

## Twitter, mesure du bien-être ?

### Résumé

Depuis sa création en 2006, le service de micro-blogging Twitter a acquis une place centrale dans la circulation et la diffusion d'informations dans de nombreux pays. Le caractère public de la plupart des messages (tweets), la présence de publics divers et la simple masse des messages en fait un révélateur des mécanismes de diffusion des nouvelles ou de l'état d'esprit des utilisateurs prisé par les chercheurs.

De nombreux travaux récents ont tenté d'extraire de Twitter des informations sur l'état de l'opinion et sur le bien-être ressenti par la population. Plutôt destinée à un environnement de recherche, cette note présente brièvement une sélection d'articles, mettant un accent sur les méthodes utilisées et les résultats obtenus. Sans prétendre à l'exhaustivité, ce survol nous semble restituer une image assez fidèle de l'état de l'art et de ses limites quant à l'utilisation qui peut être faite de Twitter comme révélateur du bien-être ou des orientations politiques d'une population.

Dylan Alezra

[dylan.alezra@cepremap.org](mailto:dylan.alezra@cepremap.org)

Assistant de recherche  
Observatoire du Bien-  
être du CEPREMAP

Iris Laugier

Assistante de  
recherche  
Observatoire du Bien-  
être du CEPREMAP

Mathieu Perona

[mathieu.perona@cepremap.org](mailto:mathieu.perona@cepremap.org)

Directeur exécutif de  
Observatoire du Bien-  
être du CEPREMAP

*Nous remercions Thomas Renault pour sa relecture attentive d'une première version de cette note. Toutes les erreurs sont de notre fait.*

Dylan Alezra, Iris Laugier, Mathieu Perona, « Twitter, mesure du bien-être ? », Observatoire du Bien-être du CEPREMAP, n°2020-09, 19 octobre 2020

## 1. Introduction

Twitter a largement popularisé le micro-blogging depuis sa création en 2006. Les utilisateurs peuvent s'exprimer dans des messages de moins de 280 caractères (*tweet*)<sup>1</sup>, partager des messages (*retweets*) et suivre des comptes d'utilisateurs, anonymes ou très connus (*follow*). Avec plus de 500 millions de tweets par jour, la plateforme représente une mine d'information, en particulier pour l'analyse du bien-être. En effet, les récentes méthodes d'analyse de données (*machine learning, natural language processing*) permettent d'analyser les émotions associées à des messages. Dès lors, de nombreux articles ont essayé d'analyser le bien-être des utilisateurs ou d'un pays, notamment dans sa dimension hédonique (le bonheur ressenti). La présente revue de littérature, non-exhaustive, essaie de montrer comment les données Twitter peuvent servir à mesurer le bien-être (I) et peuvent contribuer à son analyse tant à l'échelle microéconomique que macroéconomique (II).

## 2. Mesurer le bien-être à partir des données Twitter

### L'extraction de données sur Twitter

Il existe deux applications principales pour récupérer des données sur Twitter, à savoir la Twitter Stream API, et la Twitter Search API<sup>2</sup>. La première permet uniquement de récupérer des données en temps réel, tandis que la seconde donne au plus un accès à 7 jours d'historique. Ces applications fournissent un échantillon aléatoire de tweets du flux – filtrable – ainsi que les métadonnées associées (localisation, heure, id, nombre de followers). Il est également possible d'extraire les informations de comptes spécifiques, mais Twitter limite le nombre d'extractions par heure. Plus largement, la principale difficulté pour la constitution de base de données est la constitution d'un historique au vu des contraintes imposées par les différentes applications. Plusieurs bases de données sont dispo-

<sup>1</sup> Jusqu'en novembre 2017, la limite était de 140 caractères. Les règles de comptage des caractères ont évolué dans le temps, excluant de la limite les alias d'utilisateurs, liens hypertextes ou mots-clé (*hashtags*).

<sup>2</sup> Une API (*Application Programming Interface*) est une interface de programmation d'application. Qualitativement, il s'agit d'un ensemble de règles permettant à des programmes informatiques (ici ceux utilisés par les scientifiques) d'envoyer des demandes normées à un autre programme informatique (ici, les systèmes d'information de Twitter). Elles se différencient du *web scraping*, par lequel un programme récolte des informations conçues pour être consultées par un humain.

nibles sur des sites comme GitHub ou Kaggle. Cependant, comme chaque étude a des besoins en données spécifiques, il revient souvent aux auteurs de les constituer eux-mêmes.

Les tweets récoltés peuvent être écrits dans des langues différentes. Quel impact potentiel sur l'analyse des sentiments ? D'une part, les modèles conçus pour analyser une langue spécifique, par exemple l'anglais, auront de moins bonnes performances en moyenne appliqués à des messages traduits. (Hennessy, 2016)<sup>3</sup>. La traduction littérale fait sans doute de perdre des informations de contexte. D'autre part, le perfectionnement des traducteurs automatiques (DeepL, Google Translate) a permis en partie de répondre à cette difficulté avec des traductions plus précises, ce qui permet d'utiliser des bases de tweets multi-linguales (Balahur et Turchi, 2013)<sup>4</sup>. Ainsi, plus les traducteurs automatiques progresseront, plus l'utilisation de base multi-linguales sera efficace.

### La dimension hédonique du bien-être

Notre revue non-exhaustive de la littérature met en évidence trois méthodes de mesure du bien-être dans sa dimension hédonique (bonheur ressenti). La première consiste à catégoriser les messages sur Twitter selon qu'ils renvoient à des émotions positives ou négatives. Cette catégorisation peut être effectuée de diverses façon. Bollen et al., (2011)<sup>5</sup> et Lachanski et al., (2017)<sup>6</sup> utilisent les mots employés par les utilisateurs associés à l'expression d'un sentiment (e.g « I feel ») et le type de sentiment (« good », « bad »,etc). Xianghua et al. (2012)<sup>7</sup> utilisent un dictionnaire de plus de 6800

<sup>3</sup> Catherine M. Hennessy and others, "Social Media and Anatomy Education: Using Twitter to Enhance the Student Learning Experience in Anatomy: Use of Twitter in Anatomy Education", *Anatomical Sciences Education*, 9.6 (2016), 505–15 <https://doi.org/10.1002/ase.1610>.

<sup>4</sup> Alexandra Balahur and Marco Turchi, "Improving Sentiment Analysis in Twitter Using Multilingual Machine Translated Data", in *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013* (presented at the RANLP 2013, Hissar, Bulgaria: INCOMA Ltd. Shoumen, BULGARIA, 2013), pp. 49–55 <https://www.aclweb.org/anthology/R13-1007>.

<sup>5</sup> Johan Bollen, Huina Mao, and Xiaojun Zeng, "Twitter Mood Predicts the Stock Market", *Journal of Computational Science*, 2.1 (2011), 1–8 <https://doi.org/10.1016/j.jocs.2010.12.007>.

<sup>6</sup> Michael Lachanski and S. Pav, "Shy of the Character Limit: "Twitter Mood Predicts the Stock Market" Revisited", *Econ Journal Watch*, 14 (2017), 302–45 <https://econjwatch.org/articles/shy-of-the-character-limit-twitter-mood-predicts-the-stock-market-revisited>.

<sup>7</sup> Fu Xianghua and others, "Multi-Aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and HowNet Lexicon", *Knowledge-Based Systems*, 37 (2013), 186–195 <https://doi.org/10.1016/j.kbs.2012.12.007>.

mots, identifiés comme positifs ou négatifs. Go et al. (2012)<sup>8</sup> prennent en compte les émoticônes des messages pour pouvoir identifier le sentiment du message. Enfin, Renault (2019)<sup>9</sup> intègre également la ponctuation ce qui renforce la précision. Néanmoins, la précision de l'identification des sentiments dépasse rarement 80% pour un sentiment binaire.

À partir de cette catégorisation, la plupart des auteurs construisent des indices de sentiment agrégé par journée. Ces indices, bien qu'ils diffèrent d'un article à l'autre, représentent en règle générale la différence entre la somme des émotions positives et négatives de la journée, rapportée au nombre de tweets.

Cette catégorisation est approfondie par la seconde méthodologie qui consiste à classer les tweets en humeurs d'après les mots utilisés. Ainsi, Bollen et al. (2011) utilisent le *Profile of mood states* (POMS), qui alloue une des six humeurs suivantes à chaque tweet : *Calm, Alert, Sure, Kind* et *Happy*. Une série temporelle pour chaque humeur est ensuite constituée et leur sert de métrique. Schwartz et al. (2017)<sup>10</sup> associent les mots des tweets à l'une des cinq composantes PERMA (émotion, engagement, relation, sens, réalisation) du bien-être. Cette approche a de fortes limitations. En effet, il est déjà difficile d'identifier précisément un sentiment binaire. Or, ici, les auteurs cherchent à estimer des émotions complexes, ce qui laisse douter de la précision et de la pertinence de ce type d'analyse. Par ailleurs, la prise en compte du contexte est essentielle pour caractériser un tweet. Or, des mots traditionnellement négatifs peuvent être employés positivement, et inversement. Cette difficulté supplémentaire peut limiter la pertinence de la classification.

Enfin, la dernière méthode consiste, à partir d'une liste de mots prédéfinies, à faire attribuer par des humains un score de bien-être aux mots. Mitchell et al (2013)<sup>11</sup> font ainsi attribuer sur Mechanical Turk (Amazon) un score de bien-être à des milliers de mots. Dodds et al.

(2011)<sup>12</sup> ou encore Hennessy et al. (2016)<sup>13</sup> optent pour la même méthodologie.

### La satisfaction dans la vie

Certains articles portent sur la satisfaction dans la vie. Yang et al. (2016)<sup>14</sup> mettent ainsi en place une méthodologie innovante permettant d'inférer la satisfaction dans la vie contenue dans les tweets. En cohérence avec l'idée que la satisfaction dans la vie est une mesure plus stable du bien-être que sa définition hédonique, ils trouvent que les séries temporelles issues de Twitter de satisfaction et d'insatisfaction dans la vie présentent des fluctuations aléatoires, décorréélées des chocs politiques, économiques, ou autres. Ils mènent une analyse lexicographique au cours du temps du contenu des tweets des satisfaits et des insatisfaits et trouvent que les insatisfaits tweetent beaucoup d'émotions négatives (tristesse, colère, dépression, mort) que les satisfaits. Aussi, les individus satisfaits qui deviennent insatisfaits au cours du temps postent davantage de tweets négatifs que ceux qui restent insatisfaits tout le long. Dans le même registre, Schwartz et al. (2017)<sup>15</sup> trouvent une forte corrélation entre la satisfaction dans la vie mesurée dans les enquêtes et la prévalence sur Twitter (*i*) des termes associés aux cinq composantes du bien-être (PERMA) de Seligman et des termes du Linguistic Inquiry and Word Count (mots associés à des émotions positives, négatives, jurons...) et (*ii*) des sujets abordés dans l'environnement local<sup>16</sup> (sport, comportements pro-sociaux...). Puis, ils montrent qu'un modèle de prédiction de la satisfaction dans la vie au niveau des comtés américains a un meilleur pouvoir prédictif lorsque, en plus des contrôles sociodémographiques classiques, il inclut les mesures de prévalence de (*i*) et (*ii*). Cela donne des pistes de compréhension de la manière dont les gens explicitent et attribuent de l'importance dans leurs discours aux différentes composantes du bien-être.

knosys.2012.08.003.

<sup>8</sup> Alec Go, Richa Bhayani, and Lei Huang, "Twitter Sentiment Classification Using Distant Supervision", 6 <https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>.

<sup>9</sup> Thomas Renault, "Sentiment Analysis and Machine Learning in Finance: A Comparison of Methods and Models on One Million Messages", *Digital Finance*, 2019 <https://doi.org/10.1007/s42521-019-00014-x>.

<sup>10</sup> Aaron J. Schwartz and others, "Visitors to Urban Greenspace Have Higher Sentiment and Lower Negativity on Twitter", *People and Nature*, 0.0 <https://doi.org/10.1002/pan3.10045>.

<sup>11</sup> "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place" <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3667195/>.

<sup>12</sup> Peter Sheridan Dodds et al., « Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter », *PLOS ONE* 6, n° 12 (7 décembre 2011): e26752, <https://doi.org/10.1371/journal.pone.0026752>.

<sup>13</sup> Hennessy, Catherine M., Emma Kirkpatrick, Claire F. Smith, and Scott Border, "Social Media and Anatomy Education: Using Twitter to Enhance the Student Learning Experience in Anatomy: Use of Twitter in Anatomy Education", *Anatomical Sciences Education*, 9.6 (2016), 505–15 <https://doi.org/10.1002/ase.1610>

<sup>14</sup> Chao Chen and others, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam", *IEEE Transactions on Information Forensics and Security*, 12.4 (2017), 914–25 <https://doi.org/10.1109/TIFS.2016.2621888>.

<sup>15</sup> Voir note 10.

<sup>16</sup> Leur analyse se place au niveau du comté américain.

### 3. Analyser le bien-être : selon quelles modalités ?

#### L'impact de chocs sur le bien-être

Certains papiers offrent une description détaillée de l'évolution du bien-être sur Twitter. Ainsi, les mesures de bien-être sur Twitter évoluent significativement lors d'événements annuels, comme par exemple Noël (Dodds et al., 2011 ; Curini et al., 2015<sup>17</sup>), la fête de Thanksgiving ou le 4 juillet<sup>18</sup> (Dodds et al., 2011), ou encore lors d'événements tels que le Mariage royal ou la mort de Ben Laden en 2011. Bollen et al. (2011) décelent une corrélation positive entre les humeurs *Calm, Sure, Vital et Happy* sur Twitter et les élections américaines de 2008 et Thanksgiving. Les événements de nature économique comme le plan de sauvetage du système financier de 2008 aux US (Dodds et al. 2011) ou des variations fortes du spread entre obligations italiennes et allemandes (Curini et al., 2015), sont fortement corrélés aux mesures du bien-être sur twitter. De plus, Dodds et al. (2011) montrent que le bien-être suit un cycle au cours de la semaine (augmente de mardi à samedi et diminue du dimanche au mardi) et des heures de la journée (heures les plus heureuses de 5 à 6 heures du matin puis baisse jusqu'à 22h puis reprise de la hausse pendant la nuit). D'autres, comme Gruzd et al (2011)<sup>19</sup> décrivent comment les émotions se diffusent sur le réseau lors d'un événement particulier. Ils se concentrent sur les JO d'Hiver 2010, et montrent une certaine *contagion* sur le réseau social : les messages positifs sont davantage retweetés que les messages négatifs. Enfin, certains papiers se focalisent sur la corrélation entre les mesures des émotions /humeurs de twitter et l'évolution de variables macroéconomiques comme dans Renault (2019), Lachanski et al. (2017) ou encore Bollen et al. (2011) bien que ce dernier papier ait été très critiqué. En effet, si l'observe une corrélation entre les émotions sur twitter et des variables macro-financières, Twitter n'a pas de pouvoir prédictif sur celles-ci.

<sup>17</sup> Luigi Curini, Stefano Iacus, and Luciano Canova, "Measuring Idiosyncratic Happiness Through the Analysis of Twitter: An Application to the Italian Case", *Social Indicators Research*, 121.2 (2015), 525-42 <https://doi.org/10.1007/s11205-014-0646-2>.

<sup>18</sup> Date de la déclaration d'indépendance des États-Unis d'Amérique.

<sup>19</sup> Anatoliy Gruzd, "Discovery and Visualization of Scholarly Information Diffusion in Twitter Networks", 2 [http://mappingideas.sdsu.edu/old\\_Mappingideas/SummerWorkshop/2012/Papers/Gruzd\\_Position\\_Paper.pdf](http://mappingideas.sdsu.edu/old_Mappingideas/SummerWorkshop/2012/Papers/Gruzd_Position_Paper.pdf).

### Enrichir l'analyse par la géographie : visualiser les zones de bonheur

Du fait de la géolocalisation possible d'un certain nombre de tweets, un pan de la littérature présente des analyses géographiques de la distribution du bien-être mesuré sur Twitter. Le bien-être présente par exemple d'importantes hétérogénéités par provinces italiennes (Curini et al 2011). Aux États-Unis, Mitchell et al. (2013) retrouvent une hétérogénéité bien plus visible entre les villes qu'entre les états. Aux deux niveaux, leurs mesures sont significativement corrélées à des mesures du bien-être comme celles de Gallup.

Il est alors possible d'étudier la manière donc les mesures des conditions socio-économiques sont corrélées avec ces mesures géolocalisées du bien-être. Dans l'aire londonienne, Guo et al. (2016)<sup>20</sup> montrent que le nombre d'emplois disponibles par quartiers (*ward*) et la part d'enfant de 0 à 15 ans, sont corrélés avec le bonheur agrégé du *ward* mesuré par Twitter. L'association entre cette métrique et la qualité d'accès aux transports publics montre par exemple que les *wards heureux* sont ceux où l'accès aux transports publics est très bon, et ceux où il est si mauvais que les gens utilisent leur propre voiture pour se déplacer. Les plus malheureux sont ceux qui comptent sur les transports publics mais n'y ont pas facilement accès.

Aux niveaux des villes américaines, Mitchell et al (2013) trouvent que des caractéristiques assimilables aux *faibles positions économiques* (part de familles qui vivent sous le seuil de pauvreté, qui ont un revenu de moins de 10 000 \$, le pourcentage d'Afro-américains, et la part de ménages monoparentaux avec une cheffe de famille...) sont négativement corrélées au niveau de bonheur moyen mesuré sur Twitter. De manière symétrique, les caractéristiques associées à des positions économiques élevées (part de personnes blanches, part d'individus mariés, part d'individus ayant un diplôme universitaire, etc.) sont positivement associées à cette mesure. Pour leur part, en Italie, Curini et al. (2011) trouvent une corrélation faible entre un indice de qualité de la vie des provinces et leur niveau de bien-être. Les auteurs suggèrent que cela vient du fait que leur index est un estimateur des émotions et non de la qualité de la vie à un instant *t*. Ces émotions ne peuvent alors pas être réduites au contexte institutionnel et à des conditions objectives. Au contraire, à Londres, Quercia et al (2011)<sup>21</sup> trouvent une corréla-

<sup>20</sup> Yuan Huang and others, "Understanding U.S. Regional Linguistic Variation with Twitter Data Analysis", *Computers, Environment and Urban Systems*, 59 (2016), 244-55 <https://doi.org/10.1016/j.compenvurbsys.2015.12.003>.

<sup>21</sup> Daniele Quercia and others, "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter", 2011, pp. 180-85 <https://doi.org/10.1109/PASSAT/>

tion significative entre leurs mesures d'affects positifs issues des tweets et l'indice composite de privation : les tweets des zones plus défavorisées sont plus négatifs que ceux des zones plus favorisées.

### Déterminer les préférences politiques des utilisateurs, et faire le lien avec le bien-être

La théorie des préférences révélées décrit en économie un cadre où les actions d'un agent permettent de déterminer quelles sont ses préférences, relativement à un champ. Sur Twitter, il y a plusieurs façon de signifier ses préférences à partir des tweets, des retweets, ou alors des comptes suivis. En particulier, cette théorie s'applique relativement bien au champ politique. Estimer les positions des individus, des médias ou encore des politiciens sur le spectre idéologique est un exercice traditionnel en sciences politiques. Cependant, cet exercice est souvent fastidieux en raison de la faible disponibilité des données. Les réseaux sociaux sont donc une opportunité pour répondre à cette question. En effet, il y a de nombreuses discussions politiques sur Twitter<sup>22</sup> que l'on peut exploiter.

Bien que les utilisateurs de Twitter ne soient pas représentatifs de la population générale (Mellon et al., 2017)<sup>23</sup>, l'approche de Barbera et al. (2015)<sup>24</sup> permet tout de même de tirer des informations pertinentes. Les auteurs font deux hypothèses essentielles. La première est qu'il faut voir les utilisateurs comme des experts de la politique, qui donnent leur avis. Cette hypothèse permet de catégoriser les politiciens sur le plan idéologique. La seconde est que les utilisateurs préfèrent suivre des comptes de politiciens (ou associés) dont la position idéologique latente est similaire à la leur. En attribuant un label politique à des comptes relativement influents, ils extraient l'ensemble de leurs followers. Ils créent ainsi un modèle où la probabilité de suivre un compte est dépendant de la distance idéologique entre l'utilisateur et le compte, la popularité du compte, et l'intérêt de l'utilisateur pour la politique. Leurs données leur permettent alors d'estimer sur le plan idéologique les utilisateurs, et de confirmer la po-

sition idéologique de chaque compte Twitter politique. Actuellement, les algorithmes ont été poussés au point de catégoriser un compte Twitter politiquement seulement en regardant ses followers (Barbera et al., 2016<sup>25</sup> ; Stamatetos et al., 2020<sup>26</sup>; et pour le cas français : Briatte et al., 2017<sup>27</sup>). Cet exercice de catégorisation fonctionne aussi bien avec des systèmes bipartites, comme multipartites.

Wong et al. (2016)<sup>28</sup> mettent en avant l'intérêt d'utiliser les données sur les retweets, plus simples à extraire, que celles sur les followers. L'action de retweeter est un acte d'approbation, potentiellement politique, quand suivre un compte répond à des logiques plus complexes, et qui ne sont pas en temps réel. Ils font l'hypothèse que les tweets sur un événement politiquement clivant fera l'objet de plus de retweets, et en ce en particulier par les utilisateurs qui y sont plus sensibles. À partir des comptes Twitter de deux journaux américains aux bords politiques opposés, ils identifient les préférences des utilisateurs au moment d'événement comme par exemple l'engagement d'Obama sur le mariage homosexuel en 2013. Plus encore, Barbera et al. (2016) montrent que ce sont les individus aux extrêmes du spectre idéologique qui ont le plus tendance à retweeter les sources politiques en accord avec leurs convictions politiques. En particulier, Donald Trump au moment des élections de 2016 avait eu plus tendance à rallier des « hate-groups » (e.g anti-LGBT ou anti-immigration) sur les réseaux sociaux comme montré par Sainudiin et al. (2019)<sup>29</sup>.

L'analyse du langage sur Twitter peut aussi être un indicateur politique. Preoțiu-Pietro et al. (2017)<sup>30</sup>

SocialCom.2011.26.

<sup>22</sup> Andranik Tumasjan and others, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", 8 <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>.

<sup>23</sup> Mellon, Jonathan, and Christopher Prosser. "Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users." *Research & Politics*, July 2017, doi:10.1177/2053168017720008.

<sup>24</sup> Pablo Barberá, "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data", *Political Analysis*, 23.1 (2015), 76–91 <https://doi.org/10.1093/pan/mpu011>.

<sup>25</sup> Pablo Barbera, "Social Media, Echo Chambers, and Political Polarization", 20 <http://www.pablobarbera.com/static/echo-chambers.pdf>.

<sup>26</sup> Giorgos Stamatelatos and others, "Revealing the Political Affinity of Online Entities through Their Twitter Followers", *Information Processing & Management*, 57.2 (2020), 102172 <https://doi.org/10.1016/j.ipm.2019.102172>.

<sup>27</sup> François Briatte and Ewen Gallic, "Recovering the French Party Space from Twitter Data", 19 <https://halshs.archives-ouvertes.fr/halshs-01511384>.

<sup>28</sup> Felix Ming Fai Wong and others, "Quantifying Political Leaning from Tweets, Retweets, and Retweeters", *IEEE Transactions on Knowledge and Data Engineering*, 28.8 (2016), 2158–72 <https://doi.org/10.1109/TKDE.2016.2553667>.

<sup>29</sup> Raazesh Sainudiin and others, "Characterizing the Twitter Network of Prominent Politicians and SPLC-Defined Hate Groups in the 2016 US Presidential Election", *Social Network Analysis and Mining*, 9 (2019) <https://doi.org/10.1007/s13278-019-0567-9>.

<sup>30</sup> Daniel Preoțiu-Pietro and others, "Beyond Binary Labels: Political Ideology Prediction of Twitter Users", in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (presented at the ACL 2017, Vancouver, Canada: Association for Computational Linguistics, 2017), pp. 729–740

montrent que certains mots sont plus employés en moyenne selon l'orientation politique (déjà connue au préalable grâce à une enquête dans leur article, mais qui pourrait être retrouvé par des méthodes statistiques comme vu précédemment). Par exemple, les conservateurs américains auraient plus tendance à utiliser des mots comme « avortement », ce qui fait directement référence au débat entre les pro-life et les pro-choice aux États-Unis. Cependant, ces analyses sont moins convaincantes dans la mesure où elles nécessitent une identification politique préalable des utilisateurs, et que le faible nombre de mots utilisés sur Twitter rend difficile l'exploitation des tweets, par rapport à d'autres sources (articles de blog ou de presse).

Cette possibilité est intéressante pour l'analyse du bien-être. En effet, les récents travaux du CEVIPOF (Algan, et al., 2018)<sup>31</sup> ont montré un lien important entre les préférences politiques, et la satisfaction dans la vie. Pouvoir catégoriser le positionnement idéologique des individus et mesurer indirectement au moins leur niveau de bien-être serait une façon intéressante d'enrichir et de confirmer ces résultats.

#### 4. Conclusion

Certes, Twitter n'est pas représentatif de la population générale, et ne substitue pas à des données d'enquêtes. Cependant, le volume gigantesque d'information que l'on peut en extraire offre une opportunité intéressante de mesurer le bien-être et d'en faire l'analyse selon des modalités en lien avec les travaux de l'Observatoire du bien-être, telles que la localisation géographique ou l'orientation politique. Les mesures du bien-être issues de Twitter, sont sans doute particulièrement prometteuses pour interpréter les événements majeurs tels que le confinement.

#### Bibliographie

Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Martial Foucault, *Les origines du populisme : enquête sur un schisme politique et social* (Paris, France: Seuil, 2019)

Balahur, Alexandra, and Marco Turchi, "Improving Sentiment Analysis in Twitter Using Multilingual Machine Translated Data", in *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013* (presented at the RANLP 2013, Hissar, Bulgaria: INCOMA Ltd. Shoumen, BULGARIA, 2013), pp. 49–55 <<https://www.aclweb.org/anthology/R13-1007>> [accessed 3 July 2020]

<https://doi.org/10.18653/v1/P17-1068>.

<sup>31</sup> Yann Algan et al., *Les origines du populisme : enquête sur un schisme politique et social* (Paris, France: Seuil, 2019).

Barberá, Pablo, "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data", *Political Analysis*, 23.1 (2015), 76–91 <<https://doi.org/10.1093/pan/mpu011>>

Barbera, Pablo, "Social Media, Echo Chambers, and Political Polarization", 20 <<http://pablobarbera.com/static/echo-chambers.pdf>>

Bollen, Johan, Huina Mao, and Xiaojun Zeng, "Twitter Mood Predicts the Stock Market", *Journal of Computational Science*, 2.1 (2011), 1–8 <<https://doi.org/10.1016/j.jocs.2010.12.007>>

Briatte, François, and Ewen Gallic, "Recovering the French Party Space from Twitter Data", 19, <<https://halshs.archives-ouvertes.fr/halshs-01511384/>>

Chen, Chao, Yu Wang, Jun Zhang, Yang Xiang, Wanlei Zhou, and Geyong Min, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam", *IEEE Transactions on Information Forensics and Security*, 12.4 (2017), 914–25 <<https://doi.org/10.1109/TIFS.2016.2621888>>

Curini, Luigi, Stefano Iacus, and Luciano Canova, "Measuring Idiosyncratic Happiness Through the Analysis of Twitter: An Application to the Italian Case", *Social Indicators Research*, 121.2 (2015), 525–42 <<https://doi.org/10.1007/s11205-014-0646-2>>

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, Christopher M. Danforth, "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter", *PLOS ONE* 6, n° 12 (7 décembre 2011): e26752 <<https://doi.org/10.1371/journal.pone.0026752>>

Go, Alec, Richa Bhayani, and Lei Huang, "Twitter Sentiment Classification Using Distant Supervision", <<https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>>

Gruzd, Anatoliy, "Discovery and Visualization of Scholarly Information Diffusion in Twitter Networks", 2 <[http://mappingideas.sdsu.edu/old\\_Mappingideas/SummerWorkshop/2012/Papers/Gruzd\\_Position\\_Paper.pdf](http://mappingideas.sdsu.edu/old_Mappingideas/SummerWorkshop/2012/Papers/Gruzd_Position_Paper.pdf)>

Hennessy, Catherine M., Emma Kirkpatrick, Claire F. Smith, and Scott Border, "Social Media and Anatomy Education: Using Twitter to Enhance the Student Learning Experience in Anatomy: Use of Twitter in Anatomy Education", *Anatomical Sciences Education*, 9.6 (2016), 505–15 <<https://doi.org/10.1002/ase.1610>>

Huang, Yuan, Diansheng Guo, Alice Kasakoff, and Jack Grieve, "Understanding U.S. Regional Linguistic Variation with Twitter Data Analysis", *Computers, Environment and Urban Systems*, 59 (2016), 244–55 <<https://doi.org/10.1016/j.compenvurbsys.2015.12.003>>

Lachanski, Michael, and S. Pav, "Shy of the Character Limit: "Twitter Mood Predicts the Stock Market" Revisited", *Econ Journal Watch*, 14 (2017), 302–45 <<https://econjwatch.org/articles/shy-of-the-character-limit-twitter-mood-predicts-the-stock-market-revisited>>

Mellon, Jonathan, and Christopher Prosser. "Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users." *Research & Politics*, July 2017, doi:10.1177/2053168017720008.

Mitchell, Lewis et al. "The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place." *PloS one* vol. 8,5 e64417. 29 May. 2013, <doi:10.1371/journal.pone.0064417>

Preoțiuc-Pietro, Daniel, Ye Liu, Daniel Hopkins, and Lyle Ungar, "Beyond Binary Labels: Political Ideology Prediction of Twitter Users", in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (presented at the ACL 2017, Vancouver, Canada: Association for Computational Linguistics, 2017), pp. 729–740 <<https://doi.org/10.18653/v1/P17-1068>>

Quercia, Daniele, Michal Kosinski, David Stillwell, and Jon Crowcroft, "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter", 2011, pp. 180–85 <<https://doi.org/10.1109/PASSAT/SocialCom.2011.26>>

Renault, Thomas, "Sentiment Analysis and Machine Learning in Finance: A Comparison of Methods and Models on One Million Messages", *Digital Finance*, 2019 <<https://doi.org/10.1007/s42521-019-00014-x>>

Sainudiin, Raazesh, Kumar Yogeeswaran, Kyle Nash, and Rania Sahioun, "Characterizing the Twitter Network of Prominent Politicians and SPLC-Defined Hate Groups in the 2016 US Presidential Election", *Social Network Analysis and Mining*, 9 (2019) <<https://doi.org/10.1007/s13278-019-0567-9>>

Schwartz, Aaron J., Peter Sheridan Dodds, Jarlath P. M. O'Neil-Dunne, Christopher M. Danforth, and Taylor H. Ricketts, "Visitors to Urban Greenspace Have Higher

Sentiment and Lower Negativity on Twitter", *People and Nature*, 0.0 <<https://doi.org/10.1002/pan3.10045>>

Stamatelatos, Giorgos, Sotirios Gyftopoulos, George Drosatos, and Pavlos S. Efraimidis, "Revealing the Political Affinity of Online Entities through Their Twitter Followers", *Information Processing & Management*, 57.2 (2020), 102172 <<https://doi.org/10.1016/j.ipm.2019.102172>>

Tumasjan, Andranik, Timm O Sprenger, Philipp G Sandner, and Isabell M Welp, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", 8, <<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1441/1852>>

Wong, Felix Ming Fai, Chee Wei Tan, Soumya Sen, and Mung Chiang, "Quantifying Political Leaning from Tweets, Retweets, and Retweeters", *IEEE Transactions on Knowledge and Data Engineering*, 28.8 (2016), 2158–72 <<https://doi.org/10.1109/TKDE.2016.2553667>>

Xianghua, Fu, Liu Guo, Guo Yanyan, and Wang Zhiqiang, "Multi-Aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and HowNet Lexicon", *Knowledge-Based Systems*, 37 (2013), 186–195 <<https://doi.org/10.1016/j.knosys.2012.08.003>>

Le CEPREMAP est né en 1967 de la fusion de deux centres, le CEPREL et le CERMAP, pour éclairer la planification française grâce à la recherche économique.

Le CEPREMAP est, depuis le 1er janvier 2005, le Centre Pour la Recherche Economique et ses Applications. Il est placé sous la tutelle du Ministère de la Recherche. La mission prévue dans ses statuts est d'assurer une interface entre le monde académique et les administrations économiques.

Il est à la fois une agence de valorisation de la recherche économique auprès des décideurs, et une agence de financement de projets dont les enjeux pour la décision publique sont reconnus comme prioritaires.

<http://www.cepremap.fr>

## Observatoire du Bien-être

L'Observatoire du bien-être au CEPREMAP soutient la recherche sur le bien-être en France et dans le monde. Il réunit des chercheurs de différentes institutions appliquant des méthodes quantitatives rigoureuses et des techniques novatrices. Les chercheurs affiliés à l'Observatoire travaillent sur divers sujets, comme des questions de recherche fondamentales telles que la relation entre éducation, santé et bien-être, l'impact des relations avec les pairs sur le bien-être, la relation entre le bien-être et des variables cycliques tels que l'emploi et la croissance et enfin l'évolution du bien-être au cours de la vie. Un rôle important de l'Observatoire est de développer notre compréhension du bien-être en France: son évolution au fil du temps, sa relation avec le cycle économique, les écarts en termes de bien-être entre différents groupes de population ou régions, et enfin la relation entre politiques publiques et bien-être.

<http://www.cepremap.fr/observatoire-bien-etre>

<https://twitter.com/ObsBienEtre>

### Directeur de publication

Mathieu Perona

### Comité scientifique

Claudia Senik

Yann Algan

Andrew Clark

Observatoire du Bien-être du CEPREMAP

48 Boulevard Jourdan

75014 Paris – France

+33(0)1 80 52 13 61

Collection *Notes de l'Observatoire du Bien-être*, ISSN 2646-2834