The FAIR Metrics of Adherence to Citation Best Practices

Adam G. Craig and Carl Taswell
Brain Health Alliance

Measuring the merits of scholarly research articles only by citation counts and how often other research articles or social media messages cite a particular publication creates a perverse incentive for some authors to refrain from citing potential rivals. This dilemma has developed despite the historical publishing standard expected in peer review for citing and discussing related prior work. To encourage and support a countervailing incentive, research organizations should also consider metrics for how well and appropriately a scholarly article cites relevant prior work in the spirit of the classic phrase and metaphor *standing on the shoulders of giants*. We present a proposal for a family of such article-level metrics called the FAIR metrics and described as the FAIR Attribution to Indexed Reports or the FAIR Acknowledgment of Information Records.

Keywords: article-level metric, citation practices, FAIR metrics

Introduction

Institutions increasingly look to citation metrics to provide an objective, quantitative measure of the merits of research, researchers, and scholarly journals (Dix, 2016). However, documents such as *the Leiden Manifesto* (Hicks, Wouters, Waltman, Rijcke, & Rafols, 2015), *the San Francisco Declaration* (Cagan, 2013), and *the Metric Tide* (Wilsdon, 2016), express the widely-held view within the scientific community that such metrics fail to capture essential aspects of the quality of scholarly work. Conventional citation metrics that measure the merits of a published work according to how many other works cite it creates a perverse incentive for some authors: Citing relevant prior work improves metric values of potential rivals, not of the author. To bring balance to the system of incentives, institutions should also consider metrics of how appropriately and fairly a report of scientific work references prior reports as published in the literature and indexed in databases. We present desired properties for such a family of FAIR metrics and approaches to their formulation.

Methods

We first identified a set of essential features for a FAIR metric. The defining feature for a FAIR metric should be the ability to distinguish between the four alternative citation cases of appropriate presence or absence and inappropriate presence or absence. Allowing for additional analyses with respect to statements about literature review, results and conclusions in a scientific report yields the six different scenarios displayed in Figure 1. Beyond detecting simple errors or mistakes, a FAIR metric should most importantly differentiate an unintentional act of omission in referencing an article from an intentional act of plagiarism of that article. This requirement implies the ability to distinguish between incidental background information and the core claims of a work and to weight the penalty for failure or refusal to cite accordingly. The necessity to distinguish appropriate from inappropriate citation practices, regardless of which practices are most common, implies that the evaluation process must determine whether one work should cite another based on desired principles for citation and on intrinsic features of the two works, and thus remain resistant to any worsening prevalence of undesirable citation practices and attempts at obfuscation or plagiarism. A FAIR metric must be formulated so that scores remain stable in the face of semantically insignificant changes in writing and wording, even when imperfect lexical analysis tools perform the comparison of claims. Conventional citation metrics exhibit a bias that favors works in domains with a higher rate of publication (Tyler, 2018). In contrast, a FAIR metric must evaluate each work in a way that remains both sensitive to and robust for the context of the problem domain of scientific inquiry. A FAIR metric must also support an evaluation process that considers fairly the accepted standards for common knowledge in a problem domain. We next identified necessary tools for implementing these design features.

Results

Since the purpose of calculating the metric remains the characterization of how well a given work fits into a larger collection of works, inputs for the metric must include not only the work being evaluated but the entire set of descriptions of all potentially relevant works, including the network of citations between works. The need to consider domain-specific publication patterns and common knowledge calls for the use of data sets organized by problem domain and validated for relevance using concept-validating constraints, such as those used in the PORTAL-DOORS Project (Taswell, 2010). The first step of calculation must identify matches

between content in an article and the collection of previously published articles. The most precise approach would be to compare semantic descriptions, but the family of FAIR metrics should also permit and support use of lexical analyses. The need to define whether one work should cite another based on standards of scholarly publishing and to distinguish whether the lack of a citation results appropriately from irrelevance versus inappropriately from either an unintentional error of omission or an intentional act of plagiarism both call for the use of content analysis (a) to identify the set of statements and claims in a work and (b) to determine which claims are equivalent between the two works. Unlike citation analysis alone, content-based methods can find missing citations even when a field is split into silos (Ding et al., 2014). Similarly, content analysis can detect plagiarism when claimed original results, not background references, are plagiarized (Vani & Gupta, 2018). Plagiarism detection techniques remain vulnerable to obfuscation by changes in wording so that even state-of-the art tools for identifying equivalent claims will produce some false positives and false negatives. Consequently, as long as the matching method for the metric meets a minimum standard of accuracy, the overall score for the FAIR metric should reflect consistent patterns of appropriate or inappropriate citation. Finally, calculation must take into account the context so that the score of a work depends more heavily on correct attribution of claims central to its purpose. For example, the score of a replication study for the sake of validation should not be required to duplicate all of the background citations, nor be penalized for reporting similar results as expected in a reproducibility study.

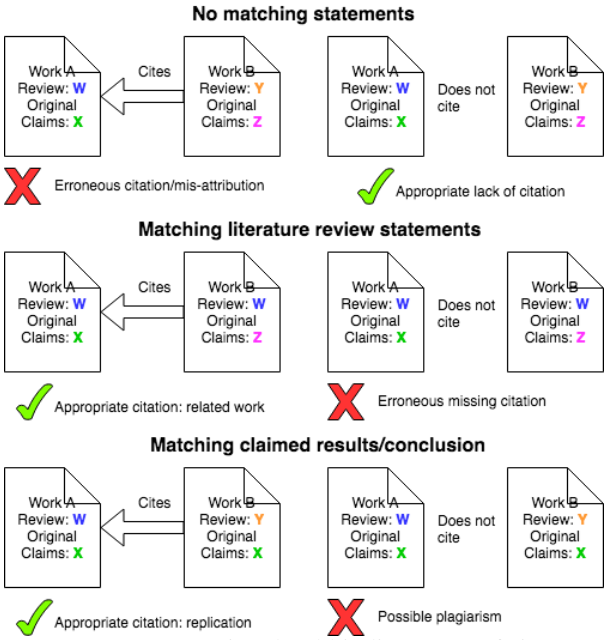

Figure 1. FAIR metrics should indicate one of six scenarios most likely probable between publications A and B.

Future Work

We plan to identify which modes of semantic and lexical similarity analysis are most suitable for the FAIR family of metrics, implement software that can evaluate them with each selected analysis method using metadata stored in Nexus-PORTAL-DOORS servers (Craig, Bae, & Taswell, 2017), and compare results for examples from the biomedical literature with ratings by human domain experts.

References

Cagan, R. (2013). The San Francisco declaration on research assessment. The Company of Biologists Ltd.

Craig, A., Bae, S. H., & Taswell, C. (2017). Message exchange between independent implementations of servers in the Nexus-PORTAL-DOORS system. In *SWAT4HCLS: Proceedings of the 10th international SWAT4HCLS conference*.

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, *65*(9), 1820–1833.

Dix, A. (2016). Evaluating research assessment: Metrics-based analysis exposes implicit bias in ref2014 results. *Impact of Social Sciences Blog*.

Hicks, D., Wouters, P., Waltman, L., Rijcke, S. d., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics.

Potthast, M., Barrón-Cedeño, A., Eiselt, A., Stein, B., & Rosso, P. (2010). Overview of the 2nd international competition on plagiarism detection. In *Proceedings of the 4th workshop on uncovering plagiarism, authorship, and social software misuse* (pp. 1–14).

Taswell, C. (2010). Concept validating methods for maintaining the integrity of problem oriented domains in the PORTAL-DOORS system. In S. Swift & K. Phillips (Eds.), *IDAMAP 2010: Intelligent data analysis in biomedicine and pharmacology*.

Tyler, T. (2018). Citation metrics and impact factors fail as measures of scientific quality, in particular in taxonomy, and are biased by biological discipline and by geographic and taxonomic factors. In *Annales botanici fennici* (Vol. 55, *1–3*, pp. 185–191). BioOne.

Vani, K., & Gupta, D. (2018). Integrating syntax-semantic-based text analysis with structural and citation information for scientific plagiarism detection. *Journal of the Association for Information Science and Technology*.

Wilsdon, J. (2016). *The metric tide: Independent review of the role of metrics in research assessment and management*. Sage.