

# Inequalities in crowdsourced knowledge

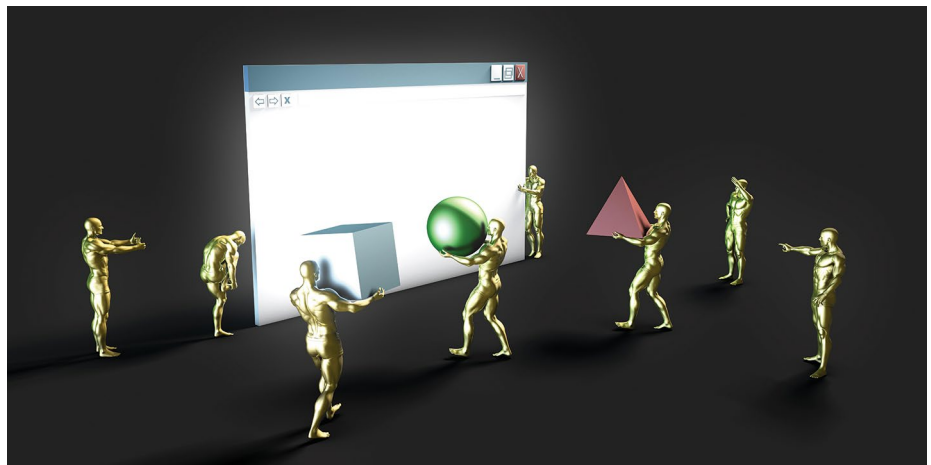
Participatory knowledge creation on platforms such as Wikipedia has revealed the enormous democratizing potential of the Internet. It has also exposed its limitations.

On October 2, 2018, Donna Strickland became the first woman to receive the Nobel Prize in physics in 55 years. At the time of the announcement, Strickland didn't have a Wikipedia page. In fact, when a Wikipedia user submitted a draft page on Strickland earlier in the year, the page was rejected for lacking sufficient evidence (in the form of mentions in secondary sources) that Strickland and her work were important enough to merit an entry in the online encyclopaedia.

A page for Strickland was quickly created an hour and a half after the Nobel Prize announcement, but the incident made headlines and generated strong criticism of gender bias on the platform.

Wikipedia and other Wikimedia projects are paradigmatic examples of 21st-century online knowledge formation communities. In the pre-internet days, knowledge repositories were under the exclusive purview of experts: an elite group with established credentials of expertise. In the Web 2.0 era, anybody with internet access and basic technology skills can contribute to the creation of new bodies of knowledge. The guiding principle is democratic and, despite widespread speculation to the contrary, the accuracy of crowdsourced content can be high. Structural inequalities in participation, however, threaten the democratic potential of these platforms and can exaggerate the biases they were created to counter.

In this issue of *Nature Human Behaviour*, Yun et al. (<https://www.nature.com/articles/s41562-018-0488-z>) examine the formation and evolution of all existing Wikimedia projects in all languages, including all versions of Wikipedia, Wiktionary, Wikibooks and Wikinews. They find that all 863 projects follow the same universal pattern of growth, regardless of their size, age, or language. Projects in different languages do differ in their size or how fast they grow, and this can be predicted from the overall size of the national economy: the richer the country, the faster and larger its Wikimedia projects grow.



Credit: Panther Media GmbH / Alamy Stock Photo

The authors' main focus, however, is the examination of structural inequalities in contributions. Is the contributor base heterogeneous or not? When do inequalities arise? And how do they arise? The authors find that a small number of super-editors have a disproportionately large role in creating and editing entries. Using a variant of the Gini coefficient, they show that several Wikimedia projects are characterized by almost complete inequality among contributors (a coefficient of 1). A hierarchy of editors develops from the very early stages of a project's creation and inequality in contributions becomes even greater as a project grows in size.

Is this pattern unique to online collaborative knowledge endeavours or does it characterize traditional collaboration systems, too? Yun and colleagues turn to research articles and patents to find an answer. It turns out that high inequality among contributors is a characteristic of scientific papers and patents, as well, even though it is less pronounced and develops at slower pace.

To identify how large disparities among contributors to communal data sets arise, the authors develop an agent-based model

that replicates the empirical results, taking into account contributors' attachment to entries they created, their memory for most recent content and declining motivation over time.

Yun et al. provide compelling evidence that the Wikimedia projects are the product of a few, not the many, from the start. Those few are overwhelmingly male, in their mid-thirties to mid-forties, from the Global North, the Wikimedia Foundation's annual community engagement surveys find year after year ([https://meta.wikimedia.org/wiki/Community\\_Engagement\\_Insights/2018\\_Report](https://meta.wikimedia.org/wiki/Community_Engagement_Insights/2018_Report)). That's a very restricted demographic that falls far short of a participatory ideal.

Strategic initiatives and active efforts—both top-down and bottom-up—are required to broaden the contributor base of communal online projects. Without concerted action, high inequality and low diversity among contributors will continue to undermine the validity, sustainability and democratic potential of online collective knowledge platforms. □

Published online: 12 February 2019  
<https://doi.org/10.1038/s41562-019-0550-5>