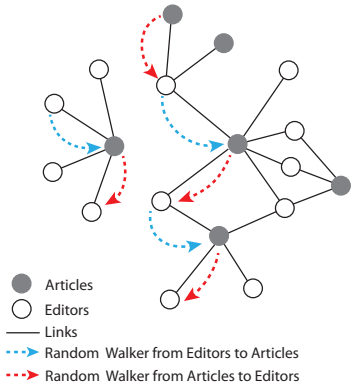
Bringing Order to Wikipedia with Bi-Partite Network Rankings

Maximilian Klein, Thomas Maillart, John Chuang

One of the very fundamental rules of peer-production in open collaboration projects is the ability for individuals to decide what tasks they want to take on [1]. For instance, on Wikipedia some editors concentrate their efforts on a few selected articles, while others make little edits on a broader set of articles, and articles are likely to benefit from edits by more – possibly expert – contributors. The outstanding question is how individual contribution strategies are likely to influence the *expertise* of each editor, the *quality* of articles [6], and ultimately the quality of the entire project [3]. 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 [4], and by Caldarelli et al. [2]. The expertise of editors is assessed from the quantity and quality of articles they have edited. Conversely, the quality of an article depends on the number and the expertise of editors who have modified the article. Each iteration, information of article quality (resp. editor expertise) is recursively incorporated until the algorithm converges. The method is a two node types version of the Google pageRank algorithm [5] and can be compared to a random walker jumping from one node to another type of node with some probability controlled by biased metrics of efferent node : α for the probability to jump from editor to article, and β for the probability to jump from article to editor (see Figure 1A).

(A)



(B)

Category	ρ	β
Bicycle parts	0.8953	0
Military history of the US	0.5767	0
Computability theory	0.7748	0.32
American male novelists	0.6729	0.4
2013 films	0.7220	0.48
Economic theories	0.7382	0.48
American women novelists	0.6304	0.64
Feminist writers	0.7036	0.72
Yoga	0.6355	1.12
Nobel Peace Prize laureates	0.9096	1.2
Counterculture festivals	0.8031	1.36
Sexual acts	0.6341	1.52

Figure 1: **(A)** Bi-partite network with Article and Editor nodes. Dashed arrows show how the random walker jumps from one node to another type of node with some 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 bias β.

The variables α and β provide a direct measure of the "value" of the number of editors per article and the "value" of the number of articles per editor, respectively. Both α and β are optimized to maximize the rank correlations of editors ($0.46 < \rho_e < 0.75$) and articles ($0.57 < \rho_a < 0.91$) between the algorithm and ground-truth metrics obtained by state-of-the-art quality metrics of editors [3] and articles [6], for 12 categories of articles on Wikipedia. We find that the best value for α is 0, while $0 \leqslant \beta \leqslant 1.52$ (c.f. Figure 1B). We find that the expertise of editors is concave increasing as a function of the number of articles they have edited for categories with $\beta > 1$ and decreasing when $0 < \beta < 1$. For two categories (*bicycle parts* and *Military history of the US*), $\beta = 0$, so the expertise of editors is a linear function of the number of articles they have edited.

Regarding articles, the quality is a linear function of the number of editors, and a function of the expertise of these editors, which in turn varies depending on the values of β. For $\beta = 0$, the expertise of editors has no influence, while for $\beta > 0$, the expertise provided by editors decreases as a function of the number of articles they have edited.

References

[1] Y. Benkler. Intellectual property and the organization of information production. International Review of Law and Economics, 22(1):81–107, July 2002.

[2] G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, and A. Tacchella. A network analysis of countries' export flows: firm grounds for the building blocks of the economy. PloS one, 7(10), 2012.

[3] R. S. Geiger and A. Halfaker. Using edit sessions to measure participation in wikipedia. Computer supported cooperative work (CSCW '13), 2013.

[4] C. A. Hidalgo and R. Hausmann. The building blocks of economic complexity. Proceedings of the National Academy of Sciences, 106(26):10570–10575, June 2009.

[5] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. 1999.

[6] C. D. Warncke-Wang, M. and J. Riedl. Tell Me More: An Actionable Quality Model for Wikipedia. Preceedings of WikiSym '13, 2013.