



**HAL**  
open science

# Comment le travail de surveillance de Wikipédia s'organise-t-il ?

Léo Joubert

► **To cite this version:**

| Léo Joubert. Comment le travail de surveillance de Wikipédia s'organise-t-il?. 2023. hal-04240207

**HAL Id: hal-04240207**

**<https://hal.science/hal-04240207v1>**

Preprint submitted on 13 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comment le travail de surveillance de Wikipédia s'organise-t-il ?

***Léo Joubert***

*DySoLab, Université de Rouen*

*LEST, Aix-Marseille Université*

**Résumé.** Ce document de travail présente un projet de recherche en cours sur l'organisation de la coécriture de Wikipédia. Tout d'abord, je rappelle les résultats de ma thèse de doctorat sur le wiki et les trajectoires des wikipédiens pour établir un schéma à deux niveaux, macro et micro. Le projet vise à compléter ce schéma en proposant une approche par le niveau « méso » des pages. Dans cette optique, j'adopte une approche basée sur la détection de « formes » sociologiquement pertinentes et j'explore comment l'analyse de réseau peut contribuer à l'analyse statistique classique sur la coécriture des notices. Enfin, je propose trois pistes pour approfondir les résultats de cette étude, qui restent encore très ouverts.

# Table des matières

<b><u>Les « communs numériques de masse » comme configuration sociologique .....</u></b>	<b><u>3</u></b>
<i>Modèle de carrière.....</i>	<i>4</i>
<i>Structure bicéphale des communs de masse .....</i>	<i>5</i>
<b><u>Analyse des réseaux de collaborations .....</u></b>	<b><u>7</u></b>
<i>Définition des réseaux et du processus abductif de fouille .....</i>	<i>8</i>
<i>Première étape : première grammaire des formes.....</i>	<i>9</i>
<i>Deuxième étape : signature quantitative des formes.....</i>	<i>10</i>
<i>Troisième étape : détection sur l'ensemble du corpus.....</i>	<i>12</i>
<b><u>Pour ne pas conclure, des pistes d'approfondissement.....</u></b>	<b><u>15</u></b>
<i>Enrichir avec de nouvelles données pour analyser les processus sociaux en jeu .....</i>	<i>15</i>
<i>Une étude des controverses par la comparaison de sous-corpus.....</i>	<i>18</i>
<i>Introduire une couche sémantique pour comprendre les formes de coopération .....</i>	<i>21</i>
<b><u>Bibliographie .....</u></b>	<b><u>22</u></b>

## Les « communs numériques de masse » comme configuration sociologique

Le projet de recherche dont ce document vise à détailler les grandes lignes prend ses racines dans la soutenance d'une thèse de doctorat en sociologie intitulée « Wikipédia, une encyclopédie à l'ère du logiciel libre. Sociologie d'un commun numérique de masse » (Joubert, 2020). L'hypothèse de travail centrale de cette première recherche est de problématiser Wikipédia comme une configuration (Elias, 1993), c'est-à-dire comme un réseau d'interdépendances évolutives. Ce réseau est saisissable soit comme une unité, « Wikipédia », soit comme une pluralité, « les wikipédiens ». Le but était alors de comprendre quelles étaient les singularités du processus social d'institution des « communs numériques de masse » comme Wikipédia.

Pour étudier cette configuration, j'ai eu recours à un dispositif méthodologique mixte composé de 4 sources :

- 12 424 réponses à un questionnaire auprès des contributeurs et des lecteurs de l'encyclopédie réalisé par le GIS Marsouin durant le mois de mars 2015 avec un bandeau sur la page d'accueil de la version francophone de Wikipédia.
- 28 entretiens de type « récit de vie » avec des contributeurs chevronnés
- Dépouillement quantitatif des « *dumps* » pour mesurer la durée de vie des contributions
- Dépouillement qualitatif d'un corpus des pages de règles et de délibération du wiki, dans l'espace de nom « Wikipédia: ».

Chaque de ces sources avait vocation à décrire une partie de la configuration wikipédienne. L'enquête par questionnaire Wikipédia 2015 avait pour but de comparer entre elles des populations de lecteurs et de contributeurs. Les entretiens visaient à saisir l'articulation entre la pratique de contribution et les autres dimensions de la vie quotidienne. Le dépouillement des *dumps* visait à positionner les contributeurs dans l'espace social des rôles de wikipédiens. Enfin, le dépouillement des archives de pages de règles visait à situer l'évolution des différents enjeux institutionnels durant les deux décennies écoulées depuis la mise en ligne de Wikipédia.

L'analyse déployée dans la thèse a permis de dégager deux résultats globaux. Le premier, un modèle séquentiel de carrière, concerne les wikipédiens. Le second, un modèle formel, vise à appréhender l'évolution institutionnelle de Wikipédia.

## *Modèle de carrière<sup>1</sup>*

Pour rendre compte des trajectoires de wikipédiens, j'ai cherché à construire un enchaînement de quelques étapes par lesquelles tous les contributeurs passaient différemment. J'ai pu ainsi dégager trois étapes : le passage à l'acte, la décision de rester et la cristallisation.

La première étape consiste dans la réalisation des premières contributions. Comparativement aux autres utilisateurs de Wikipédia, ceux ayant réalisé au moins une contribution sont bien plus fréquemment de jeunes hommes fortement diplômés. Je retrouve en cela les résultats de nombreuses études, à ceci près que l'ensemble des variables sont ici mis en concurrence là où la plupart des travaux mettaient jusqu'à présent en lumière un seul facteur, comme le genre (Hill & Shaw, 2013), où une pluralité de facteurs à l'aide de méthode qualitative sur un corpus réduit de cas (Bryant et al., 2005).

Lors de la deuxième étape, l'effet des variables sociodémographiques s'ajuste : certaines ont un effet plus fort alors que d'autres vont voir leurs effets s'inverser. C'est notamment le cas du niveau de diplôme et du genre. Alors que les diplômés d'une licence passent plus fréquemment à l'acte que les autres, plus le niveau de diplôme est haut, plus l'acteur aura tendance à rester. À l'inverse, le genre voit son effet s'inverser : si les femmes passent moins à l'acte que les hommes, celles qui franchissent tout de même la première étape vont avoir significativement plus de chances de rester.

Les entretiens montrent pour cette seconde étape le début d'une construction narrative au sujet de Wikipédia. Le moment de la décision de rester est tantôt raconté comme un lent apprentissage des règles, tantôt comme l'occasion d'un conflit aussi virulent que fondateur avec un autre contributeur, tantôt comme la rencontre de contributeurs avec lesquels un dialogue émerge à propos de sujets perçus comme passionnants.

Lors de la troisième et dernière étape, la cristallisation, l'effet des variables sociodémographiques extrêmement fortes lors des précédentes étapes n'est plus significatif. Pour prendre les mêmes exemples, le titulaire d'une licence parvenu à ce stade contribue de la même manière que le titulaire d'un doctorat et l'on n'observe pas de différence statistique entre le rôle tenu par une wikipédienne et le rôle tenu par un wikipédien. Les entretiens, cette fois-ci, montrent l'installation d'une dépendance de sentier avec la décision de rester. Les acteurs, une fois wikipédiens, tendent à tenir le rôle pour

---

<sup>1</sup> J'entends ici le mot « modèle » au sens sociologique davantage qu'au sens statistique, notamment en lien avec les travaux d'Howard Becker (1985). Bien sûr, je n'oppose pas les deux sens du mot, et d'ailleurs des modèles de régression ont été utilisés pour éprouver la robustesse statistique des résultats sociologiques permettant de construire le modèle présenté ici.

lequel ils ont décidé de rester – qui n'est pas forcément conforme à ce qu'ils souhaitaient faire au moment du passage à l'acte.

Ce modèle « passage à l'acte – décision de rester – cristallisation » pourra servir comme cadre théorique pour l'analyse des trajectoires dans d'autres communs numériques de masse. Dans ce projet de recherche, je pourrai essayer de comparer la position des contributeurs au sein des collectifs de production de page à différentes étapes de leurs carrières. Cela en continuité d'études qui ont déjà montré que les tâches réalisées n'étaient pas exactement les mêmes au début qu'à la fin des parcours (Bryant et al., 2005; Welser et al., 2011) – sachant que les contributeurs les plus actifs sont actifs dès le début (Dejean & Jullien, 2015; Panciera et al., 2009).

### *Structure bicéphale des communs de masse*

Dans l'économie des communs proposée par Elinor Ostrom, l'unité politique centrale d'un bien commun est la communauté (« *community* »). Cette communauté ne doit pas être confondue avec la communauté de la sociologie classique, dans la mesure où elle repose davantage sur la structuration d'une action collective liée à l'administration d'une ressource (*common pool resource* chez Ostrom), alors que pour des auteurs comme Weber et surtout Tönnies la communauté procède de l'hésitation d'une position sociale au sein d'un groupe de pair fortement interconnecté et socialement proche. La communauté du « faire ensemble » des biens communs s'oppose en cela à la communauté de destin des sociologues (Laval, 2016).

Dans les deux cas cependant, les frontières de la communauté sont très claires. C'est même en substance la première recommandation que formule Elinor Ostrom dans ses *design principles* sur la construction de l'architecture institutionnelle des communs<sup>2</sup>. Dans le champ de la sociologie, Norbert Elias (1997) montrera combien le sentiment d'appartenance à une communauté va de pair avec une « logique de l'exclusion » qui crée une frontière là où parfois les conditions d'existence d'un groupe d'acteur peuvent donner à l'observateur pressé le sentiment de l'existence d'une même communauté uniforme.

La rigueur avec laquelle la frontière de la communauté est déterminée n'est cependant pas la même dans les « communs fonciers », comme les forêts, que dans le cas des « communs numériques », comme Wikipédia. Dans le cas des communs fonciers, la clôture de la communauté constitue une des pièces de l'argumentaire d'Ostrom pour répondre à Hardin et sa « tragédie des communs » (1968) - pour cet auteur, les biens communs sont voués à disparaître à cause de la surexploitation. La clôture de la communauté, au sein de laquelle on trouve un mécanisme de sanctions graduées et une

---

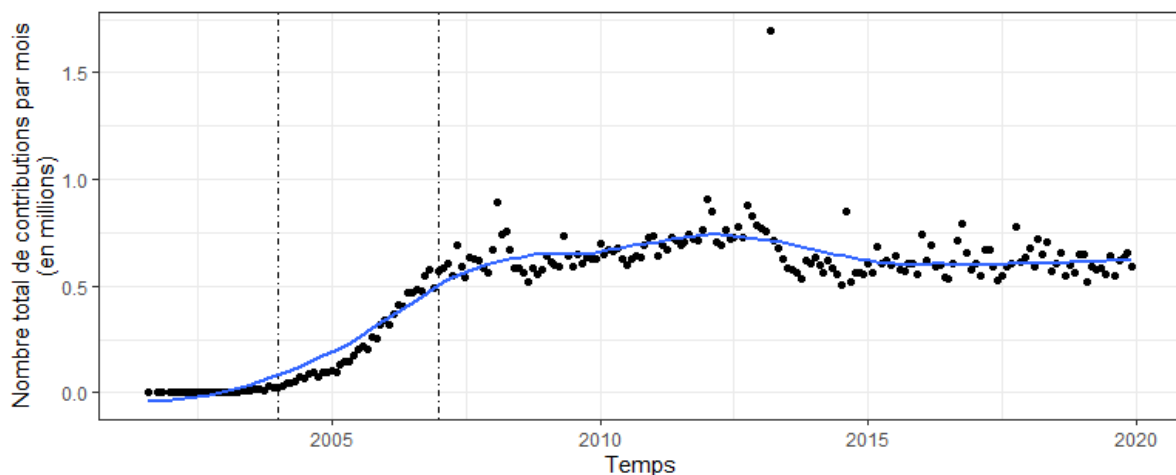
<sup>2</sup> « Les frontières de la communauté doivent être clairement définies » (*define clear group boundaries*)

délibération décentralisée, garantit l'allocation raisonnée de droits d'usage. Dans le cas d'une ressource non rivale comme peuvent l'être les notices de Wikipédia, la surexploitation ne pose évidemment pas le même problème, et dans le cas de Wikipédia l'usage de la ressource par des non-membres (les lecteurs) est même un but explicite des membres (les wikipédiens).

Les communs numériques invitent donc d'emblée leur public au sein de leur dynamique institutionnelle. Ce point permet d'ailleurs de nuancer la non-rivalité de la ressource. Si la ressource de Wikipédia est un bien non rival, la coprésence d'acteurs divers dans le commun numérique rend nécessaire une importante quantité de travail de régulation. Or, les ressources cognitives et matérielles nécessaires pour la réalisation de ce travail sont, quant à elles, bien rivales. Ce qui poussera Jullien et Roudaut (2020) à mettre en question la rivalité du commun wikipédien.

Le fait que la frontière soit poreuse peut donc avoir des effets considérables sur la dynamique normative d'un commun. Dans le cas de Wikipédia, le caractère massif du commun amplifie ces effets. Particulièrement brusque, la massification de la version française de Wikipédia a vu le nombre de contributions passer de quelques dizaines par mois à la fin de l'année 2003 (premier trait en pointillés dans la Figure 1) à un demi-million au début de l'année 2007 (deuxième trait en pointillés).

**Figure 1. Évolution du nombre de contributions par mois en fonction du temps sur la Wikipédia francophone.**



**Guide de lecture : en janvier 2008, 0,90 million de contributions ont été réalisées sur le Wikipédia francophone.**

Si la massification d'un commun fut déjà abordée sous l'angle technique – l'infrastructure technique est-elle robuste ? – et sous l'angle économique – l'équilibre entre investissement dans la contribution et la rétribution des contributeurs est-il stable ? – elle possède une dimension sociologique qui m'a semblé négligée. Pour combler ce manque, j'ai essayé de construire une réponse à la question suivante : comment la massification fit-elle évoluer la structure du lien social entre les wikipédiens ?

La massification a donné lieu à une dualisation de la structure institutionnelle de Wikipédia. D'abord, entre 2001 et 2004 s'institue une « communauté des contributeurs » en posant les premières règles cardinales de Wikipédia. C'est notamment durant cette période qu'émergeront les « principes fondateurs ». Ensuite, entre 2004 et 2007, cette communauté va se massifier et les règles vont se préciser pour encadrer la rédaction d'un corpus toujours plus vaste aux exigences toujours plus spécifiques. L'une des controverses les plus fortes, entre « inclusionnistes » et « suppressionnistes », va émerger à la fin de cette période. Enfin, entre 2008 et 2020<sup>3</sup>, va émerger une large « société des usagers » de Wikipédia. Cette société sera notamment objectivable par la survenue de controverses publiques au sujet de l'usage de Wikipédia, par exemple dans le secteur de l'éducation avec l'usage de Wikipédia en classe au lycée et à l'université.

Il reste à savoir quel effet concret cette structure a pu avoir dans la production des pages wikipédiennes. L'articulation entre communauté des contributeurs et société des usagers que la thèse a proposée en la plaçant au niveau du wiki gagne à se spécifier en fonction du contexte spécifique de chacune des pages. Par exemple, les pages liées à la Covid ont fait l'objet de processus spécifiques de production de connaissance en tant de crises (Bubendorff & Rizza, 2021), là où un texte paru dans une revue grand public a signalé les pages liées aux personnalités du mouvement antiraciste comme faisant l'objet d'un traitement spécifique (Assbague, 2022).

## Analyse des réseaux de collaborations

Appuyée sur la notion de configuration, la thèse pose les jalons d'une approche à mi-chemin entre structuralisme et constructivisme pour l'analyse des communs numériques de masse. Pour aller plus loin, ce projet de recherche vise à inspecter plus précisément comment se construit en permanence l'articulation des deux niveaux de réalité sociologique en jeu ici : le wiki et les trajectoires de contributeurs. Je cherche ici à m'inscrire dans une sociologie analytique du « coupage flou » (Goffman, 1983), partant donc du principe que l'articulation entre les niveaux ne se joue jamais à l'avance, mais au contraire se trouve toujours négociée en situation. La théorie des graphes, en ce qu'elle permet de se situer au niveau de la structure sans pour autant réifier de grands ensembles (White, 2011; Mercklé, 2016), constitue un artefact de méthode intéressant pour l'inspection détaillée de ce qui est ici le niveau « méso » de la page wikipédienne.

---

<sup>3</sup> Je stoppe la périodisation en janvier 2020, date à laquelle je considère avoir arrêté l'analyse du corpus et entamé la rédaction de la thèse. Cette dernière borne découle donc davantage de l'organisation de mon travail personnel que d'un véritable critère épistémologique.



Pour construire les graphes, j'ai fait le choix de construire les réseaux pour répondre à la question : qui efface quoi ? Ce choix découle de résultats quantitatifs obtenus par l'analyse des dumps de la version anglophone de Wikipédia, à savoir qu'une très petite élite de contributeurs (0,4 %) s'avère responsable d'une grande partie (40 %) du contenu durable (Priedhorsky et al., 2007). L'ampleur de ce travail d'effacement vient donc percuter l'idée d'un système wikipédien ouvert à la contribution du tout-venant. Au contraire, la contribution à Wikipédia suppose un long processus de socialisation (Aaltonen & Kallinikos, 2012; Bryant et al., 2005) pour pouvoir passer les fourches caudines de la « surveillance participative » (Cardon & Levrel, 2009). L'utilisation des graphes a vocation ici à modéliser cette surveillance participative telle qu'elle se réalise sur chacune des pages de l'encyclopédie.

### *Définition des réseaux et du processus abductif de fouille*

La représentation proposée aboutit à des graphes orientés pondérés adossés à une page donnée. Chaque contributeur est représenté par un nœud. Le poids du nœud est égal à la quantité d'éléments ajoutés sur la page. Par « éléments », j'entends l'ensemble des signes visibles à l'écran auxquels s'ajoutent les marqueurs du langage de balisage « wikicode » servant aux contributeurs pour mettre en forme leur contribution, en délimitant une citation, en créant un lien hypertexte, etc. Par exemple, si un contributeur ne change pas les mots d'une page, mais transforme ce qui était un simple mot en lien hypertexte, il est considéré comme ayant ajouté un élément sur la page.

Si un contributeur A a effacé du contenu déposé sur la page par un contributeur B, alors une arête unit les nœuds représentant A et B. Cette arête part de A pour aller vers B. Le poids de l'arête est proportionnel à la quantité de contenu effacé par A. À ce stade de mon travail, le poids est égal à la quantité de contenu supprimé, mais j'introduirai probablement une fonction pour normaliser ces poids. Cela notamment dans le cas où des statistiques inférentielles viendraient à être mobilisées.

Une fois l'outil technique de construction des réseaux achevé, j'applique un processus inductif de fouille pour aboutir à une première typologie de formes. Dans ce document, je présente une étude exploratoire pour prototyper la méthode sur des pages potentiellement intéressantes sur un thème singulier. Ce sont des pages liées à la politique identifiées par Jeanne Vermeirsche : *Emmanuel Macron, Marine Le Pen, Jean-Luc Mélenchon, Dora Moutot, Nationalisme en France, Les Identitaires, Extrême droite, Rassemblement national, Féminisme, Islamisme*. Le fait de respecter ce corpus de page pour un premier test de ma méthode permettra, lors de la journée du 14 avril, de confronter d'éventuels premiers résultats avec son approche plus experte sur la thématique en question. Ce fonctionnement est de plus en cohérence avec une approche de science collaborative sur laquelle je reviendrai au moment de dégager des pistes d'approfondissement.

Mon but ne sera pas d'identifier « ce qu'il y a dans les réseaux », comme s'il suffisait d'aller « voir dans les données » pour comprendre la « totalité » du processus de régulation de Wikipédia. Je chercherai plutôt à extraire l'information pertinente au vu d'une lecture sociologique du graphe. Pour cela, je vais essayer de repérer systématiquement des « formes », c'est-à-dire des sous-ensembles de nœuds et d'arrêtes correspondant vraisemblablement à des contextes sociaux particuliers.

Le processus de travail que je propose peut donc se résumer en quatre étapes :

1. **Détection inductive de formes.** À l'aide d'un outil de visualisation des réseaux, notamment en agrandissant l'image pour voir plus précisément le voisinage d'un point, je cherche des formes potentiellement intéressantes – c'est-à-dire des formes potentiellement liées à des interprétations sociologiques vraisemblables<sup>4</sup>.
2. **Définition de la signature quantitative de la forme.** À partir de l'observation, je définis un algorithme de détection de la forme à partir de ses marqueurs – par exemple un nœud plus central, une arête importante, etc.
3. **Détection sur l'ensemble du corpus.** Si la définition des formes est correcte, il devient possible de procéder à une détection sur l'ensemble du corpus. Les publications issues du projet s'accompagneront d'un *package* de R comprenant les fonctions utiles à la détection, avec un jeu de données de test.
4. **Interprétation sociologique.** Une fois la détection effectuée, je dénombre les normes sur chaque pages et je cherche à identifier qualitativement le contexte dans lequel elle s'encastre.

#### *Première étape : première grammaire des formes*

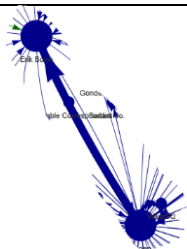

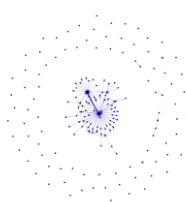
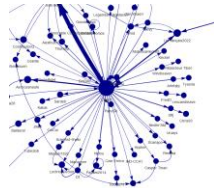

Le passage en revue du premier jeu de pages me donne l'occasion de construire une première typologie de formes récapitulée dans le tableau suivant. Chacune de ses formes est issue de l'observation du réseau d'une page, associée à l'observation liminaire des évolutions du texte de cette page. Ce double regard me permet de formuler une hypothèse vraisemblable à mettre à l'épreuve en analysant plus spécifiquement les réseaux de quelques pages pour lesquelles je dispose d'informations additionnelles<sup>5</sup>.

---

<sup>4</sup> Sans vouloir surcharger ce premier document de remarques épistémologiques, peut-être faut-il néanmoins préciser que la mise en œuvre de cette démarche relève d'une poursuite de « l'imagination sociologique » (Mills, 2006), que j'essaie ici de cultiver par la formulation tâtonnante d'hypothèses de recherche à partir d'un matériau qu'il faut prendre d'abord le risque de surinterpréter.

<sup>5</sup> Voir sur ce point la piste d'approfondissement « enrichir avec de nouvelles données » ci-dessous.

**Tableau 1. Inventaire de la première typologie de formes**

Nom de la forme	Image d'exemple	Hypothèse sociologique vraisemblable	Exemples de travaux à propos de cette forme de contribution
<i>Autoroute à double sens</i>		Conflit entre deux contributeurs importants et récurrents (« guerre d'édition »)	(Viégas et al., 2007; Yasseri et al., 2012, 2014)
<i>Dyade marginale</i>		Modification mineure à la marge de l'article (sans modification du plan, notamment), puis rectification à la marge également.	(Bryant et al., 2005; Halfaker et al., 2013)
<i>Composante centrale</i>		Petite équipe de contributeurs qui prend en main la rédaction d'un article, appuyée par une équipe de contributeurs marginaux aux ajouts systématiquement conservés.	(Anthony et al., 2007)
<i>Étoile</i>		Un seul contributeur cherche à s'appropriier la rédaction de la page.	(Halfaker et al., 2009)
<i>Étoile inversée</i>		L'autorité d'un contributeur central est très contestée.	(Kittur et al., 2007; Kostakis, 2010; Viégas et al., 2007)

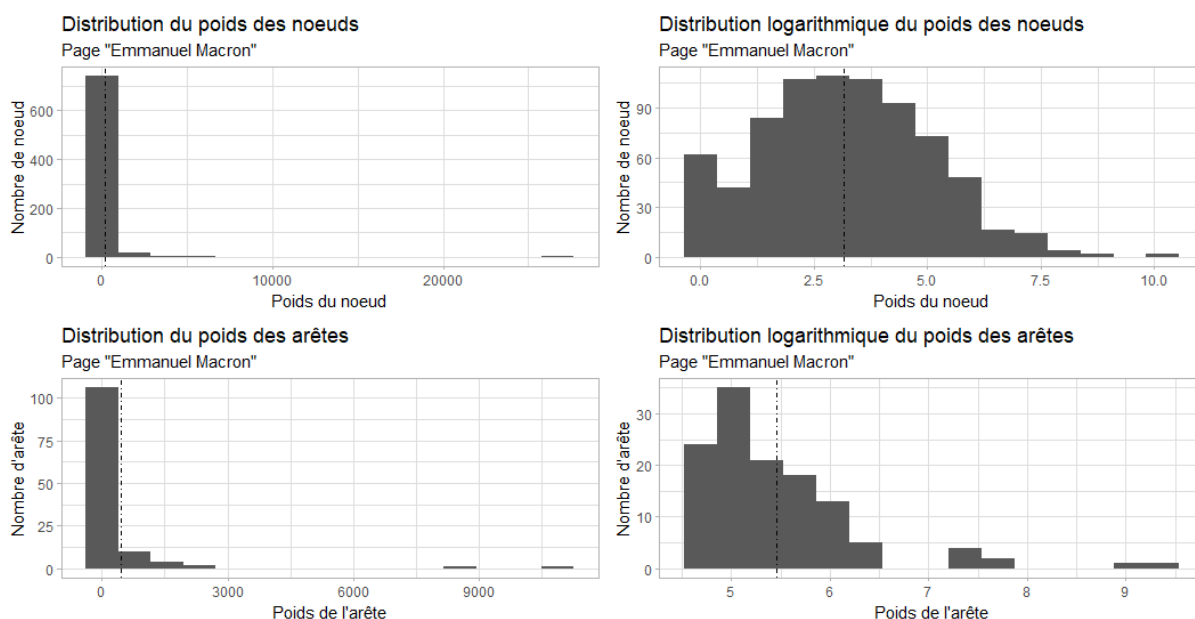
### *Deuxième étape : signature quantitative des formes*

La formulation précise de la signature quantitative des formes demande de s'atteler à un problème classique en analyse de réseaux, à savoir la définition d'un critère pour la force et la faiblesse d'un nœud et d'une arête<sup>6</sup>. Ayant défini une pondération aussi bien pour le poids des nœuds que pour celui des arêtes, je vais prendre appui sur les valeurs des poids. À ce stade du travail, j'ai choisi de me référer à un critère simple, celui de la médiane des deux distributions. Les poids des nœuds/arêtes inférieurs

<sup>6</sup> La formulation la plus célèbre de ce problème – et aussi la plus discutée – est celle de Mark Granovetter qui montrera « la force des liens faibles » sur le marché de l'emploi. Il est moins courant de restituer le contexte de ce travail, qui se situe dans une sociologie analytique foisonnante, dont le travail d'Harrison White (directeur de thèse de Granovetter) dans l'ouvrage *Identité et contrôle* constitue la matrice théorique de référence.

(respectivement, supérieurs) à la médiane des poids des nœuds/arêtes sont des nœuds/arêtes faibles (respectivement, forts).

**Figure 2. Distribution du poids des nœuds et des arêtes.**



**Guide de lecture :** le trait en pointillés délimite la frontière entre la force (partie à gauche des traits) et la faiblesse (partie à droite des traits) des nœuds et des arêtes.

La distinction entre nœuds/arêtes forts et faibles permet de définir analytiquement l'ensemble des formes proposées à l'étape précédente.

**Tableau 2. Formulation de la signature quantitative des formes**

Forme	Définition analytique
<i>Autoroute à double sens</i>	Deux nœuds forts liés par des arêtes fortes
<i>Dyade marginale</i>	Deux nœuds forts ou faibles liés entre eux par une arête forte ou faible, et tous les deux liés à aucun autre nœud fort ou faible
<i>Composante centrale</i> <sup>7</sup>	Deux indicateurs génériques : la densité de la composante et le nombre de nœuds isolés
<i>Étoile</i>	Un nœud fort lié par au moins 3 arêtes fortes ou faibles à des nœuds faibles
<i>Étoile inversée</i>	Un nœud fort ou faible vers qui pointent des arêtes fortes ou faibles

À titre indicatif, la capture d'écran ci-dessous montre le code utilisé pour la détection, dans l'ensemble du corpus, des autoroutes à double sens. Écrit dans le langage R, un travail reste à faire pour son passage à l'échelle et sa rédaction sans un style davantage conforme aux règles de l'art.

<sup>7</sup> Cette forme se mesure différemment de toutes les autres, dans la mesure où elle est présente dans tous les graphes à des degrés divers. Je retiens donc deux indicateurs qui me permettront de mesurer l'intensité de sa présence dans chaque graphe.

Figure 3. Algorithme de détection des autoroutes à double sens

```

77 autoroute <- fonction(edges_sup, nodes_sup) {
78   # Autoroute à double sens
79   ## Deux nœuds forts liés par des liens forts
80   res <- filter(edges_sup, from %in% nodes_sup, to %in% nodes_sup) %>% select(from, to)
81   sapply(1:length(res$from), function(i) {
82     test <- res[res$from == res$from[i], c("to", "from")]
83     names(test) <- c("from", "to")
84     invisible(inner_join(res, test, by = c("from", "to"))) %>% nrow()
85   }) %>% sum
86 }

```

Troisième étape : détection sur l'ensemble du corpus

Les éléments statistiques présentés ici sont exploratoires, dans la mesure où la sélection des pages n'est pas représentative de Wikipédia ou du sous-thème politique. En revanche, je les utilise comme une première mises à l'épreuve très intuitive de ma démarche : qu'apporte l'analyse de réseaux à la compréhension de l'organisation du travail de surveillance ?

Tableau 3. Tableau de contingence des formes

	Nombre de nœuds	Poids moyen des arêtes	Autoroutes	Étoiles	Étoiles inversées	Points isolés	Densité de la composante centrale	Dyades marginales
<i>Emmanuel Macron</i>	763	449	32	3	5	661	0,0110	4
<i>Marine Le Pen</i>	690	276	14	3	3	630	0,0160	8
<i>Jean-Luc Mélenchon</i>	771	202	46	10	10	689	0,0180	6
<i>Dora Moutot</i>	42	183	0	1	0	33	0,1390	0
<i>Nationalisme en France</i>	109	125	0	0	0	107	0,5000	1
<i>Les Identitaires</i>	205	213	13	3	1	183	0,0560	3
<i>Rassemblement national</i>	816	254	21	9	9	714	0,0120	8
<i>Féminisme</i>	546	286	0	2	2	508	0,0210	9
<i>Islamisme</i>	219	271	0	1	0	201	0,0590	1

*Lecture : plus la couleur se rapproche de la teinte rouge (respectivement : verte), plus la valeur est faible (respectivement : forte) au regard des autres cases de la colonne.*

Un premier regard sur les couleurs du tableau laisse entrevoir une corrélation positive entre les différents indicateurs. Une case rouge a plus de chances de se trouver sur la même ligne que d'autres cases rouges alors qu'une case verte se trouvera elle aussi plus probablement accompagnée par d'autres cases vertes sur la même ligne. Ce résultat admet une exception notable : la densité de la composante centrale semble quant à elle négativement corrélée avec l'ensemble des autres indicateurs.

Cependant, ce lien n'explique pas l'ensemble de la variation identifiable à l'aide de ce tableau. Par exemple, les pages de Marine Le Pen et de Jean-Luc Mélenchon sont toutes deux très pourvues en contributeurs puisque leurs réseaux comptent respectivement 690 et 771 nœuds. En revanche, ces deux pages très comparables thématiquement ne contiennent pas du tout le même nombre d'étoiles ou de points isolés (alors même que la densité de la composante centrale est remarquablement stable).

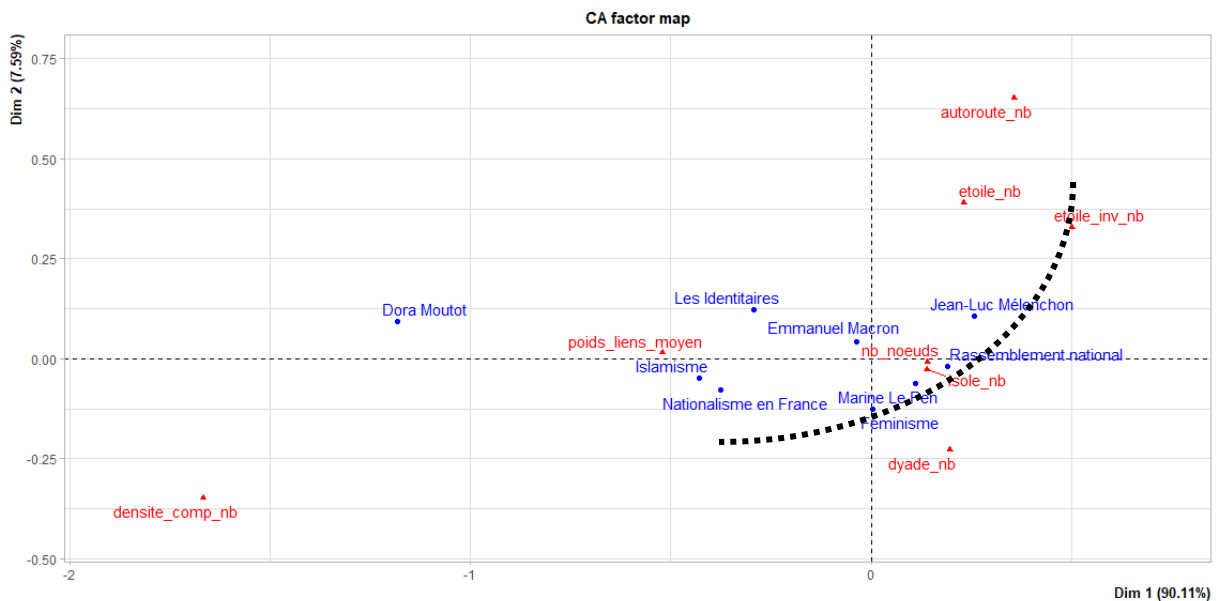
Cet exemple est un indice de l'apport de la démarche d'analyse de réseau. D'abord parce que les deux pages sont très comparables thématiquement : deux candidat.es à l'élection présidentielle de 2012, 2017 et 2022 à la tête de partis qui disposent d'importantes communautés militantes en ligne. Sur Wikipédia, les deux articles fédèrent à peu près le même nombre de contributeurs<sup>8</sup>. En revanche, l'analyse de réseau et notamment le dénombrement des étoiles les différencient fortement. Cela me permet de penser, sans pouvoir en avoir la confirmation statistique à ce stade, que l'analyse de réseau rend perceptible statistiquement des différences imperceptibles avec des indicateurs plus simples.

Au-delà de cette lecture spontanée, je réalise dans ce qui suit une analyse en composante principale sur ce même tableau pour pouvoir en isoler les points de variation les plus intéressants.

---

<sup>8</sup> Je n'utilise ici que le nombre de contributeurs, qui s'avère être un indicateur synthétique de la « qualité » des contenus produit par les collectifs de wikipédiens attachés à chaque page (Hu et al., 2007; Kittur et al., 2007; Wilkinson & Huberman, 2007; Halfaker et al., 2009). Il faudra cependant aller plus loin en identifiant d'autres indicateurs « naïfs » pour voir dans quelle mesure l'analyse de réseau apporte des finesses supplémentaires.

Figure 4. Analyse en composantes principales sur le tableau de contingence des formes.



Ce plan factoriel, expliquant  $90.11 + 7.95 = 98.05\%$  de la variance observable dans le Tableau 3, montre un « effet Gutman »<sup>9</sup> signifiant que les deux axes doivent être interprétés ensemble (Cibois, 2014). La courbure de la parabole est représentée par le trait en pointillés que j’ai représenté à la main sur ce graphe, à de seules fins de lisibilité. Lorsque le corpus de page sera stabilisé, je calculerai une courbe de tendance algébriquement pour mieux asseoir l’interprétation.

Le long de cette parabole semble ainsi dessiner un gradient de la massification. Les deux extrêmes sont occupés par les pages de Dora Moutot, non massifiée, et de Jean-Luc Mélenchon, massifiée. Les pages les moins massifiées semblent avoir des réseaux à la une composante centrale très dense et des arêtes ayant un poids moyen fort. À l’inverse, les pages massifiées voient le travail de surveillance se paralléliser en mobilisant un petit groupe relativement peu conflictuel et plusieurs petites unités séparées entre elles. Chacune de ces unités efface peu le contenu déposé par les autres. C’est ainsi que l’on peut notamment comprendre l’opposition diamétrale entre des réseaux avec une composante centrale très dense et des réseaux avec un fort nombre d’autoroutes, d’étoiles et d’étoiles inversées.

Il faut évidemment conserver ici la plus grande prudence, dans la mesure où une corrélation comme celle-ci appelle nécessairement sa généralisation sur le plus grand nombre d’unités d’analyse.

<sup>9</sup> Observable lorsque les points dessinent une parabole sur le plan.

Cependant, ce premier dépouillement laisse augurer d'une plus fine compréhension des processus de coécriture en donnant à voir une différenciation par la massification « locale » analogue à ce qui se produit de façon « globale » dans le wiki.

Cette analyse montre que les graphes semblent pouvoir permettre une approche multiniveau de la massification de Wikipédia. Cela notamment en permettant de construire un gradient de massification qui ne se limite pas à la mesure du nombre de contributeurs impliqués, mais prend aussi en compte la structure fine du collectif de contributeur. Ces réseaux ouvrent ainsi la possibilité d'une analyse de « modes de massification » des collectifs adossés aux pages. Par exemple, certaines pages comme celle des Identitaires semble se massifier avec un poids des arêtes moyen plus fort que les autres, mais un nombre d'arêtes plus faible – ce qui serait le signe d'une conflictualité davantage concentré entre quelques contributeurs.

## Pour ne pas conclure, des pistes d'approfondissement

La quatrième étape de ce travail peut pour le moment n'être présentée que sous la forme d'une série de pistes d'approfondissement. Parce que l'interprétation repose toujours, sauf à être une herméneutique gratuite, sur l'apport d'un élément supplémentaire pour « faire sens », je ne peux pour le moment dégager que des pistes qui permettront peut-être de conclure la première phase d'abduction de ce travail<sup>10</sup>.

Les trois pistes que je présente ont commun d'être rapidement explorables en l'état, mais les assumer complètement posent des problèmes pour le moment insurmontables compte tenu de l'état des outils, des données disponibles, ou d'une autre difficulté. Je me permets néanmoins de les aborder à la fois pour débattre de leur pertinence et pour tenter d'esquisser un protocole de recherche permettant de les mettre en œuvre rigoureusement.

*Enrichir avec de nouvelles données pour analyser les processus sociaux en jeu*

La première des pistes est celle de l'adjonction de données permettant soit de contextualiser soit d'expliquer les formes observables dans les graphes. Un des moyens les plus prometteurs que j'ai pu employer jusqu'à présent consiste à identifier des couplets contributeurs/pages à l'occasion d'entretiens. Ainsi, lors d'évènements wikipédiens organisés sur le modèle d'un colloque universitaire (« WikiConvention »), les contributeurs se présentent aux autres par l'affirmation « j'ai fait [telle ou

---

<sup>10</sup> Je suis en cela la préconisation épistémologique de Jean-Claude Passeron (1991) formulée au sujet de l'interprétation sociologiques des tableaux statistiques : « le raisonnement sociologique qui en dit le moins en dit plus que le raisonnement statistique qui en dit le plus ».



telle page] ». Et qu'à cette présentation suite un « ah oui, c'est toi ! » alors même que le pseudo était présent sur un badge fixé sur la poitrine au moment de l'inscription au colloque. Ce qui permet de reconnaître le contributeur ne tient pas isolément au fait d'avoir participé à l'écriture de telle ou telle page ou au fait d'exhiber tel ou tel pseudo, mais à la conjonction de ces deux éléments : « c'est toi que j'ai vu faire ça ici ».

L'analyse de réseau exposée ici permet de se demander, pour chaque couplet page/contributeur, dans quelle forme est pris le contributeur sur le réseau de telle ou telle de « ses » pages. Ainsi les deux matériaux peuvent s'enrichir pleinement, l'entretien offrant une profondeur biographique qui manque inévitablement à l'analyse formelle de la coopération, alors que la forme enrichit la dimension relationnelle de la contribution telle que décrite dans l'entretien – notamment en dépassant la limite d'un matériau déclaratif et centré sur l'acteur.

Dans cet article programmatique, je prendrai l'exemple d'un contributeur pour lequel des pages centrales sont nettement identifiables. Il s'agit de Sylvio (cf. encadré 1), professeur de physique-chimie en classes préparatoires de 41 ans au moment de l'entretien. Il commence à contribuer à Wikipédia autour de 2006, c'est-à-dire en plein dans le moment de la massification.

---

#### **Encadré 1 : biographie de Sylvio**

Sylvio a une trajectoire scolaire linéaire qui le portera vers les plus hauts diplômes de l'enseignement supérieur : baccalauréat C en 1994, DEUG, licence et maîtrise entre 1994 et 1999 à Reims. Cette deuxième étape sera ponctuée par l'obtention de l'agrégation de sciences physiques en 1999, toujours à Reims. Il souhaite ensuite faire une thèse et pour la première fois de son parcours, il devra s'y reprendre à deux fois. Une première tentative à Orsay sera infructueuse, car il n'anticipe pas le prix *exorbitant* du logement. Mais, *un très très vieux pote* lui signale que dans son établissement, à Saint-Étienne, des bourses de thèse n'ont pas été attribuées. Après avoir postulé, il obtient cette allocation et part donc faire une thèse qu'il soutiendra en 2005, après avoir pris son poste d'enseignant en classes préparatoires à Paris en 2004.

#### **Passage à l'acte**

Sylvio a son premier contact avec Wikipédia en 2005, lorsqu'il *tombe sur un papier de Charlie Hebdo qui descend le truc en flamme en disant « ouais c'est n'importe quoi, tout le monde peut tout modifier... »*. Intrigué, il *va voir, puis laisse complètement tomber le truc même s'il en rediscute une ou deux fois pendant la thèse avec des potes*.

En 2006, il *s'abîme vraiment le dos* dans un accident de ski. *Cloué au lit pendant trois mois à rien foutre*, il commence à contribuer. Ses débuts ne se font cependant pas spontanément, car *il est quelqu'un de précautionneux*. Pendant *15 jours 3 semaines*, il lit d'abord attentivement les règles avant de commencer à contribuer. Cela lui permet de *ne pas avoir la mauvaise surprise d'être mal reçu*.

#### **Cristallisation**

6 mois après, parce qu'*avant c'était beaucoup plus cool*, Sylvio devient administrateur et *patrouille à très haute dose*. Il adhère également à Wikimedia France. En 2009, Sylvio ressent *vraiment une insertion complète dans la communauté*. Il est alors membre du conseil d'administration de Wikimedia France.

Après l'année 2007, sa tendance à patrouiller diminue et notre enquêté se consacre à la rédaction d'articles. Deux périodes se succèdent : les articles de physique-chimie et les biographies de blues-man. Parmi les articles de physique chimie, *le plus emblématique et celui qui demande le plus de boulot* est sans doute celui de Newton. Puis, *arrivé à un moment où Wikipédia ressemblait trop au boulot*, Sylvio se lance dans la rédaction d'article sur le blues, à partir d'une encyclopédie papier qu'il a acheté en librairie. *Un par un, il référence en mettant la source*. Ce mode de contribution lui permet de se divertir sans que Wikipédia ne prenne *non plus trop de temps*.

Durant ces deux périodes, Sylvio rédige également *plein d'articles rigolos : la femme qui s'est mariée à la tour Eiffel, le plus grand agneau du monde. Ça lui prend cinq minutes et il se marre...* Ce dernier mode de contribution va particulièrement être percuté par la querelle entre suppressionniste et inclusionnistes à partir de 2010 jusqu'à 2012. Clairement situé dans le *camp des inclusionnistes* Sylvio vit mal que *d'anciens potes commencent à se dézinguer*. Entre fin 2012 et 2014, Sylvio s'installe avec sa copine tout en étant membre du conseil d'administration de Wikimedia France à une période où l'association recrute ses premiers salariés. *C'était beaucoup de stress, c'était très chronophage. Il arrête un peu tout*.

En 2015, il revient et fait *quelques articles quand ça l'amuse*. La crise de l'association le pousse à revenir, sans compte qu'il *s'emmerde un peu au boulot et que sa copine finit ses études de médecine, donc a beaucoup de travail*. Revenu au CA depuis septembre 2017, il s'investi pour redresser l'association qui est *par terre*. D'une manière générale, il aime *faire des choses qu'il n'a jamais fait : relever une assos, il n'a jamais fait*.

---

Je ne retiendrai ici qu'un exemple fort du travail de Sylvio sur Wikipédia, à savoir l'article consacré au physicien Issac Newton. Ces contributions interviennent peu après son passage à l'acte, et constituent la première occasion de produire un travail de rédaction du texte en application de la doctrine wikipédienne de citation des sources.

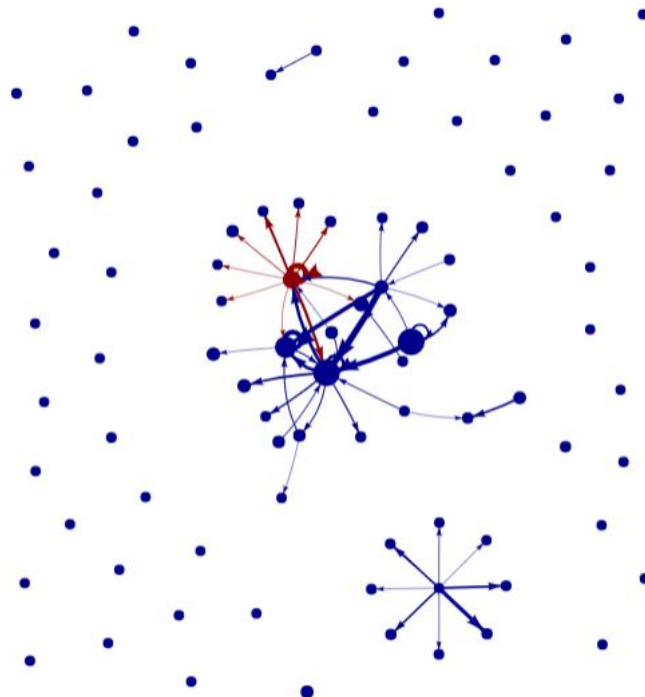
---

*« Moi le truc... Qui m'a le plus... Enfin qui m'a demandé le plus de boulot... Qui est le plus concentré en travail, c'est l'article sur Newton. Du coup c'est la partie que j'ai travaillé moi, c'est pas énorme... C'est-à-dire que c'est quelques paragraphes, mais je me suis tapé une bibliographie de ouf... C'est Newton et l'alchimie en fait. Pour essayer de condenser quelque chose dessus qui soit à peu près raisonnable, qui sortent un peu des conneries qu'on peut raconter sur l'alchimie et des conneries qu'on peut raconter sur Newton-alchimie. Et que ça ait un petit peu de sens physique, ça c'est vraiment ce qui m'a demandé le plus de temps et de travail. »*

---

La figure suivante montre la position de Sylvio dans le réseau de la page consacrée à la biographie d'Isaac Newton.

**Figure 5. Réseau de la page « Isaac Newton »**



Sylvio se trouve se trouve au confluent de deux formes : une étoile et une autoroute à double sens. Le réseau apporte deux choses à l'analyse de l'entretien la dimension conflictuelle de la contribution. Sylvio efface les contributions de contributeurs marginaux et a vraisemblablement été engagé dans un conflit avec un contributeur plus important que lui de la page Newton.

Dans la suite de mon travail de recherche, je prévois d'identifier plus rigoureusement, pour chacune des 18 biographies dont je dispose, des couplets contributeurs-pages qui permettront de questionner deux choses :

- Le mise en récit des contributions par le contributeur durant un entretien d'explicitation, et l'on voit ici que le discours de Sylvio occulte la dimension conflictuelle. Une hypothèse d'interprétation intéressante peut être de revenir au profil sociodémographique de Sylvio. Homme très fortement diplômé, il possède tous les galons du parfait « passés à l'acte ». Sa capacité à naturaliser son entrée et à passer au second plan les différents conflits (cf. encadré 1) peut provenir d'une ressource dispositionnelle (Lahire, 2018).
- L'évolution de la place du contributeur dans les multiples collectifs de contribution dans lesquels il s'inscrit. Les différents travaux mettent par exemple en évidence un engagement progressif des nouveaux contributeurs dans des trajectoires types relativement peu nombreuses (Gleave et al., 2009; Welser et al., 2011). Cependant, l'approche en termes de rôle est pour le moment limitée par une simple analyse des *namespaces* dans lesquels les contributions s'inscrivent (Arazy et al., 2015; Azary et al., 2017) ou bien à une catégorisation des tâches (Daxenberger & Gurevych, 2012, 2013; Yang et al., 2016). L'approche en termes de réseaux permet de différencier les positions occupées par un contributeur sur une même page. Cela constitue a priori un progrès dans la mesure où une même activité ou une même répartition de la contribution dans les mêmes *namespaces*, mais conduite dans une position structurellement différente fait occuper un rôle différent au contributeur.

### *Une étude des controverses par la comparaison de sous-corpus*

Une autre piste prend forme dans la comparaison de sous-corpus de pages. À ce stade de mon travail, deux critères différencient les comparaisons. Le premier est le type formel de la page : biographie, article notionnel, page consacrée à un évènement, etc. Une liste exhaustive reste à établir. Le second est l'espace thématique dans lequel se situe la page. Cette fois, les listes exhaustives sont foisonnantes et le degré de détail pourra varier au fil de la recherche. Le tableau suivant fait le point sur les différentes stratégies de comparaison possible en fonction de ces deux critères de choix.

Tableau 4. Typologie des comparaisons possibles entre corpus en fonction des thématiques et des types de pages

	Thématique identique	Thématique différente
<b>Type formel identique</b>	<p>Dans quelle mesure la configuration de la « société des usagers » influence-t-elle la « communauté des contributeurs » ?</p> <p><i>Exemple d'une comparaison entre deux biographies politiques, l'une d'un.e candidat.e à l'élection présidentielle et d'un.e non-candidat.e.</i></p>	<p>Les thématiques appellent-elles à un mécanisme de surveillance particulier ?</p>
<b>Type formel différent</b>	<p>Les types appellent-il un mécanisme de surveillance particulier ?</p>	<p>Les mêmes formes sont-elles observables dans des pages en tous points différents ?</p> <p><i>Cette question permet de prendre la mesure de la force de la régulation wikipédienne.</i></p>

La comparaison pourra donner au projet une dimension de science participative ou collaborative. Je différencie ces deux termes par l'intention qui préside à la participation au projet. Dans le cas collaboratif, l'expert est un chercheur qui se pose avec moi une question scientifique. Dans le cas participatif, l'expert joue un rôle d'alerte sur une potentialité de pages intéressantes au regard d'un jugement de valeur sur leur thématique ou leur type.

Cette distinction est bien sûr complexe à manier en pratique, tant les figures du chercheur et du militant peuvent se brouiller dans de nombreuses situations. Cependant, la contribution de l'expert sera différente selon que l'on se place dans une approche de science participative ou une approche de science collaborative. La distinction se perçoit dans le lien qui unit ces deux figures aux jugements de valeur concernant l'objet de la controverse. Là où le collaborateur instaure une « distance aux valeurs »<sup>11</sup> telle que préconisée par Max Weber, le participant entre dans le projet en assumant un jugement de valeur sur les pages dont nous devrions dessiner le réseau.

<sup>11</sup> J'utilise cette expression en lieu et place de celle de « neutralité axiologique ». Plus classique, cette dernière me semble cependant relevé d'un contre-sens. En aucun cas, Weber ne préconise une posture neutre. Au

On se rend vite compte, en le posant ainsi, que les figures du collaborateur et du participant apportent chacune une contribution potentielle très importante au projet qui est présenté ici. Formé uniquement de collaborateurs, le projet risquerait de se focaliser artificiellement sur une approche wikipédienne de Wikipédia : deux corpus ne seraient comparés que pour les différences qu'ils donnent à voir dans l'organisation du commun, au détriment des contextes sociaux dans lesquels se propagent les pages Wikipédia – une dimension pourtant cruciale si l'on veut parler d'institutionnalisation.

À l'inverse, une sélection des pages uniquement réalisée par des participants nous ferait courir le risque de surestimer artificiellement l'influence de la société des usagers dans la vie institutionnelle de la communauté des contributeurs. Dans la *Revue du Crieur*, Sihame Assbague (2022) a publié un texte sur les biais de construction observables sur les pages liées aux personnalités du mouvement antiraciste. Elle mettra notamment en évidence que les pages de ces personnalités décrivent extensivement des prises de position qualifiées de polémiques, là où la prise de parole d'éditorialistes reconnus classés à droite ne suscite pas le même type de descriptions.

La position sociale d'Assbague lui permet d'accéder spontanément à un fort degré de compréhension de la controverse. De la même manière, d'ailleurs, que l'anarchiste de Max Weber pouvait réaliser une sociologie de l'anarchisme plus percutante qu'un autre chercheur moins engagé. Dans l'enquête, Assbague joue un rôle d'alerte qui me permet de choisir deux corpus intéressants, notamment pour voir si ses constatations se retrouvent dans les formes observables dans les réseaux.

La distinction que je fais entre participant et collaborateur vise à montrer que cette alerte sera d'autant mieux mise en valeur qu'elle sera ensuite croisée avec d'autres choix méthodologiques. Ces derniers assumés non pas pour les enjeux normatifs qu'ils peuvent susciter, mais pour des raisons formelles. Par exemple, il peut être intéressant de comparer la controverse au sujet du racisme telle qu'identifiée par Assbague avec la controverse qui a fait rage au sein de Wikipédia sur le fait de savoir s'il convenait de dire « chicon » ou « endive » pour qualifier ce légume. À un militant antiraciste, cette comparaison semblera complètement absurde tant les enjeux sont différents. Pour un chercheur en revanche, cette comparaison paraîtra au contraire tout à fait pertinente, précisément pour mettre en évidence la spécificité des mécanismes de régulation ou au contraire la similarité entre deux mécanismes présents similairement sur des pages différentes.

---

contraire, il explicite très clairement combien un anarchiste peut réaliser une sociologie des mouvements anarchistes beaucoup plus fine qu'un autre chercheur. À la condition *sine qua non* de mettre à distance ses valeurs, substituant un « rapport [distant] aux valeurs » à ses « jugements de valeurs » pour produire une connaissance indépendante de ses opinions. Le fait de comprendre cette règle épistémologique comme une invitation à la neutralité bride considérablement le raisonnement sociologique, en interdisant de mettre à profit des connaissances acquises par d'autres biais que l'expérimentation neutralisée et neutralisante.

Une fois dit cela, pour ne pas tomber dans une réduction caricaturale de l'apport de la militance à l'activité scientifique, il faut rappeler que le chercheur seul n'aurait probablement pas eu l'idée d'aller puiser dans les biographies des militants antiracistes pour se donner un opérateur de comparaison pour éclairer le cas de la controverse « chicon » / « endive ». C'est en cela que les deux points de vue sont très fortement complémentaires. Et pour cette raison l'intégration d'une militance comme source d'information dans le travail de recherche ne transige en rien avec la distance aux valeurs propre au travail sociologique.

### *Introduire une couche sémantique pour comprendre les formes de coopération*

Les graphes tels que je les analyse pour le moment rendent compte du travail de suppression, sans faire grand cas des mots qui sont ajoutés ou supprimés. Si elle permet de mettre en évidence des configurations, une telle approche est nécessairement formaliste. Pour aller au-delà, une possibilité consiste à intégrer dans l'analyse des graphes non pas uniquement un poids lié au nombre d'éléments ajoutés ou supprimés, mais également une liste des mots ajoutés ou supprimés. Plus précisément, à chaque nœud serait ajoutée une liste de mots ajoutés par le contributeur correspondant et à chaque arête serait ajoutée une liste de mots supprimés par le contributeur dont part ladite arête.

Cet ajout permettrait de traiter non plus seulement l'organisation formelle du travail de surveillance, mais également de poser la question du consensus au sujet du contenu, en la décomposant en deux sous-questions :

- Existe-t-il des mots que la plupart des contributeurs ajoutent ?
- Existe-t-il des mots que la plupart des contributeurs suppriment ?

Le traitement de ces deux questions peut permettre d'étudier les deux faces d'un consensus collectif : être d'accord sur une manière de rédiger l'article et être d'accord sur ce qui ne doit pas figurer dans l'article. Plus difficilement observable, la deuxième face correspond à un phénomène assez peu étudié dans la littérature, où l'on voit des contributeurs se coaliser non pas pour « pousser un point de vue » (*POV pushing*), mais pour en écarter un autre.

S'il est simple de le proposer, la greffe de listes de mots sur chacun des objets d'un graphe, nœud comme arête, pose de nombreuses questions sur la manière de visualiser les graphes ainsi obtenus. Il peut être envisageable de construire plusieurs réseaux par types de contenu ajouté ou effacé. Par exemple, un réseau dont les arêtes ne représenteraient que l'effacement de notions liées au thème de l'article. Autre exemple, autre objectif : un réseau qui ne montrerait que la suppression et l'ajout de ponctuations mettrait en évidence l'action de contributeurs qui n'agissent que sur la structure formelle du texte au détriment de sa dimension sémantique. Ces manipulations demandent une série

de traitements lexicométriques qui, s'ils sont sur ma table de travail, n'ont pu être réalisés à cette échéance.

## Bibliographie

- Aaltonen, A., & Kallinikos, J. (2012). Coordination and Learning in Wikipedia : Revisiting the Dynamics of Exploitation and Exploration. In M. Holmqvist & A. Spicer (Éds.), *Managing 'Human Resources' by Exploiting and Exploring People's Potentials* (Vol. 37, p. 161-192). Emerald Group Publishing Limited. <http://www.emeraldinsight.com/doi/abs/10.1108/S0733-558X%282013%290000037010>
- Anthony, D., Smith, S. W., & Williamson, T. (2007). *The quality of open source production : Zealots and good Samaritans in the case of Wikipedia* (Technical Report N° TR2007-606; p. 34). Dartmouth College.
- Arazy, O., Ortega, F., Nov, O., Yeo, L., & Balila, A. (2015). Functional Roles and Career Paths in Wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (p. 1092-1105). Association for Computing Machinery. <https://doi.org/10.1145/2675133.2675257>
- Assbague, S. (2022). Wikipédia, ou la discrète neutralisation de l'antiracisme. Enquête sur la fabrique quotidienne d'un géant encyclopédique. *Revue du Crieur*, 21(2), 140-159. <https://doi.org/10.3917/crieu.021.0140>
- Azary, O., Hila, L.-A., Nov, O., Daxenberger, J., Balestra, M., & Cheshire, C. (2017). On the "How" and "Why" of Emergent Role Behaviors in Wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (p. 1066-1081). Association for Computing Machinery. <http://dl.acm.org/citation.cfm?doid=2998181.2998253>
- Becker, H. S. (1985). *Outsiders. Etudes de sociologie de la déviance*. Métailié.
- Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian : Transformation of Participation in a Collaborative Online Encyclopedia. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (Association for Computing Machinery, p. 1-10). <http://doi.acm.org/10.1145/1099203.1099205>
- Bubendorff, S., & Rizza, C. (2021). Produire collectivement du sens en temps de crise : L'utilisation de Wikipédia lors de la pandémie de COVID-19. *Communiquer. Revue de communication sociale et publique*, 32, 83-102.
- Cardon, D., & Levrel, J. (2009). La vigilance participative. Une interprétation de la gouvernance de Wikipédia. *Réseaux*, 154(2), 51-89.
- Cibois, P. (2014). Les méthodes d'analyse d'enquêtes. In *Les méthodes d'analyse d'enquêtes*. ENS Éditions. <http://books.openedition.org/enseditions/1443>
- Daxenberger, J., & Gurevych, I. (2012). A corpus-based study of edit categories in featured and non-featured Wikipedia articles. *Proceedings of COLING 2012*, 711-726.
- Daxenberger, J., & Gurevych, I. (2013). Automatically classifying edit categories in Wikipedia revisions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (p. 578-589). Association for Computing Machinery.
- Dejean, S., & Jullien, N. (2015). Big from the beginning : Assessing online contributors' behavior by their first contribution. *Research Policy*, 44(6), 1226-1239.
- Elias, N. (1993). *Qu'est-ce que la sociologie ?* Pocket.

- Gleave, E., Welsler, H. T., Lento, T. M., & Smith, M. A. (2009). A conceptual and operational definition of 'social role' in online community. *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, 1-11. <http://ieeexplore.ieee.org/abstract/document/4755505/>
- Goffman, E. (1983). The Interaction Order : American Sociological Association, 1982 Presidential Address. *American Sociological Review*, 48(1), 1-17. <https://doi.org/10.2307/2095141>
- Halfaker, A., Keyes, O., & Taraborelli, D. (2013). Making peripheral participation legitimate : Reader engagement experiments in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work* (p. 849-860). Association for Computing Machinery. <http://dl.acm.org/citation.cfm?id=2441872>
- Halfaker, A., Kittur, A., Kraut, R., & Riedl, J. (2009). A jury of your peers : Quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration—WikiSym '09* (p. 1). Association for Computing Machinery. <http://portal.acm.org/citation.cfm?doi=1641309.1641332>
- Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162(3859), 1243-1248. <https://doi.org/10.1126/science.162.3859.1243>
- Hill, B. M., & Shaw, A. (2013). The Wikipedia Gender Gap Revisited : Characterizing Survey Response Bias with Propensity Score Estimation. *PLoS ONE*, 8(6), e65782. <https://doi.org/10.1371/journal.pone.0065782>
- Hu, M., Lim, E.-P., Sun, A., Lauw, H. W., & Vuong, B.-Q. (2007). Measuring article quality in wikipedia : Models and evaluation. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 243-252.
- Joubert, L. (2020). *Wikipédia, la fabrique d'une encyclopédie à l'ère du logiciel libre. Sociologie d'un commun numérique de masse*. [Thèse de doctorat en sociologie]. Aix-Marseille Université.
- Jullien, N., & Roudaut, K. (2020). Commun numérique de connaissance : Définition et conditions d'existence. *Innovations*, 63(3), 69-93. <https://doi.org/10.3917/inno.063.0069>
- Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd : Wikipedia and the rise of the bourgeoisie. *World wide web*, 1(2), 19.
- Kostakis, V. (2010). Peer governance and Wikipedia : Identifying and understanding the problems of Wikipedia's governance. *First Monday*, 15(3). <http://pear.accc.uic.edu/ojs/index.php/fm/article/viewArticle/2613>
- Lahire, B. (2018). Les ambitions théoriques de la sociologie. *Sociologie*, 9(1), 61-71.
- Laval, C. (2016). « Commun » et « communauté » : Un essai de clarification sociologique. *SociologiesS*. <http://journals.openedition.org.lama.univ-amu.fr/sociologies/5677>
- Mercklé, P. (2016). *Sociologie des réseaux sociaux* (3ème éd.). La Découverte.
- Mills, C. W. (2006). *L'imagination sociologique*. La Découverte.
- Norbert, E., & Scotson, J. L. (1997). Logiques de l'exclusion. *Paris, Fayard*.
- Pancier, K., Halfaker, A., & Terveen, L. (2009). Wikipedians Are Born, Not Made : A Study of Power Editors on Wikipedia. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (p. 51-60). Association for Computing Machinery. <http://doi.acm.org/10.1145/1531674.1531682>
- Passeron, J.-C. (1991). *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*. Nathan.
- Priedhorsky, R., Chen, J., Lam, S. T. K., Pancier, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in Wikipedia. *Proceedings of the 2007 international ACM conference on Supporting group work*, 259-268.



- Viégas, F. B., Wattenberg, M., & McKeon, M. M. (2007). The hidden order of Wikipedia. In *International conference on Online communities and social computing* (p. 445-454). Association for Computing Machinery. [http://link.springer.com/chapter/10.1007/978-3-540-73257-0\\_49](http://link.springer.com/chapter/10.1007/978-3-540-73257-0_49)
- Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference* (p. 122-129). Association for Computing Machinery.
- White, H. C. (2011). *Identité et contrôle : Une théorie de l'émergence des formations sociales*. Editions de l'Ecole des Hautes Etudes en Sciences Sociales.
- Wilkinson, D. M., & Huberman, B. A. (2007). Cooperation and quality in wikipedia. *Proceedings of the 2007 international symposium on Wikis*, 157-164. <http://dl.acm.org/citation.cfm?id=1296968>
- Yang, D., Halfaker, A., Kraut, R. E., & Hovy, E. H. (2016). Who Did What : Editor Role Identification in Wikipedia. In *ICWSM* (p. 446-455). Association for Computing Machinery.
- Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). *The most controversial topics in Wikipedia : A multilingual and geographical analysis*. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2269392](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2269392)
- Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PLoS ONE*, 7(6), e38869. <https://doi.org/10.1371/journal.pone.0038869>