Characterizing Contribution Strategies and Quality with Bi-Partite Networks in Wikipedia

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
thomas.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 online open collaboration, knowledge is created and iteratively improved by a multitude of editors, who freely choose what should be their contributions. The *bottom-up* nature of this new kind of labor organization, makes it very difficult to systematically quantify how value is created, and quality achieved, from the complex dynamics of self-organized communities.. In this paper, we recognize that the value of each knowledge unit (e.g., article, source code file) is deeply tied to the expertise and the number of its contributors. Conversely, the expertise of contributors is a function of knowledge units contributed to. We propose a *bi-partite network random walker* model, which measures how editors expertise influences the quality of articles, how contributions to articles influence editors expertise, and so on, recursively. We calibrate the model on 12 Wikipedia categories of articles, and we find how the complex structure of knowledge production influences the quality of knowledge produced and the expertise of editors. We also find that some categories of articles are more able to pull the wisdom of crowds, while for other categories, more contributions to articles by multiple editors, create marginal dis-value. These differences might originate from the ability of communities to organize, as well as from the very nature of knowledge created, either requiring few experts, or on the contrary, a multitude of knowledge gatherers.

Author Keywords
open collaboration, bi-partite networks, contribution performance

INTRODUCTION
In online open collaboration, knowledge such as open source software, Wikipedia articles, 3D-printing designs, is 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 [14]. This new form of labor organization, underlying open collaboration, is called peer-production, and heavily relies on Internet communication systems to be effective [1]. As the Internet has become pervasive in modern societies, open collaboration has permeated to a broad variety of social contexts and industries [2].

Nonetheless, in the absence of formal rules that organize open collaboration, it has remained nearly impossible to account for individual contributions in the production of high quality and often reliable knowledge, as demonstrated for instance for Wikipedia [8]. Indeed, the most basic component of peer-production is *task self-selection*: participants in open collaboration freely choose how and when to contribute, with none or very limited vertical organization [1].

To add to the complexity of self-organized open collaboration projects, many communities exhibit critical cascades of iterative improvements, which in turn lead to super-linear productive bursts of contributions [20]. These highly non-linear, transient and intrinsically unpredictable bursts of iterative improvements are the hallmark of successfully organized communities. To enable productive bursts, a number of conditions must be met, which include transparency, self-censored clans, emergent technology, problem front-loading, distributed screening, and modularity [23]. Unfortunately, the dynamics of contributions are deeply entangled, and the ways individual inputs affect the value of collectively produced knowledge remain largely obscure.

In this paper, we recognize that the value of each knowledge unit (e.g., article, source code file) is deeply tied to the expertise and the number of its contributors. Conversely, the expertise of contributors is a function of knowledge units contributed to. We therefore examine open collaboration projects, and their evolution, as a simple network that relates knowledge units and contributors. We propose a *bi-partite network random walker* model, which measures how editors expertise influences the quality of articles, how contributions to articles influence editors expertise, and so on, recursively. We calibrate the model on 12 Wikipedia categories of articles, and we show how the structure of contribution to knowledge can be disentangled, and how this structure, genuine to categories, influences the quality of knowledge produced and the

CHI'12, May 5–10, 2012, Austin, Texas, USA.

expertise of editors.

The paper is organized as follows. The reader is first introduced to relevant literature. The method, data employed and the results are then presented and discussed. We finally present future research directions and conclude.

RELATED WORK

To the best of our knowledge, our paper is the first attempt to quantify the value of collective contribution environments from the collaboration structure alone. Our model of bi-partite random walkers follows in the lineage of bi-partite networks – networks with two node types – in the global economy of countries competing for exporting products [12, 11]. The proposed recursive *method of reflections* helps understand the competitive advantage (i.e., *fitness*) of countries from the types of products they sell, and moreover, whether other countries export similar (i.e., *ubiquitous*) products. The key insight is that most competitive countries are found to not only export non-ubiquitous products, for which they can charge higher price, but also ubiquitous ones. The model of reflections has been improved and complemented in more recent work, mainly to fix its robustness [21, 6, 22, 5]. Caldarelli et al. [3] have proposed an alternative method, based on biased stochastic Markov chains, which helps further understand the mutual influence between nodes in bi-partite networks.

Separately to networks of economic competition, collaboration networks have been studied early on in network sciences, in particular for networks of co-authorship in scientific publications [16], as well as patterns of self-organization in bi-partite networks of actors-movies[18]. Similarly, the analysis of patterns in the Wikipedia bi-partite networks confirmed the existence of overlapping cliques of densely connected articles and editors [13]. In the same study, clustering of densely connected cliques into larger modules [9] showed that editors clustered by interest with higher coordinated efforts in densely populated clusters [13].

Recent results show that the contribution dynamics of successful open source projects, stem from critical cascades of iterative improvements (commits), which in turn lead to super-linear *productive bursts* of contributions [20]. The conditions of emergence of productive bursts, include transparency, self-censored clans, emergent technology, problem front-loading, distributed screening, and modularity [23].

However, most empirical results in research on open collaboration was unable to uncover the mechanisms of value creation and performance, mainly because of the *bottom-up* emerging properties of peer-production [1]. The only notable result in this field was found during a series of Matlab contests aimed at collectively solving NP-hard problems. It was found that work shared as a public good helps individuals quickly reuse existing results, and thus, find better algorithms [10]. Source code submissions by individuals programmers were tested and benchmarked for their capacity to solve the assigned problem quickly, by executing the compiled code on a computer. Unfortunately, this approach is exclusively feasible for machine executable knowledge (i.e. software code) and in highly controlled environments. For Wikipedia, very rough and sometimes misleading metrics, such edit counts or line, are widely used [4]. However, recent efforts have been undertaken to develop stronger metrics, and to assess separately the expertise of contributors [7], as well as the quality of articles [24]. While useful, these results currently do not allow to precisely attribute the origins of value creation and performance in collective knowledge production.

METHOD

Measuring the structure of value creation by individuals is nearly impossible in most in open collaboration projects, in particular when they do not involve writing software code, which can be compiled, executed and tested on computers. Like in Wikipedia, the most common way to code knowledge is natural language (e.g., English, Chinese, Spanish, French, German), which can hardly be systematically tested for performance. Natural language is indeed the realm of subjective interpretation by humans. Here, we present a method to model value creation and performance in such environments, which neither relies on the content nor on subjective metrics, such as the number of edits, or the number of bytes changed overall or per edit. We recognize that value can be brought by each editor, regardless of the frequency, or the length of her contributions. Only the number and the expertise of editors bring value to an article. And the expertise of editors can be measured from the number and the quality of articles they have modified at least once.

For that, 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 $M_{ea} \equiv M$ of all editors having contributed to a category of articles on Wikipedia. M_{ea} takes value 1 if editor e has edited article a, and 0 otherwise. Note that M only shows which editors have ever touched an article. For the category *Feminist Writers*, as presented on Figure 1, M 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 is the only input of the *bi-partite random walker* model described thereafter, for recursively characterizing the structure and the value of contributions in open collaboration, through the evaluation of editors expertise and articles quality.

Given M, the simplest way to assess the contribution value, the *expertise*, of an editor is obtained by summing the number of articles ever edited out of a set of articles. 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" [19]. These crude expertise and quality metrics for editors and articles, respectively given by,

$$\begin{cases} u_e^{(0)} = \sum_{a=1}^{N_a} M \equiv k_e \\ u_a^{(0)} = \sum_{e=1}^{N_e} M \equiv k_a, \end{cases} \quad (1)$$

are the zero$^{\text{th}}$ order of our algorithm. They are the initial step

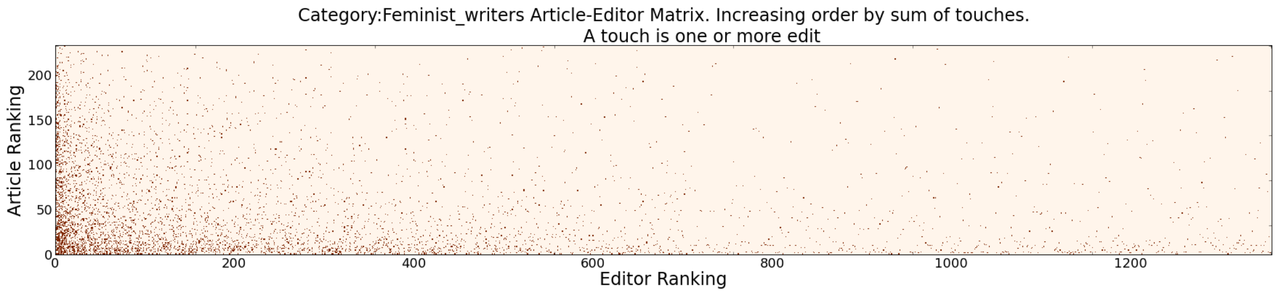

Figure 1. Typical M matrix for a Wikipedia category (here, *Feminist Writers*) ordered on both dimensions by descending order of number of articles modified by an editor (horizontal axis) and of number editors who have modified an article (vertical axis). The structure of M is triangular and shows that 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.

of the *method of reflections* proposed by Hidalgo et al. [12, 11], which derives the value of producing entities (i.e. editors) from products (i.e. articles), and *vice versa*. To help capture the intuition behind the method of reflections for open collaboration, we walk through the first and second iterations, as they are adapted to our study of Wikipedia:

- **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 changed 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 of the method of reflections are written as,

$$\begin{cases} u_e^{(n+1)} = \frac{1}{k_e} \sum_{a=1}^{N_a} M\, u_a^n \\ u_a^{(n+1)} = \frac{1}{k_a} \sum_{e=1}^{N_e} M\, u_e^n. \end{cases} \tag{2}$$

The method of reflections however suffers from a problem, which is rooted in the equal weights given to article and editor scores obtained at the previous iteration. As a result, the method converges to a fixed point with all editors (resp. all articles) having the same value, as a case of consensus dynamics [3]. It has therefore been proposed to consider giving variable weights to the scores from the previous iteration as a Markov process of random walkers on a bi-partite network, jumping with some probability from one node type to another node type [3]. 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 matrix M determines whether a jump between each pair of nodes is possible. If two nodes e and a are not directly connected $M_{ea} = 0$, and 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 [17] to two kind of nodes.

According to Caldarelli et al. [3], we reformulate the method of reflections to account for jumps of the random walker on the bi-partite network of editors and articles. We call $w_e^{(n)}$ the expertise of an editor and $w_a^{(n)}$ the quality of an article at the n^{th} iteration. We define the following Markov process on the bi-partite network of collaboration,

$$\begin{cases} w_e^{(n+1)}(\alpha,\beta) = \sum_{a=1}^{N_a} G_{ea}(\beta)\, w_a^{(n)}(\alpha,\beta) \\ w_a^{(n+1)}(\alpha,\beta) = \sum_{e=1}^{N_e} G_{ae}(\alpha)\, w_e^{(n)}(\alpha,\beta), \end{cases} \tag{3}$$

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

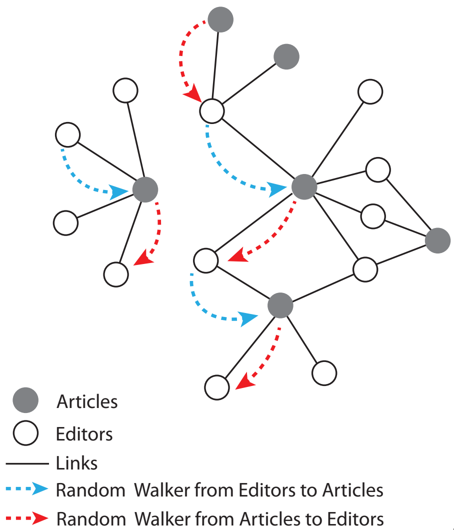

Figure 2. Representation of random walkers jumping from editors to articles (red dotted arrows) and from articles to editors (blue dotted arrows). The intuition is the following: a random walker jumps with some probability from an editor to a given article (i.e. the editor's expertise is positively influenced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

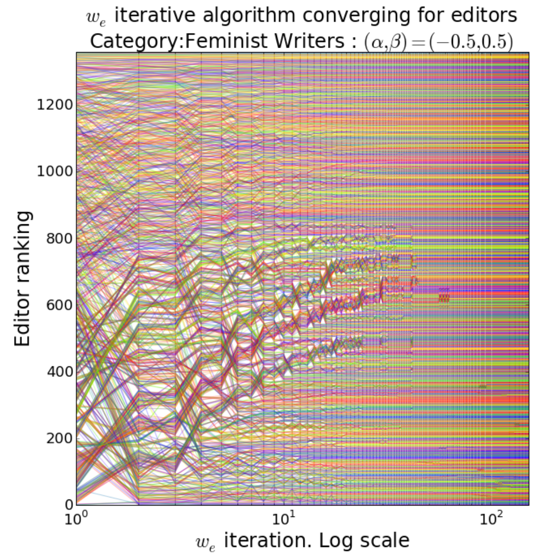

Figure 3. Convergence of the ranked expertise w_e of editors having contributed to articles in the Feminist Writers category on Wikipedia for arbitrary control parameters: $(\alpha,\beta) = (-0.5, 0.5)$. Starting from the sum of contributed articles as the initial step, we can see how the algorithm progressively 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 Wikipedia categories, the algorithm converges with clearly stable ranks, after nearly 50 iterations.

to article a also in a single step. These transition probabilities are given by,

$$\begin{cases} G_{ea}(\beta) = \frac{M_{ea}k_e^{-\beta}}{\sum_{e'=1}^{N_e} M_{e'a}k_{e'}^{-\beta}} \\ G_{ae}(\alpha) = \frac{M_{ea}k_a^{-\alpha}}{\sum_{a'=1}^{N_a} M_{ea'}k_{a'}^{-\alpha}}. \end{cases} \tag{4}$$

The transition matrices $G_{ea}(\beta)$ and $G_{ae}(\alpha)$ depend only on the initial conditions: the binary matrix M, as well as k_e and k_a given by (1). The variables α and β control the transition probabilities in a similar way. 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 zero$^{\text{th}}$ 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 the positive influence on articles. In a nutshell, β relates the amount of articles edited on the overall editor's expertise, which in turn has an influence on each edited article (along with the influence of other editors). If $\beta >> 0$, 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).

After each iteration, we have expertise and quality scores, which allow ranking editors and articles. When the rankings for both editors and articles do not change in two successive iterations we consider that the *bi-partite network random walker* model has converged. We have verified that the method converges on all our 12 Wikipedia categories. Figure 3 shows the evolution of expertise w_e ranked among 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 low rank (i.e. with few articles edited), get a higher rank as the number of iterations increases. They most probably have edited and contributed to few, but high quality articles. Similarly, some initially high ranked editors, gradually drop in the ranking. They have edited many, but low quality articles. In the case of category *Feminist Writers*, the algorithm converges with clearly stable ranks, after nearly 64 iterations.

The parameters α and β control how editors and articles influence each others at the level of the whole bi-partite collaboration network. Upon calibration with ground-truth editor expertise and article quality metrics, these control parameters thus directly inform on the structure of peer-production, and how contributions benefit to the whole open collaboration project.

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.

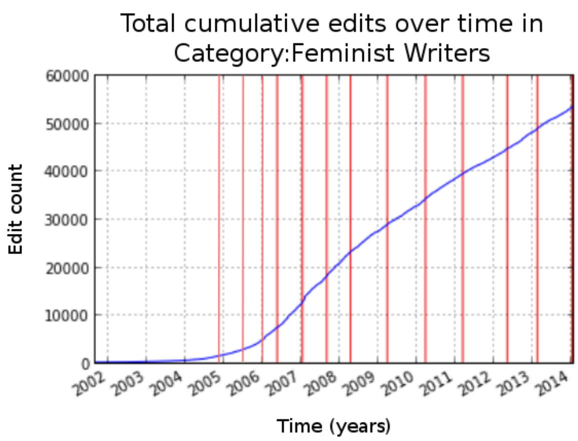

Figure 4. Cumulative edits made in Category *Feminist writers* (blue line). Vertical red lines represent the 13 snapshots taken at 2.5%, 5%, 7.5% and then, 10%, 20%, 30%, . . . , 100% of edits.

DATA

We seek to find values of α and β that minimize the distance between rankings given by the bi-partite network random walker model, which takes the matrix M as unique input, and ground-truth metrics on editor expertise and article quality, obtained independently from Wikipedia. We perform the model calibration for 13 snapshots (see Figure 4) for each of the 12 categories of Wikipedia articles presented in Table 1. For each category and snapshot, we build the binary matrix M by parsing all edit histories of all articles up to the snapshot time. We set $M_{ea} = 1$ for editor e having modified article a, and $M_{ea} = 0$ otherwise. In order to eliminate page vandals, we considered only editors, who made 5 or more edits to any article in the category. We also discarded all software robots *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 $(w_a, \bar{w}_a)$. 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,

$$\begin{cases} (\alpha^*, \beta^*) = argmax(\rho_e) \\ (\alpha^*, \beta^*) = argmax(\rho_a). \end{cases} \tag{5}$$

(α^*, β^*) characterize how value flows from editors to articles, and from articles to editors, in the bi-partite network of collaboration in Wikipedia.

According the ground-truth metrics, editor expertise is represented by labor hours, which have been found as the most representative metric [7], and was 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 in an *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 edited other categories of articles in Wikipedia. Our measure of actual article quality is performed through 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) outgoing intra-Wiki links. These metrics are used to determine quality on Wikipedia [24], and ave been widely tested [15]. We performed a 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.

RESULTS

To understand how contribution value *flows* from editors to articles, and from articles to editors, we calibrated the control parameters (α^*, β^*) of the bi-partite network random walker model on 12 Wikipedia categories (c.f. Table 1) with 13 snapshots each (Figure 4). Figure 5 shows typical optimization landscapes, which maximize the rank correlation ρ_e (upper panel) between editors expertise w_e obtained from the model and expertise obtained from state-of-the-art 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 respectively editors [7] and articles [24] 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 editors expertise exhibits a slightly concave, or at least linear, increase.

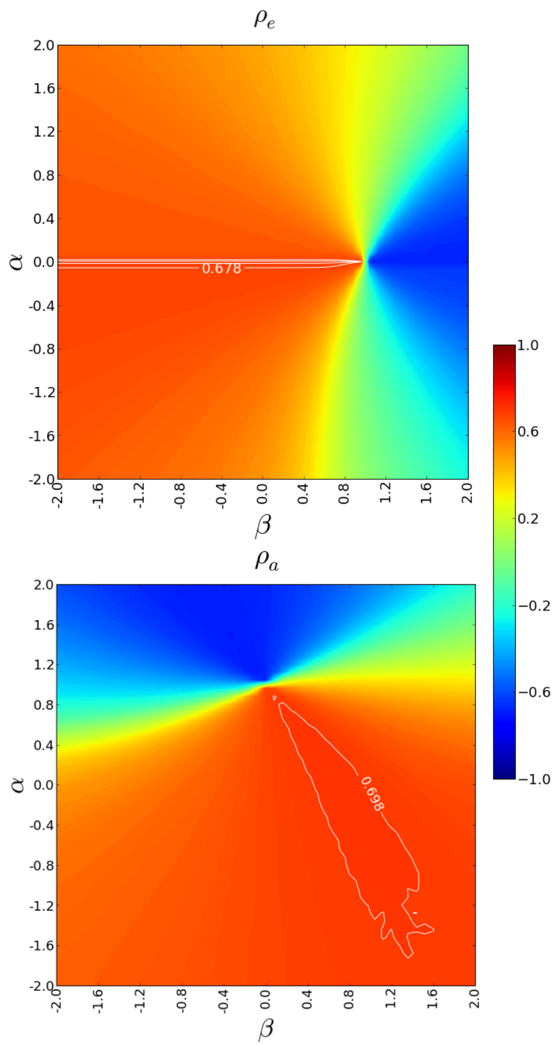

Figure 5. Typical landscape of maximum correlation as a function of α and β for articles (upper panel) and editors (lower panel). The contour line shows the 95$^{\text{th}}$ percentile of the rank correlation over the landscape. The category displayed here is *Feminist Writers*, for the last snapshot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

For the latest snapshot (i.e., the state 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 the optimal β^* (and $\alpha^* = 0$ for the sake of clarity), 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. $\alpha^* = 0$ means that editors expertise always benefits from contributions as a linear function of the number of articles edited [compounded over iterations of the recursive algorithm defined by formula (3)]. 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

	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".

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.

β is found generally stable over snapshots for categories, past the first 10% of overall contributions (i.e. the 4$^{\text{th}}$ snapshot). In fact, β switches signs only 3 times of the possible 144 measured changes. The times when $\beta < 0$ are in early history. Recall that the interpretation of $\beta < 0$ means that articles improve super-linearly with regards to number of editors - that is, very productive bursts $\rightarrow$ only if $\beta < -1 \rightarrow$ can you confirm?. The stability of (α^*, β^*) confirm that the control parameters of the bi-partite network random walker model describe a robust feature of the structure of value creation in the bi-partite network of editors contributing to articles. This additional result is also a first step towards robust predictions of editor expertise and article quality rankings, given successive inputs to new articles made by editors.

DISCUSSION

Building on the method of reflections previously used for global economic networks of production, we have applied and tested the *bi-partite network random walker* model in the context of Wikipedia open collaboration. 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 $\langle \rho_e \rangle \approx 0.72$. Moreover, the evolution of ρ_e and ρ_a of categories under editing, exhibit strong stability. In particular, the adequacy of article ranking is very high early on, and thereafter stationary, suggesting that the model can quickly capture the quality of articles. For editors expertise, the adequacy increases steadily as categories get further developed.

This difference might be due to the roughness of the actual metrics for editors $\bar{w}_e$, expressed in labor-hours, compared to the quality of $\bar{w}_a$, which is an aggregate measure of five

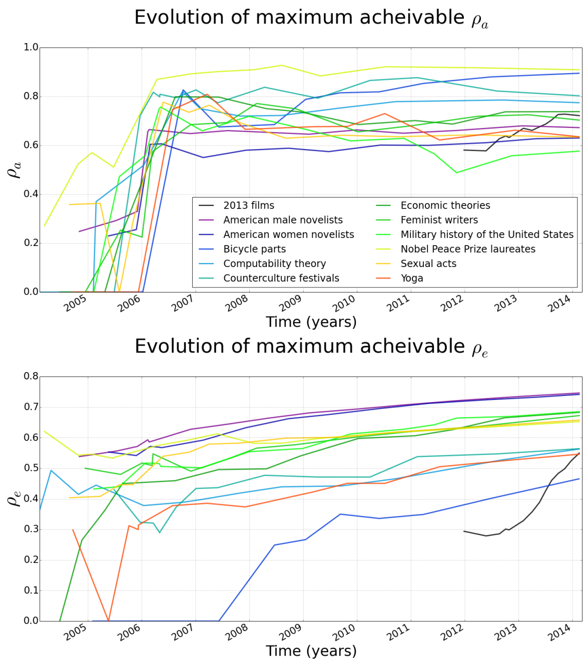

Figure 6. Evolution of Spearman ρ rank correlations between the ranking obtained from the calibrated model and the actual values for each category and for editors (upper panel) and articles (lower panel). The correlations are generally quite high : $0.46 < \rho_e < 0.75$ with $\langle \rho_e \rangle = 0.64$ 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.

precise quality metrics. Nevertheless, the correlation of editor ranking with $\bar{w}_e$ increases: ρ_e exhibits a convex increase over time, suggesting that it takes time (i.e., lots of articles edited) to capture well the expertise of editors. This is striking because the method is reflexive and the same information is incorporated on both dimensions from the input matrix $\mathbf{M}$. While it will require further investigation to explain, we interpret this result in the following way: from Figure 1 and from Table 1, we see that there are always significantly more editors than articles for each category. This means that 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. In other words, it takes more time to rightly rank editors because there are more compared to the number of articles in a given category.

We have also found $\alpha \approx 0$ for all categories, reflecting the linear influence of the number of articles edited on editors expertise. The interpretation is that an editor is generally as good as the number of articles they have edited. If we substitute "good" to our actual metric, labor-hours, we should find that labor-hours are directly related to the number of articles edited. So $\alpha \approx 0$ means that articles-touched is a reasonable proxy for labor-hours. This suggests that there are no diminishing return for value creation, in investing hours, or similarly, modifying new articles.

Finally,we find that whether editor prolificness, determines article quality or not, is category dependent. With examples from both ends of the spectrum we find footing to consider β as a proxy for "collaborativeness". Consider the highest β found, on category *Sexual acts*. We chose to include this category because it could be considered taboo or perverse to edit these articles. This category has the least gain from better editors touching more articles. An explanation could because presumably there are a lot of unmediated editors fighting, editors just overwrite each other rather than organize. Category *Military History of the US* is famous within Wikipedia for its self-organizing task-forces, and exhibits $\beta = 0$. In fact, it is also only one of two Categories that ever have consecutive snapshots with $\beta < 0$. As a result of better organization, there is probably less edit-warring in this category, and each visit to the page by good editors has a definite, productive task at hand.

Overall, it is amazing that so little input information, a binary matrix M_{ea} with only information on whether an editor as modified a given article, is sufficient to achieve a high level of fidelity with actual metrics. It suggests some kind of *less is more* mechanism, which has direct implications on the overall cost of evaluating contributions in the complex entanglement of contributions. Namely, the use of the bi-partite network random walker model requires only simple and straightforward data mining, compared with, for instance, a model that would primarily rely on the type of words written by editors.

LIMITATIONS AND FUTURE WORK

The results we have presented in this paper show that contributions – and the value they provide to the collective good – can be well disentangled with the bi-partite network random walker model. However, these results suffer from a number of limitations, which call for further work and testing.

According to the original philosophy of the method of reflections [12], an additional node type reflecting *capabilities* should connect producing and produced entities. In the current approach, capabilities are implicit in the model, mainly because they are not observable. In the context of Wikipedia and open collaboration, incorporating capabilities would be more feasible. We can for instance identify what an editor do best to improve an article, among the five metrics (ratio of mark-up to readable text, number of headings, article length, citations per article length, and outgoing intrawiki links) we have used to assess the quality of an article.

We believe there are two further directions to improve our results. First, we have followed the philosophy of the method of reflections that aims at ranking countries in the world economy. However, the bi-partite network random walker method provides absolute values, which might have a meaning in the context of open collaboration. In future work, we would like to understand further these absolute values. Second, we took the very simplest information for the input matrix $\mathbf{M}$ (i.e. whether an editor has modified a given article, or not). We wonder how the performance of the method might change if richer information is incorporated in the matrix (e.g., number of edits, number of bytes changes). Two exclusive hypotheses could be tested: either the model fits better with richer information, or on the contrary, the model is not as good. In the latter case, we would face a *less is more* scenario, which would require elucidating why less rich information accounts better for reality. Or conversely, it could help further understand what ground-truth metrics actually contain richer information, and hence, help reverse engineer most relevant direct measures of editor expertise and article quality.

As it stands, our method for finding the optimizing pair of control parameters (α^*, β^*) for the *bi-partite network random walker* model is a grid search, because gradient-based optimizers do not handle the singularities that the model produces. Given that computer time required by grid-search grows at the square of the problem (e.g. the size of a category), more work is needed to improve the efficiency of the optimizing algorithm for calibration of α and β. We shall also fix the method to obtain more smooth optimization spaces.

CONCLUSION

We have presented an recursive algorithm based on a *bi-partite network random walker* model, which jointly ranks Wikipedia editors by their expertise, and articles by their quality, from a simple input matrix recording whose editor as modified a given article. Moreover, upon calibration on 12 categories of Wikipedia articles, the input and the control parameters of the model inform directly on how value is created from the complex network of contributions. It appears that some categories of Wikipedia articles fully benefit of the multiplicity of contributors (i.e. "collaborativeness"), while for other categories, more contributors per article generate dis-value. The origins of these differences between categories could stem from limited coordination capacity between contributors. The organization of value creation in open collaboration might also be intrinsically different from one topic to another. Finally, we want to stress the generality of the method we have presented. Similarly to open collaboration in Wikipedia, the proposed algorithm can apply to a variety of situations, such as social coding (e.g., Github) , or collaborative rating (e.g., Amazon or Yelp reviews), in order to help further understand to origins of collective value creation and quality, which are the hallmark of open collaboration.

ACKNOWLEDGMENTS

One of the authors (T.M.) acknowledges support from the Swiss National Science Foundation (Grant Nr. PA00P2-145368).

REFERENCES

1. Benkler, Y. Coase's Penguin, or, Linux and "The Nature of the Firm". *The Yale Law Journal 112*, 3 (Dec. 2002), 369+.
2. Benkler, Y. *The Penguin and the Leviathan: How Cooperation Triumphs over Self-Interest*, 1 ed. Crown Business, Aug. 2011.
3. 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).
4. contributors, W.
5. 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).
6. 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).
7. Geiger, R. S., and Halfaker, A. Using Edit Sessions to Measure Participation in Wikipedia. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, ACM (New York, NY, USA, 2013), 861–870.
8. Giles, J. Internet encyclopaedias go head to head. *Nature 438*, 7070 (Dec. 2005), 900–901.
9. Guimerà, R., Sales-Pardo, M., and Amaral, L. A. Module identification in bipartite and directed networks. *Physical Review E 76*, 3 (2007), 036102.
10. Gulley, N., and Lakhani, K. R. The Determinants of Individual Performance and Collective Value in Private-Collective Software Innovation. *Social Science Research Network Working Paper Series* (Feb. 2010).
11. Hidalgo, C. A., and Hausmann, R. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences 106*, 26 (June 2009), 10570–10575.
12. 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.
13. Jesus, R., Schwartz, M., and Lehmann, S. Bipartite Networks of Wikipedia's Articles and Authors: A Meso-level Approach. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, ACM (New York, NY, USA, 2009), 1–10.
14. 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.
15. Klein, M. Kumusha Takes Wiki: Actionable Metrics for Uganda and Côte d'Ivoire, 2014.
16. Newman, M. E. J. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences 98*, 2 (Jan. 2001), 404–409.
17. Page, L., Brin, S., Motwani, R., and Winograd, T. The PageRank citation ranking: Bringing order to the web.
18. Ramasco, J. J., Dorogovtsev, S. N., and Pastor-Satorras, R. Self-organization of collaboration networks. *Physical review E 70*, 3 (2004), 036106.
19. Raymond, E. The cathedral and the bazaar. *Knowledge, Technology & Policy 12*, 3 (Sept. 1999), 23–49.

20. 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, May 2014.

21. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. A new metrics for countries' fitness and products' complexity. *Scientific reports 2* (2012).

22. 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).

23. von Krogh, G., Maillart, T., Haefliger, S., and Sornette, D. Designing organizations for productive bursts. *under review* (2014).

24. Wang, M. W., Cosley, D., and Riedl, J. Tell Me More: An Actionable Quality Model for Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration*, WikiSym '13, ACM (New York, NY, USA, 2013).