Identifying high impact scientific work using Natural Language Processing and Psychology.

Brett Buttliere
Leibniz-Institut für Wissensmedien
Knowledge Media Research Center
Schleichstraße 6, Rm 6.419
72076 Tübingen
+49 7071-979-239
b.buttliere(aт)iwm-tuebingen.de

ABSTRACT

The target article and this response focus on utilizing the available information about papers (e.g., full text, citations, mentions and discussion elsewhere on the internet) to better inform our understanding of the work's impact on the field. Whalen, Huang, Sawant, Uzzi, & Contractor (2015) found evidence that especially the number of citations from papers that utilize highly dissimilar keywords was related to more citations outside of the journal. In our response, we reply to the major aspects of their paper and suggest two further things for potential discussion.

General Terms
Algorithms, Measurement, Documentation, Performance, Theory, Experimentation, Standardization, Verification.

Keywords
Scientific Communication, Cognitive Conflict, Negativity Bias, Psychology, Social Media, Metascience, Science of Science

With the advent of the internet age in science, there is more and more data available about papers, authors, and scientific interaction on the web. Whalen et al., (2015) and I are attempting to encourage researchers to think about the potential value of this data to better understand how science is done and how it might be done more efficiently; especially by being able to better identify what makes high impact work.

1. SUMMARY WHALEN ET AL., 2015
Whalen et al. (2015) began their investigation by building a network of academic keywords from all papers published in *Web of Science*, consisting of about 85,000 keyword associations. After characterizing all papers in the journal *Social Networks* by examining how often these keywords appear in the full text of each paper, Whalen et al. (2015) computed a keyword distance score for each citation within the journal. Using this metric of distance, they counted, for each paper, the number of citations in each quantile and regressed this on citations outside of the journal.

The results suggest that the number of very close (lowest quartile distance) and especially the number of very distant citations (highest quartile of distances) were predictive of citations outside the journal. Whalen et al. (2015) suggest this is because topically distant citations indicate impact in the wider literature while topically similar citations indicate a paper influencing a particular research line. They then discuss some other potential uses of the data.

2. REACTION TO WHALEN ET AL. 2015
The goal of utilizing more of the data available online to improve our understanding of scientific communication is one we wholeheartedly endorse. This response is then an examination of how Whalen et al. (2015)'s method might be improved and expanded upon, especially when thinking about the temporality of citation counts and impact. After considering each of the major aspects of their examination, two further thoughts are proposed for discussion: using data from science sharing platforms as well, and especially applying psychology and other literatures to achieve a better theoretical understanding of the process.

2.1 The Keyword Network
This is potentially the most interesting aspect of this research for me, as I believe it has much potential. Such a network can be used, aside from examining the distance of citations, to make better predictions about what types of articles an author might like to read or to examine similar keyword distance hypotheses at the author or keyword level (i.e., those that cover more keyword distance have more impact). Such an investigation might look at the 'intellectual space' an author or keyword has covered and how this relates to the number of page views, online discussion, or citations an individual has, even examining how the most successful careers or ideas progress (watch paradigms shift).

The most pressing limitation for the keyword network for me is that it does not account for keyword popularity across time. It makes sense that older citations should be cited more in general, but it could also be important to take into account whether the paper is published when a question is gaining or losing popularity. On the one hand, a paper that utilizes very popular keywords is probably more likely to be cited, as the field is simply larger. On the other hand, truly great work might reignite a keyword or research line, or creates an entirely new research line (which is then picked up by others). Such nuances will probably need to be teased apart with some clever metrics for the novelty and popularity of the keyword combinations. One metric might count the number of times a keyword has been used (indicating popularity) while another could indicate how early it was in the series of articles using that keyword or keyword pairing.

2.2 Profiling a single paper
Examining the importance of the keywords in the paper is an excellent way to characterize the papers, and I would further urge using other metrics as well. Especially the age of the paper seems important, as older papers are more likely to be cited both inside and outside of the journal.

There are also other metrics which are very likely related to overall impact including: page views, page rank, download to citation ratio, the types of citations in the paper and the sentiment of the paper [2, 3]. Beyond the paper itself, there are many other factors that might affect impact as well, including: the author, the journal the paper is in, and especially as science sharing networks grow, online mentions and discussion. Two versions of the same paper published by a PhD student or by a Nobel Laureate will have very different impact on the field. Ultimately, all of these aspects of a paper can and even should be related to the impact a paper has on the field and science more generally, however that is operationalized (e.g., citation counts).

As the number of metrics available grows, it will become imperative to analyze all of these metrics together (potentially including traditional citation counts) to gain a more nuanced understanding of scientific impact.

2.3 Measuring good science

The potential to characterize a paper's impact in more than one way is demonstrated by the target article's differentiation between those citations within the journal and those outside of it. One could make an argument that those citations from inside the journals are the best indicator of influence in the field, especially if the target journal leads the field (like *Social Networks*). The choice of which metric to consider most informative probably comes down to what one wants to measure, and this is why it is important to nuance our view of impact.

Rather than try to compare new metrics to a metric that is widely used but known to be flawed [4], it would be better, in my opinion, to take a more nuanced view of impact, utilizing all of the available information. It is because the metrics are not perfect that we attempt to improve them; but, if we compare new metrics to those that are known to be flawed, we implicitly make that flawed metric the gold standard to meet. In even the best case scenario, it is impossible to better predict the number of citations a paper has than a simple count of the number of citations a paper has, even if another metric (or more than one) is *actually* a better indicator of impact. Instead, many metrics can be created and the relationships between them can be examined to characterize papers along one or more dimensions (e.g., methods, idea).

It makes sense to nuance our idea of impact to more than one dimension; a simple examination of the top 100 cited papers ever demonstrates plainly that some papers are cited many times because they are methodological advances while others are cited for being summative or innovative [5]. Pulling these nuances apart seems to be the next logical step. One way this might be done is by examining sentences with citations to a target article for indicators of different aspects of a paper (e.g., the methodology, results, or idea) and analyzing this further [2].

One final problem with utilizing citations as the central metric of scientific impact or as a predictor or impact is that they take a long time to accumulate. Whalen et al. (2015) looked throughout the full 36 year history of the journal *Social Networks*, which is one of the highest impact journals in the field. It is an open question whether such a journal centric analysis will hold for newer or less impactful journals (most open access journals are less than 5 years old and the majority of journals have very low citation rates). It is also important to recognize that (as noted above) the creation of the journal *Social Networks* itself signals the creation of a new field and keywords to use, again emphasizing the need to look across time at citations and keyword usage.

3. TWO MORE IDEAS FOR DISCUSSION

In the interest of discussion, I would like to broaden the conversation in two ways. The first is to suggest also using the data from science sharing platforms (e.g., Twitter, Facebook, PubPeer) to bear on the issue. The second is to more generally apply the scientific literature, especially psychological theory, to the issue.

3.1 Using social media data

Aside from the data available, for instance, from library databases, another potentially very interesting source of data about papers is the informal discussions happening in online platforms (e.g., Twitter, PubPeer). Discussion and interaction on these sites has increased rapidly in recent years and is likely to continue increasing, due to the high value they offer to users [6].

These data are available from the day the paper comes out, and contain many different types of information which should also be predictive of later impact (e.g., favorites, retweets, how many comments go to each mention, the content of each message). Unsurprisingly, researchers have examined how the simple count of social media citations relates to traditional citation metrics and found a modest but positive relationship [7,8].

The suggestion here, similar to Whalen et al., (2015) is to move toward also utilizing the actual words in each mention and response. Sticking close to Whalen et al., (2015), one could examine how much keyword space a target paper covers based upon the keywords from the network present in the discussion, or in the profiles of the researchers who take part in that discussion, or even in all of the content the researcher has posted to the site.

One initial step in this direction is our own research working toward an understanding of science communication by analyzing the sentiments expressed by the individuals engaging in these conversations [9, 10]. Using such methods, we can investigate established hypotheses from the psychological, developmental, communication, and learning literatures to better understand scientific communication and impact [11].

3.2 Applying psychology & theory

There are many hypotheses and established theories that can be applied to science in order to improve its efficiency [11]. The explanation of Whalen et al. (2015) that papers which are cited by more distant papers will have greater impact unnecessarily removes the (human) scientist from the equation. These are individuals with biases and tendencies that can be better understood to predict impact and citation practices.

One well established finding is that humans are motivated to reduce cognitive conflict (because they are motived to understand their environment) [12, 13]. Developmental research has found, for instance, that children are more interested in, play with, and learn more from toys that violate their expectations [14]; other research suggests that children who actively debate in school understand and can better explain material [15]. Similar findings from adults in online situations suggest positive relationships between: disagreement and the amount of negative words a person uses [9, 10]; discussion length, disagreement, and negative word usage in forums [9], and disagreement and the simple probability of responding [16]. It has also been shown that people read more, give more weight to, and rate as subjectively more useful negative comments and product reviews [17].

There is little reason such effects do not hold when discussing science, and even lead to some scientific work. From a general viewpoint, most scientists investigate questions which are not yet answered and it is the most conflicting questions that are researched for sustained periods of time (e.g., is evolution real?)

Theories of science [18, 19, 20] and many examples of successful science [21, 22] similarly suggest that cognitive conflict is good for science. For instance, Kuhn [18] suggests that the newest ideas are generally anomalous and generate much research as others try to fit these new findings into their existing paradigms. Similarly, Platt's strong inference [19] recommends pitting competing hypotheses against each other and Popper's falsificationism [20] tasks scientists with proving each other (and themselves) wrong.

These hypotheses can be tested, for instance, by examining whether those longer scientific discussions online are more negative (conceptually replicating [9]). Another examination could relate the sentiment in mentions of papers online. Another could examine, for instance, whether more popular keywords also contain more argumentative words. A nice example of this process is [24], the paper that basically started the open science movement and was the first PLoS paper to reach over 1,000,000 views (i.e., Ioannidis' Why most published research findings are false). It is our belief that using science to understand and fix the problems of science will lead to a better situation for all [11].

4. CONCLUSION

Whalen et al. (2015) make an excellent first endeavor toward utilizing more of the data available to inform our understanding of how science works and what good science is. In addition to the suggestions made on the explicit method of Whalen et al., (2015), we have suggested for discussion using more of the available data and utilizing more theory in framing hypotheses.

5. ACKNOWLEDGMENTS

The authors would especially like to thank the organizing committee for giving me this opportunity and Michael Schubert and Jürgen Buder for their helpful comments on earlier drafts.

6. REFERENCES

[1] Whalen, R., Huang, Y., Sawant, A., Uzzi, B. & Contractor, N. (2015). Natural language processing, article content & bibliometrics: Predicting high impact science. *Quantifying and Analyzing Scholarly Communication on the Web (ASCW'15)*

[2] Sendhilkumar, S., Elakkiya, E., & Mahalakshmi, G. S. (2013). Citation semantic based approaches to identify article quality. *In Proceedings of International Conference ICCSEA* (pp. 411-420).

[3] Teufel, S., Carletta, J., & Moens, M. (1999). An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics* (pp. 110-117).

[4] Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv preprint* arXiv:1203.4745.

[5] Van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature 514*(7524), 550-553.

[6] Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature, 512*(7513), 126-129.

[7] Costas, R., Zahedi, Z., & Wouters, P. (2014). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology.*

[8] Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research, 13*(4).

[9] Chmiel, A., Sobkowicz, P., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M., & Hołyst, J. A. (2011). Negative emotions boost user activity at BBC forum. *Physica A: Statistical Mechanics and its Applications, 390*(16), 2936-2944.

[10] Sobkowicz, P., & Sobkowicz, A. (2012). Two-year study of emotion and communication patterns in a highly polarized political discussion forum. *Social Science Computer Review*, 0894439312436512.

[11] Buttliere, B. T. (2014). Using science and psychology to improve the dissemination and evaluation of scientific work. *Frontiers in Computational Neuroscience, 8*.

[12] Piaget, J. (1954). *The Construction of Reality in the Child.* Routledge.

[13] Heidegger, M., Stambaugh, J., & Schmidt, D. J. (2010). *Being and Time*. SUNY Press.

[14] van Schijndel, T. J., Visser, I., van Bers, B. M., & Raijmakers, M. E. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology, 131*, 104-119.

[15] Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*(5977), 463-466.

[16] Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube videos: From Guatemalan rock to el big bang. *Journal of the American Society for Information Science and Technology*, *63*(3), 616-629.

[17] Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing, 21*(4), 76-94.

[18] Kuhn, T. S. (2012). *The Structure of Scientific Revolutions.* University of Chicago press.

[19] Platt, J. R. (1964). Strong inference. *Science, 146*(3642), 347-353.

[20] Popper, K. (2005). *The Logic of Scientific Discovery*. Routledge.

[21] Sternberg, R. J. (2004). Psychology 101 1/2: The unspoken rules for success in academia. *American Psychological Association.*

[22] Campanario, J. (2009). Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates. *Scientometrics, 81*(2), 549-565.

[23] Barber, B. (1961). Resistance by scientists to scientific discovery. *Science, 134.*

[24] Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 18*(4), 40-47.

Columns on Last Page Should Be Made As Close As Possible to Equal Length