

The Fork Factor: an academic impact factor based on reuse.

Ferdinando Pucci¹ and Alberto Pepe¹

¹Affiliation not available

December 21, 2020

How is academic research evaluated? There are many different ways to determine the impact of scientific research. One of the oldest and best established measures is to look at the Impact Factor (IF) of the academic journal where the research has been published. The IF is simply the average number of citations to recent articles published in such an academic journal. The IF is important because the reputation of a journal is also used as a proxy to evaluate the relevance of past research performed by a scientist when s/he is applying to a new position or for funding. So, if you are a scientist who publishes in high-impact journals (the big names) you are more likely to get tenure or a research grant. Several [criticisms](#) have been made to the use and misuse of the IF. One of these is the policies that academic journal editors adopt to boost the IF of their journal (and get more ads), to the detriment of readers, writers and science at large. Unfortunately, these policies promote the publication of sensational claims by researchers who are in turn rewarded by funding agencies for publishing in high IF journals. This effect is broadly recognized by the scientific community and represents a conflict of interests, that in the long run increases public distrust in published data and slows down scientific discoveries. Scientific discoveries should instead foster new findings through the sharing of high quality scientific data, which feeds back into increasing the pace of scientific breakthroughs. It is apparent that the IF is a crucially deviated player in this situation. To resolve the conflict of interest, it is thus fundamental that funding agents (a major driving force in science) start complementing the IF with a better proxy for the relevance of publishing venues and, in turn, scientists' work.

Research impact in the era of forking. A number of [alternative metrics for evaluating academic impact](#) are emerging. These include metrics to give scholars credit for sharing of raw science (like datasets and code), semantic publishing, and social media contribution, based not solely on citation but also on usage, social bookmarking, conversations. We, at Authorea, strongly believe that these alternative metrics should and will be a fundamental ingredient of how scholars are evaluated for funding in the future. In fact, [Authorea](#) already welcomes data, code, and raw science materials alongside its articles, and is built on an infrastructure (Git) that naturally poses as a framework for distributing, versioning, and tracking those materials. [Git](#) is a versioning control platform currently employed by developers for collaborating on source code, and its features perfectly fit the needs of most scientists as well. A versioning system, such as [Authorea](#) and [GitHub](#), empowers **forking** of peer-reviewed research data, allowing a colleague of yours to further develop it in a new direction. Forking inherits the history of the work and preserves the value chain of science (i.e., who did what). In other words, forking in science means *standing on the shoulder of giants* (or soon to be giants) and is equivalent to citing someone else's work but in a functional manner. Whether it is a "negative" result (we like to call it non-confirmatory result) or not, publishing your peer reviewed research in Authorea will promote forking of your data. (To learn how we plan to implement peer review in the system, please stay tuned for future posts on this blog.)

More forking, more impact, higher quality science. Obviously, the more of your research data are published, the higher are your chances that they will be forked and used as a basis for groundbreaking work, and in turn, the higher the interest in your work and your academic impact. Whether your projects are data-driven peer reviewed articles on Authorea discussing a new finding, raw datasets detailing some

novel findings on [Zenodo](#) or [Figshare](#), source code repositories hosted on Github presenting a new statistical package, every bit of your work that can be reused, will be forked and will give you credit. Do you want to do a favor to science? Publish also non-confirmatory results and help your scientific community to quickly spot [bad science](#) by publishing a dead end fork (Figure 1).

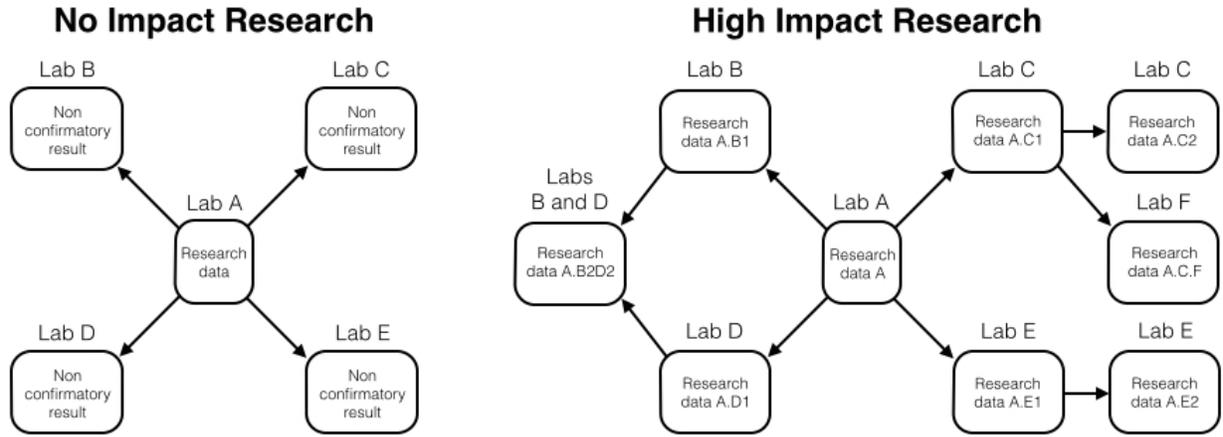


Figure 1: **Figure 1.** Left panel, Lab A publishes the results of an experiment (left). Labs B, C, D and E find it interesting and decide to fork it to confirm its conclusions and perform additional experiments. However, they soon realize that they cannot reproduce lab A’s data. Right panel, the experiment published by lab A is forked by labs B, C, D and E, who then generate more data (A.B1, A.C1, A.D1 and A.E1) that lead to further discoveries (A.C2 and A.E2), more forking (A.C.F) and new collaborations (A.B2D2). The Fork Factor can accurately measure the impact of lab A research in the two cases, in a much faster and functional way than the Impact Factor.

And now onto the nerdy part: The Fork Factor. So, we would like to imagine what academia would be like if forking actually mattered in determining a scholar’s reputation and funding. How would you calculate it? Here, we give it a shot. We define the **Fork Factor** (FF) as:

$$FF = N * (L^{\frac{1}{\sqrt{N}}} - 1) \tag{1}$$

Where N is the number of forks on your work and L their median length. In order to take into account the reproducibility of research data, the length of forks has a higher weight in the FF formula. Indeed, forks with length equal to one likely represent a failure to reproduce the forked research datum.

Anyone out there care to improve the formula above? For instance, would it be better if the FF would reach a plateau for $L \geq 3$? Let us know at hi@authorea.com or by commenting here.