

SOCWIK - Devenir contributeur sur Wikipédia

<i>Main advisor</i>	Léo Joubert	Sociologue Maître de conférences	Université de Rouen-Normandie Dysolab (UR 7476)
<i>Advisor</i>	Emilie Bourlier-Bargues	Sciences de gestion Professeure	Clermont School of Business CleRMA (UR 3849)
<i>Candidate</i>	Cécilia Calheiros	Sociologue Docteure	Université de Rouen-Normandie Dysolab (UR 7476)

Contexte, objectif et originalité de la proposition : construire des métriques de la socialisation

eNSEMBLE a placé parmi ses 5 priorités structurantes le soutien à la création de collectifs sains et durables. Aligné avec cette priorité, ce projet post-doctoral cherchera à construire des métriques d'analyse de la socialisation aux communs numériques à travers l'étude du cas Wikipédia. Au sein du PEPR, le projet ciblé CONGRATS vise à comprendre la collaboration à grande échelle, dont Wikipédia est un exemple devenu classique. Dans quelle mesure Wikipédia peut-elle garantir la mobilisation continue de ses contributeurs (Joubert 2019) ? Quels sont les mécanismes sociaux qui permettent le maintien de la qualité globale du corpus ? Comment rendre compte de l'effet de la participation sur les trajectoires sociales des contributeurs ?

L'originalité de notre projet consistera à apporter une réponse à ces questions par une analyse de la socialisation des contributeurs, par une étude des différentes étapes de l'apprentissage des normes et des valeurs propres à Wikipédia. L'idée directrice est qu'il existe un lien entre la santé d'un commun numérique (Jullien et Roudaut 2020) et les modalités de socialisation (Bourlier-Bargues 2019; Dubar 2015). Cela implique notamment de chercher à mesurer la santé du commun non pas sous la forme d'une série de métriques calculées au niveau global de la plateforme Wikipédia, mais plutôt d'une série de métriques à hauteur des acteurs.

Défi et hypothèses : caractériser la santé d'un commun numérique par des métriques de la socialisation

Notre principal défi sera la mise au point de métriques de la socialisation. Ces métriques devront d'abord être 1) fiables : leurs variations devront refléter des variations entre différentes manières de se socialiser à Wikipédia ; 2) validés statistiquement, notamment au regard de leur robustesse ; et 3) calculable efficacement à partir des données de Wikipédia. Pour construire ces métriques, nous mettrons à l'épreuve les hypothèses suivantes, portant sur la socialisation des contributeurs à Wikipédia :

- *H1 : la socialisation à Wikipédia dépend de ce que l'acteur est venu faire dans le commun.* Participer à l'écriture collective d'un texte peut prendre de multiples formes allant de la correction orthographique à la rédaction substantielle de longs paragraphes. La santé d'un commun numérique est alors d'autant plus importante qu'il est en mesure de proposer à ses membres une diversité des formes d'engagement possibles (Welser et al. 2011).
- *H2 : la participation de manière durable au fonctionnement de Wikipédia dépend de la façon dont l'acteur est intégré dans le commun.* Les raisons pour lesquelles les contributeurs restent ne sont pas nécessairement celles pour lesquelles ils commencent à contribuer (Azary et al. 2017). L'engagement sur le temps long dans le fonctionnement de Wikipédia peut s'expliquer par des motifs qui débordent le rapport à la connaissance, avec des motivations relationnelles et des enjeux de reconnaissance par exemple.
- *H3 : la socialisation à Wikipédia dépend de la manière dont l'acteur va apprendre à utiliser les sources.* La rédaction des textes wikipédiens, du fait de la prohibition du travail inédit (Reagle 2010) qu'il oblige à se référer à des sources, et promeut donc un rapport à la connaissance singulier. On peut alors raisonnablement supposer que la santé du commun numérique réside

dans la légitimité, aux yeux des contributeurs, de ce modèle épistémique. Par exemple, des controverses wikipédiennes intenses ont porté sur la question de la légitimité d'un journal donné comme source acceptable.

- *H4 : le conflit autour de la rédaction des pages est socialisateur.* Un conflit avec un autre contributeur n'est pas toujours un vecteur de découragement, mais peut paradoxalement s'avérer socialisateur. En effet, la participation à une controverse suppose soit une activité d'apprentissage de la manière la plus légitime de se faire entendre, soit un effort pour développer une posture de contestation de l'ordre wikipédien. La santé d'un commun dépend donc, ici, non pas uniquement de sa capacité à juguler le conflit, mais aussi, lorsque celui-ci éclate, de la possibilité d'en faire un moment de socialisation des contributeurs.

Ces quatre hypothèses renvoient à différentes dimensions du processus de socialisation, selon que nous le considérons comme le produit de motivations (H1), de reconnaissance au sein d'un collectif (H2), d'un rapport à la connaissance (H3) ou encore de la résultante d'un système d'interactions (H4).

Approche méthodologique et livrable : un dispositif quali-quantitatif (Venturini et al. 2017) pour concevoir des métriques de la socialisation

Étudier les processus de socialisation demande d'interroger qualitativement les contributeurs de façon à saisir les pratiques et les représentations de la contribution dans leur dynamique. Nous utiliserons deux points d'entrée complémentaires pour se donner les moyens de capter la population d'étude:

- Le premier point d'entrée se fera lors d'observations participantes d'ateliers de contributions, événements visant à enrichir l'encyclopédie, organisés par des contributeurs chevronnés autour de thèmes (les biographies de femmes scientifiques, de personnalités noires, les villes et villages d'une région données, les différents types de fromages, etc.). Ce point d'entrée présente l'avantage de rendre très facile le contact avec les contributeurs, mais nous oriente de façon exclusive vers un type de contributeur qui peut ne pas être complètement représentatif de la totalité de la communauté.
- Le second est l'envoi de messages de sollicitation sur les espaces de discussion de Wikipédia. Ce point d'entrée nous permet de cibler précautionneusement le type de contributeurs que l'on interroge, mais le taux de non-réponse à ce type de message est souvent très haut.

Ces points d'entrée seront parcourus en demandant à chaque contributeur rencontré de nous recommander quelqu'un dont il sait qu'il est actuellement un contributeur ou qu'il a déjà contribué activement. Une fois la population constituée, nous réaliserons des entretiens semi-directifs informés par des données quantitatives grâce à un dispositif méthodologique à trois instruments ayant déjà été éprouvé par l'une des enquêtes du coordinateur (Bidart et al. 2024) :

1. Un tableau de bord (*dashboard*) nous permettra de produire une fiche d'informations précises sur la trajectoire d'un contributeur donné. Cela est rendu possible grâce à l'API de MediaWiki ainsi que par la fouille des *dumps* de Wikipédia. Ce tableau de bord est déjà prototypé par le *main advisor* du projet et fera l'objet d'une mise en production dans les premiers temps du projet (cf. partie "calendrier"). La fiche d'information produite par le tableau de bord constitue la première version de la liste des métriques de la socialisation ;
2. Des entretiens semi-directifs (N approx. 30) dont les questions seront orientées par les informations de la fiche permettront d'identifier les étapes de leur socialisation afin d'identifier les différentes étapes des processus de socialisation et d'en distinguer les différents types. La réalisation et l'analyse des entretiens permettra éventuellement d'ajuster le tableau de bord pour aboutir à une deuxième liste consolidée des métriques de la socialisation ;
3. Des entretiens confirmatoires (N. approx. 10), réalisés en fin de projet, qui nous permettront de prendre la mesure des limites de notre analyse, pour aboutir à la version finale des métriques de la socialisation.

Le nombre d'entretiens est pour le moment donné approximativement, car le nombre exact dépend du moment où nous serons parvenus à saturation ; c'est-à-dire, du moment où de nouveaux entretiens ne feront que consolider les apports des précédents au regard de l'analyse des processus de socialisation.

Le tableau de bord, dans sa version finale, prendra la forme d'un package de R permettant, à partir de logs de Wikipédia, de reconstruire les métriques pour un contributeur donné. Nous réaliserons une publication dans une revue de sociologie généraliste et une publication dans une revue spécialisée, valorisant notre dispositif méthodologique et les résultats du projet à l'échelle de la population d'étude.

Calendrier

Le déroulement du projet est pensé sur 2 années séparées en six phases tuilées :

1. **Revue de la littérature (M1 - M4).** La candidate étant spécialiste de sociologie des sciences et des techniques, il sera nécessaire d'avoir un temps d'appropriation de la littérature spécialisée sur Wikipédia et en théorie des organisations. Les deux coordinateurs du post-doctorat feront un travail de guidage, étant spécialiste du numérique et de Wikipédia (Léo Joubert) et de théorie des organisations (Emilie Bourlier-Bargues).
2. **Construction du tableau de bord et de la grille d'entretien (M2 - M6).** Il s'agira de construire le tableau de bord (réalisé par Léo Joubert) et la grille adaptative (réalisée par les trois membres de l'équipe) ;
3. **Réalisation des entretiens (M7 - M12).** Principalement par la candidate, les coordinateurs pourront venir en appui.
4. **Analyse des données (M10 - M17).** Réalisée conjointement, cette analyse comprendra le traitement des entretiens grâce à une analyse de contenu et le retour réflexif sur la construction des métriques.
5. **Réalisation des entretiens confirmatoires (M18 - M20).** A ce stade, nous ajusterons les métriques avec des entretiens plus directifs visant à éclaircir des points qui seraient apparus problématiques durant la première campagne d'entretiens.
6. **Perspectives de suite (M21 - M24).** Ces trois derniers mois seront consacrés à la rédaction du rapport de recherche grâce au rassemblement des différents documents de travail produits tout au long du projet. Ce rapport contiendra, en plus de nos résultats, des pistes à destination de CONGRATS pour un travail sur les métriques de la socialisation au-delà de Wikipédia.

Bibliographie

- Azary, Ofer, Lifshitz-Assaf Hila, Oded Nov, Johannes Daxenberger, Martina Balestra, et Coye Cheshire. 2017. « On the "How" and "Why" of Emergent Role Behaviors in Wikipedia ». P. 1066-81 in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. New York: Association for Computing Machinery.
- Bidart, Claire, Lorraine Bozouls, Léo Joubert, Elisa Klüger, Patrick Pérez, et Ingrid Tucci. 2024. « Imbriquer les méthodes pour comprendre la dégradation des relations interpersonnelles en temps de crise ». *Revue française de sociologie* 65(1-2):91-128. doi:10.3917/rfs.651.0091.
- Bourlier-Bargues, Emilie. 2019. « Socialiser les nouveaux entrants dans l'organisation ; des tactiques au travail institutionnel ». Habilitation à diriger des recherches en sciences de gestion, Université Clermont-Auvergne, Université Clermont-Auvergne.
- Dubar, Claude. 2015. *La socialisation*. Paris: Armand Colin.
- Joubert, Léo. 2019. « Le parfait wikipédien. Réglementation de l'écriture et engagement des novices dans un commun de la connaissance (2000-2018) ». *Le Mouvement Social* n° 268(3):45-60.
- Jullien, Nicolas, et Karine Roudaut. 2020. « Commun numérique de connaissance : définition et conditions d'existence ». *Innovations - Revue d'économie et de management de l'innovation* 2020/3(63):69 à 93. doi:10.3917/inno.063.0069.
- Reagle, Joseph Michael. 2010. *Good faith collaboration: The culture of Wikipedia*. Cambridge: MIT Press.
- Venturini, Tommaso, Mathieu Jacomy, Axel Meunier, et Bruno Latour. 2017. « An Unexpected Journey: A Few Lessons from Sciences Po Médialab's Experience ». *Big Data & Society* 4(2). doi:10.1177/2053951717720949.
- Welser, Howard T., Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, et Marc Smith. 2011. « Finding social roles in Wikipedia ». P. 122-29 in *Proceedings of the 2011 iConference*. New York: Association for Computing Machinery.