



**HAL**  
open science

## Quelle gouvernance des taches humbles dans le contexte des communautés open-source? Le cas du wiki de la communauté WordPress.

Emmanuel Ruzé

### ► To cite this version:

Emmanuel Ruzé. Quelle gouvernance des taches humbles dans le contexte des communautés open-source? Le cas du wiki de la communauté WordPress.. AIM 2012, May 2012, Bordeaux, France. pp.Not decided for the moment. hal-00717690

**HAL Id: hal-00717690**

**<https://hal.science/hal-00717690>**

Submitted on 13 Jul 2012

**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.

# Quelle gouvernance des taches humbles dans le contexte des communautés open-source ? Le cas du wiki de la communauté WordPress.

Emmanuel Ruzé  
ISTEC,  
Enseignant à Telecom ParisTech.

## Résumé :

La documentation d'un projet open-source est un usage originel et répandu des wikis cependant négligé par la littérature. Nous étudions ici la structure de management d'une partie localisée, ancillaire d'un projet open-source connu, WordPress ; il fait usage d'un wiki pour gérer son savoir technique. Le management se fait sur une liste de discussions : il n'est pas observé qu'un tel processus structuré nécessaire à la viabilité du projet se soit concrétisé de façon notable sur les parties du wiki prévues pour la discussion. Nous proposons pour cela une approche quantitative (économétrie et analyse de réseau) qui complètent des intuitions qualitatives initiales et des travaux qualitatifs publiés ailleurs. Nous montrons dans cette étude de cas l'existence de propriétés structurelles marquées caractéristiques des systèmes dynamiques en économie cognitive. La compréhension d'un tel processus permet de montrer qu'il existe des formes d'auto-organisation se caractérisant par des régularités significatives. Cette contribution montre les particularités du processus de coordination ; celles-ci, nécessaires à la viabilité du projet, n'auraient pu se faire sur le wiki, paradoxalement, et nombre de projets pourraient se trouver dans ce cas de figure. L'étude d'un tel processus permet d'aller au-delà de la description de ce qu'on appelle parfois « structure en oignon » proposée dans la littérature pour l'ensemble des participants à un projet : nous apportons ici une description plus précise, plus localisée, en prenant en compte la dynamique *historique* d'un projet.

Mots-clefs : communautés open-source, wiki, WordPress, gestion du savoir, histoire quantitative.

## Abstract :

Open-Source project documentation is the main and initial purpose of wikis, but this is not the main theme of the scientific literature about wikis. We study here the management structure of the documentation of the WordPress community, which is supposed to provide the necessary technical knowledge to use the software and improve it. This is done mainly via a collaborative mailing-list, and not on the wiki itself. We propose to study the structure of the collaboration process via a quantitative and social network study; this shows that structural regularities are necessary to coordinate the work of the volunteers of the community.

Keywords: open-source communities, wiki, WordPress, knowledge management, quantitative history.

## 1. Introduction

L'objectif de ce travail est de comprendre davantage la nature de la *gouvernance* de la *collaboration massivement distribuée* sur des sites participatifs à contenu ouvert de type wiki. Nous proposons une étude structurelle de la liste de discussion qui coordonne l'activité sur le wiki d'une communauté open-source. C'est l'une des voies qui permet de se faire une meilleure idée du fonctionnement de la partie de la communauté choisie, WordPress qui contribue au wiki nommé Codex qui rassemble tout le savoir informel permettant de faire usage du logiciel.

La « collaboration massivement distribuée » sur un wiki se caractérise en effet par une population d'individus qui interagissent entre eux soit directement (formation de liens), soit indirectement (via un artefact comme une page de wiki). Ainsi, l'étude structurelle de la liste de discussions de coordination de la documentation (nommée « docs » puis « wp-docs ») peut paraître incongrue : l'essentiel ne serait-il pas dans le wiki ? Ce n'est pas le cas dans cette communauté, où l'analyse des archives montre que la coordination se fait principalement via une liste de discussions. L'étude d'un processus et de ses ressorts cachés est donc ici l'analyse d'une *série chronologique*, celle qui supporte la gouvernance d'un wiki. Nous nous focaliserons sur les aspects quantitatifs, incontournables pour en analyser les aspects structurels qui n'apparaissent pas forcément dans les données qualitatives, malgré l'intérêt de certains fragments. De plus, lorsqu'il est question d'étudier l'auto-organisation sur le Web, il est opportun de tenter une analyse de réseau, ce qui ne va pas techniquement de soi pour un wiki.

Voici les directions d'analyse de la gouvernance du Codex que nous proposons et quelques éléments de résultats :

Nous analysons le *processus* de coordination qui se déroule en deux phases, l'une, avec une forte activité d'une population conséquente d'intervenants, l'autre, d'activité plus faible sur ces deux points et marquée par l'émergence de petits cycles. C'est un processus aux caractéristiques fortement dissymétriques.

Nous analysons les interactions entre les deux populations de participants au processus : nous observons des phénomènes d'entraînement dits « boule de neige » entre les deux types d'intervenants, les administrateurs et les autres, les non-administrateurs.

Les administrateurs ont des caractéristiques d'intervention originales, comme d'interagir plus fréquemment entre eux lorsqu'ils parlent (« homophilie ») ou de concentrer sur des périodes significatives l'essentiel des interventions.

Le réseau formé par ces interactions qui a suffi au processus de coordination est cependant faiblement connecté, à l'exception des pôles multiples que sont les administrateurs qui sont les uns pour les autres des interlocuteurs maniant la réciprocité, ce qui n'exclut pas d'autres formes significatives de réciprocité dans le reste de la communauté.

## 2. Contextualisation et éléments d'histoire de la communauté<sup>1</sup>.

WordPress est l'histoire d'un succès du Web 2.0. , celui d'un produit open-source<sup>2</sup>, un blog conçu à partir du logiciel de forum bbpress, associé à une communauté remarquablement dynamique, associé à celui d'une « jeune pousse » (« start-up ») dont le modèle d'affaire repose sur la fourniture de services plus élaborés et contractualisés. WordPress est un blog de

---

<sup>1</sup> Cette partie auxiliaire de présentation de la communauté reprend partiellement Auteur (2009).

<sup>2</sup> [http://www.lexpansion.com/economie/actualite-high-tech/le-succes-de-wordpress-est-celui-du-modele-open-source\\_141270.html](http://www.lexpansion.com/economie/actualite-high-tech/le-succes-de-wordpress-est-celui-du-modele-open-source_141270.html) ou [http://afp.google.com/article/ALeqM5hFVImPWtuISQs5\\_NGrdIG7KOFuVA](http://afp.google.com/article/ALeqM5hFVImPWtuISQs5_NGrdIG7KOFuVA)

qualité haut de gamme dont l'usage devient un standard. La communauté est née en juin 2003<sup>3</sup>, la « jeune pousse » Automattic est lancée le 20 décembre 2005<sup>4</sup>. Le positionnement du moteur de blog est celui du haut de gamme, associé à une large diffusion, du fait du caractère open-source du moteur. Le pari a été réussi, car WordPress est devenu de juin 2003 à fin 2008 l'un des meilleurs moteurs de blogs. Le nombre de téléchargements est de plus de quatre millions, et le site rappelle que des organisations connues utilisent WordPress, comme le New York Times. L'un des facteurs clefs de succès de WordPress est l'importance de la communauté : il s'agit à la fois du nombre d'utilisateurs et des ressources qu'elle offre. Le forum et le Codex, qui est sensé rassembler tout le savoir nécessaire pour utiliser, comprendre, améliorer le logiciel, n'en sont que les éléments saillants. Un premier wiki a été mis en place le 17 décembre 2003, il fut un échec. Le second a été lancé le 30 juillet 2004, et son histoire a été contrastée.

Ecrire l'histoire de la communauté n'est pas simple, nous ne mentionnerons ici que quelques aspects synthétiques. Voici 'abord l'évolution de l'agrégation respective des activités de coordination des activités de support et de programmation (sur des listes de discussions), issue du dépouillement des archives de la communauté :

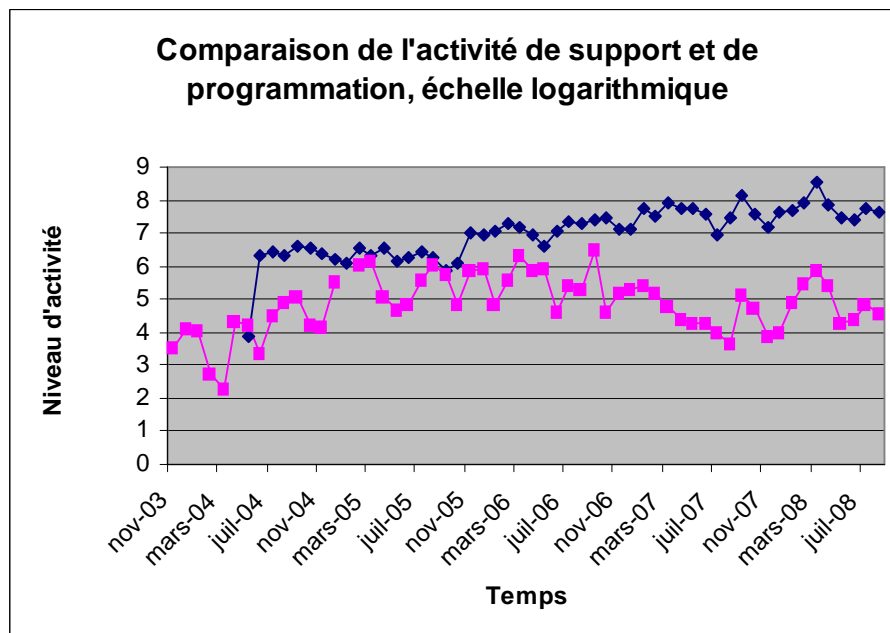


Figure 1 : Séries chronologiques des activités de programmation et de support utilisateurs.

La courbe du bas correspond à l'addition de l'activité mensuelle de régulation des activités de support. La courbe du haut correspond à l'addition de l'activité mensuelle de régulation des activités de programmation. On constate donc que les deux activités sont d'envergure différente : le pic de l'activité de programmation est de plus de 5000 courriels par mois, celui des activités de support 500. C'est pourtant la *seconde activité*, plus humble et moins impressionnante quantitativement qui nous intéresse.

<sup>3</sup> [http://web.archive.org/web/\\*/http://wordpress.org/](http://web.archive.org/web/*/http://wordpress.org/) permet de remonter le temps et de retrouver les premières captures d'écran de la communauté, ce qui permet de dater certains événements.

<sup>4</sup> <http://web.archive.org/web/20051220125353/http://automattic.com/>

L'activité de programmation et l'activité de support n'obéissent pas à une même logique. L'activité de support comporte trois phases de régulation successive : documentation, forum utilisateur, internationalisation de la documentation :

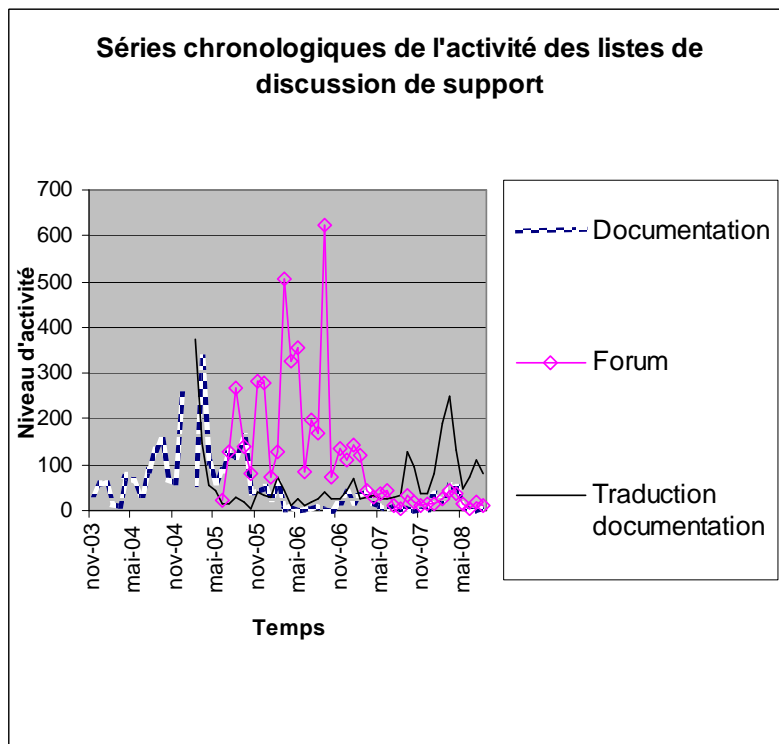


Figure 2 : comparaison des dynamiques des activités de support

L'activité de programmation se caractérise au contraire par une période de stabilité (avec une moyenne de 500 courriels), puis par l'émergence d'une activité cyclique à partir de décembre 2006<sup>5</sup>. On remarque tout de même une corrélation entre l'activité de documentation et de programmation, dans la mesure où les nouvelles versions du logiciel doivent être documentées. Voici à présent une autre vision de l'évolution comparée de la documentation et de l'activité de programmation :

<sup>5</sup> Nous ne présentons pas le graphique ici : il était trop difficilement lisible; on peut cependant deviner la présence de cycles dans le précédent graphique malgré l'échelle logarithmique.

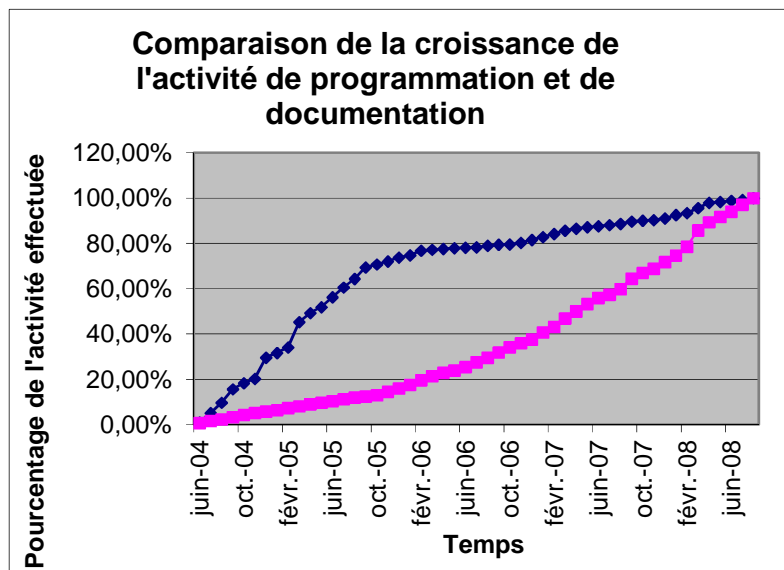
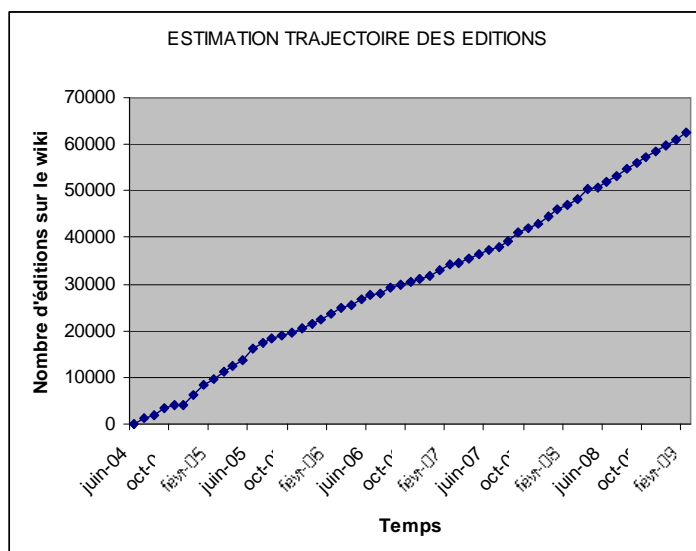


Figure 4 : comparaison de la croissance de l'activité de programmation et de documentation

La courbe du haut est une *approximation* de l'activité de la documentation sur les quatre ans de vie du deuxième wiki de la communauté. Il s'agit de la somme du nombre d'articles créés chaque mois et de l'activité mensuelle de la liste de discussion régulant la documentation. Le coefficient de corrélation entre ces deux indicateurs est de 0,64 sur quatre ans. La courbe du bas correspond à la sommation de l'activité de programmation. Ensuite, la reconstitution de l'activité totale sur le Codex suppose un travail par interpolations à partir des traces statistiques disponibles sur archives.org et du dépouillement manuel par datation du nombre de nouveaux articles créés chaque mois sur le wiki. On constate que la formation du bien collectif est un processus constant :



L'approximation de l'activité de documentation comporte des limites, mais permet cependant de corroborer à coup sûr deux hypothèses : d'une part le problème économique de *bien*

*collectif* a été résolu; d'autre part, que la documentation a été prise en compte *tôt*, et non pas après l'écriture du code<sup>6</sup>.

### 3. Quelques éléments exploratoires qualitatifs sur la coordination du projet et ses limites

L'analyse des données qualitatives, principalement des propos des volontaires intervenant dans les listes de discussions suggèrent que la coordination dans la communauté se caractériseraient par ... qu'il conviendrait d'examiner plus en avant. Une voie privilégiée est une étude des structures allant au-delà des approches qualitatives.

Voici d'abord un exemple d'incitation à la coordination des contributions volontaires, puis de réussite du processus mi-2005 :

“The docs team has never worked on the assumption that WE will tell YOU what to write, by god. If there's something you want to address, go for it. Just drop a note to the list so we can try to eliminate duplicated efforts.”<sup>7</sup>

“If you compare the process and volunteers around the Codex there is a marked contrast. There are a number of really good players working on the Codex. Things get planned and discussed in advance. It is very clearly democratic and is working brilliantly. Occasionally some discussions have floundered and ended in a stalemate, then someone has taken the decision and implemented one of the solutions. The codex is a valuable resource. The quality is on the whole outstanding. »<sup>8</sup>

Néanmoins, dès le début du projet, est apparue la nécessité de gérer, comme le montre ce fragment fort pertinent datant du deuxième mois du projet de documentation, que nous citons en entier :

“I think that for now, and perhaps forever, the documentation needs to have some tighter control. Development of code is in the hands of a small group of developers; this is necessary to keep focus and to get the job done. I think that a parallel should be drawn to documentation. Many hands makes small work, but too many hands makes chaos. Certainly there should always be a way for people to contribute their ideas, and perhaps even their own writing, but it needs to go through a smaller group of people for proofing and ultimately for approval by the developers. I'm concerned that "restriction" on access to working on documentation may be perceived by some as "elitist" or that the docs crew are somehow "special" or part of a "clique" or something.”<sup>9</sup>

---

<sup>6</sup> Il serait important de savoir dans quelle mesure un tel modèle où la documentation suit correctement la production de code et devient stratégique tranche avec les pratiques usuelles dans les communautés du Libre et si on pourrait observer là une *rupture historique* dans l'histoire du succès du logiciel libre. La diffusion supérieure d'Ubuntu par rapport à Debian où la documentation a été moins prise en compte serait un autre cas révélateur et intéressant à étudier.

<sup>7</sup> Source : <http://comox.textdrive.com/pipermail/docs/2004-April/000202.html>.

<sup>8</sup> Source : <http://comox.textdrive.com/pipermail/wp-hackers/2005-June/001430.html>.

<sup>9</sup> Source : <http://comox.textdrive.com/pipermail/docs/2003-December/000082.html>.

L'intérêt de ce fragment est de montrer l'existence de débats sur la structure relationnelle du processus de coordination. Ce n'est certes pas le seul ; pour ce qui concerne les pratiques et routines, voir Ruzé (2010).

Un tel processus de coordination n'a pas été sans créer des débats, sans pour autant que les avantages et inconvénients soient caricatures. Voici trois d'entre eux:

- Il serait plus difficile d'intervenir sur une liste centralisée et publique que sur les onglets décentralisés et moins visibles du wiki (sur ces onglets de discussions) :

"I return now to codex after a decent wiki education via wikimedia projects and find it frustrating that there is apparently zero conversation about this wiki on the wiki itself. I see that it is partially due to low edit #'s in general... but I wish there were something similar to the Village Pump here on the wiki? The email list is a much higher barrier to entry (and commitment) than simply asking a question on a meta-discussion wiki page. Further, bringing conversation into the wiki allows for easy reference within the wiki to past discussions -- and even policy.

- Il y aurait un risque de division des processus de communication:

I see potential aversion due to forking discussions between the wiki and the list -- but in general I think this is a better idea than bad. The wikimedia projects I have participated in work well with a combination of in-wiki discussion alongside fundamentally important email lists. Often individual parties will see only one -- but it all seems to work quite well. In my current situation -- I am unlikely to follow the *wp-docs* email list, but check the codex periodically."<sup>10</sup>

- On aurait selon certains contributeurs importants des risques de pouvoir excessif et de l'émergence de cliques, ce qui aurait entre autres pour effet de limiter les processus d'apprentissage sur le tas :

"I was keen on documentation. Along with Podz and Carthik and others, we got the initial wiki going (after we convinced Matt LOL) and then made the transition over to MediaWiki and Codex[...]. Soon we had lots of good folks jumping in to help out, and they were indeed wiki savvy. However, what happened was that soon these people were basically doing everything, while people like me wound up being relegated to the sidelines.

I could have learned on my own how to do a lot of things, but the point I'm trying to make is that had I been paired up with someone who was more advanced than I, we could have had two "experts" in time. Then we train two, and they train two etc. You get a CSS-effect in your documentation team development. The same would be true for support.

Right now the support and Codex are dominated by a small group. Again, this is not an attack or judgement, it's simply the reality of a lot of work to be done, and only a core group of folks with the occasional help from a lot of others.

I believe that WP has grown to the point where there needs to be some "middle management" and those folks who would so choose could become involved towards

---

<sup>10</sup> Source : [http://codex.wordpress.org/Codex\\_talk:Community\\_Portal#centralized\\_wiki\\_meta\\_discussion](http://codex.wordpress.org/Codex_talk:Community_Portal#centralized_wiki_meta_discussion) (consulté le 16 février 2006).



managing support and managing documentation almost as distinct open-source projects of their own. Craig. ”<sup>11</sup>

Qu'en est-il donc de fait ? C'est pour aller au-delà qu'il est nécessaire d'examiner des questions quantitatives, qui peuvent aussi permettre de suggérer d'autres tendances dont la communauté n'a pas forcément conscience.

#### 4. Revue de littérature

Notre propos se situe dans une littérature qui a connu une croissance importante ces dernières années, sur le Web 2.0. (Girieux, S., 2008, Koushik, S., et al. , 2009, McAfee, A., 2006, Mencarelli, R. et Pulh, M. 2009). Nous renvoyons sur ces questions à Cooke, M. (2008) pour une revue de littérature et à Ruzé (2009) pour une autre sur la gouvernance des communautés à contenu ouvert.

La littérature sur les wiki a connu une croissance significative voire exponentielle depuis les premiers articles datant de 2003, et les thématiques et la complexité des approches se sont diversifiées. Nous ne la quantifierons pas. Nous devons cependant mentionner quelques limites qui nuancent ce constat.

La première est que Wikipedia, projet gigantesque parmi les mieux référencés, mondialement connu, médiatisé et fascinant, a littéralement hypnotisé l'attention de chercheurs de plus en plus nombreux et selon des perspectives diverses. On renverra aux colloques « Wikisym » où cette focalisation est bien visible. Cette attention limitée mais orientée vers le wiki le plus connu a très probablement conduit les chercheurs à négliger les terrains originels des communautés Open-Source (Schwall, 2003), alors que les usages des wikis ont été associés depuis le début à ce type de communautés (Cubranic et al. 2003). En effet, l'articulation entre le phénomène Open-Source et celui des wikis, quoique reconnue, n'a pas été sérieusement étudiée : nombreux sont les travaux où il n'y est fait *qu'allusion*, comme UNDP (2005), qui cite les wikis comme outil de coordination des projets, de même que Bac et alii (2005) , Jullien et al. (2002), Gunnar (2005), Fromkin (2003), ou encore Kolbitsch et Maurer (2006). Les rares approches critiques (qui restent, elles encore, centrées sur Wikipedia) du phénomène « wiki » rappellent cependant à juste titre qu'il s'agit avant tout d'un « outil de geeks » :

« Wikis became popular in the free and open-source software (FOSS) community, where they were ideal for collaboratively discussing and documenting software, particularly given the loose structure of the projects » but « *these early software-focused wikis did not attract widespread attention.* », (Szybalski, 2005, c'est nous qui soulignons).

Le fait que les wikis soient utilisés seulement dans une minorité de projets il y a quelques années (Robbins, 2003) où l'outil était moins connu pourrait aussi expliquer une telle négligence<sup>12</sup>. En conséquence, la diversité des usages dans ces communautés, influencée ou non par les pratiques de Wikipedia, influence dont aucun travail de recherche ne fait mention, a été oubliée.

---

<sup>11</sup> Source : <http://comox.textdrive.com/pipermail/wp-forums/2006-October/003788.html> .Ce passage montre donc les limites de cette routine de supervision au point de poser des problèmes de gouvernance en particulier de « middle management », mais il est difficile d'aller plus loin avec ce seul fragment. Nous reviendrons sur ce point dans la conclusion générale de la thèse en développant une piste de recherche.

<sup>12</sup>Ainsi, les membres de la célèbre suite Open Office ont manifestement refusé au départ l'usage du wiki, même si ultérieurement un moteur MediaWiki a finalement été lancé.

Quelques travaux « voisins », pourrait-on dire, tentent cependant d'aborder la thématique dans le domaine de l'informatique, mais avec des limites de poids, dont la principale est le caractère exploratoire. Chau et Maurer (2005) est un travail exploratoire évoquant les wikis dans le secteur IT. Serakiotou (2008) est une courte description d'un wiki ayant eu un rôle majeur pour rendre la gestion du savoir plus efficiente dans un département IT, ce qui montre qu'il n'y a pas que des échecs, surtout en entreprise. Des références que nous ne dénombrerons pas, comme Eto et al. (2005), ont immédiatement des visées appliquées et proposent des améliorations par un nouveau *produit* informatique. Dans le secteur du logiciel, Andersen (2004) mentionne, dans un contexte propriétaire, le cas STATA (le logiciel bien connu des lecteurs statisticiens) dont la mémoire repose sur un wiki dont l'usage aurait été, d'après l'entreprise en question, un réel succès. D'autres études sont souvent un éloge de TWiki (Buffa, 2006, par exemple) qui n'est pas le meilleur outil « communautaire », mais plutôt un outil *d'entreprise*<sup>13</sup>. Martland (2003), propose une étude exploratoire de leurs usages dans *l'entreprise* Platypus Group. Plus avant, Xiao et al., (2007) abordent le sujet mais n'évoquent que la programmation, et pas la documentation. Craig (2005) propose quelques suggestions sur les wikis comme outil de management de projet dans les communautés Open-Source, mais le travail est exploratoire. Cheng et al. (2005) bien qu'abordant le rôle des wikis dans les activités de conception, est un travail qui relève plutôt du génie logiciel que des sciences sociales. Bref, le cadre original au sens strict n'a pas du tout été traité de façon approfondie. On peut penser que privilégier l'étude de « communautés internes » aux entreprises est plus rentable, en particulier dans ce contexte récent où des outils wikis « propriétaires » sont apparus. Enfin, un seul workshop en 2007<sup>14</sup> a brièvement abordé la question pour la programmation en général, mais il n'évoque pas spécifiquement le cas des communautés open-source et encore moins les activités humbles de support, comme la documentation. Il n'est en rien suffisant pour comprendre de façon exhaustive l'étendue des usages, ou la profondeur des problématiques en SI. D'une façon générale, un nombre significatif d'auteurs sont des informaticiens (en particulier ceux intervenants aux « WikiSyms ») ; or, l'expérience montre que les interrogations de ces spécialistes sont différentes de celles des sciences sociales.

*La question de la gouvernance des sites à contenu ouvert n'a été abordée que récemment, à la marge et d'une façon très exploratoire. Par exemple, le travail exploratoire de Den Besten et Dalle (2008) porte sur Simple Wikipedia, une version simplifiée pour les jeunes et qui comporte des spécificités éditoriales. Il évoque en creux les pratiques éditoriales et le management d'une telle communauté et les routines nécessaire à sa viabilité et souligne l'importance de notre objet de recherche : « We believe that the management and maintenance of collaborative knowledge production is soon to become a very important but difficult issue »<sup>15</sup>.*

*Le wiki est souvent classé parmi les « outils relationnels », ce qui justifie l'analyse de formes de collaboration.* La question des interactions entre contributeurs à un wiki, mais pas forcément uniquement *sur* le wiki, a-t-elle fait l'objet de travaux sérieux ? Klein et Iandoli (2008) soulignent l'idée d'échec des wikis en matière de délibération, précisément. L'articulation cruciale entre open-source et open-media est rarement évoquée dans la littérature. La seule évocation du sujet est Wagner et Majchrzak (2007). Crowston et Howison (2004) insistent sur la diversité des formes de centralisation/décentralisation dans les projets open-source, et on peut penser que des formes de diversité se retrouvent dans les parties des projets comportant un wiki. Comment qualifier et étudier les modes d'interaction entre agents

---

<sup>13</sup> Les utilisateurs sont par exemple Disney, Motorola, Cetelem, Google, mais aussi l'ENST, Université de Nice, etc. .

<sup>14</sup> <http://ws2007.wikisym.org/space/Wikis+for+Software+Engineering> .

<sup>15</sup> Den Besten et Dalle (2008), page 16.

sur ces outils, par ces outils, ou au-delà de ces outils ? Au vu de la nature du sujet et des phénomènes observés, la synthèse de la littérature suggère que la thématique des réseaux et de la collaboration comporte deux aspects correspondant à deux champs de la littérature soulignés ci-dessous.

*Une première littérature porte sur l'analyse quantitative de la structuration des wikis.* Les travaux en la matière comportent des limites. Quelques travaux sont plus proches d'une analyse des formes de collaboration. Korfiatis et al. (2006), un travail d'informaticiens, permet aussi d'évaluer la crédibilité de chaque contributeur en fonction de sa connectivité. De même, Korfiatis et Naeve (2006) proposent une analyse de l'activité et des contributions par une approche réseau, et tentent de conceptualiser la crédibilité de chaque auteur en mesurant la centralité de celui-ci dans le réseau des contributeurs. Cette approche reste exploratoire et localisée, et la crédibilité mesurée par une approche de cette sorte comporte des limites. Harrer (2008) aborde la construction collective du savoir. Tang (2008) est probablement l'un des travaux les plus aboutis d'un point de vue technique. Il propose un algorithme de mise en évidence des co-auteurs d'un article, ce qui est une façon pertinente d'envisager des modes de collaboration. Malheureusement, il n'est pas sûr de toute façon qu'on puisse l'adapter à notre cas. Enfin, il ne s'agit que d'un outil, et les auteurs ne proposent pas de résultats. La portée de l'analyse de réseau pour les wikis a fait l'objet de discussions (Voss, 2005) et un programme de recherche (AutoGraph) montre la difficulté d'une telle approche. Nous insistons, les résultats de ces travaux ne sont là encore valables *que pour Wikipedia*.

*Une deuxième littérature porte sur la structuration des listes de discussions.* Il n'existe pas beaucoup de travaux pertinents sur ce type de matériaux. De nombreux articles sur la question sont exploratoires, comme Conein (2004) qui propose une approche de la structuration des fils de discussions qui repose sur des théories de cognition sociale, Chen et al. (2006) qui tentent un classement des participants à une liste de discussions, ou enfin Krogh et al. (2003) pour tout ce qui touche l'analyse des fils de discussions, analysée de façon quantitative mais exploratoire. Karinbegg (1999) quoique ancien, est plus pertinent : il porte sur deux caractéristiques significatives des listes de discussions, à savoir la participation et le management. Ce type d'approche gestionnaire a été par la suite négligé d'après nos recherches, à part Matsumura et Sasaki (2005), sur le double rôle du leader (attraper les messages et en poster) et ses implications managériales dans ces listes de discussions et Sowe et al. (2007), qui étudient la question du transfert de connaissances entre développeurs et non-développeurs par une approche plutôt quantitative<sup>16</sup>. Weiss (2008) proposent des tests d'hypothèse du mécanisme d'attachement préférentiel dans la liste de discussion du projet Apache. Dorat et al. (2007), l'un des travaux les plus poussés en la matière, est une analyse raffinée des propriétés topologiques du réseau associées à la liste de discussions de Debian (loi de puissance, forme de centralité, etc.). Barcellini et al. (2005) croisent un mélange d'analyse de citations et de structure des fils de discussion. Nous signalons en particulier que l'une des limites de ces littératures est le manque d'analyses longitudinales, et donc historiques : c'est la voie vers laquelle nous allons nous orienter, tout en nous inspirant de ce qui précède. Le cas de Wikipédia, qui a centralisé pourtant l'attention des chercheurs intéressés par les communautés ouverte n'a pas été étudié (à part Ruzé, 2011, pour une étude quantitative des séries chronologiques), alors que même la fameuse encyclopédie fait aussi usage de listes de discussions pour coordonner l'activité, un phénomène qui reste également qualitativement mal connu.

---

<sup>16</sup> Cet article a aussi le mérite de découvrir d'autres phénomènes (de nature fractale) plutôt que des lois de puissance.

## 5. Eléments de méthodologie

Le cadre conceptuel de l'étude de cas se justifie pour plusieurs raisons. Les études approfondies d'un wiki restent rares, et une étude fine restait bienvenue. Etudier un seul cas s'est révélé incontournable du fait de la difficulté du traitement des sources, même s'il s'agissait d'un projet de taille moyenne. L'étude d'une situation de gestion en profondeur est aussi une tradition du laboratoire d'accueil où a été effectué le présent travail. Nous avons choisi la communauté WordPress en 2005 pour la richesse de ses sources, en particulier la présence d'une liste de discussion de régulation de la documentation qui pouvait révéler des phénomènes intéressants (et qui est notre principale source ici), ainsi que de son histoire. Nous désirions aussi éviter les grands projets comme Linux et étudier un projet de taille moyenne, où la difficulté technique en matière de programmation est de moindre importance. Enfin, nous désirions neutraliser la variable technique en choisissant une communauté faisant usage du moteur Mediawiki, comme Wikipedia. Ce choix a également été guidé à l'origine par la possibilité d'accès à des « données privilégiées » (Yin). A posteriori, le choix s'est révélé tout à fait judicieux, en particulier du fait du succès du produit en question<sup>17</sup> et du succès de la communauté. Nous devons aussi ici rappeler l'importance des blogs dans l'économie numérique, au moins la moitié des sites internet, outil fondamental de liberté d'expression, outil fondamental de communication pour beaucoup d'entreprise et d'entrepreneurs. Cependant, la philosophie de WordPress est de ne pas en faire qu'un blog, mais beaucoup d'autres façon de publier sur le web. Etudier WordPress, un outil de publication, c'est aussi important que d'étudier une manufacture d'imprimerie à l'époque de la Renaissance.

Nous avons choisi une approche historique pour ce travail reposant sur des archives numériques. Ce genre d'approche comporte des spécificités déjà décrites ailleurs (Auteur, 2009).

La liste de discussions « wp-docs » a été analysée de façon quantitative, de novembre 2003 à décembre 2008, sur plus de cinq ans. Elle comprend de novembre 2003 à décembre 2008 inclus 2872 courriers électroniques, 680 fils de discussions, 980 kb d'information, 173 discutants ayant parlé au moins une fois. Par comparaison avec d'autres travaux (Dorat et al., 2007 par exemple), c'est une liste de taille plus moyenne. Une originalité immédiatement visible, sur une période plus longue de 5 ans, est de comporter des plages significatives d'activité réduite, quoique ses contributeurs potentiels soient toujours présents, et que l'activité du wiki se poursuive.

La partie statistique de ce travail a demandé le dépouillement laborieux et systématique des archives en ligne de la liste de discussion pour créer une base de données de taille conséquente.

Pour l'analyse de réseaux, la démarche a été la suivante. La première étape a consisté à concevoir un dispositif d'extraction de données à partir des fichiers sous format « .txt » mensuels récapitulant l'activité de la liste de discussions dans les archives<sup>18</sup>. Les fichiers « .txt » permettent ainsi après nettoyage la création d'un fichier « reseau\_NomDuFichier » qui met à plat la liste des données (les intervenants sur la liste). Une fois l'extraction effectuée, une difficulté consistait à retravailler les données afin d'éliminer les pseudonymes multiples des contributeurs. Ce nettoyage des données est *absolument nécessaire* car ne pas l'effectuer peut changer *de façon significative* la forme du réseau, surtout s'il n'apparaît pas très dense à première vue. Enfin, il était nécessaire de convertir le fichier « reseau\_NomDuFichier » en

---

<sup>17</sup> Nous avons également été aidé dans ce choix par Bruno Chaves, chercheur en économie à EconomiX et expert en informatique.

<sup>18</sup> N'étant pas expert en Python, nous avons sollicité Jean-Philippe Cointet, doctorant au CREA, qui a passé quelques heures à réaliser le programme en Python. Nous le remercions.

format «.net » que le logiciel Pajek (et d'autres) permettent de lire, afin d'obtenir une visualisation et de mener à bien des analyses. C'est le rôle du logiciel « Txt2Pajek ».

## 6. Caractéristiques démographiques et structurelles de la population des participants à la liste de discussions « wp-Docs »

Cette section permet d'observer plusieurs éléments caractérisant la structuration du processus de management étudié :

- La présence d'une phase d'engouement brutale en 2005 (dont les caractéristiques seront étudiées plus loin).
- Le ratio d'encadrement entre administrateurs et non-administrateur est compris entre 2 et 4 non-administrateurs pour un administrateur.
- La présence de dissymétries marquées en ce qui concerne le nombre d'interventions (certains interviennent plus que d'autres).
- Nous mettons en évidence des régularités quantitatives (corrélation nuancées par une baisse tendancielle) simples entre le nombre d'intervention des administrateurs et non-administrateurs.

### 6.1. Statistiques démographiques dynamiques de la population qui intervient sur la liste de discussion

On observe que 173 individus parlent au moins une fois sur la liste sur une période de 5 ans. Nous précisons que d'autres personnes sont abonnées à la liste de discussion, mais restent silencieuses<sup>19</sup>. Si des personnes interviennent sur la liste, il est impossible de les manquer, et il n'y a pas de biais sur ce point. Voici l'évolution du nombre de participants par mois, que nous différencions entre administrateurs et non-administrateurs :

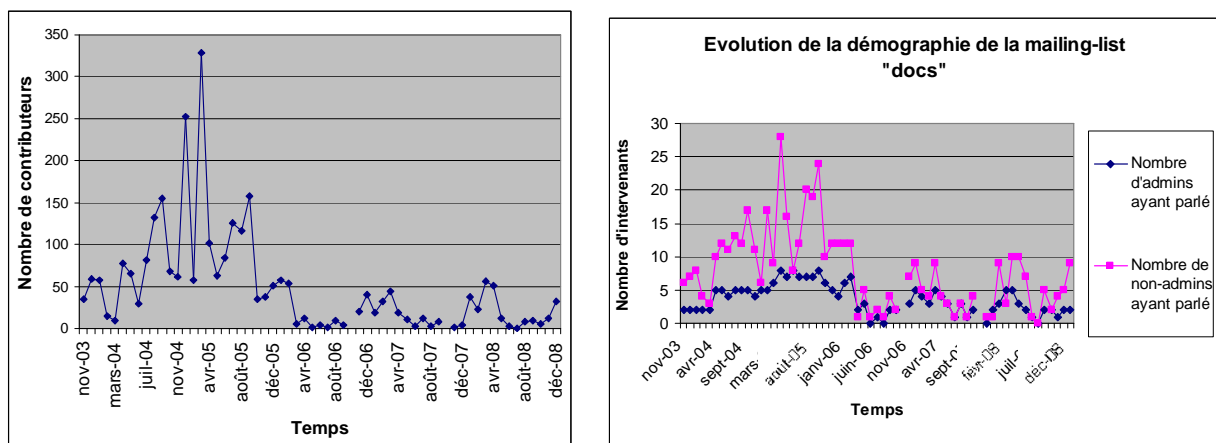


Figure 6: nombre de contributeurs, différencié entre administrateurs et non administrateurs

La démographie de la liste de discussion comporte deux phases nettement différenciées, la première allant de novembre 2003 à février-mars 2006, la seconde de mars-avril 2006 à

<sup>19</sup> On peut accéder régulièrement à la liste en devenant soi-même un « lurker », mais nous n'avons pas surveillé l'évolution de cette population, étant donné que par définition elle n'affecte pas la structuration des interactions ici. Elles peuvent le faire s'il s'agit de correspondances privées, mais l'accès à celles-ci est impossible pour des raisons éthiques.

décembre 2008. On peut discuter de la périodisation et des points de rupture à quelques mois près, mais elle différencie nettement deux périodes, visuellement parlant, l'une avec une courbe haute fortement indentée, l'autre avec des cycles plus petits mais nettement visibles. On observe une corrélation entre l'activité de discussions sur la liste et le nombre d'intervenants ou contributeurs (voir infra dans cette partie) et la courbe des interventions suit celle du nombre des intervenants.

## 6.2. Les caractéristiques en coupe (« cross-section ») de la population qui intervient sur la liste de discussion

Cette population se caractérise par des distributions en forme de *loi de puissance* autant en ce qui concerne le nombre d'interventions, le nombre de mois d'interventions (i.e. avec au minimum une intervention), la variance des interventions de chaque individu, et le nombre moyen de courriel par mois (nous ne montrons pas tous les graphiques):

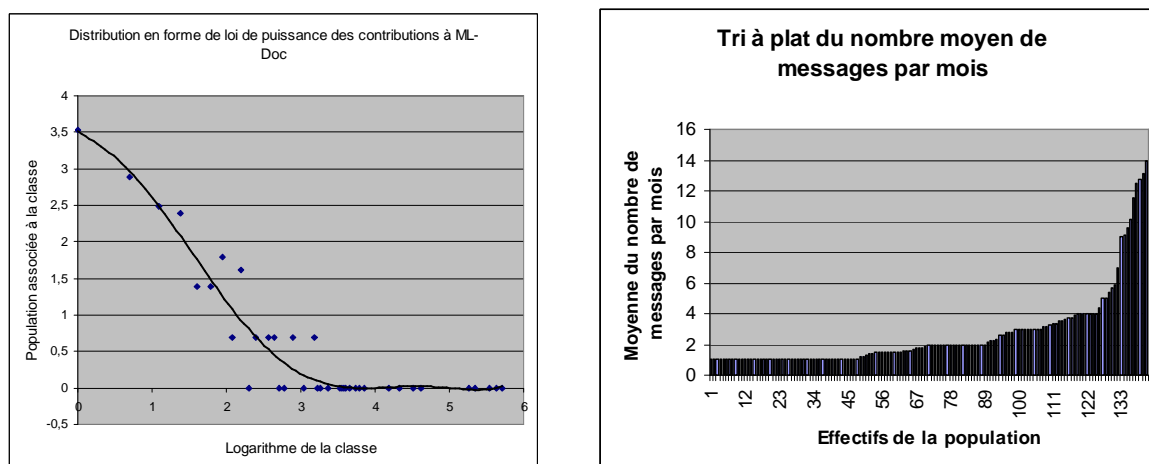


Figure 7 : tris à plats du nombre d'interventions sur la liste et du nombre moyen des interventions sur cinq ans

On peut montrer, à partir d'un classement par catégorie des intervenants en fonction du nombre de leurs interventions, que la forme des contributions à la liste de discussions suit, comme sur le wiki, une *loi de puissance* ; un petit nombre de contributeurs significatifs produit l'essentiel des interventions : l'hypothèse de l'existence d'une distribution en forme de loi de puissance est donc encore une fois pour l'essentiel vérifiée (elle n'est cependant pas « parfaite »), alors que ce type de distribution est récurrent entre autres sur le Web. Cependant, il était nécessaire de souligner qu'on l'observe *déjà* pour une population de taille réduite (173 individus). Le nombre relativement limité d'individus de la population rappelle aussi qu'observer une distribution parfaite n'est guère probable. Sur 173 intervenants sur la liste, on constate sur les données chiffrées, que l'essentiel des intervenants sont intervenus moins de 10 fois (75%) et qu'un petit noyau (5,8%) l'a fait plus de 50 fois.

### 6.3. La structure analytique des interventions sur la liste de discussions

Nous différencions à présent en matière d'interventions entre les populations des administrateurs et non-administrateurs :

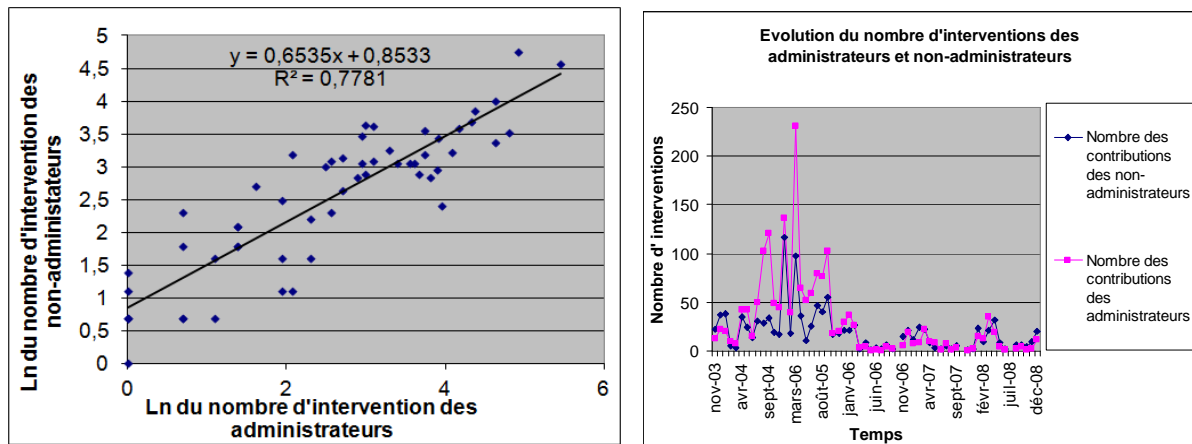


Figure 8 : structuration des interventions des administrateurs et des non-administrateurs.

Sur le graphique de gauche *ci-dessus*, on observe donc une corrélation significative entre les contributions des uns et des autres, malgré des irrégularités significatives. Une autre question est de savoir dans quel sens la relation joue. On remarque ensuite, *que les deux populations ont leur dynamique propre en matière d'interventions*; ainsi, la part d'interventions des administrateurs suit une pente décroissante après avoir dominé les échanges de novembre 2003 à mars 2006 (approximativement) ; cette pente n'était pas forcément visible sur les précédents graphiques, en particulier sur la seconde phase :

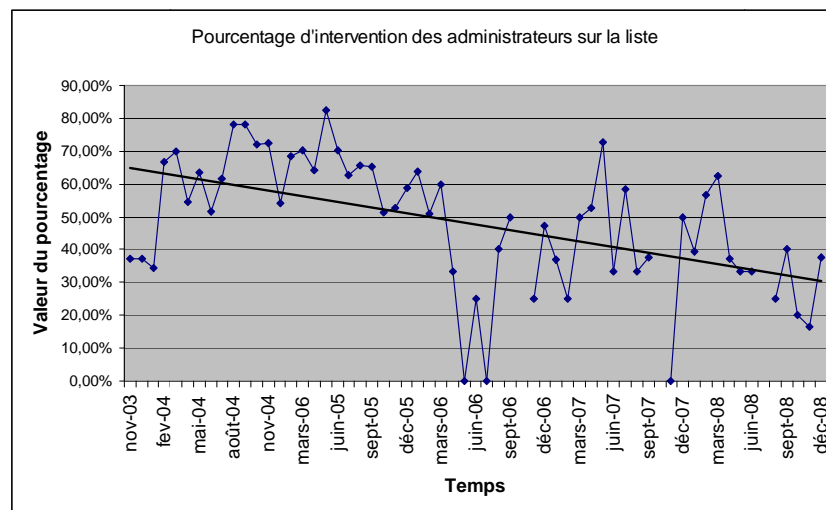


Figure 9 : pourcentage d'interventions des administrateurs sur la liste

On observe un étiage marqué entre mars 2006 et septembre 2006, mais il est difficile de lui attribuer définitivement une signification, même si cela est tentant : le débit trop bas de la liste occasionne un biais qui interdit d'attribuer la même signification aux pourcentages en fonction des périodes.

Cette distribution comporte une signification sujette à discussion. Une hypothèse prioritaire à tester dans la communauté WordPress était celle de la « montée de la

bourgeoisie » de KITTUR et al. (2007), mentionnée dans la revue de littérature, correspondant à un *renouveau de la population d'intervenants*. Ici, un renouvellement partiel de populations d'intervenants est effectivement observé dans la liste de discussions avec la chute du pourcentage d'interventions des administrateurs, mais nous ne pouvons savoir s'il permet de corroborer significativement l'hypothèse de Kittur, étant donné que la liste de discussion baisse de régime à partir de début 2006, ce qui pourrait être une modalité d'explication alternative au phénomène observé. On peut penser aussi qu'une telle hypothèse peut-être aussi due à la *maturité relative du projet* à partir de début 2006, le wiki ayant été intégré dans le projet par une uniformisation de son esthétique

De même, la baisse du trend des interventions sur les cinq ans ne permet pas forcément de voir la *baisse du nombre moyen de messages par intervenants*, présentée ci-dessous. L'origine de la baisse de cette moyenne est claire sur le graphique de droite ci-dessous, il s'agit de la baisse de la moyenne du nombre d'interventions des administrateurs, la moyenne des interventions des non-administrateurs étant d'une remarquable stabilité. C'est là l'utilité de différencier les deux variables :

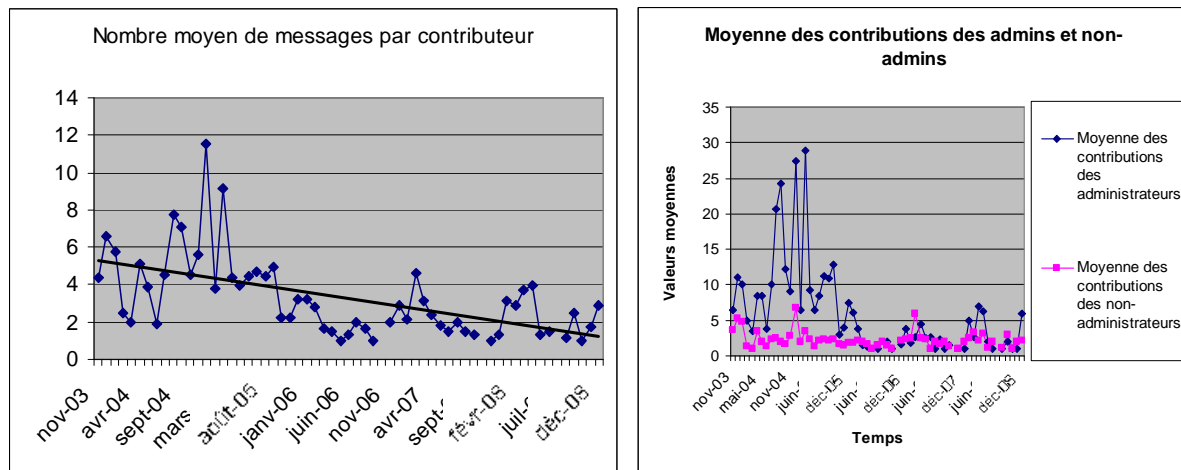


Figure 10 : séries chronologiques du nombre moyen de message par contributeur, différencié à droite entre administrateurs et non-administrateurs

On constate donc encore *deux phases* avec les valeurs moyennes : l'une partant de novembre 2003 et se terminant en mars-avril 2006, où la moyenne des contributions des administrateurs est nettement supérieure à celle des autres, et l'autre, à partir d'avril, où les deux moyennes sont semblables. On observe donc encore ici de fortes spécificités dans les comportements des deux populations

Cependant, les graphiques précédents, s'ils permettent de mieux comprendre la dynamique des interventions et les spécificités des deux sous-populations, ne sont pas suffisants. Ils cachent d'autres caractéristiques structurelles que nous allons commenter à partir des deux graphiques ci-dessous :



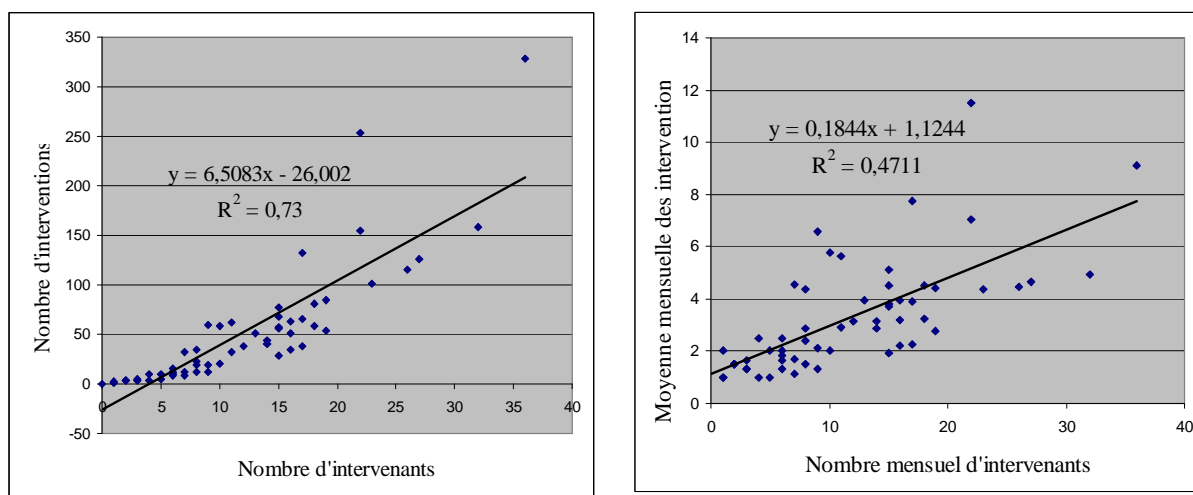


Figure 11 : le nombre d'intervenants fait augmenter le nombre et la moyenne des interventions

Le graphique de gauche ci-dessus, montre que la relation entre le nombre d'intervenants et le nombre d'interventions est linéaire et croissante, voire peut-être légèrement exponentielle<sup>20</sup>. Cependant, ce n'est pas la seule relation structurelle observée : sur le graphique de droite ci-dessus, le nombre d'intervenants fait clairement augmenter la moyenne mensuelle du nombre d'interventions, et ce, de plus, de façon linéaire.

Comment interpréter cela ? L'évolution du nombre moyen de messages par contributeurs pourrait obéir à une relation structurelle, qui dépendrait de la taille de la population mensuelle d'intervenants. Une interprétation serait que, « plus il y a de monde, plus on aurait propension à parler », ce qui correspond à un *effet d'entraînement*. Cet effet d'emballement comporte cependant quelques limites. La première apparaît en constatant l'hétéroscédasticité marquée sur les deux graphiques. La seconde est que la relation fonctionne de cette façon avec la population totale, mais *pas à l'intérieur de chaque sous-population* : sur ce point, nous n'ajoutons pas les graphiques conçus avec les mêmes variables que dans ceux commentés précédemment. *C'est donc l'interaction entre les deux populations qui crée cet effet.*

#### 6.4. Formes d'encadrement

La relation entre administrateurs et le reste de la population sur la liste, les non-administrateurs, est un élément a priori pertinent de la structuration de la collaboration. Nous avons donc différencié la *proportion* d'intervenants de chaque sous-population, qui montre une forme de *stabilité* de la relation entre administrateurs et non-administrateurs :

<sup>20</sup> Une relation non linéaire croissante améliore le R2 mais serait trop dépendante des valeurs extrêmes. Nous ne l'avons pas retenue.

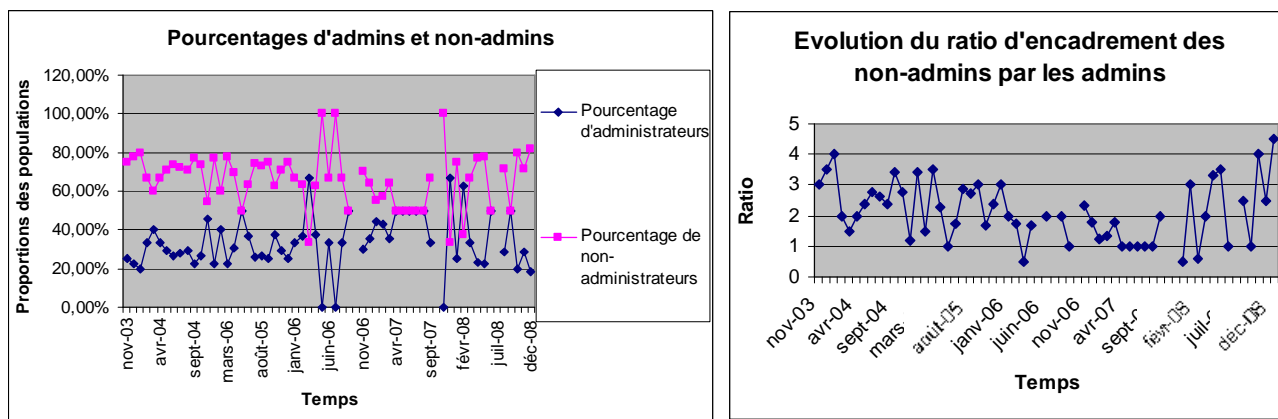


Figure 12 : présentation du rapport entre le nombre d'administrateurs et de non-administrateurs.

La moyenne et la variance du taux d'encadrement qui reste assez stable, entre 2 et 4 non-administrateurs pour un administrateur, sont de 2,13 et 0,96. La dispersion entre les effectifs mensuels des deux populations montre d'une autre façon que précédemment qu'il n'y a pas de sous-encadrement significatif. Cependant, même si le ratio comporte d'après ce qui précède une forme de stabilité, la répartition des administrateurs et non-administrateurs suit, comme le montre le graphique ci-dessous, une corrélation qui n'est pas parfaitement linéaire ; elle suggère aussi l'existence d'un « effet d'entraînement » dont le sens resterait alors à déterminer (nous n'y sommes pas parvenu) :

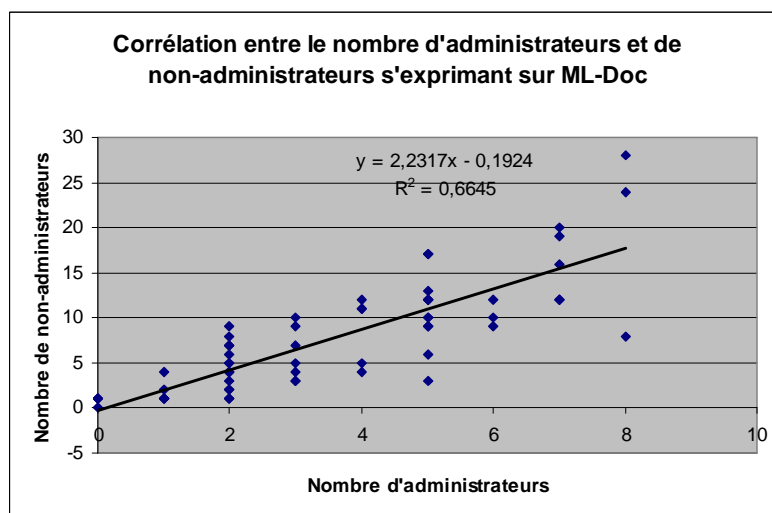


Figure 13 : relation linéaire entre le nombre d'administrateurs et de non-administrateurs

## 7. Comment comprendre les interactions internes aux fils de discussions ?

### 7.1. Premiers éléments descriptifs et alternatifs

Nous commençons par quelques éléments descriptifs rapides.

Nous disposons des chiffres requis différenciant les comportements des administrateurs (« admins ») et des autres intervenants (« non-admins ») et de l'évolution de la taille des fils de discussions :

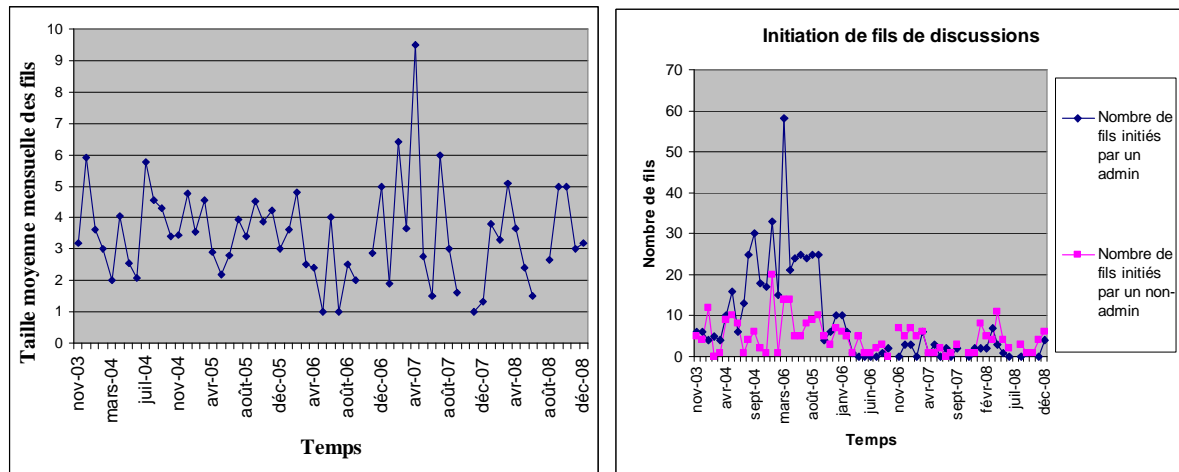


Figure 14: série chronologique de la taille mensuelle des fils et série chronologique du nombre de fils initiés par les administrateurs et les non-administrateurs.

Le premier graphique ci-dessus montre que la taille mensuelle moyenne des fils chaque mois (ou intensité des discussions) est susceptible de *variations significatives*, entre 5 et 2 courriels par fil en moyenne, à quelques pics et étiages près, et qu'il existe des formes de cycles. Le second graphique présente les 492 fils initiés par les administrateurs et 277 fils initiés par les autres. Lorsqu'on regarde les courbes, on remarque une évolution début 2006, entre janvier et avril, les courbes de prise d'initiative des administrateurs passant sous celle des « non-administrateurs », à quelques exceptions près. Une comparaison par visualisation des chiffres précédents sous forme de pourcentages aide à mettre en évidence un phénomène *d'inversion* structurelle à partir approximativement de avril 2006, inversion qui n'est remise en question qu'épisodiquement.

Avant de présenter le principal résultat auquel nous sommes arrivés, nous présentons quelques résultats moins probants concernant l'évolution des caractéristiques des interactions entre intervenants dans les fils de discussions. Il est difficile de savoir quelle est la raison de l'augmentation de la taille des fils, qui est une caractéristique de l'intensité des discussions qui varie par cycles, d'un point de vue purement quantitatif. Nous avons travaillé longuement les données sans parvenir à quoi ce se soit de vraiment significativement concluant après avoir testé de nombreuses variables possibles. Nous présentons quelques tendances seulement à partir de variables simples, et sans affirmer l'existence d'une causalité.

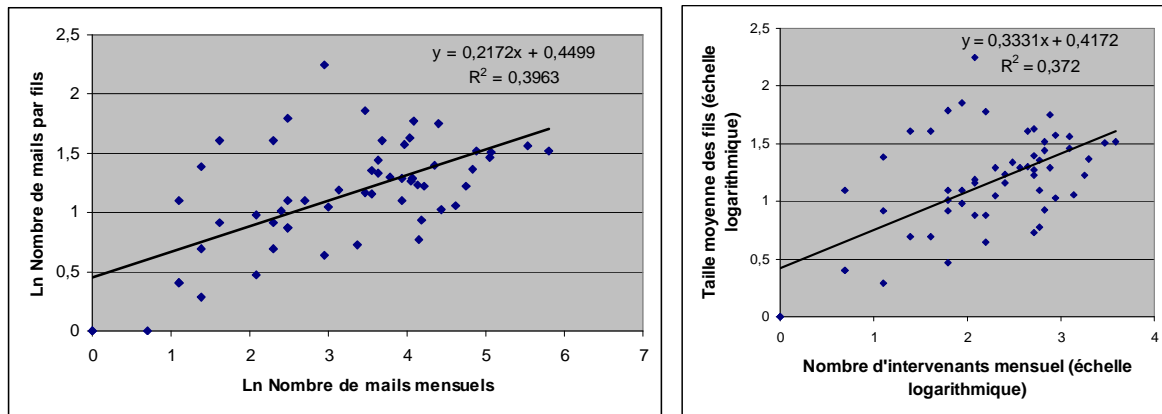


Figure 15 : mise en relation du nombre d'intervenants mensuel, du nombre de courriels par mois, et de la taille moyenne des fils

Sur le premier graphique, le nombre de mails par mois fait augmenter le nombre de mails par fil, à l'échelle logarithmique, mais il est difficile de lui attribuer une signification. Est-ce que cela signifie que les nouveaux courriels ont davantage tendance à se « loger » à l'intérieur des fils existants (i.e. renforcement de l'intensité des discussions ?). C'est une interprétation qu'on peut retenir mais qui mériterait des approfondissements. Le graphique de droite rapporte le nombre d'intervenants mensuel à la taille moyenne des fils (intensité des discussions à l'intérieur des fils). Il montre plus clairement une augmentation du nombre d'intervenants dans les fils. Il n'était nullement nécessaire qu'il en soit ainsi (le nombre d'intervenants uniques aurait pu être stable malgré cette intensification).

## 7.2. Mise en évidence d'un effet structurel d'homophilie

Raisonnement en termes de pourcentage (comme précédemment), et non plus de chiffres bruts permet de mettre en évidence la présence d'homophilie: plus la proportion de fils initiés par un administrateur augmente, plus la contribution des autres (non-administrateurs) diminue en proportion, et vice-versa. Il n'y a aucune nécessité qu'il en soit forcément ainsi :

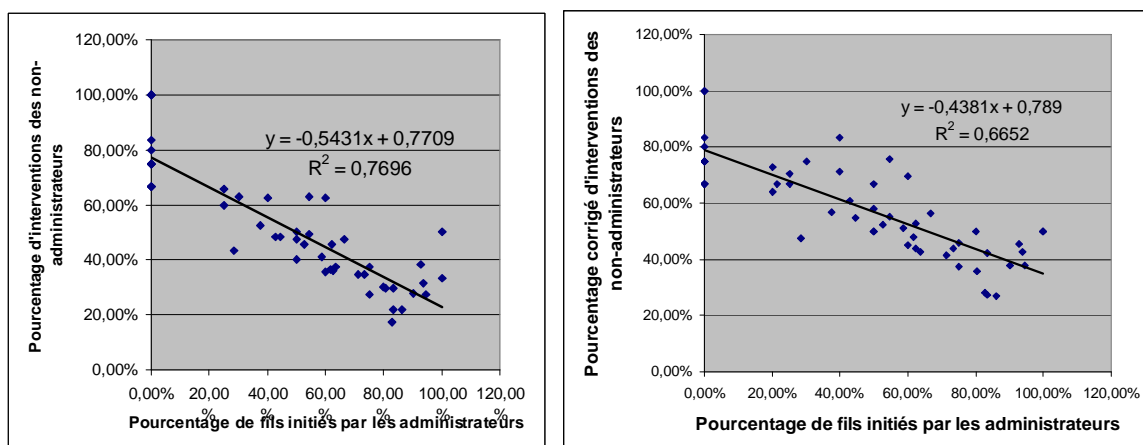


Figure 16 : régressions mettant en évidence l'éviction de la proportion d'interventions des non-administrateurs suite à l'augmentation de la proportion de prise d'initiative des administrateurs

Nous avons ici deux graphiques, l'un avec le pourcentage issu des données brutes du nombre d'interventions des non-administrateurs, l'autre avec des données d'où ont été exclues de la série chronologique en question le nombre de contributions des administrateurs correspondant à l'initiation des fils, ne conservant que « l'intérieur » des fils. En effet, utiliser les données brutes aurait *pu* diminuer excessivement le pourcentage d'interventions des non-administrateurs, étant donné que le phénomène pertinent à étudier correspondait au reste du fil, où se trouve la série chronologique pertinente. Un statisticien dira en effet que le phénomène mis en évidence pourrait se comprendre en partie, étant donné qu'un nombre significatif d'interventions des administrateurs se trouve *déjà* en début de fil. Pour donner un exemple parlant, s'il y a 100% de fils commencés par des administrateurs un mois donné, il serait probable de trouver en conséquence peu de contributions des non-administrateurs sur le reste du mois, même si ce n'est pas toujours forcément le cas. Exclure des données le nombre d'interventions des administrateurs au tout début du fil permet de considérer *uniquement* la série chronologique correspondant au reste du fil. *On constate que prendre en compte ce biais ne modifie qu'à la marge le phénomène observé, comme le montre le graphique de droite qu'on pourra comparer au graphique de gauche.* Le R2 n'a diminué que de 0,1 avec une régression linéaire standard. Ce « test de robustesse » du modèle est peut-être cependant *trop sévère*, au vu de la taille moyenne des fils (3,4). Cela mériterait d'être discuté.

Une interprétation immédiate<sup>21</sup> serait de penser que les administrateurs se parlent et se reconnaissent entre eux ; c'est un fait stylisé curieux concernant l'organisation massivement distribuée, en particulier sur la structure de régulation et de management qu'est la liste de discussions. L'explication habituellement retenue en analyse de réseaux est celle de *l'attachement préférentiel*.

*Le problème est d'expliquer, ou de mieux comprendre un tel phénomène d'attachement préférentiel*, ce qui supposerait une analyse qualitative, mais aussi une analyse de réseaux, ce à quoi nous arrivons.

## 8. Une approche de la structuration de la liste de discussions à partir de l'analyse de réseau

Voici donc une visualisation préalable du réseau des listes de discussions « docs » et « wp-docs ». Nous avons adopté la classique présentation « Kawada-Kawai » qui place les intervenants les plus connectés au centre :

---

<sup>21</sup> Au vu de la sévérité du test et pour plus de commodité, nous continuerons donc par la suite avec les données brutes.

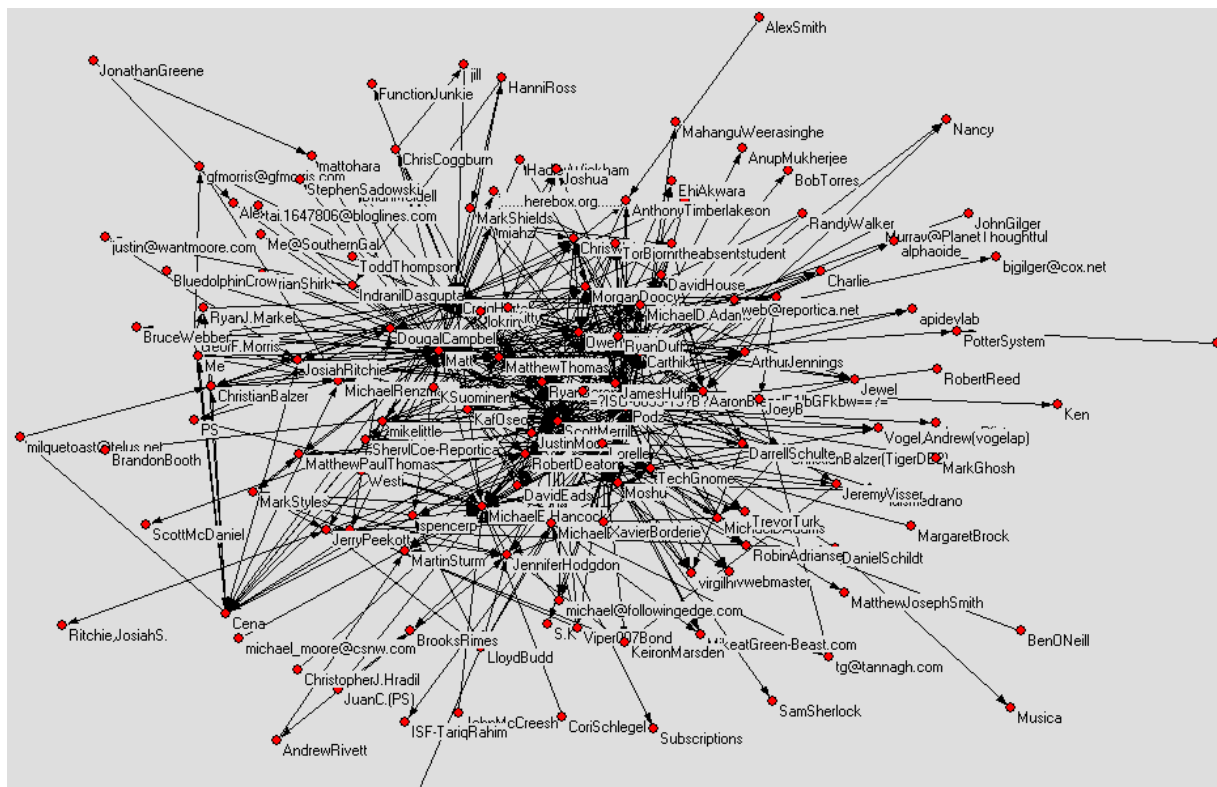


Figure 17 : visualisation du réseau des intervenants avec les plus prolifiques au centre (« kawada kawai »)

Une première observation permet de dire à vue d'œil que *la centralisation est en fait relative*, car un certain nombre d'individus sont fortement connectés. Elle ne concerne pas seulement les administrateurs mais aussi certains programmeurs : il reste en effet des zones noires denses au centre du graphe. Cependant, les administrateurs sont à première vue les individus avec le plus fort taux de connexions, une hypothèse qu'il faudra confirmer. Pour prendre une comparaison géographique parlante, on constate ici que la centralisation dans la communauté ressemble plus à la dispersion des grandes villes allemandes qu'à la centralisation française sur Paris.

On observe, cela est visible sur la *première* visualisation, peu d'interactions entre les contributeurs mineurs, ceux qui interviennent une ou deux fois; le caractère "distribué" de l'organisation est donc en fait limité sur ce point, et nous reviendrons définitivement sur la connectivité par la suite. En revanche, on observe que les interactions *se diversifient assez vite* dès que le nombre d'interventions augmente. Il y aurait quatre catégories d'intervenants, ceux qui sont centraux, interviennent de façon conséquente; une deuxième couronne de non-administrateurs mais informaticiens chevronnés (qu'on ne distinguait pas dans les statistiques qui opposaient deux populations), ceux qui ont un petit noyau d'interlocuteurs diversifiés, mais qui interviennent moins; et enfin une constellation de petits intervenants isolés.

Nous avons calculé à partir des distributions des demi-degrés entrants et sortants la somme de ces deux variables pour l'ensemble de la population pour vérifier si les administrateurs étaient effectivement les intervenants les plus connectés. C'est effectivement le cas sur ce tableau partiel; ils sont en majuscules :

1 "Cena"	76
73 "MICHAELH"	78
51 "MDAWAFFE"	95
36 "OwenWinkler"	133
18 "SKIPPY"	140
56 "LORELLE"	211
4 "CRAIG"	306
13 "CARTHIK"	376
14 "PODZ"	414
2 "MATT"	431

Figure 18 : somme des demi-degrés entrants et sortants chez les principaux intervenants

Voici des résultats descriptifs obtenus avec la commande « Network Analysis Toolkit » de Network Workbench Tool:

“Nodes: 129  
 Isolated nodes: 0  
 Average total degree: 27.410852713178212  
 Average in degree: 13.705426356589154  
 Average out degree: 13.705426356589152  
 This graph is weakly connected.  
 There are 1 weakly connected components. (0 isolates)  
 The largest connected component consists of 129 nodes.  
 This graph is not strongly connected.  
 There are 48 strongly connected components.  
 The largest strongly connected component consists of 82 nodes.  
 Did not calculate density due to the presence of self-loops and parallel edges.  
 Many algorithms will not function correctly with this graph.”

Ce qu’il faut retenir, c’est que ce graphe est intégralement connecté<sup>22</sup>, mais pas pour autant fortement connecté, ce qui confirme les premières remarques visuelles précédente sur les limites du caractère distribué de la collaboration. La taille maximale du sous-réseau fortement connecté consiste en 82 nœuds (individus). En revanche, la distance moyenne entre chaque paire de nœuds n’est pas très importante (le calcul a été effectué avec Pajek ici) :

Number of unreachable pairs: 5648  
 Average distance among reachable pairs: 2.55366  
 The most distant vertices: BenONeill (76) and Musica (84). Distance is 6.

Ensuite, l’indice de corrélation en un point («one-point correlations») indique la corrélation entre les arcs entrants (« demi-degré intérieur ») et les arcs sortants (« demi-degré extérieur ») des nœuds. Les nœuds avec nombre de degrés entrants sont dits « hubs » et les nœuds avec nombre de degrés sortants sont dits « autorités ». On peut calculer la corrélation entre les deux variables pour savoir si les « hubs » sont aussi des « autorités ». Si le coefficient de corrélation est significativement différent de 1, cela signifie qu’elle est significative. Ici, le

<sup>22</sup> Tout courriel obtient une réponse, ce qui ne va pas de soi contre l’apparence : un participant peut écrire sur la liste et ne pas obtenir de réponse.

coefficient est de 7.41, à deux décimales près, ce qui est fort significatif et appelle un approfondissement. On peut donc tenter de regarder avec profit les formes de régularités présentes ici :

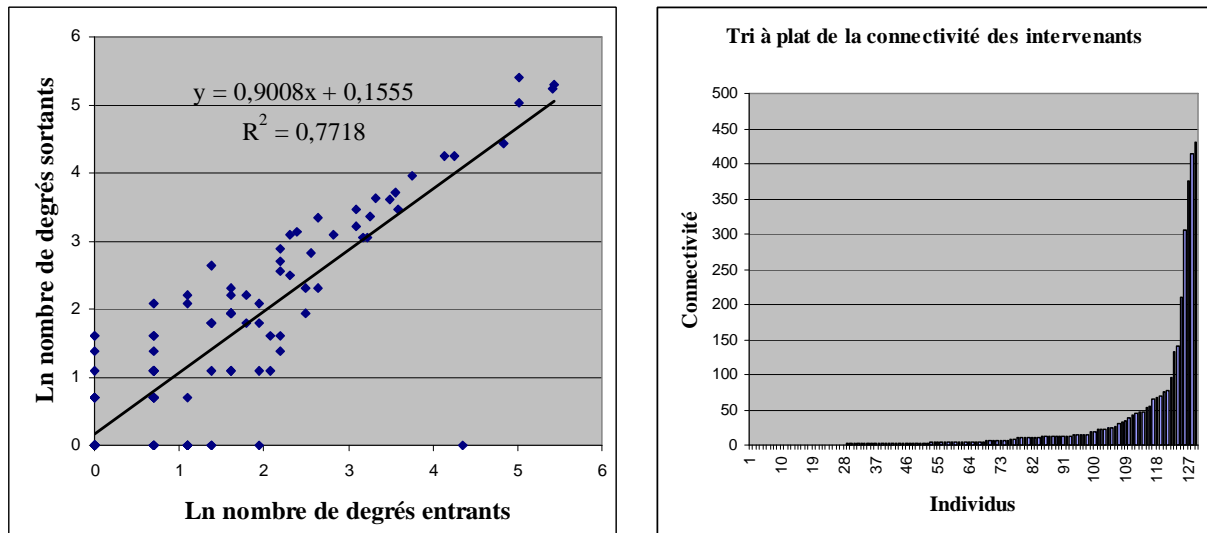


Figure 19 : mise en relation des degrés entrants et sortants et tri à plat des valeurs observées

A gauche, nous constatons que la distribution des demi-degrés intérieurs et extérieurs sont corrélés de façon significative, avec une certaine hétéroscédasticité sans conséquence pour les valeurs les plus faibles. A droite, nous avons présenté un tri à plat de la connectivité de chaque intervenant, qui forme une loi de puissance. On appelle cela une *distribution scale-free*. Les distributions des demi-degrés intérieurs et extérieurs suivent aussi des lois de puissance chacun de leur côté (nous ne présentons pas les graphiques). Précisons-le, les deux relations mentionnées sur les graphiques ci-dessus sont *indépendantes* : on peut avoir l'une sans avoir l'autre, et réciproquement. Avoir les deux est une caractéristique originale ici. L'apparition de telles caractéristiques avec une population réduite n'est pas triviale.

L'interprétation de la première relation signifie que les individus qui reçoivent des messages (« hubs ») sont aussi ceux qui en envoient (« autorités »), et ce, dans une mesure semblable. Cela signifie que l'autorité et l'expertise est distribuée de façon homogène. C'est là une composante à retenir du processus de collaboration, qui peut expliquer sa solidité : il ne dépend ni de certains « hubs », ni de certaines « autorités », si elles venaient à disparaître au point de créer des « trous » dans le réseau.

On peut pousser l'analyse un plus loin en regardant les caractéristiques de la collaboration en examinant justement la réciprocité entre intervenants. Quel est le nombre de nœuds concernés par des relations réciproques (au moins une), i.e. le nombre de dyades ? On peut le découvrir avec les commandes adéquates du logiciel. Dans le réseau, il y a 481 paires, 181 sont des dyades, ce qui fait un ratio de réciprocité de 0,376. On peut aussi examiner la prévalence de la réciprocité par le ratio d'arcs impliqués dans ces relations (362), qui est de 0,546. La collaboration peut également être comprise en examinant le ratio de triades transitives<sup>23</sup> présentes dans le réseau : 0,269.

<sup>23</sup> Le logiciel par exemple dit : "A triad is any triple of nodes (actors) {A, B, C}. In a directed network there are sixteen possible kinds of triads. A triad is said to be transitive if there is a link (tie) from A to B (AB) and a link from B to C (BC), then there is also a link from A to C (AC)."



Les statistiques descriptives montrent qu'il existe 48 nœuds fortement connectés. Nous avons par ailleurs obtenu avec la même commande du logiciel la liste des demi-degrés intérieurs de tous les intervenants (placée en annexes, car longue). Nous avons choisi un seuil de 5 qui correspond approximativement aux dyades où il y a au moins un administrateur. Cela permet d'obtenir une liste de 77 ensembles de « relations » à 5 arcs ou plus<sup>24</sup>, et donc 77 relations de collaborations. A contrario, on peut observer des relations plutôt faibles entre certains administrateurs, ce qui montre les *limites à la marge* de la collaboration dans cette sous-population. Nous dégageons ces relations dans la liste pour les plus gros intervenants où la remarque a une pertinence certaine :

“There are 2 edges between node: CRAIG and node: LORELLE  
There are 3 edges between node: LORELLE and node: MICHAELH  
There are 3 edges between node: SKIPPY and node: LORELLE  
There are 3 edges between node: SKIPPY and node: CARTHIK”

## 9. Conclusions

La gouvernance d'un système d'information (le wiki nommé Codex) est un processus, *un processus de collaboration*, qui se fait de façon incontournable sur une liste de discussions. Son étude révèle que cet objet dynamique comporte des *propriétés* marquées, au-delà des considérations sur l'existence d'un noyau central de contributeurs (comme chez les informaticiens de l'open-source par exemple). Nous avons caractérisé les dynamiques entre autres démographiques de l'encadrement par les administrateurs, l'existence d'homophilie ou de réciprocité dans les interactions. Le processus de gouvernance comporte donc des propriétés structurelles marquées caractéristiques des systèmes dynamiques en économie cognitive, et ce, sur une population pourtant assez réduite : il n'est aucunement vraisemblable qu'un tel processus nécessaire à la viabilité du projet se soit concrétisé sur les parties du wiki prévues pour la discussion. Comme l'analyse de réseau par exemple l'a montré, les liens entre les participants au sous-projet de documentation sont faibles, à l'exception d'un petit noyau. La coordination de la documentation d'un projet open-source n'avait pas été étudiée jusqu'ici.

Ce travail permettrait de nuancer le modèle d'interaction dit « tableau noir » centralisé que propose par exemple Gensollen (2005), pourtant a priori particulièrement pertinent pour les wikis : un tel modèle n'est *pas le seul* applicable au Codex et certainement à d'autres wikis ; il existerait un réseau de relations sur un autre support dont le rôle est dans une certaine mesure essentiel, et ce au minimum lors des phases initiales du projet. Il se caractérise par des dynamiques propres. Rappelons qu'on classe d'ordinaire les wikis parmi les *outils relationnels*, on entend par là que l'outil permet une mise en relation des individus. En fait, la socialisation se fait sur la liste de discussion, pas sur le « corpus », ce qui nuance ce modèle. D'un point de vue simplement factuel, l'étude d'un tel processus permet d'aller au-delà d'une description en « structure en oignon » (Crowston et al., 2006) proposée pour *l'ensemble* des participants à un projet : nous apportons ici une description plus précise, sur une *partie* du projet, et prenant en compte sa dynamique historique

---

<sup>24</sup> “There are 6 edges between node: MartinSturm and node: JenniferHodgdon.

There are 6 edges between node: JosiahRitchie and node: Cena”.

Cela signifie que la relation est orientée de MartinSturm à JenniferHodgdon, i.e ; que le premier répond six fois au second dans un ou plusieurs fils.

## 10. Annexes : archives numériques

Le wiki de la communauté appelé Codex :

[http://codex.wordpress.org/Main\\_Page](http://codex.wordpress.org/Main_Page)

Les archives des listes de discussions :

<http://lists.automattic.com/mailman/listinfo>

Les archives des anciennes listes de discussions non répertoriées dans l'entrepôt officiel :

<http://comox.textdrive.com/pipermail/docs/>

L'archivage de l'évolution de la communauté par WebArchives :

[http://web.archive.org/web/\\*/http://wordpress.org/](http://web.archive.org/web/*/http://wordpress.org/)

Le blog de WordPress dédié au développement technique du projet:

<http://wordpress.org/development/> .

## 11. Bibliographie

- Aguiton, C. et Cardon, D. (2007), "The Strength of Weak Cooperation: an Attempt to Understand the Meaning of Web 2.0", *Communications & Stratégies*, 65(1), 51-65.
- Andersen E. (2004), "Using Wikis in a Corporate Context", In Hohenstein A. and Wilbers K. (eds), *Handbuch E-Learning*, WoltersKluwer, Cologne, pp. 5-15.
- Bac C. et al. (2005), "Why and how to contribute to libre software when you integrate them into an in-house application ?", *Proceedings of the 1st International Conference on Open Source Systems*, Genova, Italy.
- Barcellini F., et al. (2005), "Thematic Coherence and Quotation Practices in OSS Design-Oriented Online Discussions", GROUP'05, Sanibel Island, Florida.
- Buffa M. (2006), "Intranet Wikis", *WWW Conference 2006*, Edinburgh.
- Chau T. et Maurer F. (2005), "A case study of wiki-based experience repository at a medium-sized software company", *Proceedings of the 3rd international conference on Knowledge capture*, Banff, Alberta, pp. 185-186.
- Chen H., et al., (2006), "Social Network Structure behind the Mailing Lists", ICT-IIIS, Pékin.
- Cheng S.-W., et al. (2005), "WICSA Wiki WAN Party: capturing experience in software architecture best practices, ACM SIGSOFT Software Engineering Notes archive", 30-1.
- Conein B. (2004), « Relations de conseil et expertise collective: comment les experts choisissent-ils leurs destinataires dans les listes de discussion? », *Recherches sociologiques*, 35-3, pp. 61-74.
- Cooke, M. (2008), "The new world of Web 2.0 research", *International Journal of Market Research*, 50(5), 569-572.
- Craig T. (2005), "Open Source Is Coming: Here's How to Deal With It", *2005 Pacific Northwest Software Quality Conference*, Portland, Oregon.
- Crowston K. and Howison J. (2005), "The social structure of Free and Open Source software development", *First Monday*, 10-2.
- Cubranic D. et al. (2003), "Tools for light-weight knowledge sharing in open-source software development", *3rd Workshop on Open Source Software Engineering*, Portland, Oregon.
- Degenne A. et Forsé M. (1994), *Les réseaux sociaux*, Armand Colin, Paris, 263 p. .
- Den Besten M. et Dalle J.-M. (2008), "Keep it simple, a Companion for Simple Wikipedia", *Industry & Innovation*, 15-2, pp. 169-178.

- Dorat R., Latapy M., Conein B. et Auray N. (2007), « Multi-level analysis of an interaction network between individuals in a mailing list », *Annales des Télécommunications*, 62, pp. 325-349.
- Eto K. et al. (2005), “Qwik-Web: integrating mailing-list and WikiWikiWeb for group communication”, *WikiSym 2005*, San Diego.
- Froomkin M. (2003), “Habermas@discourse.net:Toward a Critical Theory of Cyberspace”, *Harvard Law Review*, 116-3, pp. 751-873.
- Gensollen M. (2005), « Des réseaux aux communautés: la transformation des marchés et des hiérarchies », in Proulx S., Poissant Louise et Sénécal Michel (eds.), *Communautés virtuelles : penser et agir en réseau*, Presses de l'Université Laval, Montréal, 107-132
- Girieud, S. (2008), “Web 2.0 and social networking”, *Communications & Strategies*, 71(3), 183-188
- Gunnar R. (2005), “Case Studies of Online Deliberation: The Debian Project and Wikipedia”, *Online Deliberation 2005 Conference*, Stanford.
- Harrer A. et Zeini S. (2006), “Evaluation of communication in web-supported learning communities, an analysis with triangulation research design”, *International Journal of Web Based Communities*, 2-4, pp. 428-446.
- Jullien N. et al. (2002), *Projet RNTL: «Nouveaux modèles économiques, nouvelle économie du logiciel»*, *Rapport final*.
- Karinbegg B. (1999), “The Rewards of Managing an Electronic Mailing List”, *Library Trends*, 47-4, pp. 686-698.
- Klein M. et Iandoli L. (2008), “Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The Mit Collaboratorium”, MIT Sloan Research Paper, Cambridge.
- Kolbitsch J. et Maurer H. (2005): “Community Building around Encyclopaedic Knowledge”, *Journal of Computing and Information Technology*, 13, pp. 175-190.
- Korfiatis N. et al. (2006), “Evaluating authoritative sources using social networks: An insight from Wikipedia”, *Online Information Review*, 30-3, pp. 252-262.
- Korfiatis N. et Naeve A. (s.d.), “Evaluating Wiki Contributions using social Networks: a Case Study on Wikipedia”, Royal Institute of Technology (KTH) Working Paper, Stockholm.
- Krogh G., Haefliger S., Spaeth May S. (2003), “Collective Action and Communal Resources in Open Source Software Development: The Case of Freenet”, Working Paper.
- Koushik, S., Birkinshaw, J. et Crainer, S. (2009), “Using Web 2.0 to create Management 2.0”, *Business strategy Review*, 20(2), 20-23.
- Martland David (2003), “Building effective open source communities using software tools”, *Proceedings of the Workshop on Open Source Software Movements and Communities*, Gothenburg, Sweden.
- McAfee, A. (2006), “Enterprise 2.0: the dawn of emergent collaboration”, *Engineering Management Review*, IEEE, 34(3), 38-38.
- Matsumura Naohiro et Sasaki Yoshihiro (2005), “Leader Qualification in Managing Nonprofit Organization”, Working Paper.
- Mencarelli, R. et Pulh, M. (2009), « La communication 2.0 : un dialogue sous conditions », *Décisions Marketing*, n° 54, 71-75.
- Robbins (2003) “Adopting Open Source Software Engineering (OSSE) Practices by Adopting OSSE Tools”, in Feller J., Fitzgerald B., Hissam S. & Lakham K. (Eds.), *Making Sense of the Bazaar: Perspectives on Open Source and Free Software*, O'Reilly & Associates, Sebastopol, Canada.
- Ruzé E. (2009), « Formes de régulations et de gouvernance formelles des communautés en ligne. Ce que dit la littérature et ce qu'il faudrait savoir », rapport de recherche publié sur le site de la chaire « Innovation et Régulation de l'Economie Numérique » (Ecole polytechnique – TelecomParisTech - Orange).

- Ruzé E. (2009), « Traiter les archives de la Toile. Une histoire d'un système d'information dans une communauté : WordPress (2003-2008) », *Entreprises et histoire* 2009/2 n° 55, pp 74-89.
- Ruzé E. (2011), « Une approche quantitative des archives d'un projet numérique, Wikipedia », *Entreprises et histoire*, 63, pp 88-99.
- Schwall J. (2003), "The wiki phenomenon", Working Paper "Internet Literature seminar", Munster.
- Serakiotou N. (2008), "(Wiki + ResTechs) = (fresh documentation + organic knowledge management + training materials + good, cheap technical writers)", *Proceedings of the 36th annual ACM SIGUCCS conference on User services conference*, Portland, USA, pp. 173-180.
- Sowe S., Stamelos I., et Angelis L. (2006), "Identifying knowledge brokers that yield software engineering knowledge in OSS projects Information and Software Technology", *Journal of Systems and Software*, 81-3, pp. 431-446.
- Szybalski A. (2005), "Why it's not a wiki world (yet)", Working Paper.
- Tang L. et al. (2008), "A Method for Measuring Co-authorship Relationships in MediaWiki", *Proceedings of Wikisym 08*, Porto.
- UNDP (2005), "Building Online Communities of Practice : The International Open Source Network Model", E-Note 5.
- Voss J. (2005), "Measuring Wikipedia", *10th International Conference of the International Society for Scientometrics and Informetrics*, Stockholm.
- Wagner C. et Majchrzak A. (2007), "Enabling Customer-Centricity Using Wikis and the Wiki Way", *Journal of Management Information Systems*, 23-3, pp. 17 - 43.
- Weiss A. (2008), "Ecology and Dynamics of Open Source Communities", Working Paper.
- Xiao W. et al. (2007), "On-line collaborative software development via wiki", *Proceedings of the 2007 international symposium on Wikis*, Montreal, pp. 177 – 183.