

# PhD proposal

---

---

## 1. Context and Goals.

Mobilizing hundreds (Linux) to thousands of contributors (Wikipedia), volunteer online open projects aiming at creating new knowledge, online “communities of creation”, as named by Rullani and Haefliger [38] (i.e, communities aiming at producing a certain kind of knowledge, or “epistemic communities” [7] and doing it online), are viewed as central in the generation of new, innovative knowledge [30]. In the same time, research proved that knowledge and innovation production (patents, research papers) is increasingly produced by teams [47]. How do these online communities form the teams, what are their characteristics in terms of participants’ specificities and complementarities? These are the questions we want to address in this PhD proposal.

From an IT school (Institut Mines-Télécom) and Fondation Télécom viewpoint, understanding how these communities work and the skills their participants develop may result key for the future IT professionals we are teaching. This works aims at continuing the collaborative research done at Institut Mines Télécom on these online communities via and ANR project (CCCP-Prosodie, PI: Nicolas Jullien involving LUSI-iSchool and the team SPE (Sociologie, Psychologie et Ergonomie) from SES department at Télécom ParisTech) and a Futur & Rupture grant (data extraction in Wikipedia, visiting scientist Felipe Ortega), and with the iSchool, Syracuse University (USA, Pr Kevin Crowston).

Those open online communities are examples of “knowledge commons” [19]

and the models developed are specifications of Hess and Ostrom's framework to understand the production of such commons (p. 44). Those frameworks distinguish the characteristics of the community, or the "inputs" ("biophysical characteristics", people or "attribute of the community", "rules-in-use"), which constrain the way people interact ("the action arena", or the process), leading to "outcomes". The outcomes can be apprehended at several levels: a productive level (the various characteristics of the constructed epistemic knowledge such as completeness, creativity... as well as its utility/usability, more user-oriented), a collective level (e.g. team building, construction of rules and collective norms...), and a personal developmental level (learning and development of individuals).

In this PhD, we will focus on the impact of the attribute of the community and of the action arena on the outcomes (the construction of the knowledge), exploring one of the main questions raised in the literature and by practitioners, the characteristics of a "good" team to produce new knowledge, in terms of participants' skills, experience, number, etc. and in terms of leadership [33]. These points go beyond the simple case of open online communities: the management of those virtual teams echoes the questions of (and the growing research literature on) the collective production of knowledge and of the functioning of the groups producing it . The case study will be Wikipedia, because of the availability of the data, but also because of its connexion with firms' knowledge management and production challenges. Studying these community may help us to better understand how groups work, how (virtual) collaboration and leadership can succeed, something of growing importance for firms [8, 18].

Regarding the online production of articles (Wikipedia), and according to a recent review of the literature [20], the summary of the state of the art can be led to Arazy et al. [2], even if they focused only on a very small subset of

articles (96): "(1) diversity should be encouraged, as the creative abrasion that is generated when cognitively diverse members engage in task-related conflict leads to higher-quality articles"<sup>1</sup>; (we will just add "up to a point" here) (2) "groups should maintain a balance of both administrative- and content-oriented members, as both contribute to the collaborative process." This echoes more general findings about the efficiency of groups. As shown by Uzzi and Spiro [45] in the case of musical comedies, and Uzzi [44] in the case of a social network, for a creative group to be successful, it needs to fine tune the level of newcomers, for fresh ideas, in an already constituted group [13, 10]. One of the goal of the PhD work will be to calculate Wikipedia's "Q"-level "bliss point", which may depend on the type of article produced, general or more specialized [22].

Interestingly, this seems to confirm also certain results of personnel economics studies on team working, which show that heterogeneous teams are more productive than heterogeneous isolated workers, in the case of low level skill workers [31, 14]. But their main argument is quite different, as they argue that social pressure makes the less productive work harder, even if Hamilton et al. stressed that the internal learning process may also play a role. This

---

<sup>1</sup>The definition of the "quality" of the production is also matter of debate. For a discussion regarding open source software, see Koch [24]. As far as Wikipedia is concerned, even the internal label (Feature article) is disputable in terms of stability from one language to another [39] and in terms of respect of the criteria [29]. External criteria, mainly coming from library studies, based on [21]'s criteria (e.g., purpose, authority, scope) are also difficult to apply to Wikipedia, especially because no authorship analysis is possible [46, 11], but also, as pointed by Lewandowski and Spree [28], "due to the overall scale and the wide range of subject areas, most of the studies focus on specialized fields of knowledge". Considering this problem, two strategies have been developed to evaluate the quality of Wikipedia, defining a subset of article to be analyzed, either looking at randomly chosen articles, or looking at a sub-project or a topic. In both cases, criteria have to be defined. The most comprehensive attempt to do so may be the ones by Stvilia et al. [41], and Lewandowski and Spree [28]. The firsts proposed 11 criteria, based on more global analysis of quality in on-line projects [40], which compared the FA articles to other articles regarding these criteria. The later relied on these criteria, extended them to a list of 13, and evaluated the correlation between these criteria and the rank in search engine, with a good correlation but a strong dispersion. As stand by the authors, this does not solve the subjective aspect of the criteria (or the fact their evaluation depends on the evaluator), but the aim is to propose an evaluation grid.

debate stresses the second key explanation given and studied to explain the fact these groups function, the specificity of their leadership [33, 30], based on social interactions (processes) regarding the production of the knowledge.

Leadership can be defined as a process that results in the reinforcement, creation and evolution of ongoing structures and distinguish between two types of leadership [17]. Following them, the hypothesis this work will rely on is that there is a first-order leadership as leadership that works within and reinforces existing structures to elicit and guide effective group contributions and a second-order leadership regrouping the behaviors that effect changes in the structure that guides group action. Both are therefore action embedded, grounded in processes that define the social identity of the team, and this work plans to investigate this model by using data about contributions (and contributors) to a project, coding them as first and second order leadership, and examining the structure of these contributions.

These are the main goal and hypotheses we want to address and test in this work, in the case of the Wikipedia community of creation (the case of a more firm-centered community may be address upon time and availability of data).

## **2. Scientific work and organization.**

The three phases of the project are as follow: evaluating the structure of 1) the groups and of 2) their leadership, via data extraction and mining, regarding the characteristics of the participant Arazy et al. [2]ts and of the knowledge produced; 3) discussing the results according to the existing theories of group performance and theories of management.

### *2.1. Structure of the groups.*

In addition to the availability of the data, and to the previous works made by the team proposing this PhD on this project, there are some preliminary

results on the structure of Wikipedia teams. Kittur and Kraut [23] showed that explicit coordination (talk) is more efficient when there are few editors, when implicit coordination (few editors concentrate the main part of the edits when the majority is peripheral editor) is more efficient when there are more editors. They also found that explicit coordination is needed more at the early stage of the article. In any case, there is a core-periphery structure, similar to the one found in open source software production, and things are easier when the core team people already know each others: Nemoto et al. [35] pointed out that “the more cohesive and more centralized the collaboration network, and the more network members were already collaborating before starting to work together on an article, the faster the article they work on will be promoted or feature”. Interestingly for our question, Turek et al. [42, p. 22] showed, using Polish Wikipedia data set, that in what they call “good teams”, the level of acquaintance is higher than for normal teams (people having discussed in the talk pages) as is the level of trust (copy-pasting of existing text when rewriting an article) and of distrust (deletion of text), which can be seen as the level of creative work (if people delete more that means that the consensus is reached more slowly, after more evaluation of the proposals). This is true for Feature articles (FA, articles graded of very good quality by the Wikipedia organization), but also when articles’ quality is measured by external experts, as in Arazy and Nov [1], who estimated the impact of local inequality and global inequality on the quality of the article: having a small team, very committed (strong local inequality), improves the coordination (and thus indirectly the quality), and having strong global inequality (people very invested in Wikipedia and peripheral contributors) improves the quality of the articles (this work needing to be extended to a bigger set of articles to be confirmed). Xu et al. [48] can be seen as summarizing these findings: using an agent simulation, they retrieved these

results, showing that more agents improve the convergence and the quality of the article, especially if they are more knowledgeable, and vandalism, if increasing the number of updates, does not stop an article from being improved (it can be seen as test which allows to question the team and eventually improves its production).

This PhD work, after discussing those findings, will propose longitudinal studies to evaluate the links developed between people working on the same pieces of knowledge or the same type of knowledge (article/files or sub-projects), to measure more precisely the size of the teams and the level of connection between the members and their impact of group performance, discussing Guimera et al. [13]’s finding. This requires capacities in data extraction, data mining and especially social network analysis. This capacities are available at LUSI-iSchool department, in the DECIDE research team, which will support the student on that part. We will closely work with Felipe Ortega, who published, during his visiting period at LUSI-iSchool funded by the Futur et Rupture 2012 call, a toolkit to extract and analyze the Wikipedia data<sup>2</sup>.

## *2.2. Leadership.*

One of the most important task for the leaders or the animators of virtual communities is to involve people and to keep them involved, according to the literature on virtual management (see Crowston et al., 2010, p. 15). According to Zhu et al. [49], in Wikipedia too, leadership behaviors matter to motivate people to participate, especially when “transactional leadership and person-focused leadership” are used. Illustrating this point, Billings and Watts [3] showed that thanks to the wiki tool which allows citations, conciliators solve disputes by helping differentiate “the personal and substantive” (p. 6) in close

---

<sup>2</sup>Wikipedia Data Analysis Toolkit (WikiDAT) hosted in GitHub: <http://glimmerphoenix.github.io/WikiDAT/> this project funded also a computer to process the queries and the analyses, which will be used for the PhD work.

interaction with the disputants, and Musicant et al. [34] found "significant correlations between editor communication and article editing activity" (but without being able to evaluate the direction of the correlation). On line with our theoretical hypothesis, two levels of leadership are said to exist in Wikipedia. A project leadership, focused on content, where discussion and coordination are very linked to contribution at article level [43], with strong socialization effects [12], and a more global management, aiming at solving the cases unsolved<sup>3</sup>. Here again, an analysis of people's comparison of relative status [5, 27] may shed light on the process of interaction and leadership creation<sup>4</sup>.

Still relying on longitudinal analysis of dump files data of the career of the people (especially administrators, people responsible for sub-projects in Wikipedia) we will study the characteristics (described by the practices) of the various animators of Wikipedia. Once again data collection and analysis competences are key to this part of the work.

Finally, this organization, and the evolution of the participants from personal interest regarding the contribution toward collective interest, but also the importance of the discussion about the building of trust to become a regular contributor, echo reflexions on management, especially stewardship practices [9], and the alignment of people's interest and collective's interest [18]. This

---

<sup>3</sup>To be exhaustive, a third level of leadership should be considered, the Wikimedia foundation, and the process of election of the members designed by the participants, studied, something out of the scope of this PhD proposal.

<sup>4</sup>The admin election process gives insights of how these two levels articulate. Ortega and Gonzalez Barahona [36] showed that the admins are not the ones who contribute the more to the articles, and Burke and Kraut [4] extended this point showing that the candidate's article edits were weak predictors of success: they have to demonstrate also managerial behaviors. Diverse experiences and contributions to the development of policies and to WikiProjects are stronger predictors of RfA [Request for Adminship] success. Future admins also use article talk pages and comments for coordination and negotiation more often than unsuccessful nominees, and tend to escalate disputes less often. In addition to this, Leskovec et al. [27] showed that the voters favor people who have the same characteristics than them, i.e. who are on comparable or superior merit (and vote negatively for those who are of lower merit), especially when these people are in minority. Cabunducan et al. [6] showed that "voters tend to participate in elections that their contacts have participated in" and that "candidates who gain the support of an influential coalition tend to succeed in elections".

leads to the last part of the PhD work, an more general analysis and discussion of the process.

### *2.2.1. Theoretical study.*

The results are expected to help fuel the discussion about the structure of governance and management present a Wikipedia, as an example/model of management for virtual teams. According to the literature, Wikipedia would be a new system controlled by an oligarchy based on expertise [15, 37], creating, according to Konieczny [25, 26], and his detail discussion of the model of governance in Wikipedia, a Mintzberg’s model of adhocracy, “one closely connected to open-source development models found in the FOSS movement”. Editors at Wikipedia would ”share the adhocratic values of flat hierarchy, decentralization, little managerial control, and ad-hoc creation of informal multidisciplinary teams. However, if, according to him, ”in traditional adhocracies, individuals are bound by rules that cannot be altered; at Wikipedia, by contrast, there is no rule that cannot be altered if the community so desires”, beside the main iron rules (the five pillars), seemingly. In addition to that, “in Wikipedia’s adhocracy, the editors not only “capture opportunities,” but they also can create those opportunities, since editors can change all policies and so enjoy an unprecedented degree of empowerment”, something stressed by [15, 16, 37]. The analyses of the structure of the teams, and of the characteristics of the leadership, will help to discuss if this model is a new implementation of the global adhocracy model or if it falls into the model proposed by Mintzberg and McHugh [32].

- [1] Arazy, O., Nov, O., 2010. Determinants of Wikipedia quality: The roles of global and local contribution inequality, in: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, pp. 233–236.
- [2] Arazy, O., Nov, O., Patterson, R., Yeo, L., 2011. Information quality in



- Wikipedia: The effects of group composition and task conflict. *Journal of Management Information Systems* 27, 71–98.
- [3] Billings, M., Watts, L.A., 2010. Understanding dispute resolution online: Using text to reflect personal and substantive issues in conflict, in: *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1447–1456.
- [4] Burke, M., Kraut, R., 2008. Taking up the mop: Identifying future wikipedia administrators, in: *Conference on Human Factors in Computing Systems - Proceedings*, pp. 3441–3446.
- [5] Burt, R.S., 2009. *Neighbor Networks: Competitive Advantage Local and Personal*. Oxford University Press.
- [6] Cabunducan, G., Castillo, R., Lee, J., 2011. Voting behavior analysis in the election of Wikipedia admins, in: *Proceedings - 2011 International Conference on Advances in Social Networks Analysis and Mining*, IEEE. pp. 545–547.
- [7] Cohendet, P., Créplet, F., Dupouet, O., 2001. Interactions between epistemic communities and communities of practice as a mechanism of creation and diffusion of knowledge, in: Zimmermann, J.B., Kirman, A. (Eds.), *Interaction and Market Structure*. Springer, Londres.
- [8] Crowston, K., Heckman, R., Misiolek, N., 2010. Leadership in self-managing virtual teams. Syracuse University School of Information Studies working paper .
- [9] Davis, J.H., Schoorman, F.D., Donaldson, L., 1997. Toward a stewardship theory of management. *The Academy of Management Review* 22, 20–47.

- [10] Defélix, C., Chanal, V., Galey, B., 2005. Les personnes innovantes dans les entreprises doivent-elles faire l'objet d'une GRH spécifique ? Une étude exploratoire. *Gestion* 2000 mars-avril, 99–113.
- [11] Ehmann, K., Large, A., Beheshti, J., 2008. Collaboration in context: Comparing article evolution among subject disciplines in wikipedia. *First Monday* 13. URL: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2217/2034>.
- [12] Forte, A., Kittur, N., Larco, V., Zhu, H., Bruckman, A., Kraut, R.E., 2012. Coordination and beyond: social functions of groups in open content production, in: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, New York, NY, USA. pp. 417–426.
- [13] Guimera, R., Uzzi, B., Spiro, J., Amaral, L.A.N., 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 697–702.
- [14] Hamilton, H., Nickerson, J.A., Owan, H., Pollack, R., Koenker, R., Fitzenberger, B., Rao, A., Macdonald, G., Levitt, S., 2003. Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy* , 465–497.
- [15] Hansen, S., Berente, N., Lyytinen, K., 2009. Wikipedia, critical social theory, and the possibility of rational discourse. *Information Society* 25, 38–59.
- [16] Hartelius, E.J., 2010. Wikipedia and the emergence of dialogic expertise. *Southern Communication Journal* 75, 505–526.
- [17] Heckman, R., Crowston, K., Misiolek, N., 2007. *A structural perspective on leadership in virtual teams*, Springer, Portland, OR. pp. 151–168.

- [18] Hernandez, M., 2012. Toward an Understanding of the Psychology of Stewardship. *The Academy of Management Review (AMR)* 37, 172–193.
- [19] Hess, C., Ostrom, E. (Eds.), 2006. *Understanding Knowledge as a Commons. From Theory to Practice*. MIT Press.
- [20] Jullien, N., 2012. What We Know About Wikipedia: A Review of the Literature Analyzing the Project(s), in: Working paper, M@rsouin, Télécom Bretagne.
- [21] Katz, W.A., 2002. *Introduction to reference work, vol. 1: Basic Information Services*. 8th ed., McGraw-Hill, Boston, MA.
- [22] Keegan, B., Gergle, D., Contractor, N., 2012. Do editors or articles drive collaboration?: multilevel statistical network analysis of wikipedia coauthorship, in: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, New York, NY, USA. pp. 427–436.
- [23] Kittur, A., Kraut, R.E., 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination, in: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM, New York, NY, USA. pp. 37–46.
- [24] Koch, S., 2011. Organisation of work in open source projects: expended effort and efficiency. *Revue d'économie industrielle* 136, 17–38.
- [25] Konieczny, P., 2009. Governance, organization, and democracy on the Internet: The iron law and the evolution of Wikipedia. *Sociological Forum* 24, 162–192.
- [26] Konieczny, P., 2010. Adhocratic governance in the internet age: A case of wikipedia. *Journal of Information Technology and Politics* 7, 263–283.

- [27] Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010. Governance in Social Media: A Case Study of the Wikipedia Promotion Process, in: Proceedings of the International AAAI Conference on Weblogs and Social Media. URL: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1485>.
- [28] Lewandowski, D., Spree, U., 2011. Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science and Technology* 62, 117–132.
- [29] Lindsey, D., 2010. Evaluating quality control of wikipedia’s feature articles. *First Monday* 15. URL: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2721/2482>.
- [30] Mahr, D., Lievens, A., 2012. Virtual lead user communities: Drivers of knowledge creation for innovation. *Research Policy* 41, 167 – 177.
- [31] Mas, A., Moretti, E., 2009. Peers at Work. *American Economic Review* 99, 112–145.
- [32] Mintzberg, H., McHugh, A., 1985. Strategy formation in an adhocracy. *Administrative Science Quarterly* 30, 160–197.
- [33] Misiolek, N., Crowston, K., Seymour, J., 2012. Team dynamics in long-standing technology-supported virtual teams, in: *Academy of Management Annual Meeting, Organizational Behaviour Division*, Boston, MA. URL: <http://crowston.syr.edu/sites/crowston.syr.edu/files/ACADEMYPAPERFINAL.pdf>.
- [34] Musicant, D.R., Ren, Y., Johnson, J.A., Riedl, J., 2011. Mentoring in Wikipedia: A Clash of Cultures, in: *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, ACM, New York, NY, USA.

- [35] Nemoto, K., Gloor, P., Laubacher, R., 2011. Social capital increases efficiency of collaboration among wikipedia editors, in: Proceedings of the 22nd ACM conference on Hypertext and hypermedia, ACM, New York, NY, USA. pp. 231–240.
- [36] Ortega, F., Gonzalez Barahona, J.M., 2007. Quantitative analysis of the wikipedia community of users, in: Proceedings of the 2007 international symposium on Wikis, ACM, New York, NY, USA. pp. 75–86.
- [37] Pfister, D.S., 2011. Networked expertise in the era of many-to-many communication: On wikipedia and invention. *Social Epistemology* 25, 217–231.
- [38] Rullani, F., Haefliger, S., 2013. The periphery on stage: The intra-organizational dynamics in online communities of creation. *Research Policy* 42, 941–953.
- [39] Stvilia, B., Al-Faraj, A., Yi, Y., 2009. Issues of cross-contextual information quality evaluation—The case of Arabic, English, and Korean Wikipedias. *Library & Information Science Research* 31, 232–239.
- [40] Stvilia, B., Gasser, L., Twidale, M.B., Smith, L.C., 2007. A framework for information quality Assessment. *Journal of the American Society for Information Science and Technology* 58, 1720–1733.
- [41] Stvilia, B., Twidale, M.B., Smith, L.C., Gasser, L., 2008. Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology* 59, 983–1001.
- [42] Turek, P., Wierzbicki, A., Nielek, R., Hupa, A., Datta, A., 2010. Learning About the Quality of Teamwork from Wikiteams, in: IEEE Second International Conference on Social Computing (SocialCom), pp. 17 –24.

- [43] Ung, H., Dalle, J.M., 2010. Project management in the wikipedia community, in: Proceedings of the 6th International Symposium on Wikis and Open Collaboration, ACM, New York, NY, USA. pp. 13:1–13:4. doi:<http://doi.acm.org/10.1145/1832772.1832790>.
- [44] Uzzi, B., 2008. A social network's changing statistical properties and the quality of human innovation. *Journal of Physics A: Mathematical and Theoretical* 41, 224023, 12pgs.
- [45] Uzzi, B., Spiro, J., 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology* 111, 447–504.
- [46] Wallace, D., Van Fleet, C., 2005. The democratization of information? wikipedia as a reference resource. *Reference and User Services Quarterly* 45, 100–103.
- [47] Wuchty, S., Jones, B., Uzzi, B., 2007. The Increasing Dominance of Teams in the Production of Knowledge 316, 1036–1039.
- [48] Xu, J., Yilmaz, L., Zhang, J., 2008. Agent simulation of collaborative knowledge processing in wikipedia, in: Proceedings of the 2008 Spring Simulation Multiconference, SpringSim'08, pp. 19–25.
- [49] Zhu, H., Kraut, R., Kittur, A., 2012. Effectiveness of shared leadership in online communities, in: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, ACM, New York, NY, USA. pp. 407–416.