure 2. On constate l'importance des retweets dans ce sujet de conversation qui représentent environ 70 % des tweets émis sur la période. Seuls les contenus originaux ont été retenus pour l'analyse, ce qui représente un total de 797 374 tweets émis par 234 887 comptes. Les données ont été traitées en utilisant l'environnement R (R Core Team, 2014) et les bibliothèques du tidyverse (Taylor, 2017).

Figure 2 – Distribution des tweets dans le temps. Les lignes verticales représentent la période du confinement en France



Opérationnalisation des variables

La performance des tweets est mesurée par le logarithme du nombre de retweets obtenus pour chaque contenu original. L'expérience des utilisateurs sur la plateforme a été appréhendée par le nombre total d'actions réalisées par un utilisateur (*statuses*) et par le nombre de tweets qu'un utilisateur a mis dans ses favoris (*favorites*). L'audience d'un utilisateur est mesurée par le nombre de personnes suivant son compte (*followers*) et par le nombre de personnes suivies par l'utilisateur (*friends*). Une transformation en \log_{10} a été appliquée à ces différentes mesures. Pour rendre compte des stratégies proactives ou réactives des émetteurs vis-à-vis de la plateforme, nous ajoutons des effets d'interaction :

- Stratégie proactive de création de contenus : interaction entre le nombre de *followers* et le nombre de *statuses*.
- Stratégie proactive de création de communauté : interaction entre le nombre de *followers* et le nombre de *friends*, ainsi qu'entre le nombre de *followers* et de *favorites*.
- Stratégie réactive de consommation de contenu : interaction entre le nombre de *friends* et le nombre de *favorites*.

Les utilisateurs ont été catégorisés selon leur degré d'engagement sur le sujet étudié. Nous avons créé 5 profils d'utilisateurs, selon le nombre de posts émis sur la période et présents dans notre jeu de données (cf. tableau 1).

Tableau 1 – Description des niveaux d’engagement sur le sujet de conversation étudié

Profil	Description	N %	Tweets %
Ponctuel	1 tweet sur la période	56 %	12 %
Récidiviste	2 tweets à, en moyenne 1 tweet tous les mois	34 %	24 %
Fidèle	En moyenne, 1 tweet par mois à 1 tweet par semaine	8 %	23 %
Passionné	En moyenne, 1 tweet par semaine à 1 tweet tous les 2 jours	1,7 %	20 %
Superposteur	Plus d’1 tweet tous les 2 jours	0,3 %	21 %

Nous avons créé deux variables pour représenter le niveau de crédibilité d’un émetteur. Pour approcher la familiarité d’un utilisateur, nous nous sommes appuyés sur son comportement de dévoilement sur la plateforme, en élaborant un indicateur de transparence du profil. Nous avons appliqué un algorithme de détection des entités nommées (Rinker, 2017) sur le nom de l’utilisateur et sa description, pour identifier la présence de noms de personnes, d’organisations ou de lieux. À ces informations, nous avons ajouté la présence d’une photo de profil, le nombre de caractères utilisés dans la description du profil, ainsi que l’absence d’emojis, et si le compte a été vérifié par Twitter. Obtenir un compte vérifié par Twitter est un processus différent d’un compte officiel et suppose de faire la démarche auprès de la plateforme, et de donner à cette dernière des informations précises à même de garantir son identité et de justifier de son intérêt pour le public (il s’agit donc principalement de marques, d’organisations ou de personnalités publiques qui disposent de ce statut). Nous avons additionné ces éléments, +1 point pour la présence d’un nom propre, d’une organisation, ou d’un lieu, +1 point pour la photo de profil, +1 point pour une description de plus de 40 caractères, -1 pour la présence d’emoji, et +5 pour un compte vérifié. L’indicateur ainsi créé varie de -1 à +9. De plus, les comptes identifiés comme superposteurs ont fait l’objet d’une détection de *bot* grâce à la bibliothèque Tweetbotornot (Kearney, 2018), afin de distinguer les comptes automatisés des comptes opérés par un individu. Un compte opéré par un *bot* est considéré comme non crédible.

Les variables de composition des tweets correspondent aux éléments de forme insérés dans le contenu du tweet. Nous avons créé 4 variables binaires codées 1 si le tweet contient une mention à un autre compte, un hashtag, un lien url ou un média (photo, vidéo, gif). De plus, nous avons compté le nombre d’emojis présents dans les tweets.

Le degré de sophistication des tweets est mesuré par la création de 4 indicateurs. Pour la mesure de lisibilité, qui indique la facilité de compréhension d'un texte (plus la valeur est élevée, plus le niveau de formation nécessaire pour comprendre le texte est élevé, moins le texte est lisible), nous avons choisi l'indicateur ARI (*Automated Readability Index*) (DuBay, 2004). Nous avons choisi l'indicateur CTTR (*Caroll's Corrected Text-Type Ratio*) pour la mesure de diversité lexicale (Torruella, Capsada, 2013). Nous avons réalisé une analyse de sentiment grâce à l'annoteur du LIWC (Piolat *et al.*, 2011 ; Tausczik, Pennebaker, 2010) pour créer un indicateur de sentiment positif et un indicateur de sentiment négatif. Nous avons ajouté les effets d'interaction entre ces indicateurs de sentiments, pour rendre compte des formes d'hyper-expressivité.

Pour rendre compte de l'évolution du contexte et de l'effet d'apprentissage ou de lassitude lié à la pérennisation du sujet de conversation, nous avons réparti les tweets en trois périodes selon leur date de création : avant le confinement (jusqu'au 17 mars 2020), pendant le confinement (jusqu'au 11 mai 2020) et après le confinement (jusqu'à la fin de la période étudiée).

Le processus de collecte des données et d'opérationnalisation des variables est présenté dans la figure 3, et le tableau 2 synthétise les différentes variables mobilisées et leur opérationnalisation en fonction des hypothèses retenues.

Figure 3 - Processus de collecte des données et d'opérationnalisation des variables

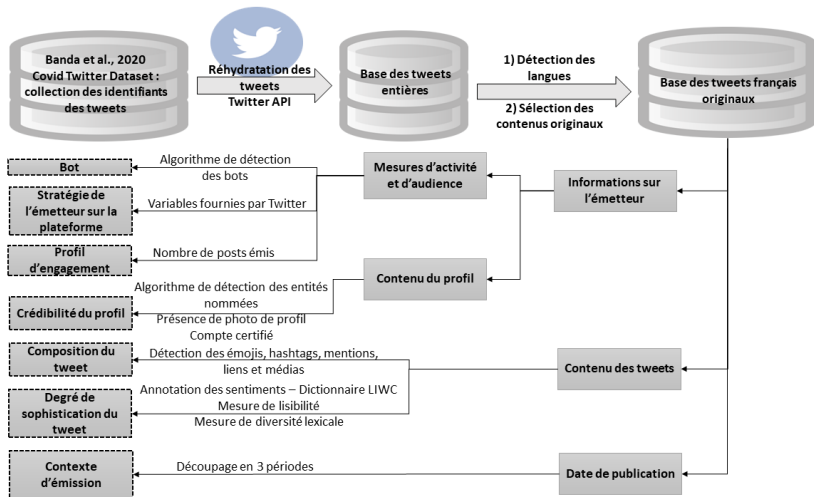


Tableau 2 – Synthèse des variables choisies et de leur opérationnalisation en fonction des hypothèses

Hypothèse	Variables	Opérationnalisation
H1 : Rôle des stratégies de l'émetteur vis-à-vis de la plateforme	Stratégie proactive de création de contenus : <i>followers:status</i> Stratégie proactive de création de communautés : <i>followers:friends</i> et <i>followers:favorites</i> Stratégie réactive de consommation de contenus : <i>favorites:friends</i>	Nombre de comptes suivis, de comptes suiveurs, d'actions réalisées et de mises en favoris données par Twitter, en log10
H2 : Rôle de la crédibilité de l'émetteur	Indicateur de transparence de profil : score agrégeant les informations concernant le profil de l'émetteur, soit la description du profil, le nom du compte, la présence d'une photo et si le compte est vérifié par Twitter ; <i>profil_transparency</i> Comptes opérés par un robot : <i>bot</i>	Algorithme de détection des entités nommées sur la description et le nom du profil de l'émetteur Comptage du nombre de caractères et du nombre d'émojis dans la description du profil Présence d'une photo de profil et comptes vérifiés par Twitter données directement par la plateforme Algorithme de détection des <i>bots</i> sur les comptes superposteurs
H3 : Rôle des éléments de composition supplémentaires	Nombre d'émojis utilisés dans le tweet : <i>emoji</i> Présence de mentions d'autres utilisateurs dans le tweet : <i>mention</i> Présence d'url dans le tweet : <i>url</i> Présence d'une image, gif ou vidéo dans le tweet : <i>media</i> Présence d'au moins un hashtag dans le tweet : <i>hashtag</i>	Comptage du nombre d'émojis dans chaque tweet Détection des mentions, url, medias et hashtags présents dans chaque tweet et création d'une variable binaire (présence versus absence)
H4 : Rôle du degré de sophistication du message	Lisibilité du tweet : <i>ARI</i> Diversité lexicale : <i>CTTR</i> Mesure de sentiments positif et négatif : <i>emopos</i> , <i>emoneg</i> , et leur interaction <i>emopos:emoneg</i>	Calcul de l'indicateur ARI Calcul de l'indicateur CTTR Annotation du texte en fonction des sentiments grâce au dictionnaire LIWC et création d'indicateurs en normalisant la présence des sentiments par la longueur des tweets

Hypothèse	Variables	Opérationnalisation
H5 : Effet modérateur du niveau d'engagement de l'émetteur	Profil d'engagement de l'émetteur : <i>ponctuels, récidivistes, fidèles, passionnés, superposteurs</i>	Répartition des émetteurs dans chaque catégorie en fonction du nombre de tweets émis sur la période
H6 : Effet modérateur du contexte d'émission et de réception	Contexte d'émission : <i>avant, pendant et post</i> confinement du printemps 2020	Répartition des tweets dans chaque catégorie en fonction de la date d'émission (avant le 17/03/2020 ; entre le 17/03/2020 et le 11/05/2020 ; après le 11/05/2020)

Méthodes

Pour tester la pertinence de notre modèle général pour expliquer le nombre de retweets, sans les effets modérateurs, nous utilisons des modèles linéaires emboîtés :

- le modèle de base comprend les variables d'expérience et d'audience des comptes émetteurs ;
- le modèle 2 ajoute les effets d'interaction entre ces variables pour rendre compte des stratégies des acteurs sur la plateforme ;
- le modèle 3 introduit les variables liées à la crédibilité de l'émetteur ;
- le modèle 4 intègre les variables de composition des tweets ;
- le modèle 5 prend en compte le degré de sophistication des tweets.

Nous comparons ensuite le modèle 5 entre les 5 profils créés en fonction de la production des utilisateurs sur la période pour mettre en évidence les effets différenciés des variables explicatives selon le degré d'engagement de l'utilisateur sur le sujet de conversation.

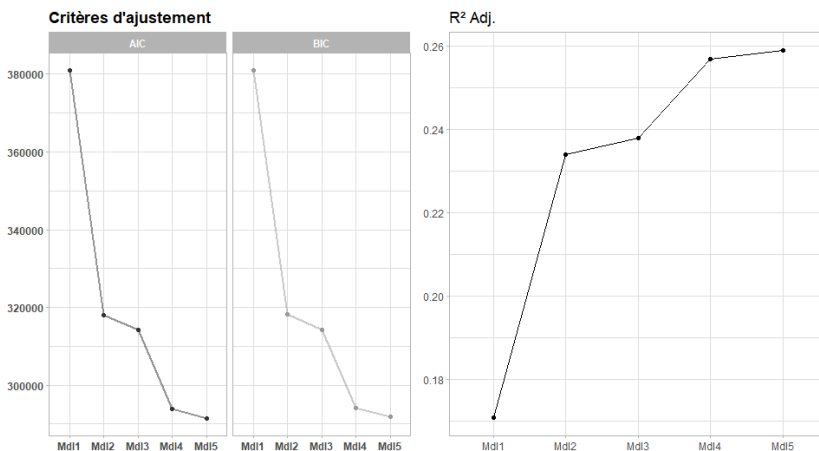
Enfin, nous comparons ces résultats en fonction des trois périodes de temps, pour tester l'effet de nouveauté ou de lassitude lié à l'évolution du contexte.

Résultats

Qualité d'ajustement des modèles

La figure 4 compare les principaux indicateurs de performance des modèles testés. Les résultats des tests ANOVA confirment l'existence de différences entre les modèles (modèle1 : modèle2, $F = 16\,364$, $p\text{-value} < 0,001$; modèle2 : modèle3, $F = 1\,973,9$, $p\text{-value} < 0,001$; modèle3 : modèle4, $F = 4\,099,2$, $p\text{-value} < 0,001$; modèle4 : modèle5, $F = 480,96$, $p\text{-value} < 0,001$).

Figure 4 - Indicateurs de qualité d'ajustement des modèles testés



Nous constatons un accroissement de la qualité d'ajustement à mesure que les modèles sont plus élaborés. Les interactions entre les variables d'expérience et d'audience améliorent la qualité de prédiction ($R^2_{\text{modèle 1}} = 0,171$; $R^2_{\text{modèle 2}} = 0,234$; gain relatif de R^2 de 37 %). Les variables de crédibilité du profil, cherchant à approcher la crédibilité accordée à la source montrent un effet positif sur les critères d'ajustement du modèle. L'ajout des variables de composition améliore sensiblement la qualité d'ajustement ($R^2_{\text{modèle 3}} = 0,238$; $R^2_{\text{modèle 4}} = 0,257$; gain relatif de R^2 de 8 %). Enfin, le degré de sophistication des tweets augmente encore un peu la qualité de la prédiction.

La figure 5 reproduit les coefficients standardisés des 5 modèles (le détail des résultats des modèles est présenté dans l'annexe 1). Les variables représentant les stratégies proactives des émetteurs sur la plateforme ont un impact positif sur le nombre de retweets ($\beta_{followers:friends} = 0,025 - p\text{-value} < 0,001$; $\beta_{followers:favorites} = 0,035 - p\text{-value} < 0,001$; $\beta_{followers:status} = 0,012 - p\text{-value} < 0,001$), tandis que celle représentant une stratégie réactive a un impact négatif ($\beta_{favorites:friends} = -0,032 - p\text{-value} < 0,001$). L'hypothèse H1 est donc acceptée.

L'impact de l'indicateur de transparence du profil a un effet positif sur le nombre de retweet ($\beta_{profil_transparency} = 0,011 - p\text{-value} < 0,001$), tandis que la variable *bot* a un effet négatif ($\beta_{bot} = -0,043 - p\text{-value} < 0,001$). L'hypothèse H2 est ainsi acceptée.

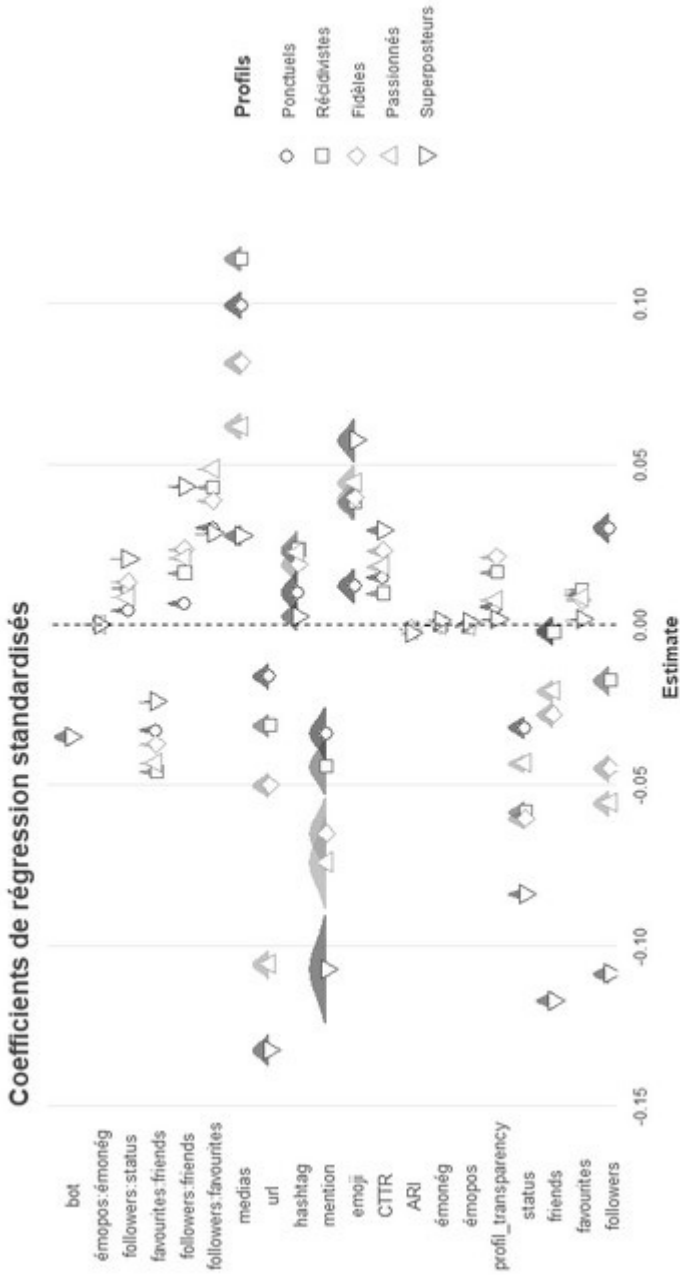
Le nombre de retweets est impacté positivement par la présence d'un média ($\beta_{medias} = 0,075 - p\text{-value} < 0,001$), par le nombre d'émojis ($\beta_{emoji} = 0,043 - p\text{-value} < 0,001$) et par la présence de hashtags ($\beta_{hashtag} = 0,019 - p\text{-value} < 0,001$). A contrario, la présence de mention et d'url ont un impact négatif sur le nombre de retweets ($\beta_{mention} = -0,066 - p\text{-value} < 0,001$; $\beta_{url} = -0,054 - p\text{-value} < 0,001$). L'hypothèse H3 n'est donc que partiellement acceptée.

Concernant les variables représentant le degré de sophistication des tweets, nous constatons un impact positif de l'indicateur de diversité lexicale ($\beta_{CTTR} = 0,019 - p\text{-value} < 0,001$), et un impact légèrement négatif de l'indicateur de lisibilité ($\beta_{ARI} = -0,002 - p\text{-value} < 0,001$). Par contre, les coefficients des variables de sentiment ne sont pas significatifs. L'hypothèse H4 n'est donc que très partiellement acceptée.

Comparaison entre les niveaux d'engagement

Le pouvoir explicatif du modèle 5 varie en fonction des niveaux d'engagement. Plus l'utilisateur a contribué au sujet de conversation sur la période étudiée, plus le R^2 est important ($R^2_{ponctuel} = 0,17$; $R^2_{récidiviste} = 0,21$; $R^2_{fidèle} = 0,23$; $R^2_{passionné} = 0,25$; $R^2_{superposteur} = 0,36$). La figure 6 représente les coefficients standardisés et leurs intervalles de confiance en fonction des différents niveaux d'engagement et l'annexe 2 présente les résultats détaillés des modèles.

Figure 6 - Coefficients de régression standardisés en fonction des niveaux d'engagement



On constate clairement un effet différencié des variables selon les niveaux d'engagement des émetteurs dans le sujet de conversation étudié. Les utilisateurs mettant en place des stratégies proactives sur Twitter (interactions entre *followers* et *status*, *followers* et *friends* et *followers* et *favorites*) ont un niveau de retweet plus élevé à mesure qu'ils sont impliqués dans le sujet de conversation. Au contraire, les acteurs aux stratégies réactives (interactions entre *favorites* et *friends*) ont moins de probabilité d'être retweeté, cet effet s'atténuant plus l'émetteur est engagé dans le sujet de conversation.

Le dévoilement de l'utilisateur concernant son identité a un effet positif sur le nombre de retweet, plus important pour les niveaux moyens d'engagement (profils récidivistes et fidèles) que pour un niveau faible ou important. Les *bots* rencontrent peu de succès dans le partage d'information (coefficient = -0,035, *p-value* < 0,001).

Les mentions ou les urls ont un impact d'autant plus négatif sur le nombre de retweets que le niveau d'engagement de l'émetteur est important. À l'inverse, la présence d'émojis favorise les retweets, impact croissant avec le niveau d'engagement. La présence de médias ou de hashtags favorisent les retweets, mais de manière différenciée selon l'implication dans le sujet de conversation : les hashtags ont un effet moins important pour les profils très actifs ou peu actifs (superposteurs et ponctuels), tandis que la présence d'un média a plus d'impact pour les profils peu actifs.

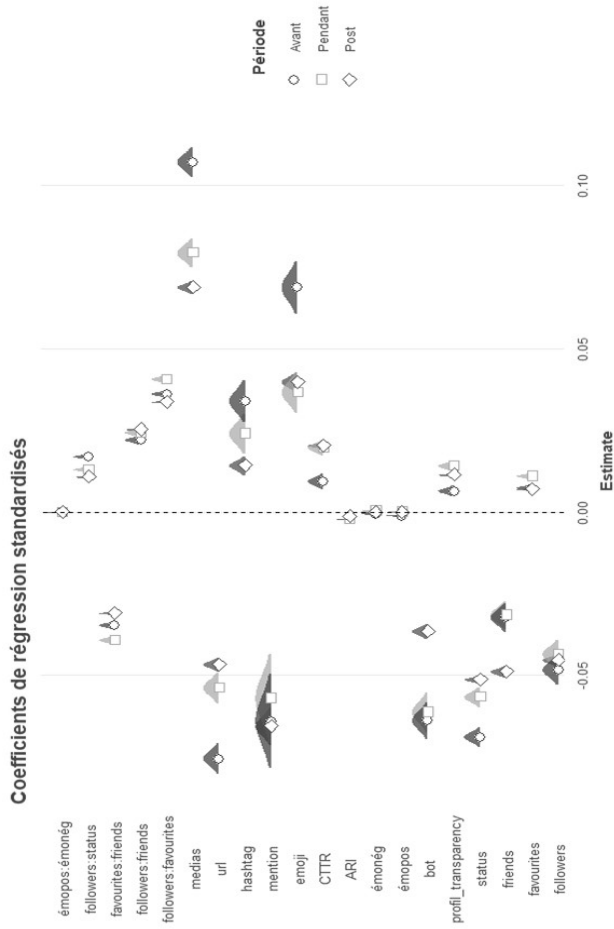
Les variables mesurant le degré de sophistication du tweet ont peu d'impact sur le nombre de retweets, notamment concernant les sentiments exprimés. L'indicateur de lisibilité (ARI) a un impact légèrement négatif pour tous les profils. A contrario, la diversité lexicale (CTTR) a un impact positif croissant avec le niveau d'engagement. Les variables de sentiment sont non-significatives sauf pour les superposteurs, mais les coefficients sont très proches de zéro.

L'examen des paramètres selon les niveaux d'engagement des émetteurs nous permet d'accepter H5.

Étude longitudinale

Le contexte d'émission et de réception des messages a un effet différencié sur le pouvoir explicatif de modèle 5 : le R^2 ajusté tend à baisser avec le temps ($R^2_{\text{Avant}} = 0.278$; $R^2_{\text{Pendant}} = 0.279$; $R^2_{\text{Post}} = 0.254$). La figure 7 représente les coefficients standardisés et leurs intervalles de confiance en fonction du contexte d'émission, avant le confinement, pendant ou après (le détail des résultats est disponible dans l'annexe 3).

Figure 7 - Coefficients de régression standardisés en fonction du contexte d'émission et de réception du message



Les coefficients des stratégies des émetteurs ont des impacts différenciés sur le nombre de retweet en fonction du contexte : la stratégie réactive de consommation de contenu voit son rôle s'affaiblir (malgré un pic négatif dans la période du confinement), comme la stratégie proactive de production de contenu, contrairement à la stratégie proactive de création de communauté (effet positif de l'interaction *followers:friends* ; effet positif pendant la période de confinement de l'interaction *followers:favourites*).

Concernant la crédibilité de l'émetteur, la variable de transparence du profil, qui représente la familiarité, voit son impact sur le nombre de retweet augmenter pendant le confinement avant de baisser dans la période suivante ; la variable *bot*, quant à elle, voit son rôle s'atténuer avec le temps.

Les variables de composition du message voient leur impact sur le nombre de retweet s'atténuer avec le temps, excepté le rôle de la présence de mention qui augmente la probabilité de ne pas être retweeté dans la période post-confinement comparé à la période pré-confinement.

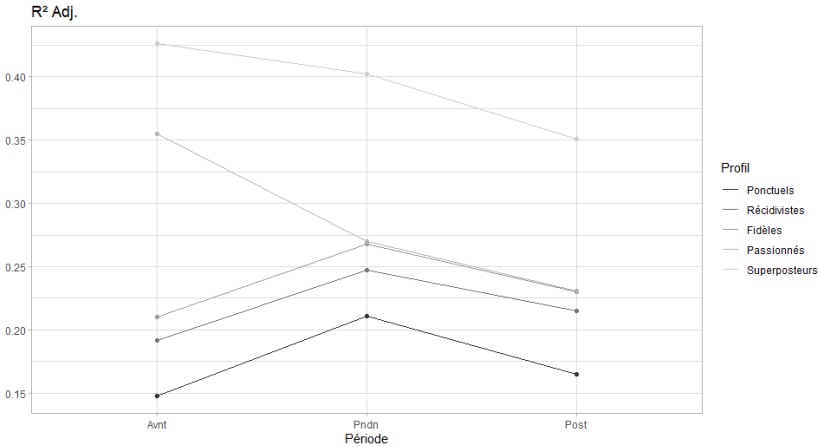
Les variables de sophistication des tweets ont peu d'impact selon les différents contextes (variables de sentiment non significatives, variable de lisibilité très proche de zéro), exceptée pour la variable de diversité lexicale qui voit son coefficient augmenter avec le temps.

Ainsi, nous pouvons accepter notre hypothèse 6, selon laquelle les effets des variables sont différents selon le contexte dans lequel le message est émis et reçu.

Comparaison de l'effet du contexte en fonction des profils d'engagement

Nous avons examiné les résultats du modèle 5 en fonction du contexte et des profils d'engagement des émetteurs, pour approfondir leurs rôles modérateurs. La figure 8 montre les différences de R^2 ajustés selon les périodes étudiées et les niveaux d'engagement. On constate deux tendances : les modèles concernant les utilisateurs très impliqués, super-posteurs et passionnés, perdent en pouvoir explicatif avec le temps, tandis que les modèles concernant les profils peu prolifiques (ponctuels, récidivistes et passionnés) ont un pouvoir explicatif bien plus important durant le confinement. Les coefficients standardisés des modèles en fonction du contexte et du niveau d'engagement sont présentés dans la figure 9.

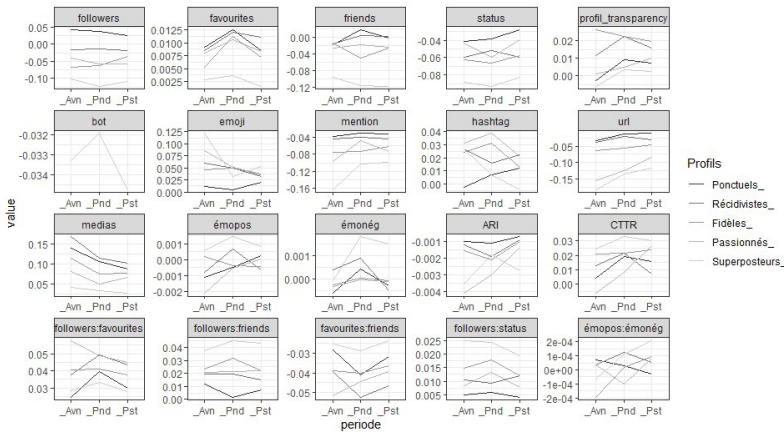
Figure 8 - Évolution du R² ajusté selon le contexte, en fonction des niveaux d'engagement



L'effet de contexte joue différemment selon les variables et les profils d'engagement. Les stratégies proactives ont globalement un impact plus fort pendant la période du confinement que durant les deux autres périodes (sauf pour l'interaction *followers:favorites* pour les passionnées ; *followers:friends* pour les ponctuels ; et *followers:status* pour les récidivistes), tandis que les stratégies réactives ont un impact plus négatif durant cette même période (sauf pour les passionnés). Les variables liées à la transparence de la source ont un impact de plus en plus positif sauf pour les profils moyennement engagés (récidivistes et fidèles). Le fait d'être identifié comme *bot* impacte moins négativement le nombre de retweet durant la période très particulière du confinement.

Concernant la forme du message, avec le temps on voit s'amoinrir l'effet positif des émojis, médias et hashtags, tandis que l'effet négatif des urls et des mentions est moins important. Au niveau des variables liées au degré de sophistication, on voit s'accroître le rôle des sentiments, aussi bien négatif que positif, véhiculés par les superposteurs. L'interaction entre ces deux variables reste très proche de zéro pour tous les niveaux d'engagement.

Figure 9 - Coefficients de régression standardisés en fonction des niveaux d'engagement et du contexte



Ces différents résultats nous permettent donc valider le cadre général du modèle proposé.

Discussion

Les tweets étudiés représentent des informations partagées sur une plateforme de micro-blogging. Nous nous sommes focalisées sur les variables traditionnellement associées à la route périphérique du modèle ELM qui favorisent la diffusion de ces informations dans un contexte C2C. Une première contribution empirique de ce travail est de s'intéresser aux comportements de partage de l'information réellement observés sur une plateforme et non à des intentions de partage mesurées en conditions expérimentales. Nous montrons l'importance des signaux extérieurs aux caractéristiques intrinsèques des messages pour favoriser leur diffusion. Une deuxième contribution empirique concerne l'échelle à laquelle nous travaillons, un corpus de tweets originaux de 797 374 tweets émis par 234 887 comptes sur 7 mois. Notre recherche fondée sur un large corpus de tweets concernant le sujet de la pandémie actuelle nous a permis de valider notre modèle général, les signaux en provenance de l'émetteur et ceux contenus dans la forme des tweets ayant un impact sur le taux de retweets, impacts différenciés en fonction du niveau d'engagement de l'émetteur dans le sujet de conversation et en fonction du contexte.

Nous avons en particulier montré le rôle important des stratégies des acteurs sur la plateforme (interaction entre les variables d'expérience et

d'audience) dans la diffusion de l'information. Les trois stratégies utilisées dans notre travail reprennent les stratégies mises en évidence par Munzel et Kunz (2014) et permettent d'approfondir leurs travaux en mettant en évidence que la stratégie proactive de création de communautés semble être la plus efficace pour la diffusion de ses messages. Bien maîtriser les règles et les normes de la plateforme est donc un pré-requis pour que les messages soient transmis de pairs à pairs, mais il faut surtout se constituer une large audience.

L'effort de dévoilement de l'émetteur sur le réseau social est aussi une stratégie payante pour voir ses messages diffusés largement, même si c'est dans une moindre mesure que l'effet d'une stratégie proactive. La crédibilité de la source, approchée par l'effort de dévoilement et par le fait d'opérer manuellement le compte, est donc un élément important dans la diffusion de l'information sur une plateforme de micro-blogging, ce qui confirme les résultats mis en évidence par Li *et al.* (2019) sur le rôle de la crédibilité de la source dans le succès commercial lié à la recommandation de produits en ligne et élargit le champ d'application au partage de l'information dans un contexte C2C et non-commercial.

Au niveau de la forme, notre étude empirique montre le rôle important de la présence de médias, d'émojis et de hashtags dans la diffusion de l'information, tandis que le fait de renvoyer à une source extérieure à la plateforme (par la présence d'url) ou de s'adresser personnellement à d'autres utilisateurs (présence de mentions) n'est pas favorable à un partage des contenus, ce qui confirme les différents résultats déjà mis en évidence dans d'autres contextes. La diffusion de l'information au sein de la plateforme est donc favorisée par la présence d'éléments internes à la plateforme et par des communications s'adressant à tous.

Par contre, les mesures employées pour représenter le degré de sophistication d'un tweet ne sont que peu concluantes : nous pouvons conclure que les messages comprenant de nombreux mots différents les uns des autres et facilement compréhensibles sont favorisés dans les comportements de diffusion de l'information. Mais nos mesures des sentiments ne montrent pas d'impact sur le partage de l'information, résultat contradictoire avec de nombreuses recherches (par exemple, Berger, Milkman, 2012). Cela peut s'expliquer par la nature des messages autorisés par la plateforme, de longueur maximale de 280 caractères, ce qui ne laisse pas beaucoup de place pour transmettre des émotions nuancées. De plus, nos outils de mesure ne sont peut-être pas les plus pertinents pour décrypter les sentiments exprimés dans les tweets. Notre approche par dictionnaire

peut être dépassée par la mise en œuvre d'algorithme de *machine learning* après annotations manuelles des sentiments des tweets, qui permettrait aussi de capturer les sentiments exprimés via les émojis.

Les variables qui contribuent le plus positivement au nombre de retweets, sont dans l'ordre la présence d'un média, la présence d'émojis et la stratégie proactive de création de communauté. A contrario, les variables qui impactent le plus négativement les retweets sont la présence de mention et d'url et le fait d'être un *bot*. Nous pouvons en déduire que la composition des tweets est le premier élément à privilégier pour favoriser la diffusion des messages. De plus, l'effet négatif du fait que le compte émetteur soit un *bot* sur la diffusion des messages est un élément intéressant concernant la diffusion des *fake news*. Même si ce sujet dépasse le cadre de cette recherche, ce résultat laisse à penser que les messages automatisés sont moins persuasifs que ceux produits par des êtres humains et entraînent moins de partage d'information. Évidemment, cet effet peut être contrebalancé par la large diffusion de ces messages automatisés et le fort taux d'impression qu'ils peuvent avoir, leur permettant de toucher une très large audience.

Nous avons constaté également un plus grand pouvoir de persuasion des éléments périphériques lorsque les émetteurs sont fortement engagés dans le sujet de conversation concerné, c'est-à-dire lorsqu'ils communiquent fréquemment sur ce thème. Ceci peut s'expliquer par une meilleure maîtrise des codes liés au sujet de conversation et donc une meilleure crédibilité de la source, comme montré par Luo *et al.* (2013). A contrario, on constate des effets d'aubaine pour les membres peu engagés, pour qui le bon média (photo, gif, vidéo) au bon moment a un effet très nettement plus important que pour les membres plus engagés.

Au niveau du contexte dans lequel les messages sont émis et reçus, il apparaît un effet d'atténuation avec le temps, les modèles étant moins performants pour la période post-confinement, une fois que l'épidémie est entrée depuis plusieurs mois dans la vie quotidienne. La lassitude face à cette crise semble s'être installée sur Twitter, comme dans la population. La plupart des conversations Twitter ont un temps court (quelques jours) durant lequel l'engouement est à son maximum, avant de perdre en audience et diffusion. Le sujet du Covid-19 ne semble pas faire exception. La période pré-confinement est celle de la nouveauté du sujet de conversation et d'une inquiétude croissante face à cette nouvelle maladie. On constate que les membres les plus engagés sont ceux pour qui les éléments périphériques ont un rôle plus important dans la diffusion de leurs

messages, ce qui peut s'expliquer par la nouveauté du sujet et leur maîtrise des codes de la plateforme. Au contraire, c'est la période du confinement qui a le meilleur pouvoir explicatif pour les membres moins engagés dans le sujet de conversation, période caractérisée par une situation inédite pour les Français forcés de rester chez eux et où l'activité économique s'était quasiment arrêtée, permettant à ces membres sinon peu actifs sur la plateforme de passer plus de temps à l'élaboration de leurs messages.

Conclusion

Les travaux présentés ici s'inscrivent dans le cadre conceptuel de l'ELM et élargissent ce cadre à l'étude de la persuasion des messages publiés sur les réseaux sociaux et participant à la création de bouche-à-oreille électronique. Cette recherche participe également au développement des travaux sur les comportements de mimétisme dans un environnement digital, en particulier celui du partage de l'information. Des travaux récents se sont intéressés aux formes prises par le *herding behaviour* sur Twitter (Langley *et al.*, 2014) ou concernant les notes émises par les membres d'une plateforme, en s'intéressant particulièrement à l'influence du groupe social d'appartenance (Sunder *et al.*, 2019). Ici, nous avons mis en évidence des caractéristiques clés favorisant la reprise d'information sur Twitter. En particulier, nous montrons que les éléments visuels de forme (notamment les émojis, les médias et les hashtags) et les éléments liés à la crédibilité de la source (expertise de la plateforme à travers les stratégies proactives et familiarité à travers l'effort de dévoilement) jouent un rôle important dans la persuasion et la diffusion de l'information. Nous montrons également que ces éléments, traditionnellement associés à la route périphérique du modèle ELM, ont un poids différent selon le niveau d'engagement de la source et selon le contexte d'émission-réception du message. Enfin nous validons un modèle conceptuel intégrant ces différents éléments et leur impact sur la diffusion de l'information.

D'un point de vue méthodologique, le principal apport de ce travail repose sur l'approche mixte que nous avons employée pour traiter les données. Nous avons utilisé différents outils de *text mining* pour extraire des indicateurs de sentiments exprimés, repérer les émojis, détecter les comptes automatisés et créer un indicateur de transparence des profils grâce notamment aux algorithmes de détection d'entités nommées. Le deuxième apport majeur de ce travail d'un point de vue méthodologique

est d'avoir conduit cette recherche sur des données de comportementales réelles et massives.

Les apports managériaux de cette recherche s'adressent principalement aux *community managers* qui utilisent Twitter, à qui nous pouvons conseiller de bien choisir les sujets de conversation sur lesquels ils interviennent pour diffuser plus largement leur message. Les messages seront d'autant plus persuasifs qu'ils apparaissent dans un contexte où leurs émetteurs connaissent et maîtrisent les codes de la conversation et les contextes dans lesquels ils seront reçus (recherche d'information, moments de détente, engagement sociétal, etc.). Dans ces différents contextes, les images ou gif pertinents, ainsi que l'intention associée à des émojis peuvent varier. Nous pouvons aussi leur conseiller de mettre en œuvre une stratégie proactive de création de communauté sur la plateforme, favorisant les messages fréquents et la création de relations bi-directionnelles avec leur audience, tout en affichant clairement leur identité. Pour la forme des messages émis, la présence d'un média et l'utilisation des émojis et des hashtags (à bon escient) sont des éléments pertinents pour être remarqué et diffusé. Enfin, nous pouvons conseiller à Twitter d'intégrer un aperçu des liens présents dans les messages pour favoriser leur visibilité et le partage, ce qui s'apparenterait à la présence d'un média, et ce qui est fait sur de nombreuses autres plateformes de réseaux sociaux.

Ce travail comporte également des limites, la première étant liée au fait de ne pas avoir trouvé d'impact du sentiment exprimé dans la diffusion de l'information. L'analyse du sentiment s'est ici appuyée sur une approche par dictionnaire, en utilisant celui élaboré par le LIWC. De futures recherches pourraient recourir à d'autres dictionnaires de sentiment même si ceux disponibles en français sont encore peu nombreux. Il serait aussi possible d'ajouter des indicateurs liés aux émotions exprimées, comme l'anxiété ou la joie, pour réussir à mieux capturer l'effet du contenu émotionnel. Une autre manière de traiter les sentiments seraient de s'intéresser aux émojis et aux sentiments qu'ils expriment en s'appuyant sur les travaux de Novak *et al.* (2015). Enfin, nous encourageons de futurs travaux à développer des algorithmes de *machine learning* pour approcher au plus près de la réalité le rôle des sentiments et émotions exprimés. Une telle approche nécessite des ressources importantes et permettrait de prendre en considération à la fois les indices de sentiment textuels et visuels (émojis, gifs, en particulier). Il pourrait aussi être judicieux d'utiliser d'autres types de modèles que le modèle linéaire pour mieux capturer l'impact des différentes variables, en particulier celles concernant les sentiments.

De plus, nous nous sommes intéressés à un sujet de conversation particulier, favorisant les communications C2C. De futures recherches pourraient reproduire cette étude à un sujet de conversation impliquant des marques, pour voir si les éléments périphériques étudiés ici ont un rôle similaire. Enfin, nous nous sommes focalisés spécifiquement sur les éléments périphériques des tweets dans le partage de l'information. Dans l'avenir, il serait pertinent de voir leur impact lorsque l'on tient compte également du sujet spécifique que traitent ces messages, en réalisant une analyse de *topics* et de la qualité des arguments exprimés.

RÉFÉRENCES

- ARAUJO, T., NEIJENS, P., Vliegenthart, R. (2017), Getting the Word Out on Twitter: The Role of Influentials, Information Brokers and Strong Ties in Building Word-of-mouth for Brands, *International Journal of Advertising*, 36(3), 496-513.
- ASSELIN, C. (2021), Twitter : les chiffres essentiels France et Monde pour 2021. Available at Digimind : <https://blog.digimind.com/fr/tendances/twitter-chiffres-essentiels-france-monde-2020#sources>
- BANDA, J. M., TEKUMALLA, R., WANG, G., YU, J., LIU, T., DING, Y., CHOWELL, G. (2020), A Large-scale COVID-19 Twitter Chatter Dataset for Open Scientific Research, An International Collaboration, *arXiv:2004.03688 [cs]*.
- BERGER, J., MILKMAN, K. L. (2012), What Makes Online Content Viral?, *Journal of Marketing Research*, 49(2), 192-205.
- BERMAN, R., MELUMAD, S., HUMPHREY, C., MEYER, R. (2019), A Tale of Two Twitterspheres: Political Microblogging During and After the 2016 Primary and Presidential Debates, *Journal of Marketing Research*, 56(6), 895-917.
- BÉRAUD, P., CORMERAIS, F. (2011), Économie de la contribution et innovation sociétale, *Innovations*, 34(1), 163-183.
- BOYD, D., GOLDER S., LOTAN G. (2010), Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, *2010 43rd Hawaii International Conference on System Sciences*, 1-10.
- CHENG, C., EBRAHIMI, O. V., LAU, Y. C. (2021), Maladaptive Coping with the Infodemic and Sleep Disturbance in the COVID-19 Pandemic, *Journal of Sleep Research*, 30(4), e13235.
- DE BARNIER, V. (2006), Le modèle ELM : bilan et perspectives, *Recherche et Applications en Marketing*, 21, 61-83.
- DUBAY, W. (2004), *The Principles of Readability*, ERIC Clearinghouse.
- ERKU, D. A., BELACHEW, S. A., ABRHA, S., SINNOLLAREDDY, M., THOMAS, J., STEADMAN, K. J., TESFAYE, W. H. (2021), When Fear and Misinformation Go Viral: Pharmacists' Role in Deterring Medication Misinformation during

- the “Infodemic” Surrounding COVID-19, *Research in Social and Administrative Pharmacy*, 17(1), 1954-1963.
- HERNANDEZ-SUAREZ, A., SANCHEZ-PEREZ, G., TOSCANO-MEDINA, K., PEREZ-MEANA, H., PORTILLO-PORTILLO, J., SANCHEZ, V., GARCÍA VILLALBA, L. J. (2019), Using Twitter Data to Monitor Natural Disaster Social Dynamics: A Recurrent Neural Network Approach with Word Embeddings and Kernel Density Estimation, *Sensors*, 19(7).
- JENDERS, M., KASNECI, G., NAUMANN, F. (2013), Analyzing and Predicting Viral Tweets, *Proceedings of the 22nd International Conference on World Wide Web*, 657-664.
- KEARNEY, M. W. (2018), *R Package: Tweetbotornot*.
- LAHUERTA-OTERO, E., CORDERO-GUTIÉRREZ, R., DE LA PRIETA-PINTADO, F. (2018), Retweet or Like? That is the Question, *Online Information Review*, 42(5), 562-578.
- LANGLEY, D. J., HOEVE, M. C., ORTT, J. R., PALS, N., VAN DER VECHT, B. (2014), Patterns of Herding and their Occurrence in an Online Setting, *Journal of Interactive Marketing*, 28(1), 16-25.
- LI, Q., LIU, Y. (2017), Exploring the Diversity of Retweeting Behavior Patterns in Chinese Microblogging Platform, *Information Processing and Management*, 53(4), 945-962.
- LI, S., WANG, Y., XUE, J., ZHAO, N., ZHU, T. (2020), The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users, *International Journal of Environmental Research and Public Health*, 17(6), 2032.
- LI, X., WU, C., MAI, F. (2019), The Effect of Online Reviews on Product Sales: A Joint Sentiment-Topic Analysis, *Information & Management*, 56, 172-184.
- LUO, C., LUO, X., SCHATZBERG, L., SIA, C. L. (2013), Impact of Informational Factors on Online Recommendation Credibility: The Moderating Role of Source Credibility, *Decision Support Systems*, 56, 92-102.
- MCCRACKEN, G. (1989), Who is the Celebrity Endorser? Cultural Foundations of the Endorsement Process, *Journal of Consumer Research*, 16(3), 310.
- MUNZEL, A., KUNZ, W. H. (2014), Creators, Multipliers, and Lurkers: Who Contributes and Who Benefits at Online Review Sites, *Journal of Service Management*, 25(1), 49-74.
- NESI, P., PANTALEO, G., POALI, I., ZAZA, I. (2018), Assessing the ReTweet Proneness of Tweets: Predictive Models for Retweeting, *Multimedia Tools and Applications*, 77, 26371-26396.
- NOVAK, P. K., SMAILOVIĆ, J., SLUBAN, B., MOZETIČ, I. (2015), Sentiment of Emojis, *PLoS ONE*, 10(12), 1-22.
- PETTY, R. E., CACIOPPO, J. T. (1986), The Elaboration Likelihood Model of Persuasion, *Advances in Experimental Social Psychology*, 19(C), 123-205.

- PETTY, R. E., CACIOPPO, J. T., SCHUMANN, D. (1983), Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement, *Journal of Consumer Research*, 10(2), 135.
- PIOLAT, A., BOOTH, R. J., CHUNG, C. K., DAVIDS, M., PENNEBAKER, J. W. (2011), La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation, *Psychologie Française*, 56(3), 145-159.
- PRESSGROVE, G., MCKEEVER, B. W. (2018), What is Contagious? Exploring why Content goes Viral on Twitter: A Case Study of the ALS Ice Bucket Challenge, *International Journal of Nonprofit and Voluntary Sector Marketing*, 23, 1-8.
- QUESENBERRY, K., COOLSEN, M. (2019), Twitter Posts that are Engaging: A Content Analysis of Twitter Brand Post Text That Increases Retweets, Replies and Favorites in Twitter Brand Posts To Influence Organic Viral Reach, *American Academy of Advertising Conference Proceedings*, (1), 120.
- R CORE TEAM (2014), *R: A Language and Environment for Statistical Computing*, Vienna, Austria, R Foundation for Statistical Computing.
- RINKER, T. (2017), *R Package: Entity*.
- RUI, H., LIU, Y., WHINSTON, A. B. (2010), Chatter Matters: How Twitter Can Open the Black Box of Online Word-of-Mouth, *Thirty First International Conference on Information Systems*, Saint-Louis.
- SOHN, K., GARDNER, J. T., WEAVER, J. L. (2013), Viral Marketing: More than a Buzzword, *Journal of Applied Business and Economics*, 14(1), 21-42.
- SOULARD, O. (2015), La crédibilité des avis en ligne : une revue de littérature et un modèle intégrateur, *Management & Avenir*, 82(8), 129.
- SUNDER, S., KIM, K. H., YORKSTON, E. A. (2019), What Drives Herding Behavior in Online Ratings? The Role of Rater Experience, Product Portfolio, and Diverging Opinions, *Journal of Marketing*, 83(6), 93-112.
- TAUSCZIK, Y. R., PENNEBAKER, J. W. (2010), The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods, *Journal of Language and Social Psychology*, 29(1), 24-54.
- TAYLOR, A. (2017), A Tidy Data Model for Natural Language Processing Using CleanNLP, *The R Journal*, 9(2), 1-20.
- TENG, S., KHONG, K. W., GOH, W. W. (2014), Conceptualizing Persuasive Messages Using ELM in Social Media, *Journal of Internet Commerce*, 13(1), 65-87.
- TORRUELLA, J., CAPSADA, R. (2013), Lexical Statistics and Typological Structures: A Measure of Lexical Richness, *Procedia - Social and Behavioral Sciences*, 95, 447-454.
- VAN DE VELDE, B., MEIJER, A., HOMBURG, V. (2015) Police Message Diffusion on Twitter: Analysing the Reach of Social Media Communications, *Behaviour and Information Technology*, 34(1), 4-16.
- YOO, E., GU, B., RABINOVICH, E. (2019), Diffusion on Social Media Platforms: A Point Process Model for Interaction among Similar Content, *Journal of Management Information Systems*, 36(4), 1105-1141.

Annexes

Annexe 1 – Résultats des modèles emboîtés

	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
(Intercept)	-0.0529*** (0.0012)	0.2389*** (0.0022)	0.2280*** (0.0022)	0.2488*** (0.0023)	0.2129*** (0.0026)
followers	0.1262*** (0.0003)	-0.0509*** (0.0008)	-0.0454*** (0.0008)	-0.0448*** (0.0008)	-0.0449*** (0.0008)
friends	-0.0310*** (0.0004)	-0.0371*** (0.0008)	-0.0426*** (0.0008)	-0.0444*** (0.0008)	-0.0445*** (0.0008)
favourites	0.0240*** (0.0002)	0.0096*** (0.0002)	0.0087*** (0.0002)	0.0074*** (0.0002)	0.0074*** (0.0002)
status	-0.0337*** (0.0003)	-0.0672*** (0.0005)	-0.0606*** (0.0006)	-0.0542*** (0.0006)	-0.0542*** (0.0006)
followers:favourites		0.0375*** (0.0002)	0.0364*** (0.0002)	0.0349*** (0.0002)	0.0349*** (0.0002)
followers:friends		0.0227*** (0.0003)	0.0231*** (0.0003)	0.0247*** (0.0003)	0.0246*** (0.0003)
favourites:friends		-0.0337*** (0.0002)	-0.0335*** (0.0002)	-0.0322*** (0.0002)	-0.0323*** (0.0002)
followers:status		0.0143*** (0.0002)	0.0133*** (0.0002)	0.0119*** (0.0002)	0.0119*** (0.0002)
profil_transparency			0.0104*** (0.0003)	0.0115*** (0.0003)	0.0113*** (0.0003)
bot			-0.0559*** (0.0010)	-0.0433*** (0.0010)	-0.0430*** (0.0010)
emoji				0.0461*** (0.0013)	0.0426*** (0.0013)
mention				-0.0629*** (0.0025)	-0.0656*** (0.0025)
hashtag				0.0165*** (0.0012)	0.0193*** (0.0012)
url				-0.0705*** (0.0008)	-0.0536*** (0.0009)
medias				0.0659*** (0.0008)	0.0750*** (0.0009)
émopos					-0.0002 (0.0001)
émonég					0.0000 (0.0001)
ARI					-0.0015*** (0.0000)
CTTR					0.0186*** (0.0004)
émopos:émonég					0.0000 (0.0000)
R ²	0.1709	0.2338	0.2376	0.2567	0.2589
Adj. R ²	0.1709	0.2338	0.2375	0.2567	0.2589
Num. obs.	797374	797374	797374	797374	797374

*** p < 0.001; ** p < 0.01; * p < 0.05

Statistical models

Annexe 2 – Résultats du modèle 5 en fonction du niveau d'engagement des émetteurs dans le sujet de conversation

	Ponctuels	Récidivistes	Fidèles	Passionnés	Superposteurs
(Intercept)	0.03*** (0.00)	0.14*** (0.01)	0.17*** (0.01)	0.23*** (0.01)	0.53*** (0.01)
followers	0.03*** (0.00)	-0.02*** (0.00)	-0.04*** (0.00)	-0.06*** (0.00)	-0.11*** (0.00)
favourites	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.00*** (0.00)
friends	-0.00 (0.00)	-0.00 (0.00)	-0.03*** (0.00)	-0.02*** (0.00)	-0.12*** (0.00)
status	-0.03*** (0.00)	-0.06*** (0.00)	-0.06*** (0.00)	-0.04*** (0.00)	-0.08*** (0.00)
profil_transparency	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.01*** (0.00)	0.00*** (0.00)
émopos	-0.00 (0.00)	-0.00** (0.00)	-0.00 (0.00)	-0.00 (0.00)	0.00*** (0.00)
émonég	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	0.00*** (0.00)
ARI	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
CTTR	0.01*** (0.00)	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.03*** (0.00)
emoji	0.01*** (0.00)	0.04*** (0.00)	0.04*** (0.00)	0.04*** (0.00)	0.06*** (0.00)
mention	-0.03*** (0.00)	-0.04*** (0.00)	-0.06*** (0.01)	-0.07*** (0.01)	-0.11*** (0.01)
hashtag	0.01** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.00 (0.00)
url	-0.02*** (0.00)	-0.03*** (0.00)	-0.05*** (0.00)	-0.11*** (0.00)	-0.13*** (0.00)
medias	0.10*** (0.00)	0.11*** (0.00)	0.08*** (0.00)	0.06*** (0.00)	0.03*** (0.00)
followers:favourites	0.03*** (0.00)	0.04*** (0.00)	0.04*** (0.00)	0.05*** (0.00)	0.03*** (0.00)
followers:friends	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.04*** (0.00)
favourites:friends	-0.03*** (0.00)	-0.05*** (0.00)	-0.04*** (0.00)	-0.04*** (0.00)	-0.02*** (0.00)
followers:status	0.00*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.02*** (0.00)
émopos:émonég	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00** (0.00)
bot					-0.03*** (0.00)
R ²	0.17	0.21	0.23	0.25	0.36
Adj. R ²	0.17	0.21	0.23	0.25	0.36
Num. obs.	90835	162373	158164	156102	229900
RMSE	0.24	0.29	0.30	0.31	0.28

***p < 0.001; **p < 0.01; *p < 0.05

Statistical models

Annexe 3 – Résultats du modèle 5 en fonction du contexte d'émission et de réception des messages

	Avant	Pendant	Post
(Intercept)	0.2661*** (0.0071)	0.1867*** (0.0067)	0.2040*** (0.0030)
followers	-0.0482*** (0.0023)	-0.0434*** (0.0023)	-0.0453*** (0.0010)
favourites	0.0075*** (0.0005)	0.0111*** (0.0005)	0.0070*** (0.0002)
friends	-0.0322*** (0.0022)	-0.0314*** (0.0022)	-0.0488*** (0.0009)
status	-0.0688*** (0.0015)	-0.0566*** (0.0015)	-0.0512*** (0.0006)
profil_transparency	0.0066*** (0.0008)	0.0142*** (0.0008)	0.0114*** (0.0003)
bot	-0.0637*** (0.0029)	-0.0612*** (0.0031)	-0.0365*** (0.0012)
émopos	-0.0009*** (0.0002)	0.0002 (0.0002)	-0.0000 (0.0001)
émonég	-0.0004 (0.0002)	0.0005 (0.0003)	-0.0000 (0.0001)
ARI	-0.0020*** (0.0001)	-0.0020*** (0.0001)	-0.0013*** (0.0001)
CTTR	0.0095*** (0.0012)	0.0199*** (0.0012)	0.0205*** (0.0005)
emoji	0.0690*** (0.0042)	0.0368*** (0.0032)	0.0399*** (0.0016)
mention	-0.0638*** (0.0073)	-0.0570*** (0.0070)	-0.0655*** (0.0029)
hashtag	0.0341*** (0.0033)	0.0241*** (0.0031)	0.0143*** (0.0014)
url	-0.0753*** (0.0023)	-0.0537*** (0.0025)	-0.0466*** (0.0011)
medias	0.1073*** (0.0024)	0.0795*** (0.0023)	0.0689*** (0.0010)
followers:favourites	0.0363*** (0.0006)	0.0408*** (0.0006)	0.0338*** (0.0003)
followers:friends	0.0222*** (0.0007)	0.0244*** (0.0007)	0.0254*** (0.0003)
favourites:friends	-0.0346*** (0.0006)	-0.0391*** (0.0007)	-0.0309*** (0.0003)
followers:status	0.0171*** (0.0004)	0.0131*** (0.0004)	0.0108*** (0.0002)
émopos:émonég	-0.0000 (0.0001)	0.0000 (0.0001)	0.0001* (0.0000)
R ²	0.2779	0.2796	0.2539
Adj. R ²	0.2778	0.2794	0.2539
Num. obs.	117440	109434	570500
RMSE	0.3020	0.2969	0.2860

***p < 0.001; **p < 0.01; *p < 0.05

Statistical models

