The Virtuous Circle of Wikipedia Recursive Measures of Collaboration Structures

Maximilian Klein
OCLC Research
777 Mariners Island Blvd
San Mateo, CA, 94404
kleinm@oclc.org

Thomas Maillart
School of Information
University of California,
Berkeley, 102 South Hall
Berkeley, CA 94720
maillart
@ischool.berkeley.edu

John Chuang
School of Information
University of California,
Berkeley, 102 South Hall
Berkeley, CA 94720
chuang
@ischool.berkeley.edu

ABSTRACT
In open collaboration, knowledge is created and iteratively improved by a multitude of editors who freely choose what should be their contributions. The quality of knowledge artifacts (e.g. article, source code file) is deeply tied to their individual expertise, and to their ability to collaborate well. Conversely, the expertise of contributors is a function of artifacts contributed to. Building upon a large stream of literature on the measurement of article quality and contributor expertise, we propose a recursive algorithm to measure how editor expertise influences the quality of articles, and how contributions to articles influence editor expertise. This *bi-partite network random walker* metric reveals the specific structure of cooperation and how the quality of articles is achieved through coordination. We show that while the wisdom of crowds is well pulled in some categories, more editors per article can also create disvalue.

ACM Classification Keywords
H.5.3 [Information Systems]: Group and Organization Interfaces: computer-supported collaborative work

Author Keywords
open collaboration; bi-partite networks; performance; coordination

INTRODUCTION
In online open collaboration knowledge, artifacts such as open source code, Wikipedia articles, and 3D-printing designs, are usually produced and improved collectively by a multitude of contributors. Some people devote numerous hours of labor improving existing content and adding new features, while most contributors only make minor changes. Yet, in addition to the power of the few, a mass of small changes can make the difference as a form of emergent collective intelligence [25]. As the Internet has become pervasive in modern societies, open collaboration has permeated to a broad variety of social contexts and industries [4].

Despite *bottom-up* self-organization, participants in open collaboration can collectively achieve the production of high quality and reliable knowledge, as demonstrated for instance in Wikipedia [14]. This form of labor organization is called peer-production and it usually heavily relies on Internet communication systems. Peer-production is based on *task self-selection* and *peer-review* [3]: participants decide to contribute according to their skills, and in turn, skills are improved as they contribute more, and so on, following a virtuous circle.

Because open collaboration enjoys horizontal organization, the dynamics of contributions are contingent to the heterogenous motivations and incentives of participants [41], and some knowledge artifacts enjoy various attention from the community, with time localized bursts for hot topics [24]. These highly non-linear, transient and intrinsically unpredictable bursts of iterative improvements are the hallmark of successfully organized communities [42]. They can be rationalized by critical cascades of both individual contributions and interactive community-based iterative improvements [37]. *Individual* versus *interaction-based* mechanisms are hard to disentangle, and therefore, understanding the structure of collaboration remains a difficult challenge. For larger groups concentrating on precise problems (e.g., in open collaboration), interactions typically magnify coordination problems [15].

To understand the origins of cooperation structures and quality in open collaboration, we posit that the value of each knowledge artifact (e.g., source code file, article) is deeply tied to the expertise and the number of its contributors, who can witness potential mistakes or outdated information. Conversely, the expertise of contributors is a function of artifacts contributed to, and so on, recursively.

To measure how artifacts benefit from a larger number of editors with a given expertise, and how editors benefit from having contributed to more artifacts of some quality, we propose a *bi-partite network random walker* algorithm, which is a two

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CSCW 2015, March 14–18, 2015, Vancouver, BC, Canada.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2922-4/15/03 ...$15.00.
http://dx.doi.org/10.1145/2675133.2675286

node type extension of the recursive *pageRank* algorithm [33, 29]. We calibrate the algorithm on 12 Wikipedia categories of articles, and we show, at the level of each category, how articles do (or do not) benefit from the intervention of more editors and their expertise.

The rest of this paper is organized as follows. We first expose the reader to the large literature on Wikipedia, measuring article quality, editor expertise and their mutual interplay. We then introduce the intuition behind the *bi-partite network random walker* algorithm, as well as its implementation for the present study. Data employed and the results are then presented and discussed. We finally conclude with limitations and future research directions.

RELATED WORK

The structure and dynamics of individual and collective contributions have long since been recognized by researchers as primary factors for the achievement of high quality content, starting with scientific publications [31] and open collaboration projects [6]. In the meantime, some of these open collaboration projects have tremendously increased their size and the number of their contributors, making it hard to assess the value of each knowledge artifact, even by intensive peer-review. The Wikipedia community, as well as researchers, have tried to find ways to determine article quality and editor expertise in a systematic way. These approaches have systematically faced criticism. Many quality article metrics have been proposed from methods based on word count [5], revision history [18], general structure of articles [43], patterns of changes between article versions [45], and combinations of type and volume of edits and editor expertise [22]. Editor expertise has also been investigated by considering total number of words written, number of edits made, longevity of edits [2], time spent in edit sessions [13], and number of *barnstars* collected [30]. Other editor features that have received the attention of researchers include creative editing [19], how power editors differ from normal editors [34], and the influence of the type of contributors on the quality of articles [38].

The effort to measure article quality and editor expertise has extended to predicting the quality of contributions [12, 46], developing reputation systems for editors [1], and identifying editor candidates for promotion [7].

We believe that the general skepticism about these metrics and reputation systems is grounded in their inability to capture and make sense of coordination between contributors. Coordination, defined as an on-going process that produces other measurable outcomes, is in general hard to understand in societies [32]. CSCW researchers have been specifically concerned with coordination viewed as a feature of a community, i.e., the effect of more peers on output quality. While collaboration and additional reviews by peers are generally perceived as positive, depending on the type of tasks and their required coordination, performance can also be undermined by inadequate coordination processes [27]. On the contrary, the effects of diversity on group productivity seem to increase group productivity [9]. It was also found that editors cluster by interest, with higher coordinated efforts in densely populated clusters [20]. In particular, Wikipedia has been a heavily explored *field* for social scientists, starting with concerns on the effects of peer-review and whether single or repeated contributions by editors would help improve the quality of articles [18, 44].

As an hybrid example, the specific problem of coordination in *featured* Wikipedia articles, which are heavily contributed over short time periods, has raised concerns on implicit versus explicit coordination processes and the limited positive quality it can bring when editors are too numerous [26].

The connection between article quality and editor expertise is present in nearly all literature aiming to understand the effects of the coordination process on the value of Wikipedia articles. The typical structure of networks with edges that connect uniquely two kinds of nodes is called *bi-partite* [31]. The analysis of patterns in Wikipedia bi-partite networks, with editors being one node type and articles the other, confirmed the existence of overlapping cliques of densely connected articles and editors [20]. A more detailed analysis of medical and health-related articles on Wikipedia, showed that the position of articles in the bi-partite network of articles and editors significantly influenced its quality [21].

Recent developments in the science of bi-partite networks has shown the feasibility to rank entities of each type through a recursive algorithm called *method of reflections*.This method has been tested on the bi-partite network of countries exporting products [17, 16]. The method of reflections has been improved and complemented in more recent work, mainly to improve its robustness [39, 11, 40, 10]. Caldarelli et al. [8] have proposed an alternative method, based on biased stochastic Markov chains, which helps further understand the mutual influence between nodes in bi-partite networks.

METHOD

We present a comprehensive method to *reverse-engineer* coordination as a feature of categories in Wikipedia. We expect that categories of articles exhibit more or less coordination, which in turn can be captured by the fundamental structure of the *bi-partite* network of articles and editors. The underlying idea of our model is to account for the recursive flow of *value* circulating between editors and articles, with editors benefiting from having edited higher quality articles, and articles having been edited by more expert editors. If coordination brings "more than the sum of its parts", then articles benefit from more editors, and primarily from expert editors. Conversely, if coordination is not efficient, *disvalue* is generated by more editors editing one article, or by an editor contributing to many articles in the category. A typical example of disvalue is vandalism [13].

We now turn to explaining the formalism of the *bi-partite random walker* method, and we show how the structure of collaboration can be encapsulated and measured with a single parameter. We consider a simple input, which is a representation of the bi-partite network of editors and their contributions to articles. Namely, let us consider a matrix M_{ea} of all editors having contributed to a Wikipedia category of articles. M_{ea} takes value 1 if editor e has edited article a, and 0 otherwise. For simplicity and because mixed results have been

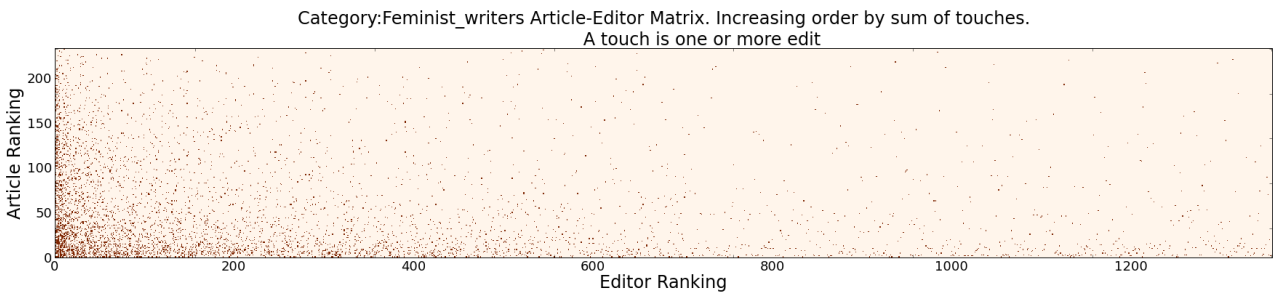

Figure 1. Typical M_{ea} matrix for a Wikipedia category (here, *Feminist Writers*) ordered on both dimensions by descending order of number of articles modified by an editor (horizontal axis) and of number editors who have modified an article (vertical axis). The structure of M_{ea} is triangular and shows that some editors have a pervasive activity over articles, while most editors edit only a few. Similarly, some articles receive widespread attention by editors, while most articles are modified only by a few editors.

previously reported in the literature [44], we consider only if editors have ever touched an article, rather than incorporating a more fine grained metric, such as the count of edits made by an editor on a specific article. As a robustness check, we show later that using edit counts reduces drastically the fitness of the method. For the category *Feminist Writers*, as presented on Figure 1, M_{ea} exhibits a triangular structure in which editors (resp. articles) are sorted (max on the bottom-left corner) by the number of articles they have touched (resp. by the number of editors who have touched each article). M_{ea} is the only input of the *bi-partite random walker* model.

Given M_{ea}, the simplest, and arguably naive, way to assess the contribution value (i.e., the *expertise* thereafter) of an editor is obtained by summing the number of articles ever edited out of all articles in a category. Similarly, a simple *quality* measure for an article is the sum of editors who have ever modified it, following the famous adage on open source development: "Given enough eyeballs, all bugs are shallow" [35]. These crude expertise and quality metrics for editors and articles, respectively given by,

$$\begin{cases} w_e^{(0)} = \sum_{a=1}^{N_a} M_{ea} \equiv k_e \\ w_a^{(0)} = \sum_{e=1}^{N_e} M_{ea} \equiv k_a \end{cases} \tag{1}$$

are the zeroth order of our algorithm. They are the initial step of the *method of reflections* proposed by Hidalgo et al., which derives the value of producing entities (i.e., editors) from products (i.e., articles), and *vice versa* [17, 16]. To help capture the intuition behind the method of reflections for open collaboration, we walk through the first and second iterations:

- **1st order iteration,**
  - **Articles**: if an article has been edited by higher expertise editors, it is of higher quality. That is, quality is a function of expertise calculated from zeroth iteration expertise scores.
  - **Editors**: conversely, if an editor has contributed to higher quality articles, her expertise is higher. That is, expertise is a function of quality calculated from zeroth iteration quality scores.
- **2nd order iteration,**
  - **Articles**: if an article has been edited by higher expertise editors who have edited higher value articles, which in turn have been edited by higher expertise contributors, the article quality is higher. That is, quality is a function of expertise calculated from 1st iteration expertise scores.
  - **Editors**: conversely, if an editor has edited higher quality articles, which have been edited by better editors who have edited higher quality articles, then expertise is higher. That is, expertise is a function of quality calculated from 1st iteration quality scores.
- **And so on, recursively.**

Although interpretation is difficult past the very first iteration steps, at each iteration, the algorithm incorporates additional information on the quality of the articles and expertise of editor from the neighboring nodes in the bi-partite network. The higher order iterations can be modeled as a Markov process of random walkers on a bi-partite network, jumping with some probability from one node type to another node type [8]. A schematic representation of the random walk process on a bi-partite network is depicted in Figure 2. The intuition is the following: a random walker jumps with some probability from an editor to a given article (i.e., the editor's expertise is positively influenced by the article's quality), and with another probability from an article to a given editor (i.e. the value of the article is positively by the editor's expertise). The binary matrix M_{ea} determines whether a jump between each pair of nodes is possible: if two nodes e and a are not directly connected ($M_{ea} = 0$), the transition probability is 0. Conceptually, the *bi-partite network random walker* model is an extension of the single node type (i.e. Web pages) *Page Rank* Google search algorithm [33, 29] to two types of nodes.

We call $w_e^{(n)}$ the expertise of an editor and $w_a^{(n)}$ the quality of an article at the nth iteration, and we define the following Markov process on the bi-partite network of collaboration,

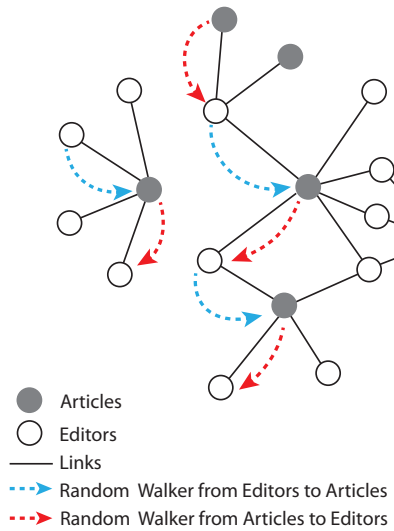

Figure 2. Representation of random walkers jumping from editors to articles (red dotted arrows) and from articles to editors (blue dotted arrows). The intuition is the following: a random walker jumps with some probability from an editor to a given article (i.e., the editor's expertise is positively influenced by the article's quality), and with another probability from an article to a given editor (i.e., the value of the article positively influences the editor's expertise).

$$\begin{cases} w_e^{(n+1)}(\alpha,\beta) = \sum_{a=1}^{N_a} G_{ea}(\beta)\, w_a^{(n)}(\alpha,\beta) \\ w_a^{(n+1)}(\alpha,\beta) = \sum_{e=1}^{N_e} G_{ae}(\alpha)\, w_e^{(n)}(\alpha,\beta) \end{cases} \tag{2}$$

with G_{ea} the probability to jump from article a to editor e in a single step, and the probability G_{ae} to jump from editor e to article a also in a single step. These transition probabilities are given by,

$$\begin{cases} G_{ea}(\beta) = \frac{M_{ea} k_e^{-\beta}}{\sum_{e'=1}^{N_e} M_{e'a} k_{e'}^{-\beta}} \\ G_{ae}(\alpha) = \frac{M_{ea} k_a^{-\alpha}}{\sum_{a'=1}^{N_a} M_{ea'} k_{a'}^{-\alpha}}. \end{cases} \tag{3}$$

The transition matrices $G_{ea}(\beta)$ and $G_{ae}(\alpha)$ depend only on the initial conditions: the binary matrix M_{ea}, as well as k_e and k_a given by (1), and are controlled only by parameters α and β. We shall therefore explain only how β influences the probability to jump from an article to an editor (i.e. the value of the article positively influences the editor's expertise). For $\beta = 0$, we recover the zeroth order iteration (1). For $\beta > 0$, the probability to jump from article a to editor e is a power law function $\sim 1/k_e^{\beta}$ of the sum of articles k_e modified by editor e. Hence, the larger k_e, the lower the probability to jump from a to e relative to other editors. On the contrary, if $\beta < 0$ the probability to jump from an article to an editor is a positive function of the sum of articles modified by the editor. For $-1 < \beta < 0$, the function is concave, while for $\beta < -1$, the function is convex, which means that the more articles have been edited by the editor, the even more

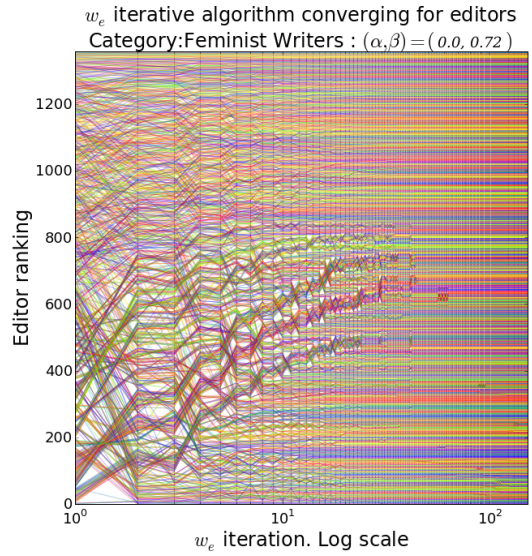

Figure 3. Convergence of the ranked expertise w_e of editors having contributed to articles in the Feminist Writers category on Wikipedia for arbitrary control parameters: $(\alpha,\beta) = (0, 0.72)$. Starting from the sum of contributed articles as the initial step, we can see how the algorithm progressively ranks editors: some editors with initial lowest rank, i.e., with few articles edited, get a higher rank as the number of iterations increases. Similarly, some initially high ranked editors, gradually drop in the ranking. In the case *Feminist Writers*, the algorithm converges after 64 iterations.

the positive influence on articles. In a nutshell, β relates the amount of articles edited on the overall editor's expertise. If $\beta \gg 0$, the positive influence of the number of contributed articles on the editor's expertise decreases. If β close to 0, the number of contributed articles increases linearly the editor's expertise. The same considerations hold for α and the probability $G_{ae}(\alpha)$ to jump from an editor to an article (i.e. the expertise of the editor positively influences the quality of an article).

Figure 3 shows the evolution of expertise w_e ranked among editors having contributed to articles in the *Feminist Writers* category on Wikipedia for the set of control parameters $(\alpha,\beta) = (0, 0.72)$. We can see how the algorithm progressively ranks editors: some editors with initial low rank (i.e. with few articles edited), get a higher rank as more information is incorporated from neighboring nodes as the number of iterations increases. In that case ($\beta > 0$), higher ranked editors have edited and contributed to fewer, but higher quality articles (i.e. articles edited by more editors who have edited less articles). Similarly, some initially high ranked editors, gradually drop in the ranking. They have edited many, but lower quality articles.

Upon calibration of the bi-partite random walker model with ground-truth metrics of article quality and editor expertise, the parameters α and β directly inform how coordination generates value (i.e. more articles edited by more editors brings value), or on the contrary, if value is created by small clusters of highly experienced editors. This latter scenario implies less coordination among large crowds of contributors.

Category	Articles	Editors	Edits
American male novelists	2,460	9,946	224,783
2013 films	1,896	5,215	150,956
American women novelists	1,936	5,968	138,716
Nobel Peace Prize laureates	104	4,165	91,522
Sexual acts	93	2,190	45,901
Economic theories	212	1,145	28,658
Feminist writers	233	1,357	25,738
Yoga	123	730	25,315
Military history of the US	180	854	20,172
Counterculture festivals	66	578	10,515
Computability theory	92	272	7,117
Bicycle parts	70	210	4,981

Table 1. Size statistics of investigated Wikipedia categories sorted by total edits.

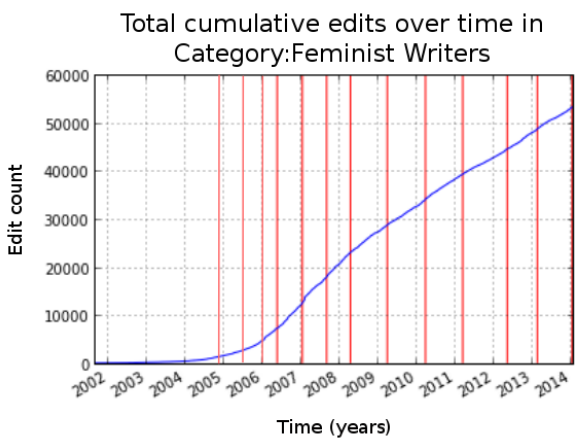

Figure 4. Cumulative edits made in Category *Feminist writers* **(blue line). Vertical red lines represent the 13 snapshots taken at 2.5%, 5%, 7.5% and then, 10%, 20%, 30%, . . . , 100% of edits.**

DATA

To uncover the coordination features of Wikipedia categories, we seek to calibrate the bi-partite random walker model with empirical data. For that, we aim to find values of α and β, which minimize the distance between rankings, of both article quality and editor expertise, given by the model on the one hand, and on the other hand, by ground truth metrics obtained independently. We performed the model calibration for 13 snapshots (see Figure 4) for each of the 12 categories of Wikipedia articles presented in Table 1. To account as much as possible for collaboration structures, we have selected a spectrum of categories ranging from anarchy and edit-warring (e.g., *Sexual Acts*) to acknowledged high organization level (e.g., *Military history of the US*).

For each category and snapshot we have built the binary matrix M_{ea} by parsing all edit histories of all articles in the main namespace up to the snapshot time. We set $M_{ea} = 1$ for editor e having modified article a, and $M_{ea} = 0$ otherwise. We considered only editors who made 5 or more edits to any article in the category. We also discarded all software robots (i.e., *bots*) that programmatically edit Wikipedia.

To calibrate α and β, we resorted to state-of-the-art ground truth evaluations for editor expertise $\bar{w}_e$ and article quality $\bar{w}_a$. From these exogenous evaluations, we ranked editors and articles according to their expertise and quality respectively. We then performed a grid search for values of α^* and β^*, which maximize the Spearman rank-correlation ρ_e and ρ_a between rankings obtained from the bi-partite random walker model (w_e, w_a) and from exogenous metrics $(\bar{w}_e, \bar{w}_a)$. Actually, (α^*, β^*) must maximize both ρ_e and ρ_a, even though ρ_e and ρ_a might actually be different. The optimization function of (α^*, β^*) is given by,

$$\begin{cases} (\alpha^*, \beta^*) = \text{argmax}_{\alpha,\beta}(\rho_e) \\ (\alpha^*, \beta^*) = \text{argmax}_{\alpha,\beta}(\rho_a). \end{cases} \tag{4}$$

The set (α^*, β^*) characterizes how the structure of collaboration creates values in each Wikipedia category. To calibrate the model, we have used ground truth metrics for article quality and editor expertise.

A variety of techniques for measuring article quality have been proposed, from a collection of word-count related metrics [5] to analyzing persistent and transient contributions throughout revisions [45]. We have selected metrics used on Wikipedia [43, 28] which have also been used in the CSCW literature in different combinations [22, 23]. Our measure of actual article quality is a combination of 5 text analysis metrics: (i) ratio of mark-up to readable text, (ii) number of headings, (iii) article length, (iv) citations per article length, (v) number of outgoing intra-Wiki links. We performed principal component analysis (PCA) for each category and snapshot in order to reduce dimensionality from 5 metrics to a single one (i.e., the principal component). The variance explained by the principal component varied between 0.5 and 0.72, confirming the dominance of the axis of maximum variance. Even though these five article quality metrics do not directly incorporate information from the bi-partite network (e.g. number of contributors, number of edits), they might indirectly be related, as some editors specialize in some types of editing, such as adding citations or systematically improving the structure of articles.

Editor expertise is even more difficult to address. As each article is a blend of edits by several contributors, disentangling the value of individual contributions remains a challenge, which has occupied Wikipedia researchers long before us. Techniques ranging from parsing the revision history to measuring text survival rate [2] have been used. Although they are sophisticated, these metrics pose a variety of problems. For instance, some articles are likely to evolve not only because former editors introduced wrong statements, but simply because of new information brought to public attention.

We decided to use the *labor hours* metric proposed by Geiger and Halfaker [13], which is calculated for each editor by taking contribution history up to the snapshot point. All edits made within 1 hour of a previous edit are counted as one *edit session*. If more than one hour separates two edits, a new period of edits starts. The expertise expressed in labor hours is the sum of edit sessions. For the calculation of ground

truth expertise, we only consider edits for a given category, although the same editor might have simultaneously edited other categories of Wikipedia. This metric purposefully does not tell how this time is spent in the number (resp. size) of edits actually made during a period, or whether the effort has been spent on one or multiple articles. In other words, we do not distinguish a single minded user spending 100 hours on a single article trying to get it to "feature article status" from a user making 100 stub articles for 1 hour each. However, it is clear that a highly contributing editor has more chance to touch more articles over time, but the metric does not distinguish if editors had a dispersed contribution or concentrated on a single article.

How this effort is distributed and brings quality is precisely what the *bi-partite random walker* model can say that other metrics cannot. In a nutshell, parameters α and β describe the most likely structure of collaboration given calibration of the model to ground truth quality and expertise metrics. The higher the correlation between the model and the exogenous metrics, the better the collaboration structure is captured by the model.

RESULTS

To understand how contributions by editors to articles shape the structure of collaboration in Wikipedia, we have performed a calibration of the bi-partite network random walker model on 12 Wikipedia categories (c.f., Table 1) with 13 snapshots each (Figure 4). For each category and snapshot, we found the set of parameters (α^*, β^*), which maximize the fitness of the model to ground truth metrics of article quality and editor expertise. Figure 5 shows typical optimization landscapes, which maximize the rank correlation ρ_e (upper panel) between editor expertise w_e obtained from the model and expertise obtained from ground truth measures $\bar{w}_e$. The same is done for rank correlation ρ_a between w_a and $\bar{w}_a$ (lower panel).

The maximum achievable rank-correlation with ground truth expertise and quality metrics for editors [13] and articles [43] shows that the bi-partite network random walker model accounts particularly well for both quality of articles ($0.58 < \rho_a < 0.91$) and expertise of editors ($0.46 < \rho_e < 0.75$) at the last snapshot. Actually, the model reproduces very well, and very early the ranking of editors and articles according to the ground truth metrics as shown on Figure 6. In particular, the quality of articles is very well accounted for, while the level of correlation with the ground truth of editor expertise exhibits a slightly concave, or at least linear, increase.

For the latest snapshot (i.e., the state of contributions in February 2014), we find that the best possible α^* is 0 in all circumstances, while β^* varies considerably across categories. Table 2 shows the categories ordered by β^* (and $\alpha^* = 0$ for the sake of completeness), as well as the corresponding maximum rank correlations ρ_e and ρ_a. Since there is no single optimal value for (α^*, β^*), but rather a space of optimal values for ρ_e and ρ_a separately, we have searched for a set of values that jointly maximizes both ρ_e and ρ_a. The optimal parameter $\alpha^* = 0$ means that editor expertise always benefits from contributions as a linear function of the

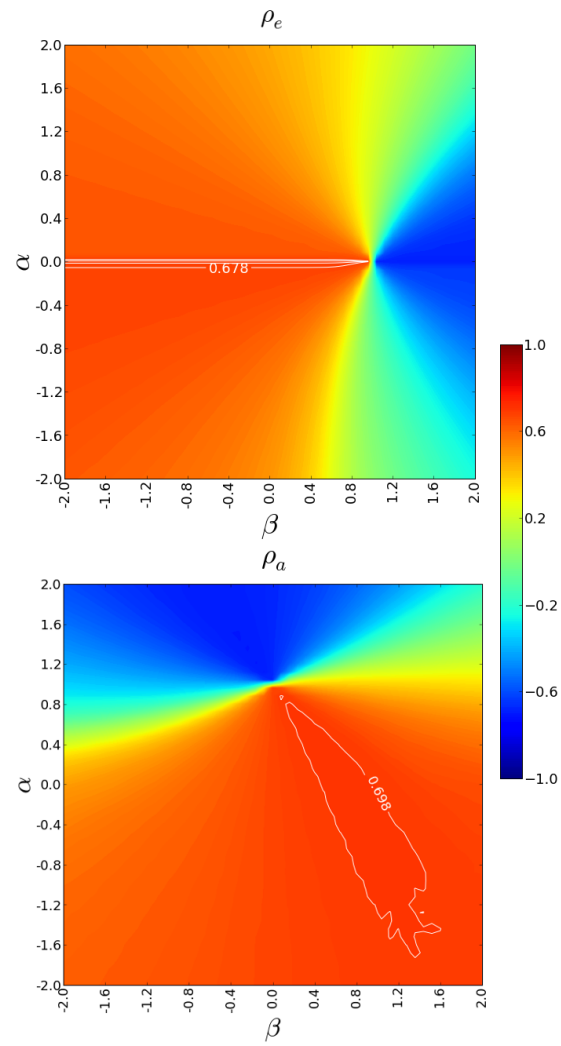

Figure 5. Typical landscape of maximum correlation as a function of α and β for articles (upper panel) and editors (lower panel). The contour line shows the 95th percentile of the rank correlation over the landscape. The category displayed here is *Feminist Writers*, for the last snapshot ending February 2014.

number of articles edited [compounded over iterations of the recursive algorithm defined by formula (2)]. However, β^* exhibits a continuum of values between 0 (*Bicycle parts* and *US Military History*) and 1.52 (*Sexual Acts*). β controls the influence of the number of editors on the quality of a given article. When $\beta \approx 0$, the quality of articles increases as a linear function of the number of editors who have modified them. For $\beta \gg 0$, the marginal gain of having more editors for a given article decreases. So, in that case, when the number of editors touching an article increases, the marginal quality improvement decreases.

The evolution of β^* over snapshots as shown on Figure 6 exhibits large variations for early snapshots corresponding to the early 10% of overall contributions per category (i.e., the 4th snapshot). While β^* exhibits a tendency to more stability afterwards, large variations within the range 0 to 1.5 can

	Category	ρ_a	ρ_e	α^*	β^*
1	Bicycle parts	0.90	0.46	0.00	0.00
2	Military history of the US	0.58	0.70	0.00	0.00
3	Computability theory	0.77	0.56	0.00	0.32
4	American male novelists	0.67	0.75	0.00	0.40
5	2013 films	0.72	0.55	0.00	0.48
6	Economic theories	0.74	0.70	0.00	0.48
7	American women novelists	0.63	0.75	0.00	0.64
8	Feminist writers	0.70	0.69	0.00	0.72
9	Yoga	0.64	0.57	0.00	1.12
10	Nobel Peace Prize laureates	0.91	0.66	0.00	1.20
11	Counterculture festivals	0.80	0.61	0.00	1.36
12	Sexual acts	0.63	0.66	0.00	1.52

Table 2. Categories ordered by increasing β^* obtained from best rank-correlation ρ_a and ρ_e of the bi-partite network random walker with the ground truth. As shown on the upper panel of Figure 5, highest rank-correlation is always obtained for $\alpha^* = 0$ suggesting that editors are experts in direct proportion to the number of articles they edit. The different values of β^* show the effect of marginal editors on a article. As β^* grows larger having more editors shows diminishing returns on article quality - "too many cooks spoil the broth".

be observed for some categories, suggesting that organization and coordination level changes can occur as categories develop.

DISCUSSION

To understand how the structure of collaboration influences article quality, we have applied and tested the *bi-partite network random walker* model for a variety of categories in Wikipedia. Our results show that the model accounts well for the quality of articles $\langle \rho_a \rangle \approx 0.64$ and for the expertise of contributors $\langle \rho_e \rangle \approx 0.72$, and overall exhibits a high degree of fitness. Moreover, ρ_a remains stable over time, while ρ_e increases, suggesting that the model better reflects editor expertise as more contributions to a broader set of articles occur, i.e., when the bi-partite network gets more densely connected. This suggests that loosely connected entities, either articles and editors, cannot be ranked accurately. From Figure 1 and from Table 1, we see that there are always significantly more editors than articles for each category. Hence, the probability for an article to get contributions early on is higher than the probability to find editors who have contributed to a lot of articles early.

To account for single-minded editors who have concentrated on only one or few articles, we have tested the *bi-partite random walker* model with a different input, namely the matrix of edit counts (instead of a binary matrix). As shown on Figure 7, the model using the *edit counts* input matrix accounts nearly as well for article quality, while it does a much worse job ranking editor expertise compared to a *binary* input matrix. Counter-intuitively, we observe a *less is more* situation: the number of articles ever touched by an editor better reflects the structure of collaboration and value creation, compared to edit counts, a much richer information input. Also, the labor-hour ground truth metric for editors is more a proxy of number of edits rather the number of articles ever touched [13]. Nevertheless, the model does not perform as well with *edit counts* as an input. This suggests that what really counts for

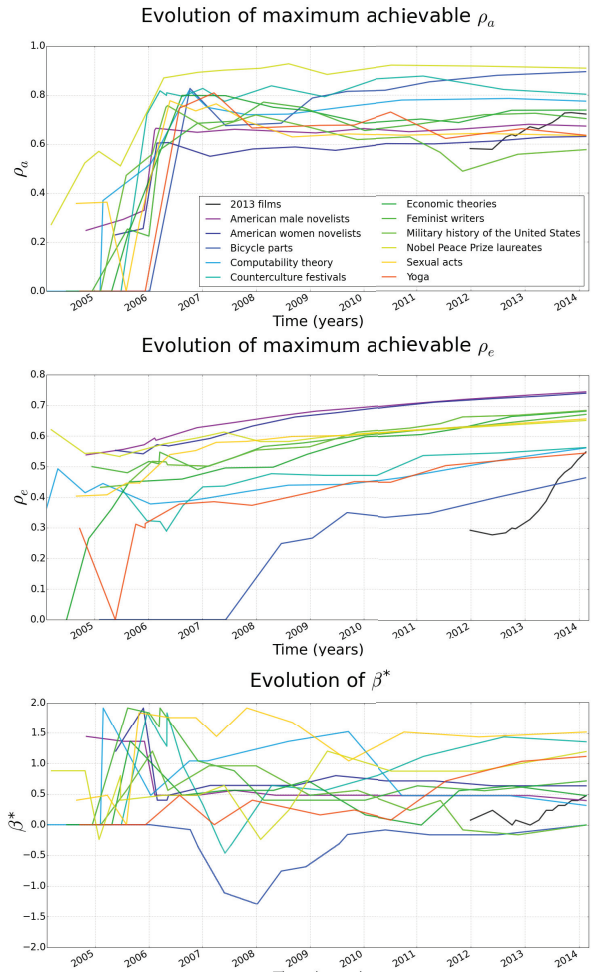

Figure 6. Evolution of Spearman ρ rank correlations between the ranking obtained from the calibrated model and the actual values for each category and for editors (upper panel) and articles (middle panel). Corresponding β^* values are also shown for interest (lower panel). The correlations are generally quite high : $0.46 < \rho_e < 0.75$ with $\langle \rho_e \rangle = 0.64$ for editors and $0.57 < \rho_a < 0.91$ with $\langle \rho_a \rangle = 0.72$. ρ_a is stable over time, which means that the quality of articles can be well captured early on by the model. However, ρ_e exhibits a convex increase over time, suggesting that it takes time (i.e., lots of edits) to capture well the expertise of editors.

assessing the expertise of an editor is the number of articles touched, rather than the number of edits per article.

We now discuss how the fitted parameters α^* and β^* inform on the structures of collaboration in Wikipedia categories. On the one hand, we have found $\alpha^* \approx 0$ for all categories, reflecting the positive influence of the number of articles edited on editor expertise. This result is compatible with previous results by Keegan et al. [23]. On the other hand, β^* varies across categories with values ranging from 0 to 1.52 at the last snapshot. β can be considered as a measure of the collaboration structure: the smaller β, the more articles benefit from more editors. On the contrary, the larger β, the more articles benefit from less editors. If we consider for instance *Sexual acts*, a category that could be considered taboo or perverse with articles being the least collaboratively edited: $\beta > 1$

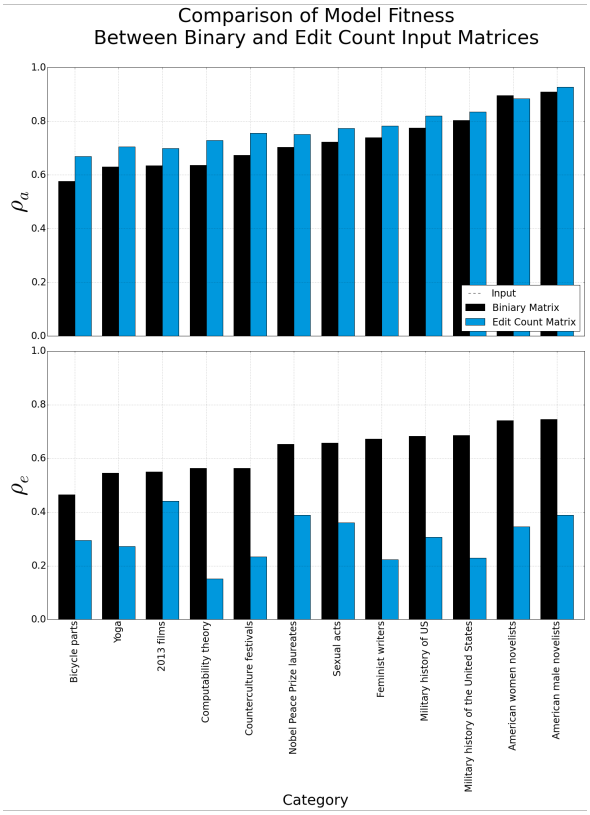

Figure 7. Comparison between ρ_a (upper panel) and ρ_e (lower panel) for two input matrices: binary (black) and edit counts (blue) for each category numbered according to Table 2. While taking edit counts as the input matrix only marginally increases ρ_a, it drastically reduces ρ_e.

reflects that an editor with edits to many articles will see her ranking drop. Indeed, it should not be taken for granted that contributions are necessarily positive. Because of their socially-sensitive nature, articles about sexual acts are particularly prone to attracting vandals or edit-warring behaviors across the whole category. Therefore, editors making the most edits are not necessarily the ones improving article quality, as we would typically expect. This result is again compatible with previous research, which has shown that, in some circumstances, most active editors exhibit deleting behaviors that lower metric-based article quality ratings [22].

Conversely, the category *Military History of the US* is famous for its self-organized task-forces. At the latest snapshot $\beta = 0$, it is the one of only a few categories we have analyzed, which exhibits β consistently negative over time. Accordingly, the marginal quality of articles is positively influenced by the number of editors touching the article. Unsurprisingly, *Military History of the US* is literally a *WikiProject* with a hierarchy of coordinators, an active IRC channel, and a mailing list. As a result of better coordination, there is less edit-warring and more efficient contributions: editors edit articles with well-defined task at hand.

LIMITATIONS AND FUTURE WORK

As a two node extension of the *pageRank* algorithm [33, 29], the *bi-partite network random walker* model is an efficient approach to recursively traverse the complex network of articles and editors in Wikipedia. Our results show that model calibration accounts well with ground-truth metrics, and can help characterize how more contributors for each article and better (resp. less) coordination create value (resp. destroy value) in open collaboration. Its very simple input (a binary matrix of contributions) makes it computationally affordable, though not cheap. While applying this algorithm to the entire Wikipedia would be a challenge, it is straightforward to use on small wikis or most open source software projects.

Our results show a first attempt to understand the structure of cooperation and how value is created with a unique model, which can be fully rationalized. The pertinence of the bi-partite network random walker for the study of open collaboration shall be confirmed by future work, to examine in a systematic way some of the results reported in this paper.

Namely, we would have expected that all categories, or at least each category, would exhibit a typical set (α^*, β^*) of explanatory parameters, which in turn would help gain better understanding of the general structure of collaboration in Wikipedia. Not only our results show that β^* varies across categories, but can also vary significantly over time for some of the categories we have analyzed. These results require further scrutiny on the evolution of contribution structures and coordination processes, in particular in these specific categories.

Future work shall also be devoted to further validation, in order to bring quantitative evidence that the model can systematically account for the influence of the coordination feature on value generated by contributions. We have only indirect evidence that coordination is efficient in some categories, like *Military History of the US*. An orthogonal way for testing the model would require measuring specifically the level of constructive (resp. destructive) interactions between editors, on articles (e.g. revert actions), and on usual communication channels used by the community of a specific category (e.g., discussion page, IRC channel, mailing list). A negative relationship between β^* and the amount of positive interactions would further demonstrate the validity of the model.

The structure of the input matrix (i.e., its dimensions and sparsity) requires further scrutiny. We aim to know the sensitivity of β^* to the total number of editors versus the total number of articles in a category. Presumably coordination problems are more likely to occur if there are more editors per article. To thoroughly perform these types of tests, we need to investigate more categories of Wikipedia.

The progressive validation process we have described will help gain trust in the model [36], and will perhaps allow meaningful out-of-sample predictions of article quality and editors experience rankings, given the structure of cooperation characterized by β^*. Conversely, the *bi-partite network random walker* model could be used in the future to set incentives for a reward system that would specifically encourage cooperation. It could also be used as a *Suggestbot*[1] to help new editors find friendly Wikipedia categories to start

[1] https://en.wikipedia.org/wiki/User:Suggestbot

their on-boarding process. This is a reverse approach from current on-boarding practices, where an interest topic is first chosen and then an edit is made in basically a random-chosen environment.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation under award CCF-0424422 (TRUST). One of the authors (T.M.) acknowledges support by the Swiss National Science Foundation (Grant Nr. PA00P2-145368).

REFERENCES

1. Adler, B. T., and de Alfaro, L. A Content-driven Reputation System for the Wikipedia. In *Proc. WWW '07*, ACM (New York, NY, USA, 2007), 261–270.
2. Adler, B. T., de Alfaro, L., Pye, I., and Raman, V. Measuring Author Contributions to the Wikipedia. In *Proc. WikiSym '08*, ACM (New York, NY, USA, 2008), 1–10.
3. Benkler, Y. Coase's Penguin, or, Linux and "The Nature of the Firm". *The Yale Law Journal 112*, 3 (Dec. 2002), 369+.
4. Benkler, Y. *The Penguin and the Leviathan: How Cooperation Triumphs over Self-Interest*, 1 ed. Crown Business, Aug. 2011.
5. Blumenstock, J. E. Size Matters: Word Count As a Measure of Quality on Wikipedia. In *Proc. WWW '08*, ACM (New York, NY, USA, 2008), 1095–1096.
6. Bryant, S. L., Forte, A., and Bruckman, A. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proc. GROUP '05*, ACM (New York, NY, USA, 2005), 1–10.
7. Burke, M., and Kraut, R. Taking up the mop: identifying future wikipedia administrators. In *CHI'08 extended abstracts on Human factors in computing systems*, ACM (2008), 3441–3446.
8. Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., and Tacchella, A. A network analysis of countries' export flows: firm grounds for the building blocks of the economy. *PloS one 7*, 10 (2012).
9. Chen, J., Ren, Y., and Riedl, J. The Effects of Diversity on Group Productivity and Member Withdrawal in Online Volunteer Groups. In *Proc. CHI '10*, ACM (New York, NY, USA, 2010), 821–830.
10. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., and Pietronero, L. Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products. *PloS one 8*, 8 (2013).
11. Cristelli, M., Tacchella, A., Gabrielli, A., Pietronero, L., Scala, A., and Caldarelli, G. Competitors' communities and taxonomy of products according to export fluxes. *The European Physical Journal-Special Topics 212*, 1 (2012).
12. Druck, G., Miklau, G., and McCallum, A. Learning to Predict the Quality of Contributions to Wikipedia. *Wikipedia and Artificial Intelligence: An Evolving Synergy* (2008), 7.
13. Geiger, R. S., and Halfaker, A. Using Edit Sessions to Measure Participation in Wikipedia. In *Proc. CSCW '13*, ACM (New York, NY, USA, 2013), 861–870.
14. Giles, J. Internet encyclopaedias go head to head. *Nature 438*, 7070 (Dec. 2005), 900–901.
15. Halfaker, A., Geiger, R. S., Morgan, J. T., and Riedl, J. The Rise and Decline of an Open Collaboration System. *American Behavioral Scientist 57*, 5 (May 2013), 664–688.
16. Hidalgo, C. A., and Hausmann, R. The building blocks of economic complexity. *Proc. National Academy of Sciences 106*, 26 (June 2009), 10570–10575.
17. Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. The Product Space Conditions the Development of Nations. *Science 317*, 5837 (July 2007), 482–487.
18. Hu, M., Lim, E. P., Sun, A., Lauw, H. W., and Vuong, B. Q. Measuring Article Quality in Wikipedia: Models and Evaluation. In *Proc. CIKM '07*, ACM (New York, NY, USA, 2007), 243–252.
19. Iba, T., Nemoto, K., Peters, B., and Gloor, P. A. Analyzing the Creative Editing Behavior of Wikipedia Editors. *Procedia - Social and Behavioral Sciences 2*, 4 (Jan. 2010), 6441–6456.
20. Jesus, R., Schwartz, M., and Lehmann, S. Bipartite Networks of Wikipedia's Articles and Authors: A Meso-level Approach. In *Proc. WikiSym '09*, ACM (New York, NY, USA, 2009), 1–10.
21. Kane, G. C. It's a Network, Not an Encyclopedia: A Social Network Perspective on Wikipedia Collaboration. *Academy of Management Proceedings 2009*, 1 (Aug. 2009), 1–6.
22. Kane, G. C. A Multimethod Study of Information Quality in Wiki Collaboration. *ACM Trans. Manage. Inf. Syst. 2*, 1 (Mar. 2011).
23. Keegan, B., Gergle, D., and Contractor, N. Do Editors or Articles Drive Collaboration?: Multilevel Statistical Network Analysis of Wikipedia Coauthorship. In *Proc. CSCW '12*, ACM (New York, NY, USA, 2012), 427–436.
24. Keegan, B., Gergle, D., and Contractor, N. Hot Off the Wiki: Structures and Dynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist* (Jan. 2013), 0002764212469367+.
25. Kittur, A., Chi, E., Pendleton, B. A., Suh, B., and Mytkowicz, T. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web 1*, 2 (2007), 19.

26. Kittur, A., and Kraut, R. E. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In *Proc. CSCW '08*, ACM (New York, NY, USA, 2008), 37–46.
27. Kittur, A., Lee, B., and Kraut, R. E. Coordination in Collective Intelligence: The Role of Team Structure and Task Interdependence. In *Proc. CHI '09*, ACM (New York, NY, USA, 2009), 1495–1504.
28. Klein, M. Kumusha Takes Wiki: Actionable Metrics for Uganda and Côte d'Ivoire, 2014.
29. Kleinberg, J. M. Authoritative Sources in a Hyperlinked Environment. *J. ACM 46*, 5 (Sept. 1999), 604–632.
30. Kriplean, T., Beschastnikh, I., and McDonald, D. W. Articulations of Wikiwork: Uncovering Valued Work in Wikipedia Through Barnstars. In *Proc. CSCW '08*, ACM (New York, NY, USA, 2008), 47–56.
31. Newman, M. E. J. The structure of scientific collaboration networks. *Proc. National Academy of Sciences 98*, 2 (Jan. 2001), 404–409.
32. Ostrom, E. *Governing the commons : the evolution of institutions for collective action*. Cambridge University Press, Nov. 1990.
33. Page, L., Brin, S., Motwani, R., and Winograd, T. The PageRank citation ranking: Bringing order to the web. *Technical Report* (1999).
34. Panciera, K., Halfaker, A., and Terveen, L. Wikipedians Are Born, Not Made: A Study of Power Editors on Wikipedia. In *Proc. GROUP '09*, ACM (New York, NY, USA, 2009), 51–60.
35. Raymond, E. The cathedral and the bazaar. *Knowledge, Technology & Policy 12*, 3 (Sept. 1999), 23–49.
36. Sornette, D., Davis, A. B., Ide, K., Vixie, K. R., Pisarenko, V., and Kamm, J. R. Algorithm for model validation: Theory and applications. *Proc. National Academy of Sciences 104*, 16 (Apr. 2007), 6562–6567.
37. Sornette, D., Maillart, T., and Ghezzi, G. How Much is the Whole Really More than the Sum of its Parts? 1 + 1 = 2.5: Superlinear Productivity in Collective Group Actions. *PloS ONE 9*, 8 (2014).
38. Stein, K., and Hess, C. Does It Matter Who Contributes: A Study on Featured Articles in the German Wikipedia. In *Proc. HT '07*, ACM (New York, NY, USA, 2007), 171–174.
39. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. A new metrics for countries' fitness and products' complexity. *Scientific reports 2* (2012).
40. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. Economic complexity: Conceptual grounding of a new metrics for global competitiveness. *Journal of Economic Dynamics and Control* (2013).
41. von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M. W. Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. *MIS Quarterly 36* (2012).
42. von Krogh, G., Maillart, T., Haefliger, S., and Sornette, D. Designing organizations for productive bursts. *in revision* (2014).
43. Wang, M. W., Cosley, D., and Riedl, J. Tell Me More: An Actionable Quality Model for Wikipedia. In *Proc. WikiSym '13*, ACM (New York, NY, USA, 2013).
44. Wilkinson, D. M., and Huberman, B. A. Cooperation and Quality in Wikipedia. In *Proc. WikiSym '07*, ACM (New York, NY, USA, 2007), 157–164.
45. Wöhner, T., and Peters, R. Assessing the Quality of Wikipedia Articles with Lifecycle Based Metrics. In *Proc. WikiSym '09*, ACM (New York, NY, USA, 2009), 1–10.
46. Zeng, H., Alhossaini, M. A., Ding, L., Fikes, R., and McGuinness, D. L. Computing Trust from Revision History. In *Proc. PST '06*, ACM (New York, NY, USA, 2006), 1.