# Terms and sentiments may apply: Examining sentiments in tweets and epistemic terms in scientific abstracts

Timothy D. Bowman* & Fereshteh Didegah
Research Unit for the Sociology of Education
University of Turku, Finland
*timothy.bowman@utu.fi

It is important to understand how scholars communicate across various online platforms as they are increasingly using these contexts to present themselves and to consume and disseminate various types of information. Bibliometrics, scientometrics, and altmetrics are fields dedicated to the examination of scholarly communication events including face-to-face encounters, within analog and digital research objects, and across platforms with hopes of learning how science is shared and referenced, how it might make an impact outside of academia, how scholars represent themselves in online environments, and how information is consumed and disseminated across large networks, to name just a few.

There are many methods used to examine scholarly communication, two of which are sentiment analysis and discourse analysis. Sentiment analysis examines the text of corpora in order to routinely examine changes of sentiment polarity (positive, neutral, negative) in the same document, in the same conversation, or across time. Sentiment analysis has been used to examine various online texts including the sentiment on stock market bulletin boards (Das & Chen, 2001), review websites (Dave, Lawrence, & Pennock, 2003), Facebook (Ortigosa, Martin, & Carro, 2014), YouTube (Thelwall, Sud, & Vis, 2012), TED Talks (Tsou, Thelwall, Mongeon, & Sugimoto, 2014), and Twitter (Friedrich, Bowman, Stock, & Haustein, 2015). Discourse analysis is an umbrella term for methods that examine written, spoken, or vocalized text. One recent branch of discourse analysis research is referred to as discourse epistemetrics (Demarest & Sugimoto, 2015) and examines differences in the use of epistemic terms across scientific disciplines. Recent work (Demarest, Larivière, & Sugimoto, 2015) in this area found that there are differences in words and phrases used in scientific abstracts across 13 disciplines. Each of these computational methods provides insight into the ways in which scholars communicate and present themselves to other academics and the society at large both similarly and differently by academic fields and disciplines.

For this work in progress, 165,746 tweets from 300 Twitter accounts belonging to American scholars from four disciplines in the social sciences and humanities (Philosophy (n=39), English (n=112), Anthropology (n=49), Sociology (n=100)) were harvested. The first phase of this work will include a sentiment analysis of the tweets using SentiStrength[1] by individual, department, gender, academic title, and real age. In the second phase, articles of the Twitter account holders will be identified from the Web of Science (WoS) using manual author name disambiguation and a machine-learning method for discourse epistemtetric analysis will be performed comparing the abstracts across disciplines to identify words and phrases commonly used in each discipline. The final phase will include harvesting the epistemic words and phrases and searching for these by discipline in the tweets to see if similarities exist between academic abstracts and tweets.

This work will be of value to those studying scholarly communication, altmetrics, and bibliometrics in that it will shed light on the differences in epistemic terms used across disciplines both in scientific abstracts and in tweets and point to various social norms and rules that may have developed in different contexts. In addition, this work will shed insight on how scholars present themselves across time on Twitter by tweeting (mostly) positively, negatively, or neutrally. Finally, this work will inform those examining altmetrics as it might provide new intuitions regarding the ways in which scholars might be able to detect scientific consumption and dissemination and to create new measures of societal impact. The results from this work will be important to a number of audiences including the scholars themselves, organizations funding research, university tenure and promotion evaluators, librarians, governments, and to members of the general public, as scholars are now asked to provide evidence of impact outside of the academy.

**References:**

Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific Finance Association annual conference (APFA)*. doi: 10.2139/ssrn.276189

---

[1] http://sentistrength.wlv.ac.uk/

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of World Wide Web* (pp. 519–528), Budapest, Hungary. doi: 10.1145/775152.775226

Demarest, B. and Sugimoto, C. R. (2015), Argue, observe, assess: Measuring disciplinary identities and differences through socio-epistemic discourse. *Journal of the Association for Information Science and Technology, 66*, 1374–1387. doi: 10.1002/asi.23271

Demarest, B., Larivière, V., and Sugimoto, C. R. (2015). Coming to Terms: A Discourse Epistemetrics Study of Article Abstracts from the Web of Science. In Salah, A.A., Y. Tonta, A.A. Akdag Salah, C. Sugimoto, U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*. Istanbul, Turkey: Bogaziçi University Printhouse. Retrieved from: http://www.issi2015.org/files/downloads/all-papers/1079.pdf

Friedrich, A., Bowman, T.D., Stock, M., & Haustein, S. (2015). Adapting sentiment analysis for tweets linking to scientific papers. In Salah, A.A., Y. Tonta, A.A. Akdag Salah, C. Sugimoto, U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*. Istanbul, Turkey: Bogaziçi University Printhouse. Retrieved from: http://www.issi2015.org/files/downloads/all-papers/0107.pdf

Ortigosa, A., Martin, J.M., & Carro, R.M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Beahvior. 31*, 527-571. doi: 10.1016/j.chb.2013.05.024

Tsou, A., Thelwall, M., Mongeon, P., & Sugimoto, C. (2014). A Community of Curious Souls: An Analysis of Commenting Behavior on TED Talks Videos. *PLoS ONE, 9*(4): e93609. doi: 10.1371/journal.pone.0093609

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. doi: 10.1002/asi.21679