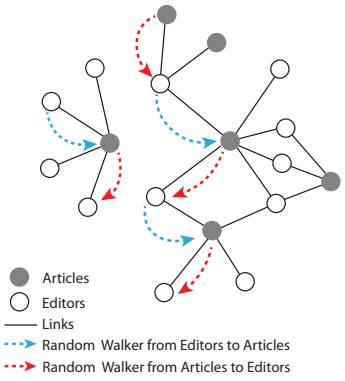
Bringing Order to Wikipedia with Bi-Partite Network Rankings

Maximilian Klein, Thomas Maillart, John Chuang

From open source software projects to online encyclopedias, open collaboration has become one of the greatest success of the Internet, with millions of individuals sharing effort and knowledge as a collective good for their own interest and for the advancement of society [5]. Unfortunately, assessing quickly the value of knowledge produced on open collaboration platforms remains nearly impossible : software code must be compiled and executed beforehand, and the value of written text remains subjective. Moreover, large amounts of untangled contributions by heterogenous editors prevents proper capture of editors' expertise. Here, we tackle the problem of ranking the expertise of editors and the quality of articles on subsets of Wikipedia with a minimum information input, namely whether an editor has ever modified a given article or not. The approach, called *wikiRanks*, is an extension of the *pageRank* algorithm [4] to bi-partite networks of relations between two kinds of nodes [3]: 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, quality (resp. expertise) information is recursively incorporated until the algorithm converges. As shown on Figure 1A, the wikiRanks method can be assimilated to a random walker jumping from one node to another type of node with some probability controlled by a biased metric of efferent node connectivity, with the bias β putting more or less emphasis on quality ($\beta > 1$) or on quantity ($\beta < 1$) of efferent links [1].

(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.

Despite the very limited amount of input information, wikiRanks can achieve very high levels of rank correlation with exogenous ground-truth state-of-the-art article quality [6] and editor expertise [2] as shown on Figure 1B. Interestingly, when we look at the evolution of wikiRanks accuracy as a category of articles gets more contributions, we find that it takes more time for the algorithm to achieve high levels of correlation with the ground-truth (not shown). Actually there are much more editors than articles in a Wikipedia category. Therefore, it take more contributions in the category for the algorithm to get sufficient information to assess the expertise of an editor. We also find that the calibration of β informs us on the importance of editor's

expertise versus the quantity of editors who have contributed to the quality of articles in selected Wikipedia categories (c.f. Figure 1B).

References

[1] 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.

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

[3] 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.

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

[5] E. von Hippel and G. von Krogh. Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. Organization Science, 14(2), 2003.

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