

SUPPLEMENTARY INFORMATION:

Dataset

Citations to a certain article can be studied using Scopus database (available from: <http://www.scopus.com>). In case of each article, the year-by-year citation statistics were determined.

In the present work, “Systems Science” category was chosen as an emerging interdisciplinary field with several subcategories included in Wikipedia (available from: http://en.wikipedia.org/wiki/Category:Systems_science). Three subcategories were chosen for this study: “Social Systems” as a proxy for the humanities; “Chaos Theory” as a proxy for physical sciences; and “Systems Biology” as a proxy for biological sciences.

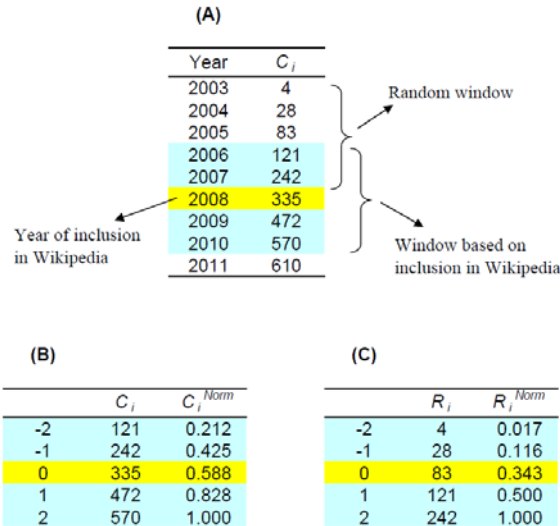
Methodology

In the present work, the main focus is to study the effect of inclusion of an article in Wikipedia on the citation counts. Therefore, the year of inclusion in Wikipedia is defined as the “origin of time”, or “year zero”. Citations to each article were studied in a five-year period, i.e., in two years before and two years after the inclusion in Wikipedia. We considered only those articles which were included in Wikipedia in 2009 or before, and which were included in Wikipedia at least two years after their publication.

Since the citation counts in our dataset range from very few citations to hundreds of citations in each year, a comparable measure for evaluating the citation counts is required. In case of each article, we normalize the citation count in year i (C_i) to the maximum of the citation counts in the corresponding five-year period (C^{max}):

$$C_i^{Norm} = C_i / C^{max}$$

where C_i^{Norm} is the normalized citation count in year i (see suppl. Figure 1B). Obviously, C_i^{Norm} is always a value between zero and one.



Suppl. Figure 1: Example of the process of computing C_i^{Norm} and R_i^{Norm} for an article which is published in 2003 and cited by Wikipedia in 2008.

In each year, more and more scientific articles are indexed in Scopus. This increase may gradually add to the number of citations to an article, independent of whether the article is included in Wikipedia entries. To correct for this increase, we recommend using the propensity of citations (P_i) in a certain year. The P_i values can be computed as follows.

We firstly compute R_i^{Norm} , which is the normalized “random citation count” in year i . Suppose that in the list of citation counts, we randomly select a five-year time window (suppl. Figure 1C). We define the third year of this window as the “origin of time”. Then, R_i^{Norm} can be computed by dividing the random citation count in year i (R_i) to the maximum of the random citation counts in the corresponding five-year period (R^{max}):

$$R_i^{Norm} = R_i / R^{max}$$

Propensity of citations in year i (P_i) is defined as:

$$P_i = C_i^{Norm} / R_i^{Norm}$$

Please note that P_i is equal to 1 when the articles are cited as frequently as expected by chance. $P_i > 1$ (resp. $P_i < 1$) means that articles are cited more than (resp. less than) what is expected by chance.