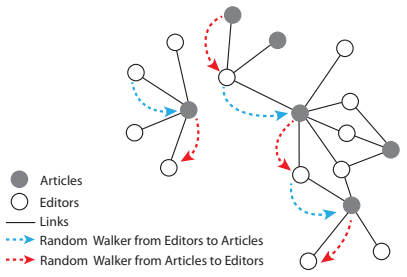
Characterizing Contribution Strategies in Wikipedia with Bi-Partite Network Rankings

Maximilian Klein, Thomas Maillart, John Chuang

One of the very fundamental principles of peer-production in open collaboration projects is the ability for individuals to decide what tasks they want to take on [**?**]. For instance, in Wikipedia editors touch varying numbers of articles, at different levels of development. The outstanding question is how individual contribution strategies are likely to influence the *investment* of each editor, the *quality* of articles [**?**], and ultimately the quality of a project [**?**]. One way to tackle the problem consists in considering a project – here a category of Wikipedia articles – as a bi-partite network of articles and their related editors [**?**]. Assessing the value of the two types of components of such a network has previously been undertaken in the context of macro-economics (ranking countries by the type of products they export) by Hidalgo and Haussmann [**?**], and by Caldarelli et al. [**?**]. The investment of editors is assessed from the quantity and quality of articles they have edited. Conversely, the development of an article depends on the number and the investment of editors who have modified the article. The method is a two node-type version of the Google pageRank algorithm [**?**] and can be compared to a random walker jumping between editors and articles. Walking is controlled by biased parameters α for the probability to jump to more developed articles, and β for the probability to jump more invested editors (see Figure 1A).

(A)



(B)

	Category	Articles	Editors	ρ_a	ρ_e	α	β
1	Bicycle parts	70	210	0.90	0.46	0.00	0.00
2	Military history of the United States	180	854	0.58	0.70	0.00	0.00
3	Computability theory	92	272	0.77	0.56	0.00	0.32
4	American male novelists	2460	9946	0.67	0.75	0.00	0.40
5	2013 films	1896	5215	0.72	0.55	0.00	0.48
6	Economic theories	212	1145	0.74	0.70	0.00	0.48
7	American women novelists	1936	5968	0.63	0.75	0.00	0.64
8	Feminist writers	233	1357	0.70	0.69	0.00	0.72
9	Yoga	123	730	0.64	0.57	0.00	1.12
10	Nobel Peace Prize laureates	104	4165	0.91	0.66	0.00	1.20
11	Counterculture festivals	66	578	0.80	0.61	0.00	1.36
12	Sexual acts	93	2190	0.63	0.66	0.00	1.52

Figure 1: **(A)** Bi-partite network with Article and Editor nodes. Dashed arrows show how the random walker jumping between articles and editors with probability controlled by the appropriately biased connectivity of the each node. **(B)** Table shows the best rank-correlation ρ_a and ρ_e of the algorithm with the ground truth for each Wikipedia category, as well the value of the divestment bias β.

The variables α and β provide a direct measure of the nonlinear "importance" of the number of highly developed articles, and the nonlinear importance of highly invested editors, respectively. Both α and β are optimized to maximize the rank correlations of editors ($0.46 < \rho_e < 0.75$) and articles ($0.58 < \rho_a < 0.91$) between the algorithm and ground-truth metrics obtained by state-of-the-art quality metrics of editors [**?**] and articles [**?**], for 12 categories on Wikipedia. We find that the best value for α is 0, while $0 \leqslant \beta \leqslant 1.52$ (c.f. Figure 1B). Looking at 1B, we find telling extremes. The best editors in *Category:Military history of the US* - a category known for being very competitive - are characterized by emphasizing investment (number of articles edited in the category). On the other end, the editors in *Category:Sexual acts* - a taboo subject where much editing could be considered perverse - are characterized by divesting in touching many articles in the category.

Category	Editor divestment β	ρ_e	ρ_a
Bicycle parts	0.00	0.46	0.90
Military history of the US	0.00	0.70	0.58
Computability theory	0.32	0.56	0.77
American male novelists	0.40	0.75	0.67
2013 films	0.48	0.55	0.72
Economic theories	0.48	0.70	0.74
American women novelists	0.64	0.75	0.63
Feminist writers	0.72	0.69	0.70
Yoga	1.12	0.57	0.64
Nobel Peace Prize laureates	1.20	0.66	0.91
Counterculture festivals	1.36	0.61	0.80
Sexual acts	1.52	0.66	0.63