Audience Based View of Publication Impact

Robert M. Patton
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831
pattonrm@ornl.gov

Drahomira Herrmannova
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831
herrmannovad@ornl.gov

Christopher G. Stahl
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831
stahlcg@ornl.gov

Jack C. Wells
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831
wellsjc@ornl.gov

Thomas E. Potok
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831
potokte@ornl.gov

Abstract

With increasing pressure to demonstrate a return on investment, scientific user facilities and government funded research projects must strive to demonstrate positive impact of their research on others. Historically, citation counts and journal impact factor have been widely used as a single quantitative measure of impact. However, with the expansion of various modes of communication, citation counts no longer hold as much merit as a single measure of impact. This work seeks to expand on citation counts by considering the audience to which a publication draws attention. Based on communication mode, there are various audiences that may be "impacted" but those audiences are not fully captured by citation count. Consequently, we propose that impact measures for publications should be oriented around the audience with quantifiable measures based on the various communication modes.

CCS Concepts

Applied computing → Digital libraries and archives; • Information systems→Information retrieval;

1. Introduction

Scientific research continues to expand both human understanding of our world and solve societal problems through technical progress. Accomplishing progress in scientific research typically requires a) scientific user facilities, b) funding to cover expenses, and c) people capable of pursuing this progress. Unfortunately, the limiting factor of these three is often funding. On one hand, the people who need the funding are those who understand and can accomplish the technical or scientific hurdles. On the other hand, the people who control the -ow of funding must balance the need for scientific discovery with meeting the practical needs of society. This conflict has driven the desire for a single metric that can accurately, objectively, and quickly evaluate proposed scientific research or the results of scientific research already performed. Funding decisionmakers desire to use this evaluation in order to help decide where to invest funding and ensure a return on that investment while the scientific individual desires to use this evaluation to establish their credibility as an expert in their respective field. Metaphorically, an evaluation metric that is a two-edged sword.

The origins of scientific publications can be traced back to a small group of scientists whose work would have been published in a physical book, most likely in a limited distribution [18]. Publications within this small group would most likely be reference by those members in future publications. In this case, a citation count makes practical sense under the assumption that if someone were to reference a prior work, then it must have in some way shaped or influenced the theory or paradigm behind the current work being published. Furthermore, the audience for these publications would be sufficiently small such that awareness or visibility of prior works would not likely have been an issue. In this case, there is small audience of scientists with at most two modes of communication (written documents or oral discussion via face to face meetings). Consequently, citation counts would be an accurate, objective, and quick metric for evaluating publication impact for both funders and scientists.

In sharp contrast, today's science communities can reach multiple audiences whose sizes, educational backgrounds, cultural backgrounds, geographic locations as well as other characteristics can vary widely. Further, the mediums by which to reach these audiences can vary widely: hard copy books, soft copies available via the Internet, blogs, news, social media, etc. Consequently, citation counting becomes severely diluted in its ability to accurately, objectively, and quickly evaluate publication impact. Is a low citation count indicative of low quality, or is it that the work was published in a closed access journal that is not widely subscribed? Or, is a low citation count a result of the impacted audience not even capable of citing the work (e.g., policy makers, librarians)? Despite this dilution of citation counting, it is still used as a means to measure publication impact, and for good reason as highly cited works continue to demonstrate influence in moving an audience from one way of thinking to a different way of thinking.

We are not proposing that citation counting must be replaced or ignored. Instead, we must realize that for a specific audience (e.g., scientific community) and specific modes of communication (e.g., conferences and journals), citation counting can still be an accurate, objective, and quick assessment of a publication impact despite its flaws and its ability to be manipulated. However, what of the other audiences and modes of communication outside this scope? Are there metrics appropriate for those audiences and communication modes that could be used in conjunction with citation counts that can measure impact?

Altmetrics provides a means of measuring interest from audiences beyond the scientific community and beyond the normal mode of communication within the science community (e.g., conference and journals) [17]. In this context, this work explores the comparison of citations with altmetrics and show how they may be used together to view publication impact with respect to audiences.

2. Related Works

A recent survey by Waltman [22] provides an overview of current developments in the area of bibliometrics. This form of analysis has historical been used to evaluate research quality and impact [8], and, despite the problems with citation analysis identified in [14], this approach to evaluating research continues today. As will be discussed, our work differs in that the paradigm adopted here is that citation analysis is a measure of the interests of the audience, and not necessarily a measure of quality of the researcher or research performed.

In [5], a description of altmetrics and an overview of its advantages and disadvantages is provided. As highlighted in that work, there continues to be mixed reviews as to the value that altmetrics provides. While it captures information regarding research from different communication mediums, the context of [5] is from the perspective of evaluating the particular research or researcher. As will be discussed, our work differs in that our context is not to evaluate the research or researcher, but rather the opposite: to evaluate the interest of the audience in the published research, specifically the general public. In fact, the work of [16] investigated whether altmetrics could be used as a proxy for bibliometrics. The author shows that this is not the case and that altmetrics "could describe an alternative dimension of the academic uses, close to science popularization, networking abilities and social skills."We extend on this premise that altmetrics is another dimension and suggest that altmetrics should be considered as a measure of audience interests in a particular research work.

3. Pasteur's Quadrant

In the traditional, linear view of innovation [20], scientific research ranges from basic to applied. This model suggests that innovation starts with basic research, then adds applied research, development and production. In this view, basic research intends to understand fundamental phenomena within the universe regardless of its applicability to solving societal problems. At the other end of the spectrum is applied research, which intends to directly solve societal problems without necessarily expanding our fundamental understanding of natural phenomena. This model has been much studied over the past decade [3, 9, 20].

Stokes [19] challenges this traditional view of innovation. He motivates this by the growing need to harness science for societal benefits and technological race on the one hand and the inappropriateness of the linear model for describing "oriented basic research" on the other hand. He introduces the phrase "Pasteur's Quadrant", and describes this quadrant as being research that spans both basic and applied research. The author states that the research of Louis Pasteur, a renowned biologist and chemist whose use-inspired studies led to many fundamental contributions to science and laid the foundations of microbiology, exemplifies this dual-purpose research. When using the linear model, the first instinct might be to place Pasteur's research at a point between basic and applied research. However, Stokes points out Pasteur's research should be placed at both ends, thanks to his quest for understanding the microbiological processes he discovered and the inventions he made which had direct effects on society. Stokes therefore proposes to use a two-dimensional rather than a linear view of science. In this view, the axes describe the degree to which the research focuses on fundamental understanding (vertical axis) or immediate applications (horizontal axis), and which therefore enables representing Pasteur's research as a single point. Aside of the Pasteur's quadrant, Stokes goes on to define a Niels Bohr quadrant that represents more purely basic research, and a Thomas Edison quadrant that represents more purely applied research. These quadrants are shown in Figure 1. The remaining quadrant, which represents research, which is neither focused on fundamental understanding nor on specific societal problems, is not empty, but includes "research that systematically explores particular phenomena without having in view" specific objectives. Such research may be driven by the curiosity of the researcher.

4. Approach

In this work, we consider the context of the people who control the flow of funding must balance the need for scientific discovery with meeting the practical needs of society. As such, we define broadly that a scientific publication will simply have two audiences: the scientific community and the general public. In addition, we define that the modes of communication for the scientific community audience would be conference and journal publications and we assume that citation counts continue to accurately, objectively, and quickly measure at a minimum, the interest of the audience in the work, and at most, measure the influence of a publication to move the audience from one way of thinking to a different way of thinking. Further, we define that the modes of communication for the general public audience would be news and social media and that an altmetrics count provides an accurate, objective, and quick measure of the general public interest in the publication, which we also assume would be a result of the publication addressing a societal problem.

Further, we adopt the paradigm described by Stokes [19]. Our approach extends this paradigm to leverage citation counts as a measure of interest in a publication by the scientific community. The assumption here is that a high citation count tells us that the science community is very interested in the work, and vice versa. Likewise, our approach leverages altmetrics as a measure of interest in a publication by the general public; with a high altmetric count telling us that the general public is very interested in the work, and vice versa.

With this quantifiable paradigm of "Pasteur's Quadrant" in place, we applied this to a set of 1,117 papers spanning publication years 2012 to 2017. These papers specifically acknowledge having used the resources provided by the Oak Ridge Leadership Computing Facility and are recent enough that altmetric values for the publications could exist. The results are shown in Figure 2. Each dot of the graph represents one of the publications, with publications from different years being distinguished by color. The X-axis is the *immediate application* represented by altmetric value (collected from altmetric.com) and the Y-axis is the *fundamental understanding* represented by citation count (collected from Web of Science). The dashed gray lines in the figure represent mean value, which is what we used to separate the quadrants.

4.1 Bohr's Quadrant

According to Stokes [19], the top left quadrant represents purely basic research with the goal of fundamental understanding without thinking about practical use. Based on our quantified version of this paradigm, strong examples of this quadrant have very high citation counts and no altmetric counts. Examples include [6], [7], [12], and [13]. In reading these examples, it is clear that these papers are simply studying and reporting on experiments and research of very specific natural phenomena, and that there does not appear to be an implied or explicit claim to solving a specific societal problem.

4.2 Edison's Quadrant

According to Stokes [19], the bottom right quadrant represents purely applied research (research, which is guided by applied goals without considering basic understanding). Based on our quantified version of this paradigm, strong examples of this quadrant have almost no citation counts and very high altmetric counts. Examples include [1], [2], [11], and [15]. In contrast to Bohr's quadrant, these papers are specifically addressing a societal problem, so it does not surprise us that they have high altmetric counts. Likewise, these papers are not necessarily researching a basic understanding of some phenomena in the universe. Consequently, these papers have low, if any, citation counts.

There does appear to be some exception in this quadrant. In particular, the work of [1]. This work focuses on a societal problem of developing drought resilience in crops. However, the paper is discussing natural phenomena of the dynamics of drought resistant plants, in order to better understand how they work. At the time of our study, this paper was published only 5 months prior. As a result, the altmetrics count is significantly higher than the citation count since altmetric counts accumulate significantly faster than citations. Consequently, we expect that some of the publications in the Edison quadrant such as [1] will move into the Pasteur quadrant as time allows the citation counts to accumulate. In contrast, we do not think that this will hold true for the Bohr quadrant. Since citation counts take time to accumulate, if a publication has a high citation count but a low altmetric count, then there is little, if any, chance that it would move into the Pasteur quadrant.

4.3 Pasteur's Quadrant

According to Stokes [19], this quadrant represents use-inspired basic research. Based on our quantified version of this paradigm, strong examples of this quadrant have very high citation counts and very high altmetric counts. Examples include: [4], [10], and [21]. Unlike the Bohr and Edison quadrant, the Pasteur quadrant is distinctly different based on these examples. In the work of Bhimanapati et al. [4], the authors state that one of their objectives for the work is to "discuss the newest families of 2D materials, including monoelement 2D materials (i.e., silicene, phosphorene, etc.) and transition metal carbide- and carbon nitride-based MXenes." This clearly positions the publication as basic research. In addition, the authors also state that another objective of their work is to "discuss the doping and functionalization of 2D materials beyond graphene that enable device applications, followed by advances in electronic, optoelectronic, and magnetic devices and theory." This clearly shows that the authors recognize that their work is use-inspired basic research in that while it's basic, it has a clear translational path into technology that will directly apply to society.

In the work of Hashim et al. [10], the authors state that "Detailed elemental analysis revealed that the "elbow" junctions are preferred sites for excess boron atoms, indicating the role of boron and curvature in the junction formation mechanism, in agreement with our first principle theoretical calculations." In other words, the authors are performing basic research into the role of boron in forming 3D macroscale nanotube elastic solids. In addition, the authors recognize the direct translational path of their research to solve a specific societal problem when they state that "the strongly oleophilic sponge-like solids are demonstrated as unique reusable sorbent scaffolds able to efficiently remove oil from contaminated seawater even after repeated use."

In the work of Tessum et al. [21], the authors state that their approach includes “a state-of-the-science mechanistic meteorology and chemical transport model.” In other words, the authors are directly leveraging the very latest in basic research in order to accomplish their objective of assessing “the life cycle air quality impacts on human health of 10 alternatives to conventional gasoline vehicles.” Clearly, this is a significant societal problem, but the approach taken was to leverage the latest advancements in chemical and meteorological basic research.

5. Summary and Future Work

Despite the challenges and issues associated with citation counts and altmetrics individually, when used together in a complimentary way, they provide insight into the dynamics of the interactions between a publication and its corresponding audience. These two metrics are often used together in a way where one is thought to replace or predict the other (in other words, both metrics are thought to express the same or a similar concept). In this paper we have shown a different view of the two metrics that can be used for identifying publications representing use-inspired basic research. There are several things we plan on investigating next. First, we have observed a pattern in the use of language by papers in the different quadrants. While papers in the Bohr quadrant use very domain specific language that may be hard to understand for non-experts, papers in the Edison quadrant use more generalized language. Therefore, we may be able to predict the quadrant in which a paper will fall strictly based off the language used in the papers at the time of publication. Another avenue of research is to use sentiment analysis on altmetrics and citations as an additional metric to measure the impact of a particular research paper. Our current research assumes that all mentions are equal, but that may not be the case. References to a research paper that counter or contradict its content should weigh differently than references that reinforce or build on its content. We believe that incorporating this metric can bring additional insight into the way that research is used.

Acknowledgements

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract

References

[1] Paul E. Abraham, Hengfu Yin, Anne M. Borland, DeborahWeighill, Sung Don Lim, Henrique Cestari De Paoli, Nancy Engle, Piet C. Jones, Ryan Agh, David J.Weston, Stan D. Wullschleger, Timothy Tschaplinski, Daniel Jacobson, John C. Cushman, Robert L. Hettich, Gerald A. Tuskan, and Xiaohan Yang. 2016. Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. Nature Plants 2 (2016), 16178. DOI:http://dx.doi.org/10.1038/nplants.2016.178

[2] Marcella Alsan. 2014. The effect of the tsetse -y on African development. The American Economic Review 105, 1 (2014), 382–410.

[3] Margherita Balconi, Stefano Brusoni, and Luigi Orsenigo. 2010. In defence of the linear model: An essay. Research Policy 39, 1 (2010), 1–13.

[4] Ganesh R Bhimanapati, Zhong Lin, Vincent Meunier, Yeonwoong Jung, Judy Cha, Saptarshi Das, Di Xiao, Youngwoo Son, Michael S Strano, Valentino R Cooper, and others. 2015. Recent advances in two-dimensional materials beyond graphene. Acs Nano 9, 12 (2015), 11509–11539.

[5] Lutz Bornmann. 2014. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. Journal of Informetrics 8, 4 (2014), 895 – 903. DOI:http://dx.doi.org/10.1016/j.joi.2014.09.005

[6] StephenWBruenn, Anthony Mezzacappa,WRaphael Hix, Eric J Lentz, OE Bronson Messer, Eric J Lingerfelt, John M Blondin, Eirik Endeve, Pedro Marronetti, and Konstantin N Yakunin. 2013. Axisymmetric Ab Initio Core-collapse Supernova Simulations of 12-25 MâŸL Stars. The Astrophysical Journal Letters 767, 1 (2013), L6.

[7] Peter Ferrin, Shampa Kandoi, Anand Udaykumar Nilekar, and Manos Mavrikakis. 2012. Hydrogen adsorption, absorption and diffusion on and in transition metal surfaces: A DFT study. Surface science 606, 7 (2012), 679–689.

[8] E. Gareld. 1979. Is citation analysis a legitimate evaluation tool? Scientometrics 1, 4 (1979), 359–375. DOI:http://dx.doi.org/10.1007/BF02019306

[9] Benoît Godin. 2006. The linear model of innovation: The historical construction of an analytical framework. Science, Technology, & Human Values 31, 6 (2006), 639–667.

[10] Daniel P Hashim, Narayanan T Narayanan, Jose M Romo-Herrera, David A Cullen, Myung Gwan Hahm, Peter Lezzi, Joseph R Suttle, Doug Kelkhoff, Emilio Munoz-Sandoval, Sabyasachi Ganguli, and others. 2012. Covalently bonded three-dimensional carbon nanotube solids via boron induced nanojunctions. Scientific reports 2 (2012), 363.

[11] Aixue Hu, Samuel Levis, Gerald A Meehl, Weiqing Han, Warren M Washington, Keith W Oleson, Bas J van Ruijven, Mingqiong He, and Warren G Strand. 2016. Impact of solar panels on global climate. Nature Climate Change 6, 3 (2016), 290–294.

[12] Jun Kang, Jingbo Li, Shu-Shen Li, Jian-Bai Xia, and Lin-Wang Wang. 2013. Electronic structural Moiré pattern effects on MoS2/MoSe2 2D heterostructures. Nano letters 13, 11 (2013), 5485–5490.

[13] Xiaodong Li, Jeffrey T Mullen, Zhenghe Jin, Kostyantyn M Borysenko, M Buongiorno Nardelli, and Ki Wook Kim. 2013. Intrinsic electrical transport properties of monolayer silicene and MoS 2 from first principles. Physical Review B 87, 11 (2013), 115418.

[14] Michael H. MacRoberts and Barbara R. MacRoberts. 1989. Problems of citation analysis: A critical review. Journal of the American Society for Information Science 40, 5 (1989), 342–349. DOI:http://dx.doi.org/10.1002/(SICI)1097-4571(198909)40: 5<342::AID-ASI7>3.0.CO;2-U

[15] Gerald A Meehl, Claudia Tebaldi, and Dennis Adams-Smith. 2016. US daily temperature records past, present, and future. Proceedings of the National Academy of Sciences 113, 49 (2016), 13977–13982.

[16] JosÃl Luis Ortega. 2015. Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. Journal of Informetrics 9, 1 (2015), 39 – 49. DOI:http://dx.doi.org/10.1016/j.joi.2014.11.004

[17] Jason Priem, Dario Taraborelli, Paul Groth, and Cameron Neylon. 2010. Altmetrics: A manifesto. (2010). http://altmetrics.org/manifesto

[18] Sally Shuttleworth and Berris Charnley. 2016. Science periodicals in the nineteenth and twenty-first centuries. Notes and Records: the Royal Society journal of the history of science 70, 4 (oct 2016), 297–304. DOI:http://dx.doi.org/10.1098/ rsnr.2016.0026

[19] Donald E Stokes. 2011. Pasteur's quadrant: Basic science and technological innovation. Brookings Institution Press.

[20] William J Sutherland, David Goulson, Simon G Potts, and Lynn V Dicks. 2011. Quantifying the impact and relevance of scientific research. PLoS One 6, 11 (2011), e27537.

[21] Christopher W Tessum, Jason D Hill, and Julian D Marshall. 2014. Life cycle air quality impacts of conventional and alternative light-duty transportation in the United States. Proceedings of the National Academy of Sciences 111, 52 (2014), 18490–18495.

[22] Ludo Waltman. 2016. A review of the literature on citation impact indicators. Journal of Informetrics 10, 2 (2016), 365–391. 5No. DE-AC05-00OR22725.

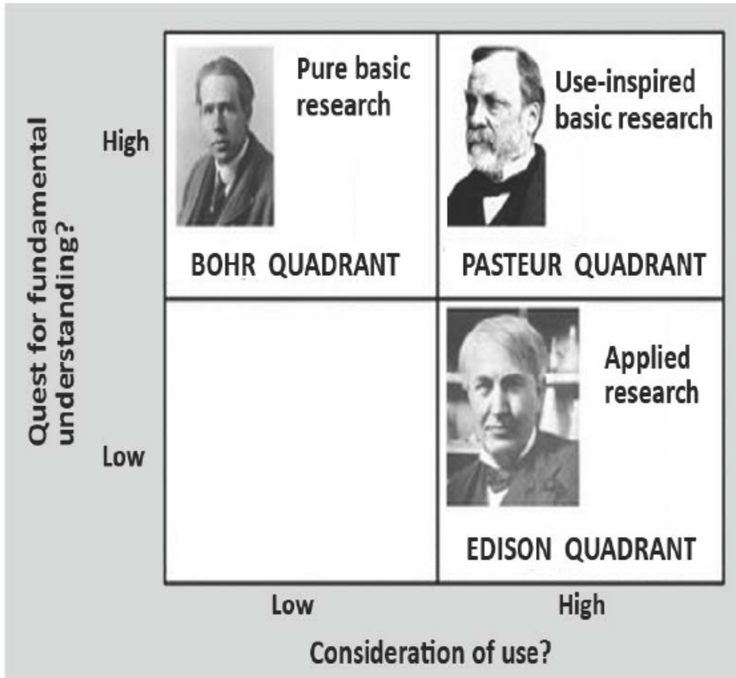

Figure 1: As defined in [19], there are different quadrants of research depending on purpose: purely basic, purely applied, and use-inspired basic research.

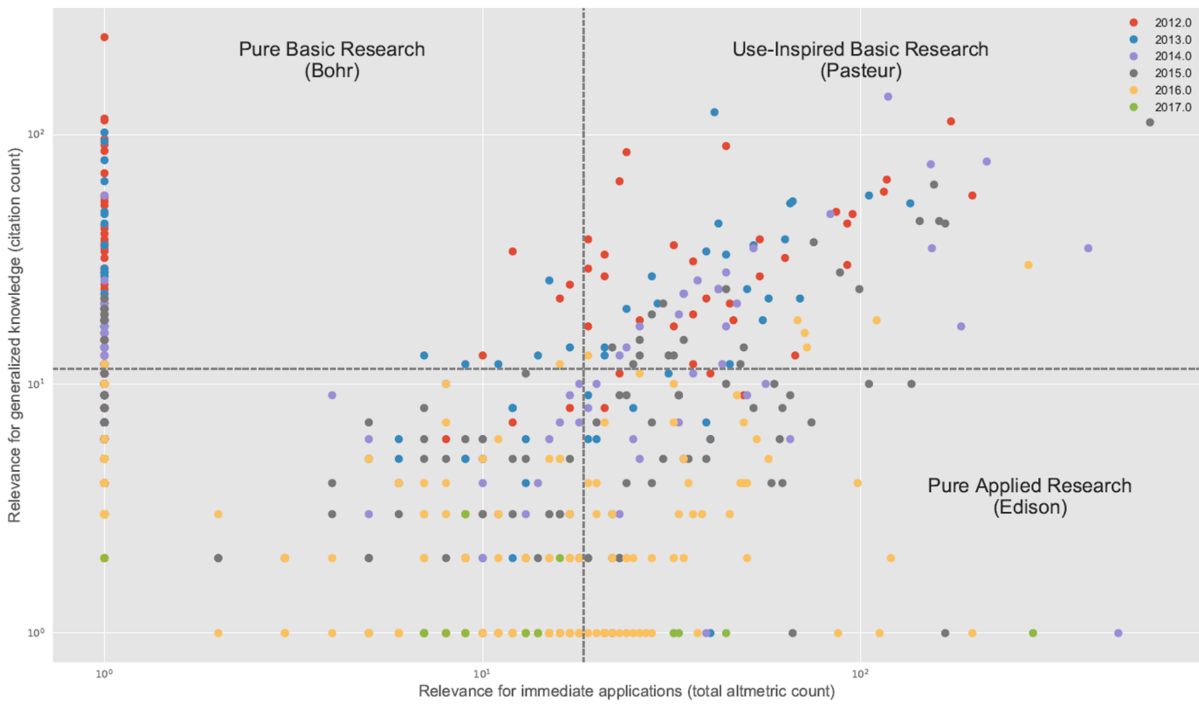

Figure 2: A scatter plot of publications that acknowledge having utilized the Oak Ridge Leadership Computing Facility and their corresponding citation (Y-axis) and altmetric (X-axis) values. Each dot represents a single publication.