Les chiffres de l’influence, après réanalyse

Tyler Kohler . 25 août 2026 . note 04

En septembre et octobre 1968, Maxwell McCombs et Donald Shaw ont interrogé cent électeurs indécis de Chapel Hill et dépouillé les neuf supports d’information qui desservaient la ville. Ils ont comparé le classement des sujets par les rédactions et celui des électeurs. La corrélation publiée atteint +0,967.

Le chiffre a fondé une littérature entière. Il porte sur deux listes d’une quinzaine de catégories d’enjeux, agrégées, et sur un nuage aussi court un coefficient élevé s’obtient sans peine. Le protocole est transversal, une seule prise de mesure, et les auteurs ne prétendaient établir aucune causalité. L’enquête de terrain a duré dix-neuf jours.

De l’agenda à l’amorçage

La démonstration expérimentale est venue quatorze ans plus tard. Shanto Iyengar, Mark Peters et Donald Kinder ont monté des journaux télévisés où ils avaient inséré de vrais sujets, à l’insu des participants, en faisant varier le problème mis en avant, la défense, la pollution ou l’inflation. Le problème inséré gagnait en importance perçue. Il pesait aussi davantage dans le jugement porté sur le président en fonction, second effet qui relève de l’amorçage, mécanisme distinct de la fixation de l’agenda. Les effectifs étaient faibles et les volontaires recrutés localement.

On rapproche souvent ces travaux de la méta-analyse de Joshua Kalla et David Broockman, dont le résultat central est un effet moyen nul du contact de campagne sur le choix du candidat, sur quarante essais de terrain menés en élection générale. Le rapprochement demande de la prudence. Kalla et Broockman mesurent du porte-à-porte, du courrier, des appels et de la publicité ; la saillance d’un enjeu se mesure autrement. Ils relèvent d’ailleurs des effets persuasifs non nuls dans les primaires, sur les objets soumis au vote, et lorsqu’un candidat défend une position impopulaire.

Comment grandit une cascade

Sharad Goel, Ashton Anderson, Jake Hofman et Duncan Watts ont analysé environ un milliard d’événements de diffusion sur Twitter, entre juillet 2011 et juin 2012, en se limitant aux liens pointant vers une liste fixe de sites très fréquentés. Ils ont introduit la viralité structurelle, distance moyenne entre toutes les paires de nœuds d’un arbre de diffusion. La taille moyenne d’un arbre est de 1,3, près de 99 % des adoptions viennent du compte émetteur ou de ses abonnés directs, et une cascade sur 4'000 environ atteint cent nœuds, seuil retenu pour l’analyse structurelle.

Au-dessus de ce seuil, la diversité des formes devient grande, au point que les auteurs la disent surprenante. À popularité égale, la viralité varie beaucoup. Un même niveau d’audience s’obtient par une émission unique vers un très large public, par une chaîne qui traverse plusieurs générations d’intermédiaires, ou par des régimes intermédiaires entre ces deux extrêmes. Le volume observé ne renseigne donc pas sur le mécanisme qui l’a produit.

Le constat vaut pour les dossiers de réputation. La structure de l’arbre reste inaccessible de l’extérieur, et je n’ai jamais réussi à la reconstituer. Demeurent observables les horodatages de publication et la reprise de formulations d’un site à l’autre.

Cent vingt-six mille cascades

En mars 2018, Science a publié l’étude de Soroush Vosoughi, Deb Roy et Sinan Aral. Environ 126'000 cascades de rumeurs diffusées entre 2006 et 2017, reprises quelque 4,5 millions de fois par près de trois millions de comptes, avec un statut de vrai ou de faux établi par six organisations de vérification dont les jugements concordaient dans plus de 95 % des cas. Une information fausse avait environ 70 % de chances de plus d’être retransmise qu’une information vraie, et il fallait à la vérité six fois plus de temps pour atteindre 1'500 personnes. Le retrait des comptes automatisés n’a pas déplacé ces écarts. L’échantillon ne retient que des énoncés dont des vérificateurs se sont saisis, et une seule plateforme.

En 2021, Jonas Juul et Johan Ugander ont réanalysé ces données en appariant les cascades sur leur taille. À taille égale, les différences de profondeur, de largeur, de viralité structurelle et de vitesse entre vrai et faux s’effacent presque entièrement, alors qu’elles subsistent entre types de contenus, vidéos, images, articles ou pétitions. La lecture qui a circulé ensuite, celle d’un Vosoughi réfuté, déforme le papier. En agrégat, le faux atteint toujours plus de monde et plus vite ; la propriété qui disparaît une fois la taille contrôlée est la forme de la diffusion, ce qui oriente vers une contagiosité initiale plus élevée. Aral a répondu dans un billet en ligne, sans passer par une revue. Aucune de ces séries ne descend au niveau d’une personne privée.

Où les gens lisent l’actualité

Seth Flaxman, Sharad Goel et Justin Rao ont travaillé sur les historiques de navigation de 50'383 internautes américains, retenus par filtrage au sein d’un panel volontaire de 1,2 million d’utilisateurs d’une barre d’outils de navigateur, entre mars et mai 2013. Moteurs et réseaux sociaux y sont associés à une ségrégation idéologique plus élevée et, simultanément, à une exposition accrue aux contenus du bord opposé. L’ampleur reste modeste, de l’aveu des auteurs, l’indice passant pour l’information descriptive de 0,11 en accès direct à 0,12 par les réseaux sociaux. La navigation directe domine la consommation, 79 % pour l’information descriptive et 67 % pour l’opinion.

Jennifer Allen et ses coauteurs ont estimé que l’information de toute nature occupe 14,2 % du régime médiatique quotidien des Américains, et les fausses nouvelles au sens strict 0,15 %. Ce second chiffre dépend d’une liste fermée d’une centaine de domaines, et son dénominateur inclut la télévision, qui écrase tout le reste. Ces travaux portent sur des agrégats nationaux. Je n’ai aucune source sérieuse sur le nombre de personnes qui consultent les résultats attachés à un nom propre peu recherché.

Vingt pour cent, puis deux à quatre

Robert Epstein et Ronald Robertson ont annoncé, dans les PNAS en 2015, des déplacements de préférence de vote de 20 % ou davantage chez des indécis exposés à des classements de résultats favorisant un candidat, le biais échappant à environ trois participants sur quatre. Cinq expériences randomisées en double aveugle, 4'556 participants, aux États-Unis et en Inde.

Quatre de ces cinq expériences faisaient juger à des Américains l’élection australienne de 2010, avec des candidats dont ils n’avaient jamais entendu parler, situation où déplacer une préférence coûte peu. Katharina Zweig a montré ensuite que la mesure retenue additionne les déplacements obtenus pour chaque candidat au lieu de les moyenner. Sur l’étude indienne, la plus proche d’une situation réelle, elle situe l’effet entre 2 et 4 % de l’ensemble des électeurs, et entre 3 et 10 % parmi ceux qui étaient acquis au départ à un autre candidat. La réplication citée à l’appui de ces résultats est signée par Epstein lui-même et porte sur 1'137 résidents américains, sur trois sujets sans rapport avec un scrutin.

Documenter pendant que c’est visible

Ces travaux portent sur des campagnes électorales et sur des plateformes publiques. Leur transposition à la réputation d’une personne demande de la prudence, et je ne connais pas de série expérimentale solide sur ce terrain ; l’assignation aléatoire y est exclue pour des raisons éthiques.

Depuis 2023, les dossiers qui me sont soumis arrivent presque tous avec le contenu litigieux, capturé ou imprimé, et sans la liste de résultats qui l’entourait au moment des faits. Chaque élément pris isolément peut être parfaitement licite, ce qui déplace la description au niveau de l’ensemble, soit le nombre de points de publication, leur répartition dans le temps, la reprise littérale de formulations d’un site à l’autre, la position tenue dans une liste de résultats à une date donnée. Ces éléments portent une date et s’effacent vite. Une page modifiée suffit à rendre un relevé non reproductible, et les archives publiques ne couvrent qu’une fraction du web.

Le règlement européen sur les services numériques impose aux très grandes plateformes une évaluation périodique des risques systémiques liés à la conception de leurs services, puis des mesures d’atténuation, sous audit indépendant. La Suisse n’est pas liée par ce texte, à ceci près que ces plateformes n’exploitent pas d’architecture distincte pour les utilisateurs suisses.

Reste la partie ennuyeuse, un relevé horodaté de ce qui est visible sous le nom, repris à intervalles fixes même quand rien ne bouge. Après le litige, il ne se reconstitue qu’imparfaitement, à partir d’archives dont la couverture est aléatoire.

Maxwell McCombs et Donald Shaw, « The Agenda-Setting Function of Mass Media », Public Opinion Quarterly, 36(2), 1972, p. 176-187. Cent entretiens auprès d’électeurs indécis de Chapel Hill, du 18 septembre au 6 octobre 1968, neuf supports dépouillés, corrélations de +0,967 et +0,979 selon la place accordée aux articles.

Shanto Iyengar, Mark Peters et Donald Kinder, « Experimental Demonstrations of the “Not-So-Minimal” Consequences of Television News Programs », American Political Science Review, 76(4), 1982, p. 848-858 ; Shanto Iyengar et Donald Kinder, News That Matters. Television and American Opinion, University of Chicago Press, 1987, qui rassemble la série d’expériences menées au début des années 1980.

Joshua Kalla et David Broockman, « The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments », American Political Science Review, 112(1), 2018, p. 148-166. Quarante essais recensés et neuf expériences originales des auteurs.

Sharad Goel, Ashton Anderson, Jake Hofman et Duncan Watts, « The Structural Virality of Online Diffusion », Management Science, 62(1), 2016, p. 180-196. L’analyse de viralité structurelle porte sur les cascades d’au moins cent nœuds, soit 0,025 % du corpus, lui-même limité aux liens vers une liste fixe de sites populaires.

Soroush Vosoughi, Deb Roy et Sinan Aral, « The Spread of True and False News Online », Science, 359, 2018, p. 1146-1151. Vérificateurs retenus : factcheck.org, hoax-slayer.com, politifact.com, snopes.com, truthorfiction.com et urbanlegends.about.com.

Jonas Juul et Johan Ugander, « Comparing Information Diffusion Mechanisms by Matching on Cascade Size », Proceedings of the National Academy of Sciences, 118, 2021, article e2100786118. Les auteurs ont rappelé publiquement que leur réanalyse confirme le résultat agrégé de l’étude de 2018.

Sinan Aral, « Fake News about our Fake News Study Spread Faster than its Truth », billet publié en ligne, 2021. Prise de position publique, sans évaluation par les pairs.

Seth Flaxman, Sharad Goel et Justin Rao, « Filter Bubbles, Echo Chambers, and Online News Consumption », Public Opinion Quarterly, 80, numéro spécial, 2016, p. 298-320. Filtrage retenu : au moins dix articles d’actualité et deux éditoriaux entre mars et mai 2013.

Jennifer Allen, Baird Howland, Markus Mobius, David Rothschild et Duncan Watts, « Evaluating the Fake News Problem at the Scale of the Information Ecosystem », Science Advances, 6(14), 2020, article eaay3539.

Robert Epstein et Ronald Robertson, « The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections », Proceedings of the National Academy of Sciences, 112, 2015, p. E4512-E4521.

Katharina Zweig, « Watching the Watchers: Epstein and Robertson’s Search Engine Manipulation Effect », document de travail, AlgorithmWatch, 7 avril 2017.

Robert Epstein et Ji Li, « Can Biased Search Results Change People’s Opinions About Anything at All? A Close Replication of the Search Engine Manipulation Effect (SEME) », PLOS ONE, 19(3), 2024, article e0300727. Trois expériences, 1'137 participants américains, sur l’intelligence artificielle, la fracturation hydraulique et l’origine de l’homosexualité. Epstein publie depuis son propre institut de recherche.

Règlement (UE) 2022/2065 relatif à un marché unique des services numériques, articles 34 et 35.

Kohler, T. (2026). Les chiffres de l’influence, après réanalyse. Notes, tylerkohler.ch.
https://tylerkohler.ch/notes/les-chiffres-de-l-influence/