

## Guest editorial:

### IMPACT OF WIKIPEDIA ON CITATION TRENDS\*

Sayed-Amir Marashi\*\*, Seyed Mohammad Amin Hosseini-Nami, Khadijeh Alishah, Mahdieh Hadi, Ali Karimi, Saeedeh Hosseinian, Rouhallah Ramezani-fard, Reihaneh Sadat Mirhassani, Zhaleh Hosseini, Zahra Shojaie

Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran

\*All authors contributed equally to the manuscript.

\*\* corresponding author: e-mail address: [marashi@ut.ac.ir](mailto:marashi@ut.ac.ir)

#### ABSTRACT

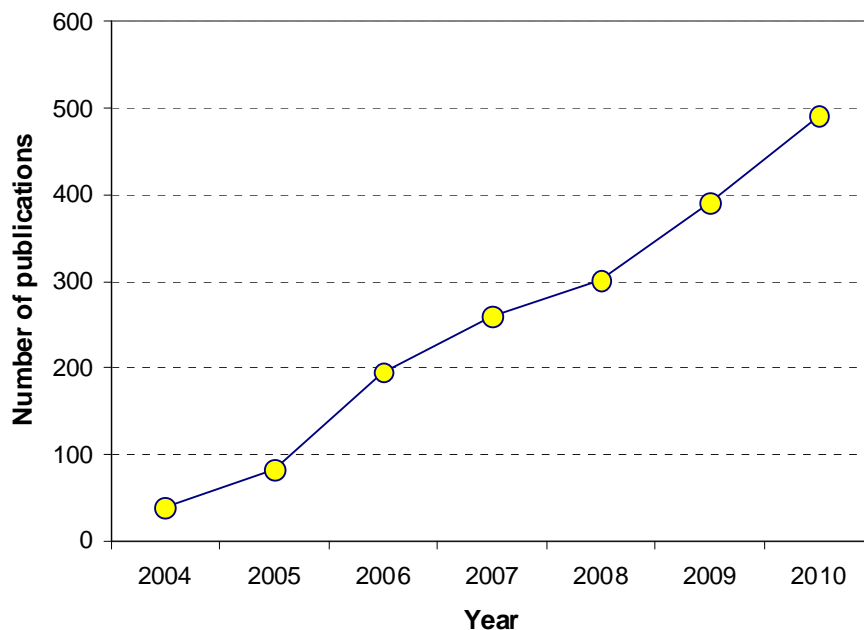
It has been suggested that the “visibility” of an article influences its citation count. More specifically, it is believed that the social media can influence article citations. Here we tested the hypothesis that inclusion of scholarly references in Wikipedia affects the citation trends. To perform this analysis, we introduced a citation “propensity” measure, which is inspired by the concept of amino acid propensity for protein secondary structures. We show that although citation counts generally increase during time, the citation “propensity” does not increase after inclusion of a reference in Wikipedia.

Citation analysis is of central importance in research evaluation (Adam, 2002; Bird, 2008; Bornmann et al., 2008; Fava et al., 2004; Garfield, 1972; Gisvold, 1999), in spite of its shortcomings (Sancho, 1992; Seglen, 1997). Moreover, almost all bibliometric measures, including impact factor (Garfield, 2006; Whitehouse, 2002) and *h*-index (Glänzel, 2006; Hirsch, 2007), are functions of article citations. As a result, it is also important to study factors and variables which can influence citation counts.

It is generally believed that the “visibility” of an article can influence its citation count. Lay media are previously suggested to have a role in increasing the citation counts of a scientific article (Callaham et al., 2002; Phillips et al., 1991). Some authors have reported that, on average, open access articles are cited more than non-open access articles (Lawrence, 2001; Norris et al., 2008), although this claim is disputed by others (Calver and Bradley, 2010; Davis, 2011; Davis et al., 2008; Lansingh and

Carter, 2009). On the other hand, it is believed that many authors do not necessarily read the articles to cite them (Braun et al., 2010; Marashi, 2005), but rather adopt citations from the reference list of some subsequent papers (Braun et al., 2010). All these findings suggest that citations to an article might strongly depend on the visibility, rather than the merit of the article.

With the popularity of Web 2.0 in recent years, it has been suggested that the social media can also influence scientific article citations (Bar-Ilan et al., 2012; Eysenbach, 2011; Li and Thelwall, 2012). One of these social websites is the free online encyclopedia Wikipedia (<http://en.wikipedia.org>). Although scientific entries in Wikipedia are typically written by anonymous authors, many scientists may still try to find information in this website, because of its high visibility and the simplicity of the language used (Figure 1).



**Figure 1:** Number of publications citing Wikipedia from 2004 to 2010. The data are obtained from Scopus (<http://www.scopus.com/>) by searching for the word “Wikipedia” in “source title”.

Figure 1 suggests that Wikipedia entries have high visibility to the authors of scholarly articles, especially in recent years. In the present work, we investigate whether the opposite is also true, i.e., inclusion of scholarly references in Wikipedia significantly affects the citation trends.

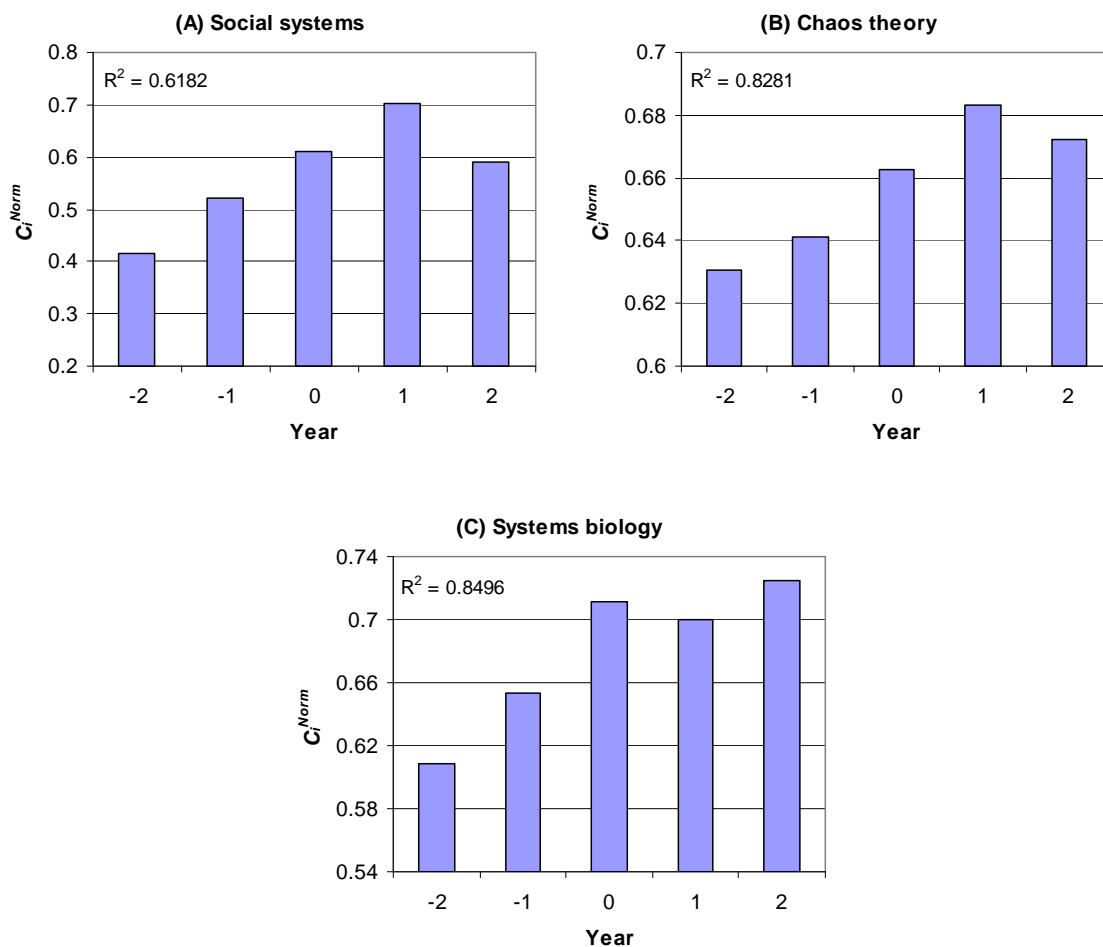
The Materials and Methods of this work are presented in the *Supplementary Information*.

In the first part of this study, we investigated the effect of inclusion of articles in Wikipedia on their citation counts. The inclusion time was defined as the origin of time, i.e., year zero. Citation analysis was performed on the five-year period of  $[-2, -1, 0, +1, +2]$ . Normalized citation count in year  $i$ ,  $C_i^{Norm}$ , was computed for each of the analyzed articles. Average  $C_i^{Norm}$  values are shown in Figure 2.

From this figure, it is obvious that relatively high coefficients of determination ( $R^2$  values) in the linear regression of data are

observed. In a closer look, the analyzed articles, on average, have an increasing trend in the years  $[-2, -1, 0]$ . However, this increasing trend does not continue in years  $+1$  and  $+2$ . In all three cases in Figure 2, namely “Social systems”, “Chaos theory” and “Systems biology”, either in year  $+1$  or in year  $+2$ , a decrease in the citation trend is observable. This observation may question the increasing trend of the citations.

One may argue that the observed increase in the number of citations occurs independently of inclusion of article in Wikipedia. For example, such a trend is expected because of the inclusion of an increasing number of articles in Scopus during time. Therefore, the observed increase in Figure 1 may not be more than what is expected by chance.



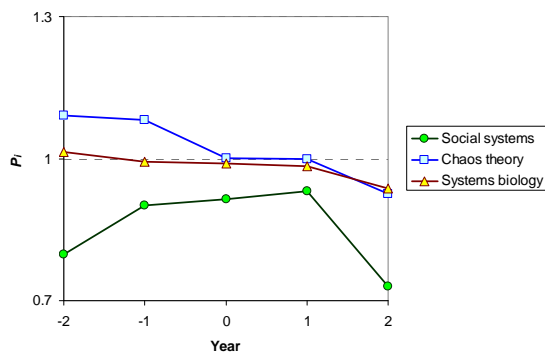
**Figure 2:** Normalized citation count before and after inclusion in Wikipedia. The reported  $R^2$  values are related to fitting of the data to the best-fit linear model.

In order to study the citations while correcting for the intrinsically increasing citation counts, we computed the citation propensity of the articles (see the Materials and Methods section). The citation propensity in year  $i$ ,  $P_i$ , measures the tendency of an article to be cited more than what is expected by chance.  $P_i > 1$  means high citation tendency in year  $i$  compared to a random year.  $P_i < 1$  means low citation tendency in year  $i$  compared to a random year. Finally,  $P_i = 1$  shows a citation tendency in year  $i$  equal to a random year. This concept is comparable to the amino acid propensity for protein secondary structures, which is the tendency of an amino acid to be includ-

ed in  $\alpha$ -helix or  $\beta$ -sheet structures (Chou and Fasman, 1974; Marashi et al., 2007).

Figure 3 shows the results of the propensity analysis. In years 0 and +1, in all of the three cases, especially in case of “Systems biology”, citation propensity is almost constant and close to 1. However, in year +2 in all of the cases we observe a considerable decrease in the propensities.

Figure 3 suggests that the increase in the number of citations after inclusion in Wikipedia (Figure 2) can be simply due to the general increase in the number of citations during time.



**Figure 3:** Propensity of citations ( $P_c$ ) before and after inclusion of the reference in Wikipedia

In the present study, we show that inclusion of articles in Wikipedia does not increase the propensity of articles to be cited. Interestingly, the reverse is reported to be true, i.e., Wikipedia selectively lists high impact articles shortly after their publication (Evans and Krauthammer, 2011).

It has been previously observed that visibility of an article, e.g., Mendeley user counts (Bar-Ilan et al., 2012; Li and Thelwall, 2012), bookmarks in CiteULike (Bar-Ilan et al., 2012) and tweets in Twitter (Eysenbach, 2011) are reported to be correlated to the citation counts of scholarly articles. Does the article “impact” increase these visibility measures, or the visibility is the cause of the increase in the scholarly citation counts of the article? We believe that this interesting question is yet to be answered.

## REFERENCES

Adam D. The counting house. *Nature* 2002;415:726-9.

Bar-Ilan J, Haustein S, Peters I, Priem J, Shema H, Terliesner J. Beyond citations: scholars’ visibility on the social web. In: Proceedings of 17th International Conference on Science and Technology Indicators, Montréal 2012, pp. 98-109.

Bird SB. Journal impact factors, h indices, and citation analyses in toxicology. *J Med Toxicol* 2008;4:261-74.

Bornmann L, Mutz R, Neuhaus C, Daniel HD. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics Sci Environ Politics* 2008;8:93-102.

Braun T, Glänzel W, Schubert A. On sleeping beauties, princes and other tales of citation distributions. *Res Eval* 2010;19: 195-202.

Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *J Am Med Assoc* 2002;287:2847-50.

Calver MC, Bradley JS. Patterns of citations of open access and non-open access conservation biology journal papers and book chapters. *Conserv Biol* 2010;24: 872-80.

Chou PY, Fasman GD. Conformational parameters for amino acids in helical,  $\beta$ -sheet, and random coil regions calculated from proteins. *Biochemistry* 1974;13:211-22.

Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *FASEB J* 2011;25: 2129-34.

Davis PM, Lewenstein BV, Simon DH, Booth JG, Connolly MJ. Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ* 2008;337: a568.

Evans P, Krauthammer M. Exploring the use of social media to measure journal article impact. *AMIA Annu Symp Proc* 2011;2011:374-81.

Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res* 2011; 13:e123.

Fava GA, Guidi J, Sonino N. How citation analysis can monitor the progress of research in clinical medicine. *Psychother Psychosom* 2004;73:331-3.

Garfield E. Citation analysis as a tool in journal evaluation. *Science* 1972;178:471-9.

Garfield E. The history and meaning of the journal impact factor. *J Am Med Assoc* 2006;295:90-3.

Gisvold SE. Citation analysis and journal impact factors – is the tail wagging the dog? *Acta Anaesthesiol Scand* 1999;43: 971-3.

Glänzel W. On the *h*-index – A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics* 2006;67:315-21.

Hirsch JE. Does the *h* index have predictive power? *Proc Natl Acad Sci USA* 2007;104: 19193-8.

Lansingh VC, Carter MJ. Does open access in ophthalmology affect how articles are subsequently cited in research? *Ophthalmology* 2009;116:1425-31.

Lawrence S. Free online availability substantially increases a paper's impact. *Nature* 2001;411:521.

Li X, Thelwall M. F1000, Mendeley and traditional bibliometric indicators. In: *Proceedings of 17th International Conference on Science and Technology Indicators*, Montréal 2012, pp. 451-551.

Marashi SA. On the identity of "citers": are papers promptly recognized by other investigators? *Med Hypotheses* 2005;65: 822.

Marashi SA, Behrouzi R, Pezeshk H. Adaptation of proteins to different environments: a comparison of proteome structural properties in *Bacillus subtilis* and *Escherichia coli*. *J Theor Biol* 2007;244:127-32.

Norris M, Oppenheim C, Rowland F. The citation advantage of open-access articles. *J Am Soc Inf Sci Tech* 2008;59:1963-72.

Phillips DP, Kanter EJ, Bednarczyk B, Tastad PL. Importance of the lay press in the transmission of medical knowledge to the scientific community. *N Engl J Med* 1991;325:1180-3.

Sancho R. Misjudgments and shortcomings in the measurement of scientific activities in less developed countries. *Scientometrics* 1992;23:221-33.

Seglen PO. Citations and journal impact factors: questionable indicators of research quality. *Allergy* 1997;52:1050-6.

Whitehouse GH. Impact factors: facts and myths. *Eur Radiol* 2002;12:715-7.