The Rise and Fall of an Online Project. Is Bureaucracy Killing Efficiency in Open Knowledge Production?

Nicolas Jullien
LUSSI-iSchool, ICI-M@rsouin,
Télécom Bretagne
Nicolas.Jullien@telecom-bretagne.eu

Kevin Crowston
School of Information Studies,
Syracuse University
Crowston@syr.edu

Felipe Ortega
Dept. Statistics and
Operations Research,
University Rey Juan Carlos

ABSTRACT

In this article we ask whether it is possible to evaluate the efficiency of an online knowledge production project to determine its position in the life-cycle of projects and to identify factors that affect efficiency. To assess efficiency, we used the Data Envelopment Analysis (DEA) modeling methodology, applying it to data from Wikipedia, studying more than 30 Wikipedia language projects over three years. We showed that if the main Wikipedia projects were indeed less efficient that smaller ones, corrected by the decreasing return to scale phenomenon, they were quite as efficient as the other projects.

1. INTRODUCTION

The Olson paradox [61] suggests that large groups are less able than small ones to promote their common interest because individual incentives to contribute are assumed to diminish with group size. However, countering this assumption, many online open projects, in various contexts, have demonstrated their ability to develop selective incentives and institutions, allowing them to develop and protect their "commons" even as they grow [64, 38, 29]. Even so, those same projects seem to encounter difficulties in assuring their sustainability in the long run. Evidence suggests that as they age, projects find it harder to recruit and retain newcomers (e.g., [80], in the case of open source software communities, [30], for Wikipedia), and their organization is said to become increasingly bureaucratic [9]. In that respect, these online open projects appear to follow a trend common to other organizations, i.e., a natural tendency toward structural inertia when they get bigger, leading to a growing difficulty to adapt [32]. Such inefficiencies have a direct impact on the output of the project, and an indirect impact by making the work less rewarding [68], and so reducing the number of active contributors[1]. For example, though Wikipedia is still progressing in terms of scope and quality, it is hard to deny the decreasing number of Wikipedia editors [30]. We also note that the rules imposed in 2007, which are said to be at the origin of the increase of bureaucracy, were enforced to address the increasing inefficiency in editing due to edit wars (ibid).

There are several possible alternative explanations for this phenomenon that this paper explores. First, the decrease in the number of participants in online open projects may be simply due to the fact that the reservoir of potential new contributors is drying up. For example, the number of Wikipedia contributors seems to be correlated with the number of Internet users [**?**] and the percentage of new Internet user is reaching a plateau in many developed countries.

Second, the decrease in the growth in participants may simply be the result of, or the signal that, the project has entered a mature phase in which it needs fewer additions and thus fewer contributors [35][2]. In this regard, the slowdown may even be a good sign regarding the management and the health of the project [56]. As Heckerthorn explained (ibid), the controls put in place by the projects over newcomers would be justified by the fact that fewer participants are needed, so the barriers to entry have to be raised.

However, a final and more disturbing possibility is that there is something about the current structure of the project that is unattractive to potential participants. For example, in the case of Wikipedia, [65] proposed a global model to estimate the evolution of Wikipedia participants and to evaluate the impact of management on people's willing to participate, and on the production. They used Douglas' grid/group framework [1978], where people's behavior is more or less constraint by their commitment to the group (high/low) and by the structure of the organization (high/low). Their model shows that "while an open environment accelerates the growth of an online network at the early stage, openness may negatively impact quality and subsequently the attractiveness of the network, so that users will be less inclined to join or to participate in the network" (p. 346).

In this article we ask whether it is possible to evaluate the efficiency of a knowledge production project to determine its position in the life-cycle of projects and to identify factors that affect efficiency. Despite the strong theoretical bases for the question of the explanation for the slowing growth of large collective project, and its importance in terms of practical consequences, few studies have looked at the question of the efficiency of the production, and its variation, something key in the discussion regarding the right size of the community of producers. We further discuss possible reasons for the observed level of efficiency, in particular, is the administra-

[1]See also the article published by the MIT TechReview, October 22, 2013, `http://www.technologyreview.com/featuredstory/520446/the-decline-of-Wikipedia/`

[2]In that regard, the slowdown in investment sounds like the effort level reaches a plateau in online projects, something quite known in software production projects, where the cost of production follows an "S" curve, modeled by [59] in the general case of R&D project, and called the Norden-Rayleigh curve (see [48] for a short review of the literature on that topic, stressing on the open-source software case).

OpenSym 2015, San Francisco, USA

tive superstructure, incarnated in the development of the administrator team, as an explanation for an observed inefficiency. To assess efficiency, we used the Data Envelopment Analysis (DEA) modeling methodology, applying it to data from Wikipedia, studying more than 30 Wikipedia language projects over three years.

The article is organized as follows. Section 2 presents a review of the literature used to construct our framework of investigation; section 3, the data analysis approach and formulation of our hypotheses; section 4, the data collection strategy; and section 5, the results. We discuss the consequences of this work, its limits and future research, before concluding, in section 6.

2. THEORY DEVELOPMENT: MEASURING THE PRODUCTIVITY OF AN (ONLINE) KNOWLEDGE-PRODUCING PROJECT

We suggest that three phases can be identified in the life of a collective action project: the initial phase, where the returns on investment are less than the investment but increasing; the diffusion phase, where it is increasingly rewarding to participate in the collective action; and the mature phase, when the returns on contribution are decreasing and thus fewer contributors are needed. Following Marwell and Olliver's argument (ibid), we characterize this mature phase as a phase where there is a decreasing return on participation and so a decreasing number of participants.

There is a lot of literature on the premises, on the construction of the collective action[3], but as far as we know, very few studies have considered the mature phase of a project (beside [36] on the negotiation of labor standards).

To assess the phase of a particular project, we can examine the projects' production function, that is, the relation between the quantity of inputs consumed by the project and the outputs produced [79]. Entry into a mature phase can be measured by the fact that the project is in a decreasing return on scale regarding the production function. Description of the project production function can further be used to study whether some projects are more efficient than others, and if the structure of the project can explain this efficiency.

Regarding the definition of the production function, the inputs and the outputs of such communities are rather straightforward: as framed by [38, p. 44], it is about turning the people into "outcomes". The outcomes can be apprehended at several levels: a productive level (the various characteristics of the constructed epistemic knowledge such as completeness, creativity), a collective level (e.g. team building or construction of rules and collective norms), and a personal developmental level (learning and development of individuals). This production is affected by the "biophysical characteristics" of the community (such as the information system used) and the "rules-in-use", which constrain the way people interact ("the action arena", or the process).

Crowston et al. [18], who proposed one of the first lists of the measures of an open online project success (open source software), stressed the variety and the non-convergence of the possible measures. For instance, since the pioneering work of [19], scholars have stressed the difficulties to assess the impact of information system from other measures (especially re-organization, 8) on firm's performance, and even to precisely define what performance is [58].

Fortunately, in our case, the problem can be simplified, as we are looking at the level of production of the communities of creation, communities which have a defined goal, are even defined by their goal, unlike other online "communities" around Facebook or Twitter. They are (virtual) epistemic communities, or task-oriented groups, bringing experts together around a common goal [70], here the building of (new) knowledge available to other people (explicit, published online knowledge, being programs or encyclopedic articles). Project contributors are often evaluated by their capacity to produce these pieces of knowledge [14]. Therefore, in defining a project production function, we are trying to measure the efficiency of the system in turning project inputs into explicit published online pieces of knowledge.

The focus of the projects means that we will not attempt to measure the whole outcomes of such collective projects. Focusing on the pieces of knowledge produced, we are leaving out most of the fallouts for the participants, such as social connexions, fun, training [10], or perceived knowledge satisfaction [13]. Actually, when looking at the production function of a company, these kind of outcomes are also usually not taken into account, even if they are very important to explain why people start and stay participating in such projects (or why a company succeeds). Instead, regarding the outputs, the type of knowledge produced and the 'quality' of this knowledge are the key indicator [24, 85], whose precise definition depends on the type of knowledge produced and thus of the criteria in that field, but also of the goal of the organization, here the community.

A similar simplification applies also to the inputs to the production function. Even though open online projects rely on an information system infrastructure (Wikipedia's wiki editor MediaWiki or Linux's source code management system, Git), making them, as pointed by [37], sociotechnics systems [1], as long as the system remains stable during the evaluation the efficiency of the tool itself can be considered as constant and so omitted from the analysis. As a result, we will not need to look at the abundant literature on the impact of information systems on firm's productivity and/or performance. The same can be said for the rules organizing the community, which we suppose remain stable during the study.

The inputs we consider are the efforts put by the people to realize the communities' tasks, and the number of people participating. Unfortunately, it is impossible to determine precisely the effort made by each contributor (in terms of time spent, for instance). Effort is difficult to measure in a classical organization, but even harder in the case of online voluntary work because this information is subject to self-estimation biases [48]. Most of the work in the literature on effort assessment in online communities approximate efforts by the number of contributions made a certain point of time (e.g. ibid, 27), an approach we will adopt.

We propose to go a step further here, introducing the heterogeneity of the participants as a second factor. The critical mass theory of the construction of collective action [60, 56], the theory analyzing the construction of the (knowledge) commons [64, 37], but also the studies of groups' creativity and efficiency [for instance, 77], stress the fact that these projects are made possible by the aggregation of various motivations and level of involvement.

To rephrase our research question, we ask here which projects are efficient in turning the input of participants and participant contributions into knowledge products, and whether this efficiency is due to a distribution of participants among the very involved (i.e., the administrators), and the occasional contributors, to the projects' stage in its life cycle, or to other external variables.

3. ANALYSIS APPROACH AND HYPOTHESES

[3] See [33] for a review of the literature, in the context of technology innovation, and [73] for the study of the emergence of a market. Simulating Marwell and Oliver's proposals, [72] stressed the importance of participants' heterogeneity (the "status" of the participants in their article) in the starting of a collective action.

3.1 Analysis approach: DEA modeling, and t-tests.

Our goal is to evaluate projects' efficiency (synonymous to productivity, here) in turning contributors (and contributions) into content, and second to test our hypotheses regarding the explanation of the variation of efficiency. We use DEA modeling to evaluate the productivity of the projects and we then run t-test and regressions to evaluate the correlation between our explained variable (the projects are efficient) and the explanatory variables we proposed in the previous sections.

3.1.1 DEA modeling.

Economists formalize the link between inputs and output as a production function [79]. To be efficient is to reach the maximum possible outputs for a given amount of inputs; a project that used the same amount of inputs but produced less is less efficient. In our case, the form of this production function is unknown, as are the coefficients relating its components. However, we are not trying to propose a characterization of the Wikipedia production function, but rather to evaluate if some projects are more (or less) efficient than the others. Since [28], this can be done by looking at the "frontier production function", which describes, for various combination of inputs and outputs, the producers who are efficient, i.e., the ones for which none of the outputs can be increased, without either or several of the inputs increasing or other outputs being reduced, and vice versa. Note that as different mixes of input or outputs are possible, there can be multiple efficient producers. Other producers can then be compared to these to determine their level of efficiency, e.g., what fraction of outputs they produce with the same inputs or how much more input they require for the same outputs.

There are several techniques for estimating the frontier production function, and a detailed comparison is out of the scope of this paper[4]. In this paper, we use Data Envelopment Analysis, a 'data-oriented' approach for evaluating the performances of a set of peer entities, called Decision Making Units, or DMUs in the original source, in this article, each project. DEA was first proposed by [12] as a technique to compare projects and to estimate which projects are "efficient" without assuming a form for the production function. According to the definition of relative efficiency, a DMU "is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs" [17, def. 1.2, chapter 1, p. 3]. A further advantage of DEA for our analysis, as pointed out by [47], "DEA can account for economies or diseconomies of scale, and is able to deal with multi-input, multi-output systems in which the factors have different scales".

The basic DEA models suffer from a limitation which may be problematic in some cases: being linear optimization models, they implicitly suppose that the production function is made of substitutable 'variables', for both the input and the output sides. Regarding open online projects, where an addition of independent participants' contributions represent the input, and where we want to test the hypothesis that this people are not perfectly substituable, the linearity makes sense to be able to perform the test. Regarding the outputs, we speak here of pieces of knowledge (characters, photos) which, if not substituable, add all to the knowledge base. So, in our specific case, this limitation does not seem to matter very much.

[4]The reader interested by this discussion will be happy to look at [44]'s study in the case of software production.

Koch [47] used the DEA approach in studies of the efficiency of FLOSS communities. As he pointed out, the outcomes of a community have to be measured along several dimensions, without a clear ranking of the importance of each factor, leading him to use Data Envelopment Analysis (DEA) techniques [47]. This approach permitted the comparison of projects efficiency in turning contributors into lines of code, but also into bug fixing, for instance [46, b].

Two main criteria have to be taken into account regarding the choice of a DEA model: the orientation of the model (input-oriented or output-oriented) and the return to scale in the production process. Regarding the first criteria, as in [47], an output-orientation seems to be more appropriate, as, for a period of time, the inputs (the volunteers in an open online project), are more or less fixed and the goal is to maximize the output. Considering the second criteria, based on the study of [56] on collective action, on the analysis of software projects, and on the discussion above, it seems rather difficult to assume a constant return to scale. Instead, these projects seem to have a increasing return to scale in a first phase, and then a decreasing one. So we will choose a model that includes a variable that captures a project's return to scale, the BCC model, and more specifically the BCC-O (output oriented) model [5].

For the data analysis, we used [71]'s macro under SAS, with non-constant return on scale constraint. The original program is an input-oriented one, so we had to change the equations into an output-oriented one (equations 4.27 to 4.30 in [16, p. 89]).

3.2 Hypotheses: efficiency in the organization of the input factors

Given the analysis approach adopted, we can now state more specific hypotheses that we will test in this paper.

3.2.1 Size of the project. Hypothesis 1: bigger projects are less efficient.

Following Marwell and Olliver (ibid), we characterize the mature phase as a phase where there is a decreasing return on participation. This can be measured by the fact that the project is in a decreasing return on scale phase, regarding the production function (which has to be defined). Our hypothesis is that:

- H1. Big project are less productive than small ones (i.e., projects exhibit decreasing return to scale).

3.2.2 Structure of the team. Hypothesis 2: under and over heterogeneous projects are less efficient.

Once this decreasing efficiency is controlled, we explore explanations for the differences between the projects regarding their productivity. The first hypothesis regarding this point is that the projects have to have a 'good' structure in terms of team building. This leads to two hypotheses, concerning the structure and the management of the projects.

- H2.1. Following Uzzi (ibid), we hypothesize that the efficient projects (in terms of productivity) are heterogeneous, but not too much, regarding the variety of the participants, between big and small contributors.

- H2.2. Following [32, 9] on the bureaucratic aspect, we hypothesize that the efficient projects have neither a too heavy, nor too light techno-structure.

4. DATA COLLECTION

4.1 Case selection: Wikipedia

As a setting for our study, we choose the most visible open knowledge project, the encyclopedia editing project Wikipedia. Defining itself as an *"online encyclopedia"*[5], incorporating *elements of general and specialized encyclopedias, almanacs and gazetteers*", Wikipedia covers a large scope. For each entry or article, it aims at *"explain[ing] the major points of view in a balanced impartial manner*", with *"verifiable accuracy"* and *"references"*.

There were three reasons for the choice of this research setting. First, Wikipedia has become one of the most successful knowledge production projects ever, with almost 5 million articles for the English version and more than one million visits per day, and is seen as a model for knowledge management theory [57, 34]. But even this successful project has recruiting problems, as already mentioned. [76] even showed that "both the rate of page growth and editor growth has declined" in the English Wikipedia. Secondly, Ostrom and Hess pointed out is that online communities lowered the boundary between those who are in and those who are out. Wikipedia, which does not require programming competences from its participants, seems to be one of the communities where the boundary is the lowest.

Finally, focusing the analysis of the inputs on people's involvement is facilitated by the structure of Wikipedia. Wikipedia is not a single project, but rather a multitude of independent sub-projects: for each language there is a separate version of the encyclopaedia with its own editor community and collection of articles. Importantly, the projects are at different levels of maturity, some quite mature, others still getting started and others somewhere in between. However, the projects are still comparable, as is required for our analysis. They all share the same tool for collaborative edition (MediaWiki) and the same basic rules for collaboration, the "five pillars" of Wikipedia[6], meaning that the only one of Hess and Ostrom's inputs that vary are people's involvement, which is what we are looking for. Thus the process structures seem quite the same, in contrast to studies on open source software (see for instance 18, 47) that suggest that different project that use various technologies, programming languages and collaborative tools. Comparison of the output of the projects provides further evidence for their comparability. The global structure of the projects, measured as a network, the nodes being the articles and the links the links between the articles, seems to be about the same in terms of "degree distributions, growth, topology, reciprocity, clustering, assortativity, path lengths, and triad significance profiles", at least for the main projects [86]. This uniformity may help us to better understand, in their difference, what differences are due to the collective action phases.

4.2 Data Collection

As did prior studies of Wikipedia (see [63] for discussion), we relied on internal Wikipedia data to estimate the number of contributors involved, their characteristics and level of activity. To compute these variables, we obtained the complete database dump with all edits performed in 37 Wikipedias in different languages. These dump files include all required data to trace the creation of new articles and individual changes on any page in these Wikipedia projects. In Wikipedia terminology these edits are known as revisions. The dump data directly provide the number of revisions and the length of the text added. By examining which user performed each revision, we obtained counts of the number of editors actually active in the period of the dump[7]. For each language we also retrieved data about any special privileges granted to certain Wikipedia users, such as the administrators.

Table 1: Wikipedia languages parsed & used in the analysis.

Japanese	ja	Czech	cs	Catalan	ca
Spanish	es	Indonesian	id	Bulgarian	bg
German	de	Thai	th	Croatian	hr
French	fr	Arabic	ar	Greek	el
Rusian	ru	Korean	ko	Slovak	sk
Italian	it	Hebrew	he	Serbian	sr
Portuguese	pt	Norwegian	no	Lithuanian	lt
Polish	pl	Hungarian	hu	Slovenian	sl
Chinese	zh	Vietnamese	vi	Estonian	et
Dutch	nl	Ukrainian	uk	Malaysian	ms
Swedish	sv	Danish	da		
Turkish	tr	Farsi	fa		
Finnish	fi	Romanian	ro	English	en

The data extraction has been implemented as a software program written in Python to automate this process. This program is part of WikiDAT (Wikipedia Data Analysis Toolkit)[8], a multi-purpose framework aiming at facilitating Wikipedia data analysis for any of the 280 languages currently available in the free encyclopedia. The use of Python lxml[9], an efficient library for XML parsing, and multiple sub-processes, let us speed up significantly data retrieval and extraction[10], and to develop more precise data than those presented by the Wikimedia Foundation[11] as far as the edits and the contributors are concerned, and to include new, original data, the number of Feature Articles and of new Featured Article (FA, or articles of exceptional quality). The languages analyzed with their correspondence in Wikipedia are presented in table 1.

4.3 Construction of the variables

4.3.1 The Outputs. Scope, accuracy, quality

The main outputs we consider are contributions to Wikipedia (individual edits) and the articles created. For the former, as a first approach, we examine simply the length and basic quality of the writing for edits (result already stressed by [7]). Prior research suggests that these simple measures are good indicators of a contribution, at least for the English version [20], are they are for open source contribution[39]. The number of new characters serves to measure the increase of the content available For the later, we will consider the number of new articles to measure the increase in scope of the encyclopedia.

We will also consider measures of output quality, as projects may be prioritizing improvements in quality over quantity. As explained by [31, p. 3], Wikipedia article quality can be assessed in different ways. One approach seeks to rate articles against external metrics, e.g., those coming from information system studies on the qual-

[5]All the citations of this paragraph come from the Wikipedia page describing its fundamental principals, or "five pillars": `http://en.Wikipedia.org/wiki/Wikipedia:Fivepillars`. See also [69] for a discussion of Wikipedia as a model for collaboration.

[6]`http://en.Wikipedia.org/wiki/Wikipedia:Fivepillars`.

[7]We elided all revisions undertaken by bots, as they do not represent work done directly by editors, which is the production process we wanted to study.

[8]`http://libresoft.es/node/564`

[9]`http://lxml.de/`

[10](For instance, as far as the English Wikipedia is concerned, 444,946,704 revisions in 27,023,430 pages were analyzed in approximately 44 hours.

[11]`http://stats.wikimedia.org/EN//`

ity of the process 75, 84, bibliographic studies (on the product), based on [43]'s criteria (e.g., purpose, authority, scope) or on subjective metrics (user experience, being reader or producer). Such an approach is arguably more valid and would be more comparable across projects. However, [82, 26] concluded that these criteria are difficult to apply to Wikipedia, especially because there is no authorship analysis possible, but also "due to the overall scale and the wide range of subject areas, most of the studies focus on specialized fields of knowledge" [51], rather than performing a comprehensive analysis. A further problem would be finding raters able to assess contributions in multiple languages.

A simpler approach to article quality relies on internal measures, as in [67, 11]. Wikipedia has several levels of quality for its articles, from article needing to be improved to featured article, see `http://en.Wikipedia.org/wiki/Wikipedia:Article_development`. In light of the difficulties of applying external measures, we relied on these internal measures for this study. Specifically, we count the number of new featured articles (FA), which is how every project signals its best articles. To augment this measure, we also examine the number of new internal redirects, as a proxi for the improvement of the internal organization and of the number of the (internal) references.

4.3.2 *The Inputs. Measuring participants' efforts, defining the participants by their efforts.*

As argued, different level of participation and of participants are important for a collective action to emerge and bloom. But the characterization of those different contributors is far from being simple. For simplicity, and according to our orientation to take internal measures if existing, we relied on Wikimedia Foundation's, and Wikipedia Statistics pages[12] regarding the different type of contributors. Following them, we named (and counted) "active Wikipedians" those "who contributed times or more in the month", "very active Wikipedians" those "who contributed 100 times or more in the month", and "other contributors"the others. To test the complementary (or competitive) explanations (H2.1 about the structure of the team and H2.2 about the size of the administrative team), we counted the number of administrators each month.

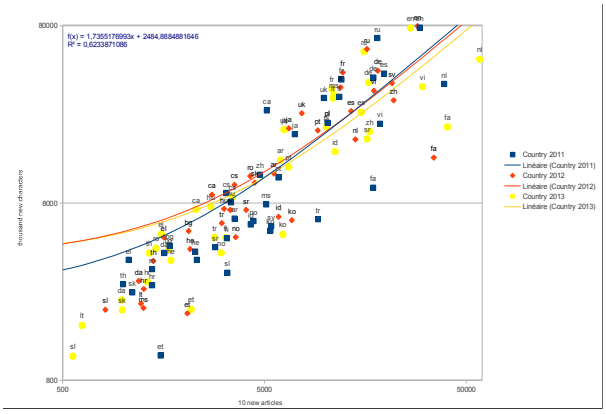
Finally, a specificity of Wikipedia is that anonymous contribution is possible, even if not encouraged. If most of the best contributions in terms of quality are done by registered users and by a small subset of the whole contributors, a significant number of anonymous users also provide quality content [41]. The share of such contributions varies dramatically between the different languages (between 26% for the Japanese project and less than 2% for the Slovenian one), a variation which could explain, in part, the differences in efficiency between the projects, because the management of these anonymous contributions may be time consuming.

4.3.3 *Control variables: cultural factors.*

While the discussion above has focused simply on inputs, variation in the productivity of editors might be explained by the cultural background of the project participants. According to [65], criterion related to the commitment to the group may impact efficiency. For instance, and referring to the cultural studies and indicators proposed by [40], individualism is directly connected to the willing to participate in collective projects. Various works, especially looking at Wikipedia [49, 15], pointed out the gender gap in number of editors, and explained this gap by the conflicting (thus female unfriendly) atmosphere of such projects. We will do so using the four first Hofstede's cultural dimensions: Power distance (PDi) Individualism (Ind), Masculinity / Femininity (Masc), and Uncertainty

[12] `http://stats.wikimedia.org/EN//`

Figure 1: Number of new characters versus new articles by project, 2011 to 2013.



avoidance (UncAv). . For project that are supported by several countries (such as the Spanish one), we calculated a preponderated value of these dimensions taking into account the number of Internet users in the main countries participating.

The second set of control variables regards the linguistic vs national aspect of the projects: some projects, because they include people from various countries, possibly with different agendas, may have to spend more time solving national / cultural conflicts than contributing[13]. We constructed a dummy variable indicating whether the linguistic project is multi-countries (1) or supported mainly by one country (0) (e.g., the Finish project).

5. RESULTS

5.1 Some descriptive statistics

When speaking of the decrease in output, as already said, most prior research points at the decrease of the number of contributors, at least in the case of the English Wikipedia [30]. [**?**] extended this analysis to the other Wikipedias and showed that the biggest projects are less efficient in recruiting new contributors, which could be explained by the fact they already exhausted the reservoir of potential contributors.

However, regarding the efficiency in producing new knowledge, which is what we are looking at here, simple statistics seem to describe a more complex situation where projects may be producing different mixes of outputs. For instance, plotting the number of new characters produced against the number of new articles created on a scatter plot (figure 1; NB, log scales) makes clear that, in general, and even if there is quite a large dispersion ($R^2 = 0.62$), smaller projects are all below the linear trend line (relatively more articles than new characters), while the larger projects are mostly above it (relatively more characters than new articles). We interpret this result as suggesting that the larger projects seem to have been more often adding to the content of the articles (number of characters) rather than increasing coverage (number of articles) in the past three years, when smaller projects are still caching up on coverage.

Still from a scatter plot view, this focus on new articles seems to imply more edits than the average (figure 2), as most of the big Wikipedias are below the regression curve[14]. But there are also a

[13] Such as the Chinese Wikipedia, which has had to solve the conflict between different writing forms [52].

[14] On this graphic and on the followings, we did not include the English Wikipedia, which is far bigger that the other projects. As

Figure 2: Number of new characters versus edits by project, 2011 to 2013.

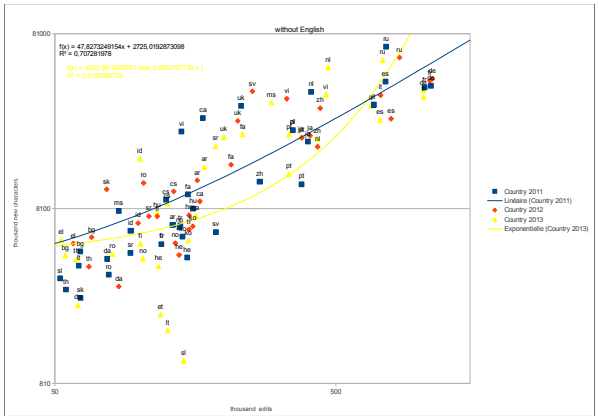
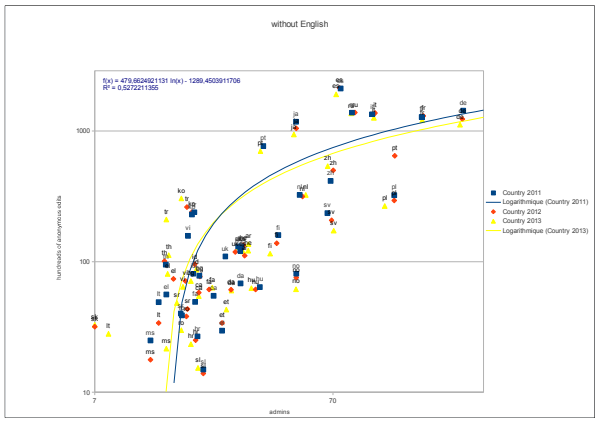


Figure 3: Number of anonymous edits versus the number of admins by project, 2011 to 2013.



lot of small projects below this curve, and those big projects are very close to the curve.

We compared the number of admin with the number of anonymous edits (figure 3) or the number of occasional contributors (figure 4) (which are both said to have to be closely monitored). However, while there is a quite good linear correlation between the number of occasional contributors and the number of admins, the relation is logarithmic between the anonymous and the admin, indicating that the biggest projects are growing slowerly in the number of anonymous contributors than in the number of administrative staff (but, still, they are above the curve).

An explanation offered for the seeming inefficiency of Wikipedia is that that projects have become inefficient because they have added too much administration in order to deal with problematic edits. However, our examination of the data does not seem to provide any consistent support to this postulate. DEA modeling analysis allowed us to refine the analysis of the hypotheses.

5.2 Measuring the efficiency of the projects

As the three years we studied do not fundamentally differ, we will focus in this analysis of the results of Year 2013 in Figure 5, for clarity.

an outlier in the distribution, it impacts too much the shape of the curve and the coefficients

Figure 4: Number of occasional contributors versus the number of admins by project between 2011 and 2013.

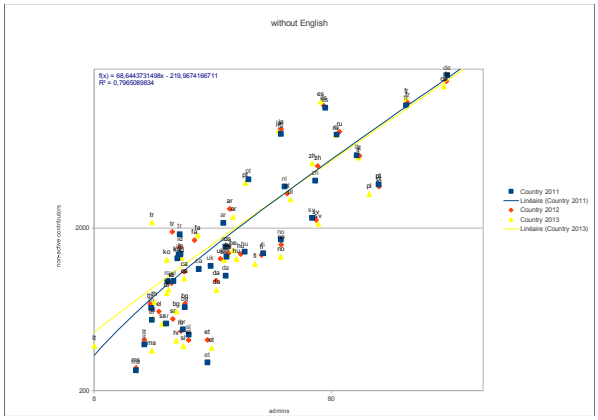
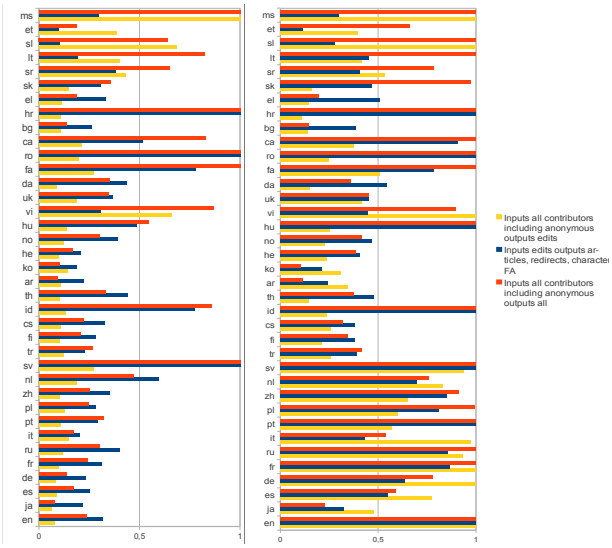


Figure 5: Efficiency in production edits and new knowledge for the 37 Wkipedia language projects, 2013 (left, without taking into account the return to scale, right taking into account the return to scale)



5.2.1 *Hypothesis 1*

The results of the three years together are available in the Annex.

The results show that the big projects are globally less efficient than the small ones. Considering the discussion in section 2, we evaluated if this inefficiency is due to a decreasing return to scale (it is harder to find new article to write, or to improve the already written articles) or, even taking this into account, if the big projects are still less efficient

In a second time, and still using DEA modeling, we evaluated if this lack of efficiency is due to a lack of edits (people in big projects edit less than people in small projects) or if this is due to inefficiency in turning the edits in articles.

The main result is that, as expected, the big projects are less efficient, in terms of productivity. But this has to be mitigated, as, when taking into account a decreasing return to scale effect, the English project turns to be efficient again, and the other big projects, but the Japanese one, are quite close to efficiency, at least closer that smaller projects. This is particularly true when looking at the

conversion of contributors into edits, less in the process of converting the edits into new knowledge, something we will address in the discussion. These results remain the same with of without taking into account the FA, and with or without taking into account the anonymous contributions, which seems to indicate that the production of FA articles as the production of anonymous, remain of marginal impact on the trajectory of the projects.

5.2.2 Hypothesis 2

To test hypothesis 2, we performed a linear regression of the assessed efficiency in turning contributors into knowledge (with and without return to scale) on the variables proposed above to explain differences in efficiency, specifically ratio of administrators over (hundred) anonymous contribution, ratio of administrators over contributors (to test Hypothesis 2.2), active and very active contributors over contributors (to test Hypothesis 2.1), Hofstede's cultural dimensions, and whether or not the language project concerns more than one country. We used the efficiency measured by the DEA model with return to scale to control from the size of the projects. Model 1 in Table 2 introduces the Control variables and Model 2 adds the administrative structure variables and the structure-of-the-team variables.

The regression found no statistically significant relations, but one, the link between the ratio of administrator to anonymous contributions and the efficiency of the projects[15]. The efficient projects are significantly more administrated, as this ratio of administrator over (hundred) anonymous edits is around 46% for the efficient projects (standard dev 29) versus 28% for the inefficient projects (standard dev 24). This somehow proves Hypothesis 2.2 wrong, when Hypothesis 2.1 (structure of the team) is not proved, as the percentage of active or very active contributors over contributors does not seem to impact the efficiency of the teams.

This drives us to the discussion of our results.

6. DISCUSSION AND CONCLUSION

The validity of our analysis is dependent on the quality of the data used. We are quite confident in the data extracted from the Wikipedia dumps, e.g., regarding number of characters and edits or number of Featured Articles. The most significant limitation to our estimate of the efficiency of the edit and article production process is that we have only a partial information on the input: the number of people involved, but not their effort, as we do not know, for instance, how many hours a contributor spent developing an edit, which could vary quite a lot, from a few seconds to fix a spelling error to hours spent researching a topic. As we want to compare Wikipedia language projects, we can only assume that from one project to another, the mean time spent is the same for each type of contributor and has not varied through time. Violations of this assumption will affect our measure of the relative efficiency of the projects.

With regard to outputs, a limitation is that we had only the count of FA projects to assess the quality dimension of the articles. We note that there are concerns about the validity of FA status as a quality measure. The concrete rules of the process of granting FA status to a Wikipedia article varies between the different language projects, but in general it involves voting on the quality of the article: the article should receive a substantial proportion of positive votes to be granted the FA label. However, [74] showed that the argument of quality to qualify an article as FA varies from one language to another. Furthermore, when external expertise is mobi-

[15]We tested variation of the borders of the low / high values without impacting the results.

Table 2: Regression explaining the **non-efficiency** of the projects (Hypothesis 2)

		Model 1			Model 2		
Parameters		Estim. Values	Stand. Dev.	Pr > Khi-2	Estim. Values	Stand. Dev.	Pr > Khi-2
Intercept		1.0799	0.7017	0.1238	0.4545	1.2719	0.7209
Power distance	Low	0.118	0.6260	0.8498	-0.012	0.9530	0.9901
	Medium	0.647	0.4191	0.1226	1.297	0.5947	0.0291
	High	ref.			ref.		
Individualism	Low	-0.180	0.660	0.7845	-1.551	1.024	0.1299
	Medium	-1.220	0.4877	0.0123	-1.697	0.7290	0.0199
	High	ref.			ref		
Mascul. / Feminin.	Low	0.074	0.4137	0.8571	-0.123	0.6028	0.8375
	Medium	0.252	0.4394	0.5656	-0.423	0.6141	0.4904
	High	ref.			ref.		
Uncert. avoidance	Low	-0.672	0.4835	0.1642	-0.042	0.8437	0.9602
	Medium	0.104	0.3502	0.7660	0.9947	0.5897	0.0917
	High	ref.			ref.		
Multi-country project	MIXV	-0.166	0.3209	0.6037	-0.7300	0.5381	0.1749
% admin ov. contrib.	Low				-1.617	1.1609	0.1636
	Medium				**1.288**	**0.6462**	**0.0462**
	High				ref.		
% admin ov. anonym.	**Low**				**4.068**	**1.2877**	**0.0016**
	Medium				0.699	0.5837	0.2305
	High				ref.		
% activ. contrib. ov. contrib.	Low				0.557	0.9293	0.5489
	Medium				-0.451	0.7598	0.5520
	High				ref.		
% very activ. contrib. ov. contrib.	Low				-0.777	0.8238	0.3452
	Medium				-0.545	0.6769	0.4204
	High				ref.		

Table 3: Distribution of the percentage of active contributors in different Wikipedia projects

	Means	Stand. dev.	Min	Max
Active contributors over contributors	32.98	4.86	42.93	20.37
Very active contributors over contributors	5.26	1.29	8.63	2.58

lized to evaluate the quality of FAs, as in [54], they show strong variations regarding their assessed level of quality. Finally, as said before, our results seem to indicate that the production of FA articles is of marginal impact on the global (productivity) of the project as our results do no change if we take into account (or not) these FA in the evaluation of the productivity. On the other hand, and as stressed before, we proposed to evaluate the projects by their internal objective, and as far as quality is concerned, the objective is to create Featured Articles, even if other levels of quality are proposed in some projects.

Thanks to the DEA modeling technique, we proved it possible to compare projects in different phases of development and with various agenda, some being more focused on adding articles, other in filling the existing ones, for instance. Regarding the argument of efficiency, we showed that if the main Wikipedia projects were indeed less efficient that smaller ones, corrected by the decreasing return to scale phenomenon proposed by these authors, they were quite as efficient as the other projects. This is especially true when considering that this supposed lack of efficiency may be due to the elements measured: some actions of improvement, like the rephrasing of an article, or the adding of a picture, or the adding of templates, do not increase the number of characters, and thus are not measured here. These big projects may also be more targeted by vandalism or more strict in the creation of new articles than smaller projects. This would explain the fact that the big projects are as efficient in converting the contributors into edits, but less in the process of converting the edits into new knowledge.

We cannot deny the fact that there are less contributors and less new articles produced. But, considering the results of this article, we wonder if this fact is not an indication that they try to control their eventual slowdown, their entrance into the mature phase by raising the boundary in order to not manage unneeded manpower, and to keep being efficient. Comparing years 2011 to 2013 with older years, such as 2006 to 2008 may help precise this point. Regarding the other hypothesis, the impact of the structure of the team on the efficiency, our results are inconclusive. this may be due to the too small number of cases (111), making harder to have statistically significant results. This may also be due to the fact that, despite the international coverage of Wikipedia, the structure of the teams is similar, the open online culture being stronger that the local culture in such projects. An argument in favor of this point is the stability of the percentage of active or very active contributors over the contributors in all the projects we analyzed: even if there are extreme cases, the standard deviation is low (Table 3).

To conclude, beside its limitations, this article is one of the first attempts to measure the supposed decrease in efficiency of the main Wikipedia projects and, more generally of the theory of the collective action by Marwell and Oliver. We advocate for developing the studies on online collective action, using the DEA modeling to refine our results.

References

[1] M. Akrich, M. Callon, and B. Latour. *Sociologie de la traduction : textes fondateurs*. Les presses des Mines de Paris, Paris, 2006.

[5] R. D. Banker, A. Charnes, and W. W. Cooper. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092, 1984.

[7] J. E. Blumenstock. Size matters: Word count as a measure of quality on wikipedia. In *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*, pages 1095–1096, 2008.

[8] E. Brynjolfsson and L. M. Hitt. Computing productivity: Firm-level evidence. *Review of Economics and Statistics*, 85 (4):793–808, 2003.

[10] B. Butler, L. Sproull, S. Kiesler, and R. Kraut. Community effort in online groups: Who does the work and why? In S. Weisband, editor, *Leadership at a distance: Research in Technologically Supported Work*. Lawrence Erlbaum, Mahwah, NJ, 2007.

[9] B. Butler, E. Joyce, and J. Pike. Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 1101–1110, New York, NY, USA, 2008. ACM. doi: http://doi.acm.org/10.1145/1357054.1357227.

[11] K. Carillo and C. Okoli. Generating quality open content: A functional group perspective based on the time, interaction, and performance theory. *Information & Management*, 48(6): 208–219, 2011.

[12] A. Charnes, W. Cooper, and R. E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2:429–444, 1978.

[13] S. Y. Choi, H. Lee, and Y. Yoo. The impact of information technology and transactive memory systems on knowledge sharing, application, and team performance: A field study. *MIS quarterly*, 34(4):855–870, 2010.

[14] P. Cohendet, F. Créplet, and O. Dupouet. Interactions between epistemic communities and communities of practice as a mechanism of creation and diffusion of knowledge. In J.-B. Zimmermann and A. Kirman, editors, *Interaction and Market Structure*. Springer, Londres, 2001.

[15] B. Collier and J. Bear. Conflict, criticism, or confidence: An empirical examination of the gender gap in Wikipedia contributions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 383–392, New York, NY, USA, 2012. ACM.

[16] W. W. Cooper, L. M. Seiford, and K. Tone. *Introduction to data envelopment analysis and its uses: with DEA-solver software and references*. Birkhq̈user, 2006.

[17] W. W. Cooper, L. M. Seiford, and J. Zhu, editors. *Handbook on Data Envelopment Analysis*. Springer, 2nd edition, 2011.

[19] K. Crowston and M. E. Treacy. *Assessing the impact of information technology on enterprise level performance*. Citeseer, 1986.

[18] K. Crowston, J. Howison, and H. Annabi. Success in free and open source software development: Theory and measures. *Software Process Improvement and Practice*, 11:123–148, 2006.

[] K. Crowston, N. Jullien, and F. Ortega. Is Wikipedia Inefficient? Modelling Effort and Participation in Wikipedia. In *Forty-sixth Hawai'i International Conference on System Sciences (HICSS-46)*. IEEE Computer society, January 2013. URL `http://ssrn.com/abstract=1960696`.

[20] D. H. Dalip, M. A. Gonçalves, M. Cristo, and P. Calado. Automatic quality assessment of content created collaboratively by web communities: A case study of wikipedia. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, pages 295–304, 2009.

[22] M. Douglas. *Cultural Bias*. Royal Anthropological Institute, London, 1978.

[24] P. F. Drucker. Knowledge-Worker Productivity: The Biggest Challenge. *California Management Review*, 41(2):79–94, 1999.

[26] K. Ehmann, A. Large, and J. Beheshti. Collaboration in context: Comparing article evolution among subject disciplines in wikipedia. *First Monday*, 13(10), 2008. URL `http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2217/2034`.

[27] Y. Fang and D. Neufeld. Understanding Sustained Participation in Open Source Software Projects. *Journal on Management Information Systems*, 25(4):9–50, Apr. 2009.

[28] M. J. Farell. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, 120(3):250–290, 1957.

[29] A. Forte, V. Larco, and A. Bruckman. Decentralization in wikipedia governance. *Journal of Management Information Systems*, 26(1):49–72, 2009.

[30] A. Halfaker, R. S. Geiger, J. T. Morgan, and J. Riedl. The rise and decline of an open collaboration system how Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688, 2013.

[31] R. Hammwöhner. Interlingual aspects of wikipedia's quality. In *Proceedings of the 12th International Conference on Information Quality, MIT, Cambridge, MA, USA*, pages 477–488. MIT, 2007.

[32] M. T. Hannan and J. Freeman. Structural inertia and organizational change. *American sociological review*, pages 149–164, 1984.

[33] T. J. Hargrave and A. H. Van De Ven. A collective action model of institutional innovation. *Academy of Management Review*, 31(4):864–888, 2006.

[34] H. Hasan and C. Pfaff. The wiki: an environment to revolutionise employees' interaction with corporate knowledge. In *Proceedings of the Australasian Computer-Human Interaction Conference*, Sydney, 2006. OZCHI 2006.

[35] D. D. Heckathorn. The dynamics and dilemmas of collective action. *American Sociological Review*, 61:250–277, Apr. 1996.

[36] M. Helfen and J. Sydow. Negotiating as Institutional Work: The Case of Labour Standards and International Framework Agreements. *Organization Studies*, 34(8):1073–1098, 2013.

[38] C. Hess and E. Ostrom, editors. *Understanding Knowledge as a Commons. From Theory to Practice*. MIT Press, december 2006.

[37] C. Hess and E. Ostrom. Introduction: An Overview of the Knowledge Commons. In [38], editor, *Understanding Knowledge as a Commons. From Theory to Practice*, pages 3–26. 2006.

[39] P. Hofmann and D. Riehle. Estimating Commit Sizes Efficiently. In *Open Source Ecosystems: Diverse Communities Interacting (IFIP 2.13)*, volume 299/2009 of *IFIP Advances in Information and Communication Technology*, pages 105 – 115. Springer, Springer, 2009. URL `http://flosshub.org/sites/flosshub.org/files/EstimatingCommitSizesEffciently.pdf`.

[40] G. Hofstede. *Cultures and organizations: Software of the mind*. McGraw-Hill, London, 1991.

[41] S. Javanmardi, Y. Ganjisaffar, C. Lopes, and P. Baldi. User contribution and trust in Wikipedia. In *5th International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2009.

[43] W. A. Katz. *Introduction to reference work, vol. 1: Basic Information Services*. McGraw-Hill, Boston, MA, 8th edition, 2002.

[44] B. A. Kitchenham. The question of scale economies in software - why cannot researchers agree? *Information and Software Technology*, 44(1):13 – 24, 2002.

[46] S. Koch. Measuring the efficiency of free and open source software projects using data envelopment analysis. In S. Sowe, I. Stamelos, and I. Samoladas, editors, *Emerging Free and Open Source Software Practices*, pages 25–44. Hershey, PA: IGI, 2008.

[47] S. Koch. Exploring the effects of sourceforge.net coordination and communication tools on the efficiency of open source projects using data envelopment analysis. *Empirical Software Engineering*, 14:397–417, 2009.

[48] S. Koch. Organisation of work in open source projects: expended effort and efficiency. *Revue d'économie industrielle*, 136:17–38, 2011.

[49] S. K. Lam, A. Uduwage, Z. Dong, S. Sen, D. R. Musicant, L. Terveen, and J. Riedl. WP:Clubhouse? An Exploration of Wikipedia's Gender Imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, New York, NY, USA, Oct. 2011. ACM.

[51] D. Lewandowski and U. Spree. Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science and Technology*, 62(1):117–132, 2011.

[52] H.-T. Liao. Conflict and consensus in the chinese version of wikipedia. *IEEE Technology and Society Magazine*, 28, 2009.

[54] D. Lindsey. Evaluating quality control of wikipedia's feature articles. *First Monday*, 15(4), 2010. URL `http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2721/2482`.

[56] G. Marwell and P. Oliver. *The Critical Mass in Collective Action: A Micro-Social Theory*. Cambridge University Press, Cambridge, 1993.

[57] A. P. McAfee. Enterprise 2.0: The Dawn of Emergent Collaboration. *Management of Technology and Innovation*, 47 (3), 2006.

[58] B. L. Myers, L. A. Kappelman, and V. R. Prybutok. A comprehensive model for assessing the quality and productivity of the information systems function: toward a theory for information systems assessment. *Information Resources Management Journal (IRMJ)*, 10(1):6–26, 1997.

[59] P. V. Norden. On the anatomy of development projects. *IRE Transactions on Engineering Management*, 7(1):34–42, 1960.

[60] P. Oliver, G. Marwell, and R. Teixeira. A theory of critical mass interdependence, group heterogeneity, and the production of collective action. *American Journal of Sociology*, 91 (3):522–556, November 1985.

[61] M. Olson. *The logic of Collective Action*. Harvard University Press, Cambridge Mass., 1965.

[63] F. Ortega, D. Izquierdo-Cortazar, J. M. Gonzalez-Barahona, and G. Robles. On the analysis of contributions from privileged users in virtual open communities. In *42th Hawaii International Conference on System Sciences (HICSS)*, page 110, Waikoloa, Big Island, Hawaii, USA, Jan. 2009. IEEE Computer Society. doi: http://dx.doi.org/10.1109/HICSS.2009.843.

[64] E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

[65] P. Otto and M. Simon. Dynamic perspectives on social characteristics and sustainability in online community networks. *System Dynamics Review*, 24(3):321–347, 2008.

[67] G. Poderi. Comparing featured article groups and revision patterns correlations in Wikipedia. *First Monday*, 14(5), 2009. URL `http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2365/2182`.

[68] S. Ransbotham and G. C. Kane. Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia. *MIS Quarterly*, 14 (9):541–545, September 2011.

[69] J. M. Reagle Jr. *Good Faith Collaboration. The Culture of Wikipedia*. MIT Press, Cambridge, Massachusetts, 2010.

[70] F. Rullani and S. Haefliger. The periphery on stage: The intra-organizational dynamics in online communities of creation. *Research Policy*, 42(4):941–953, 2013.

[71] S. Sadiq. The Final Frontier: A SAS Approach to Data Envelopment Analysis. In *SAS Global Forum 2011 Operations Research*, 2011. URL `http://support.sas.com/resources/papers/proceedings11/198-2011.pdf`. Paper 198-2011.

[72] B. Simpson, R. Willer, and C. L. Ridgeway. Status Hierarchies and the Organization of Collective Action. *Sociological Theory*, 30(3):149–166, 2012.

[73] J. Struben and B. Lee. Market Formation: Examining the Coordination of Heterogeneous Contributions. Technical report, DRUID, Copenhagen Business School, Department of Industrial Economics and Strategy/Aalborg University, Department of Business Studies, 2012.

[75] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser. Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6):983–1001, 2008.

[74] B. Stvilia, A. Al-Faraj, and Y. Yi. Issues of cross-contextual information quality evaluation–The case of Arabic, English, and Korean Wikipedias. *Library & Information Science Research*, 31(4):232–239, 2009.

[76] B. Suh, G. Convertino, E. H. Chi, and P. Pirolli. The singularity is not near: Slowing growth of wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration, WiKiSym 2009*, 2009.

[77] B. Uzzi. A social network's changing statistical properties and the quality of human innovation. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224023, 12pgs, June 2008.

[79] H. R. Varian. *Intermediate microeconomics: a modern approach*. W. W. Norton, New York, NY, 2nd edition, 2005.

[80] G. Von Krogh, S. Spaeth, and K. R. Lakhani. Community, joining, and specialization in open source software innovation: A case study. *Research Policy*, 32(7):1217–1241, 2003.

[82] D. Wallace and C. Van Fleet. The democratization of information? wikipedia as a reference resource. *Reference and User Services Quarterly*, 45:100–103, 2005.

[84] T. Wöhner and R. Peters. Assessing the quality of Wikipedia articles with lifecycle based metrics. In *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 1–10, New York, NY, USA, 2009. ACM.

[85] S.-H. Yu, Y.-G. Kim, and M.-Y. Kim. Do we know what really drives km performance? *Journal of Knowledge Management*, 11(6):39–53, 2007.

[86] V. Zlatić and H. Stefancić. Model of Wikipedia growth based on information exchange via reciprocal arcs. *EPL*, 93(5): 58005, 2011.

Annex 1. Extra Figures.

Figure 6: Efficiency in production edits and new knowledge for the 37 Wkipedia language projects, 2011 to 2013 (left, without taking into account the return to scale, right taking into account the return to scale)