

Qui devient contributeur à Wikipédia ?

Analyser les pratiques de Wikipédia dans la durée (2023)

Léo Joubert¹ et Nicolas Jullien²

Cette enquête a bénéficié du soutien de la Région Bretagne et du PEPR eSEMBLE (projet ciblé CONGRATS), qui travaille sur le futur de la collaboration numérique.

Introduction

Comment devient-on contributeur à Wikipédia ? Il est maintenant établi que la participation à Wikipédia obéit à une série de corrélations qui définissent un profil sociodémographique précis. Par exemple, les contributeurs sont plus jeunes et plus masculins que l'ensemble des lecteurs de Wikipédia (Hargittai et Shaw, 2015 ; Bear et Collins, 2016). Qu'il y ait une toute petite minorité qui contribue à des projets en ligne en général, et à Wikipédia en particulier n'est pas surprenant. Par contre, que cette minorité soit constituée seulement d'hommes de plus de 40 ans avec au moins une licence interroge sur l'objectif d'universalité, tant sur l'encapacitation des contributeurs ("que tous peuvent éditer") que sur la couverture des sujets traités (cela est illustré par le mouvement des "[sans pages](#)" qui a montré comment il s'agissait surtout de femmes "sans pages", et qui a donné lieu à des projets pour y remédier, comme le projet [Women in red](#)).

Cependant, et malgré son utilité pour comprendre la contribution, l'inventaire raisonné de ces corrélations reste prisonnier d'une approche binaire « non-contributeurs » *versus* contributeurs (sans que ne soit, bien souvent, précisé le niveau de régularité nécessaire pour être considéré comme « contributeur »). Cela au mépris de la multitude d'étapes par lesquelles passent un individu avant de devenir, éventuellement, contributeur régulier. On sait, par exemple, que la façon dont la ou les premières contributions sont reçues par le projet est essentielle dans la probabilité qu'une personne devienne une contributrice régulière (Warncke-Wang, Morten, et al. 2023). Mais pour se jeter à l'eau et contribuer au moins une fois, il faut s'en sentir la capacité, ou même le droit (Lockett, 2020). Enfin, et ces enquêtes l'oublie trop souvent, ce n'est pas parce qu'on utilise Wikipédia qu'on sait comment il est construit, ou même si c'est le cas comment, concrètement, il est possible de contribuer.

¹ LPED, Université de Rouen Normandie

² LEGO, IMT Atlantique

Si, logiquement, on ne peut pas être contributeur régulier sans avoir d'abord essayé, ou essayé sans avoir précédemment appris comment on pouvait le faire, il n'est pas dit que ces différentes étapes dépendent de ressors socio-économiques identiques. Par exemple, les contributeurs réguliers sont-ils (majoritairement) des hommes parce que les hommes se sentent plus légitimes que les femmes à contribuer ? Ou parce que la plate-forme, numérique, renvoie à un imaginaire informatique dont les femmes sont socialement écartées ? Ou encore parce qu'ils ont plus de temps pour le faire régulièrement ? Ou est-ce que ces trois effets existent, parmi d'autres, et se renforcent mutuellement ?

Arrivant à la suite de moments d'un processus d'apprentissage, d'essais-erreurs, la contribution à Wikipédia suppose une socialisation, comme tous les usages et notamment les usages numériques (Proulx, 2002). C'est ainsi que ces activités sont corrélées à des caractéristiques socio-démographiques. L'exemple de la sous-représentation des femmes parmi les contributeurs en est une illustration. L'hypothèse que nous faisons ici, et que nous voulons étudier, est que si toutes ces étapes sont corrélées à des variables sociodémographiques, elles ne le sont peut-être pas de la même façon. Si tel s'avérait être le cas, alors les efforts pour augmenter la diversité des contributeurs à Wikipédia devraient tenir compte de la population à laquelle on s'adresse. Cherche-t-on à expliquer Wikipédia à « ceux qui ne savent pas » comment contribuer ou à convaincre « ceux qui savent » de se lancer ? Les actions à mener dans l'un ou l'autre des cas sont différentes.

Pour explorer cette question, nous avons utilisé des données collectées à l'aide d'une enquête en ligne réalisée entre juin et juillet 2023 auprès des lecteurs (et parfois contributeurs) de Wikipédia (Cruciani et al., 2023 présente l'enquête et ses données, et sur [une page du wiki](#) [wikimedia.org](#)). Le lien vers le questionnaire a été distribué via une bannière publiée en 8 langues sur la page Wikipédia (le questionnaire en anglais est accessible [ici](#)). L'enquête comprenait 200 questions sur : ce que les gens faisaient sur Wikipédia avant de cliquer sur le lien vers le questionnaire ; la manière dont ils utilisent Wikipédia en tant que lecteurs (utilisations professionnelles et personnelles) ; leur opinion sur la qualité, la couverture thématique et l'importance de l'encyclopédie ; la création de Wikipédia (comment ils pensent qu'elle est créée, s'ils y ont déjà contribué et comment) ; leurs activités sociales, sportives, artistiques et culturelles, en ligne et hors ligne ; leurs caractéristiques socio-économiques, y compris leurs convictions politiques, et leur propension à faire confiance à l'encyclopédie. Plus de 200 000 personnes ont ouvert le questionnaire, 100 332 ont commencé à y répondre, et 10 576 l'ont terminé.

Sur ces 10 704 réponses, et après traitement des valeurs manquantes, nous en retenons 7 827. Le retrait de 27% des répondants se justifie par le fait que notre modélisation nécessite

l'utilisation d'un grand nombre de variables indépendantes et, par conséquent, la sélection de la sous-population des répondants ayant répondu au questionnaire de la manière la plus complète possible.

Il faut préciser que notre échantillon est ce qu'on appelle « un échantillon de convenance », c'est-à-dire sans plan de sondage défini *a priori*. De ce fait, il ne nous dit rien sur la répartition effective dans la population qui utilise Wikipédia dans les huit langues de l'enquête. En effet, il ne prétend pas à la représentativité. En revanche, en formant l'hypothèse que les répondants représentent bien l'ensemble des profils et pratiques d'usage de l'encyclopédie, il devient possible d'analyser les relations statistiques entre des pratiques (ici contribuer) et des caractéristiques socio-démographiques, culturelles, professionnelles, etc. des personnes ayant ces pratiques. Autrement dit, sans pouvoir estimer de proportions (« il y a X % de femmes parmi les contributeurs »), notre échantillon nous laisse cependant mesurer des corrélations (par exemple, ce n'est qu'une illustration, pas une affirmation : « les femmes contribuant à Wikipédia sont plus diplômées que les hommes contribuant à Wikipédia », toutes choses étant égales par ailleurs).

Présentation du modèle

La variable à expliquer

Pour construire notre variable à expliquer, nous avons utilisé la question suivante : « Avez-vous déjà modifié une page Wikipédia ? », à laquelle nos enquêtés pouvaient répondre : « Non, et vous ne savez pas comment on fait » ou « Non, mais vous savez comment on fait » ou « Oui, rarement » ou « Oui, quelques fois » ou « Oui, souvent ».

Cette question nous permet de modéliser nos trois étapes : savoir comment on contribue (ou non) ; l'avoir fait (ou non) ; le faire régulièrement, selon le Tableau 1 ci-dessous. Par construction les individus qui ont répondu « Non, et vous ne savez pas comment on fait » n'ont pas franchi l'étape 1 et donc ne seront pas pris en compte dans l'analyse des éléments qui expliquent qu'on franchisse l'étape 2 (avoir contribué). De même, seuls les individus ayant franchi l'étape 2 (n'ayant pas répondu « Non, mais vous savez comment on fait ») sont gardés pour étudier les éléments qui expliquent qu'on franchisse l'étape 3 (contribuer régulièrement)³.

³ D'un point de vue statistique, cela veut dire qu'il faut s'assurer que la modélisation de l'étape N garde en mémoire le franchissement de l'étape N-1. Cela peut se faire grâce à la « procédure de Heckman », de façon à ce que chacune des modélisations garde la mémoire du franchissement des éventuelles étapes antérieures, sous la forme d'un coefficient évaluant la significativité d'un « biais de sélection ».

Question :	Avez-vous déjà modifié une page Wikipédia ?					Total
Modalités	Non, et vous ne savez pas comment on fait	Non, mais vous savez comment on fait	Oui, rarement	Oui, quelque fois	Oui, souvent	
Étape 1 : savoir comment contribuer	non	oui				100%
Étape 2 : contribuer au moins une fois		non	oui			100%
Étape 3 : contribuer beaucoup				non	oui	100%

Tableau 1 : modélisation des trois étapes selon la réponse à la question

Nous avons réalisé des régressions du type probit à chaque étape. Parce qu'il faut s'assurer de réponses complètes pour pouvoir modéliser ces étapes, la population de notre modèle est plus faible que la population de l'ensemble des répondants ayant fini le questionnaire : nous avons gardé seulement 7 827 réponses, parmi lesquelles 3 754 répondants ont indiqué savoir comment on contribue à Wikipédia (et donc 4 073 ne le savaient pas). Ces 3 754 répondants correspondent à notre population gardée pour l'étape 2, et dans cette population, 1 857 ont indiqué avoir déjà contribué à Wikipédia (et donc 1 897 n'ont jamais contribué). Ces 1 857 répondants, dont 487 ont indiqué avoir contribué souvent, correspondent à la population gardée pour l'étape 3.

Les variables explicatives

Les variables explicatives comprenaient des variables liées à l'individu interrogé à savoir, tout d'abord, des variables socio-démographiques (sexe, âge, diplôme). Ensuite, une série de variables comprenaient les compétences perçues dans le numérique, le fait d'utiliser un ordinateur, de réaliser des documents complexes « professionnellement » (dans son travail ou ses études), ou dans ses activités privées. Puis, d'autres variables explicatives concernaient le temps libre disponible, le fait d'avoir des activités culturelles, sportives, etc. ou encore des activités pro-sociales (comme donner son sang, avoir fait des dons à des associations, à Wikipédia, etc.). Plus loin, des variables étaient liées à l'environnement social de l'individu interrogé : comment est perçue Wikipédia (est-ce qu'il y a au moins un contributeur dans l'entourage ?). Et enfin, l'enquête comprenait des variables liées à la pratique de Wikipédia du répondant : le fait d'utiliser Wikipédia (professionnellement et personnellement), ou les caractéristiques de la première contribution (pour ceux qui ont déjà contribué), ou les raisons pour lesquelles on ne contribue pas plus. Ces dernières variables permettent de souligner que

si certaines variables sont utilisées à chaque étape (comme les variables socio-démographiques), d'autres ne le sont que pour ceux qui ont déjà contribué, par exemple et servent à expliquer le fait de contribuer régulièrement, ou non (étape 3 du modèle).

Première étape : savoir comment contribuer

Ce premier modèle vise à rendre compte de l'existence d'une compétence : savoir comment on contribue à Wikipédia, compétence dont le répondant va se sentir autorisé à faire état dans un questionnaire. La variable mobilisée prend comme valeur « non » si les gens ont répondu « non et vous ne savez pas comment on fait » (4 073 personnes) et « oui » sinon (3 754 personnes) – voir la première ligne du tableau ci-dessus.

Trois ensembles d'effets sont visibles pour cette première étape.

Le premier ensemble d'effet désigne un profil sociodémographique. Les répondants déclarant savoir contribuer sont – de façon significativement plus fréquente – des hommes, jeunes adultes (toutes les tranches au-dessus de 35 ans ont une probabilité significativement forte de moins savoir comment on contribue que la tranche 18-34 ans). Il est ici notable que le niveau de diplôme ne soit pas significativement corrélé, toutes choses égales par ailleurs, au fait de savoir comment contribuer.

Le deuxième ensemble d'effet nous montre que savoir comment contribuer va avec une série de pratiques sociales. Ainsi les répondants déclarant savoir contribuer sont aussi ceux qui utilisent souvent un ordinateur (au travail ou chez eux) et ceux qui possèdent des compétences numériques fortes.

Le troisième et dernier ensemble est celui des facteurs plus interactionnels. D'abord, connaître au moins un contributeur est significativement et positivement lié au fait de savoir contribuer, comme le fait de discuter avec l'environnement de Wikipédia (surtout si la vision qu'a cet environnement du projet est controversée).

Ces trois ensembles d'effets dessinent un portrait sociologique : celui d'hommes jeunes adultes appartenant à un milieu social où Wikipédia est un sujet de discussion et qui sont familiers des usages numériques. L'idée de l'appartenance à un milieu social spécifique est confortée par l'idée que ces répondants utilisent régulièrement un ordinateur. Tout ceci suggère que la connaissance de la manière de contribuer est diffusée dans un milieu social particulier, où l'encyclopédie fait partie du paysage de l'information, plutôt que d'être apprise par l'enseignement à l'ensemble de la population. À cet égard, l'effet négatif du genre nous place ici dans une longue liste d'études qui ont montré une tendance à l'exclusion des femmes des usages les plus techniques du numérique.

Deuxième étape : avoir contribué au moins une fois

Parmi ceux qui déclarent savoir contribuer, qui sont les répondants qui contribuent au moins une fois à Wikipédia ? Il faut bien comprendre que notre nouvelle population est constituée uniquement des personnes ayant répondu « oui » à l'étape 1, soit 3 754 personnes. Nous construisons une nouvelle variable « avoir contribué » qui prend comme valeur « non » si les gens ont répondu « Non, mais vous savez comment on fait » à la question de la contribution (1 897 personnes) et « oui » sinon (1 857 personnes).

Là aussi, trois ensembles d'effets se dégagent.

Le premier ensemble dessine un profil sociodémographique légèrement différent de celui perceptible en modélisant le franchissement de la première étape. Les femmes sont là encore significativement moins susceptibles de contribuer. Le niveau de diplôme est, contrairement à la première étape, discriminant selon une corrélation positive : plus le répondant est diplômé, plus la probabilité de s'être essayé à la contribution est élevée. En revanche, l'âge n'a pas d'effet significatif alors même qu'il s'agit généralement d'une variable traditionnellement très discriminante au regard des usages numériques. On voit ici l'intérêt de notre modèle en plusieurs étapes. Les jeunes adultes savent plus que les autres tranches d'âge comment on contribue à Wikipédia – c'est ce que nous avons modélisé à l'étape 1. Mais parmi ceux qui savent, tous les âges ont une même probabilité de contribuer.

Un deuxième ensemble d'effets renvoie aux pratiques qui vont avec la contribution à Wikipédia. Parmi les activités sociales que nous avons évaluées, ce sont celles qui sont liées au numérique qui se trouvent positivement corrélées avec le fait de contribuer au moins une fois : le fait d'avoir commenté des articles d'actualité en ligne et le fait d'avoir une expérience d'administrateur ou de webmaster... Ou d'avoir donné de l'argent à la Wikimedia Foundation.

Un troisième ensemble d'effets se dessine en regroupant les coefficients liés aux pratiques culturelles, aux obstacles à la contribution et à l'importance de Wikipédia. Les répondants ayant contribué au moins une fois se distingue par des pratiques culturelles moins fréquentes que les autres et une importance plus forte accordée à Wikipédia. Si l'on pense au fait que ce sont également des individus fortement diplômés, il semble que Wikipédia joue ici le rôle de la principale activité culturelle.

La deuxième étape (avoir contribué, parmi ceux qui savent) semble indiquer que le fait de contribuer au moins une fois est très lié avec le fait de s'en sentir légitime, tant sur les compétences (numériques), que sur le fait d'avoir des connaissances à apporter.

Troisième étape : contribuer beaucoup

Parmi ceux qui déclarent avoir contribué au moins une fois, qui sont les répondants qui contribuent régulièrement à Wikipédia ? Notre troisième population est donc constituée uniquement des personnes ayant répondu « oui » à l'étape 2, soit 1 857 personnes. Nous construisons une nouvelle variable « avoir contribué beaucoup » qui prend comme valeur « non » si les gens ont répondu « Oui, rarement » ou « Oui, quelque fois » à la question de la contribution (1 370 personnes) et « oui » s'ils ont répondu « Oui, souvent » (487 personnes).

La troisième étape permet de distinguer deux ensembles d'effets.

Le premier ensemble montre que la contribution régulière est un choix lié au contexte dans lequel le néo-contributeur évolue et relève de caractéristiques idiosyncratiques, plus que d'appartenance des grands habitus sociaux.

D'abord, les contributeurs réguliers ne se demandent pas si leur manque de compétences ou le fait qu'ils aient autre chose à faire est un problème ; la question elle-même n'est pas pertinente, comme le montre le fait que les gros contributeurs invoquent très souvent le manque de temps libre. Si l'on ajoute à cela l'effet négatif de penser que contribuer n'est pas une activité qui prend du temps, il devient clair que la probabilité d'être un grand contributeur dépend du degré d'ancrage de la pratique de la contribution dans un contexte social, mais est liée à la situation personnelle plutôt qu'aux caractéristiques démographiques : le fait de connaître au moins un contributeur a un effet positif, tandis que le sexe, le diplôme ou l'âge n'ont pas d'effet.

Le second ensemble regroupe des effets liés à Wikipédia. Il est d'abord manifeste que deux effets positifs divergeant soient discernables pour la variable « connaître un autre contributeur ». Il est significativement plus probable qu'on contribue beaucoup si on sait ne connaître aucun contributeur ou connaître au moins un contributeur, par rapport à une modalité de référence qui est de ne pas savoir qui parmi ses proches est contributeur. De même, le fait de ne pas connaître l'année où l'on a contribué pour la première fois a un effet négatif significatif sur la probabilité d'être un gros contributeur.

Autrement dit, contribuer beaucoup est lié à une sociabilisation individuelle (on sait identifier si les personnes qu'on connaît sont contributrices ou non) et à l'attachement à des épisodes marquants de cette sociabilisation (on se souvient de quand ça a commencé), plutôt que l'accès à une information véhiculé par la société ou les groupes sociaux auxquels on appartient.

Conclusion

Notre modélisation permet de mesurer avec plus de précision l'effet du profil social sur la contribution. On peut être surpris, par exemple, par la disparition de l'importance du genre dans la dernière étape. Celle-ci doit être comprise en relation avec ce que nous avons dit des deux premières étapes.

Cette enquête ne nie pas l'effet du genre sur les trajectoires de contribution, mais il montre que cet effet agit comme une forte barrière à l'entrée dans les deux premières « étapes » : par rapport aux hommes, les femmes ont moins de chances d'appartenir aux cercles sociaux où se diffuse l'apport de connaissances, et même si elles appartiennent à ce milieu, elles ont moins de chances de franchir au moins une fois le pas de la contribution. Un résultat similaire peut être observé pour l'âge, qui a un effet très concentré sur la première étape. L'âge est surtout un facteur de connaissance du fonctionnement de Wikipédia, mais peu un facteur de pratiques contributives. Bien entendu, cela doit être compris dans le contexte où le fait de savoir comment contribuer est une condition a priori pour contribuer.

Notre stratégie de modélisation nous a permis de décomposer l'effet bien connu des variables sociodémographiques sur les pratiques de contribution et de formuler des recommandations plus précises pour diversifier le profil des wikipédiens. Les trois types d'actions dépendent du public cible : il s'agit soit d'apprendre aux gens à contribuer, soit de les encourager à contribuer au moins une fois, soit de les encourager à contribuer davantage. Par exemple, pour toucher les personnes âgées, il faut d'abord leur expliquer comment faire. Une fois qu'elles savent, elles se sentiront probablement justifiées de contribuer (souvent). Les jeunes, en revanche, savent généralement comment contribuer mais ne franchissent pas le pas, et c'est là qu'un effet de levier est nécessaire. Une intervention destinée à un public féminin devra combiner deux leviers : montrer comment contribuer et tenir un discours d'encouragement - dans lequel l'activisme peut jouer un rôle, comme le montre l'expérience du projet « [Women in red](#) ».

Références

Bear JB, Collier B (2016) Where are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia. *Sex Roles* 74(5–1): 254–265.

Cruciani C, Joubert L, Jullien N, Mell L, Piccione S et Vermeirsche J (2023) Surveying wikipedians: a dataset of users and contributors' practices on wikipedia in 8 languages. arXiv preprint arXiv:2311.07964 URL <https://hal.science/hal-04279803>.

Hargittai, E., & Shaw, A. (2015). Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, communication & society*, 18(4), 424-442.

Lockett, A. (2020). Why do I have authority to edit the page? The politics of user agency and participation on Wikipedia. *Wikipedia @ 20* Reagle and Koerner (edit.), MIT Press: 205-220.

Proulx, S. (2002). Trajectoires d'usages des technologies de communication : les formes d'appropriation d'une culture numérique comme enjeu d'une « société du savoir ». *Ann. des Télécommunications*, 57(3-4), 180-189.

Warncke-Wang, Morten, et al. "Increasing Participation in Peer Production Communities with the Newcomer Homepage." *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW2 (2023): 1-26.