

STI 2018 Leiden

*23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"*

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at <https://hdl.handle.net/1887/64521>

ISBN: 978-90-9031204-0

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Tweeting Library and Information Science: a socio-topical distance analysis¹

Philippe Mongeon^{*}, Shenmeng Xu^{**}, Timothy D. Bowman^{***} and Rodrigo Costas^{****}

^{*}*p.mongeon@cwts.leidenuniv.nl*

Centre for Science and Technology Studies, Leiden University, P.O. Box 905 2300 AX Leiden, The Netherlands

^{**}*shenmeng@email.unc.edu*

University of North Carolina at Chapel Hill, 216 Lenoir Drive, Chapel Hill, NC 27517, United States

^{***}*timothy.d.bowman@wayne.edu*

School of Information Sciences, Wayne State University, 106 Kresge Library, Detroit, MI 48202, United States.

^{****}*rcostas@cwts.leidenuniv.nl*

Centre for Science and Technology Studies, Leiden University, P.O. Box 905 2300 AX Leiden, The Netherlands
DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, South Africa.

Introduction

As a real-time microblogging network, Twitter has been studied extensively in various contexts, among which the mentioning and discussion of scholarly outputs on Twitter and how scholars use Twitter are two important topics. As governments and funding agencies are increasingly taking an interest in a broader view of impact (Dinsmore, Allen, & Dolby, 2014), altmetrics is valued due to its ability to show a fuller picture of research impact (Das & Mishra, 2014; Priem, Piwowar, & Hemminger, 2012; Priem, Taraborelli, Groth, & Neylon, 2010). On Twitter, the general presence of mentions of scholarly outputs is found to vary across different scientific disciplines (Costas, Zahedi, & Wouters, 2015a; Haustein, Bowman, Holmberg, Peters, & Larivière, 2014; Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014; Holmberg & Thelwall, 2014); indicating the existence of different thematic interests of research topics among Twitter users.

Scholars' use of Twitter shows some distinct patterns. For instance, scholars tend to share more links and retweet more than the average Twitter users (Holmberg & Thelwall, 2014). Across personal and professional tweets, affordance use on Twitter has been shown to vary based on department, gender, academic age, age, and Twitter activity (Bowman, 2015). In a study of Twitter user profiles, it was reported that users who tweet academic articles describe themselves more factually by emphasizing their occupational expertise (Vainio & Holmberg, 2017). Although most academic tweeters provide their full name and identity professionally in the profile descriptions (Bowman, 2015; Chretien et al., 2011; Hadgu & Jäschke, 2014), a large share of their activity is personal as opposed to professional (Bowman, 2015; Haustein et al., 2014; Van Noorden, 2014). When using Twitter for professional purposes, scholars

¹ This work was supported by the Social Science and Humanities Research Council of Canada and partially funded by the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP).

discuss research-related topics and communicate with others in the field (Van Noorden, 2014). Scholarly tweets tend to contain links to both recent journal articles (Eysenbach, 2011; Holmberg & Thelwall, 2014; Priem & Costello, 2010) and blogs (Letierce et al., 2010; Priem & Costello, 2010). However, the content of the tweets tends to be limited to the title, or part of the title of the scientific article being tweeted (Friedrich et al., 2015; Thelwall et al., 2013) and the level of engagement of Twitter users with publications is generally low (Robinson-García et al, 2017).

In any case, the use of Twitter has effects on the dissemination of research papers. According to Jose Luis Ortega (2016), articles authored by Twitter users are more tweeted than those of non-Twitter users. In addition, the number of followers on Twitter is found to indirectly influence the citation impact (Ortega, 2016). By investigating the referral data of the links of scholarly articles posted on Twitter, Wang and colleagues (2017) found that exposure of scholarly articles on Twitter increases their dissemination.

In this study we use the concepts of *topical distance* and *social distance* to analyse tweets of scientific papers, the tweeting behaviour of scholars, and the relationship between tweets and citations. Social distance is defined as the distance in a given social network between the tweeter and the authors of the tweeted content, while topical distance is the cognitive distance between the tweeted paper and the research area of the scholar who tweets it. Our analytical framework is thus a two-dimensional space where the position of a tweet is determined by its' social distance and topical distance coordinates.

Methods

All tweets between 2012 and 2016 containing a link (with a DOI) were collected using the altmetrics.com API. Only links to papers containing a DOI are included. We then identified the tweets that linked to a paper in the information science & library science Web of Science (WoS) subject category, using the WoS database hosted by the Centre for Science and Technology Studies (CWTS). We used a method developed by Costas, van Honk and Franssen (2017) to identify the Twitter accounts that belonged to a researcher with at least one publication indexed in WoS. The algorithm uses information on the authors extracted from bibliographic records of their publications and compares with information on the Twitter account obtained through the Twitter API. We could identify a total of 8,222 tweets made to 3,478 distinct papers by 3,770 distinct tweeters. In 754 cases, we found that the tweeted paper was also cited by the tweeter.

For each tweeter, we retrieved all articles published between 1980 and 2016 indexed in the WoS. These publications were obtained from the CWTS database in which authors are already disambiguated using the Caron and van Eck (2014) algorithm.

We calculate the topical distance in two ways: the cosine distance between the references list of each of tweeter's papers and the tweeted paper, and the cosine distance between noun phrases extracted from the title and abstract of each of the tweeter's papers and the tweeted paper. The topical distance between a tweeter and a tweeted paper is the minimum value of those two cosine distances, which ranges between 0 and 1. A value of 0 indicates that the tweeted publication is identical to one of the user's own publications, either in terms of references or in terms of noun phrases used in the title and abstract (it is most likely a self-tweet), and a value of 1 indicates that none of the references and noun phrases of the tweeted paper can be found in any the tweeter's own papers.

Like the topical distance, the social distance between the tweeter and the tweeted paper is the minimum value of the social distances between the tweeter and any of the authors of the tweeted paper. It is calculated in four steps:

1. We verify if the tweeter is an author of the tweeted paper. If so, the social distance is 0 (i.e., it is a self-tweet).
2. We create the co-authorship network of the tweeter (network A). The social distance is 1 if an author of the tweeted paper is found in network A.
3. We create the co-authorship networks of all the scholars in network A (network B). The social distance is 2 if an author of the tweeted paper is found in network B.
4. We create the co-authorship networks of all the authors of the tweeted paper (network C). The social distance is 3 if a scholar in network B is found in network C.
5. We create the co-authorship networks of all the scholars in network C (network D). The social distance is 4 if a scholar in network C appears in network D.

The social distance thus ranges from 0 to 5, where 0 indicates a self-tweet and 5 indicates that the shortest path between the tweeter and the authors of the tweeted paper in the global collaboration network is longer than 4. We stopped the process there because the number of nodes in the network tends to increase exponentially at each step, making the process computationally intensive.

Results

Figure 1 presents the distribution of the 8,222 tweets in terms of topical distance and social distance. We see that about one thousand tweets are self-tweets (they have a topical and a social distance of 0). The rest of the distributions show that the frequency of tweets increases as the topical or social distance increases, suggesting that most tweets are made by scholars that are not closely related neither to the research topic of the tweeted paper, nor to their authors. One should however note that these are raw counts and that the population of scholars (the potential tweeters) increases exponentially with the distance.

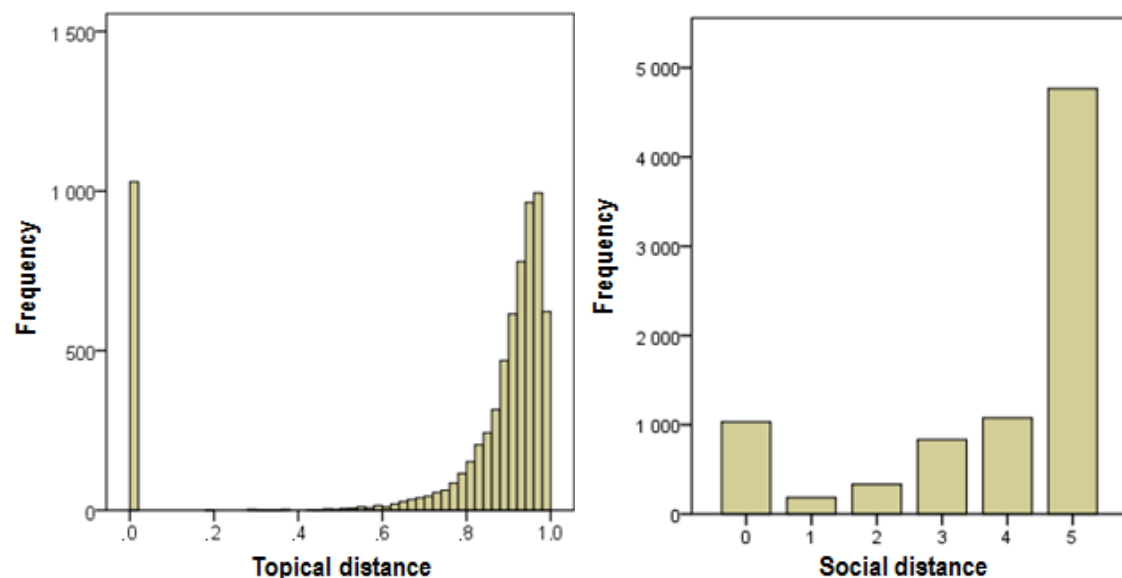


Figure 1. Distribution of the 8,222 tweets of LIS papers by topical distance (left) and social distance (right)

Figure 2 displays the 300 tweets of the most prolific tweeter (left) and the 112 tweets of the most tweeted paper (right) in our dataset. We see that the researcher on the left has tweeted papers from close collaborators (the few dots on the bottom right of the left-hand side graph), several papers that closely relate to his or her research topic (the dots on the top left of the left-hand side graph), as well as many papers that are both further away from both his or her research topic and collaboration network. In the right panel of Figure 2, we see that the most tweeted paper in our dataset was tweeted by at least one of the authors, and by a few of their collaborators, and then by many other scholars at higher social distance levels. This paper thus seemingly was both the object of self-tweeting but also of interest to other scholars in the field, as well as others in the broader research community.

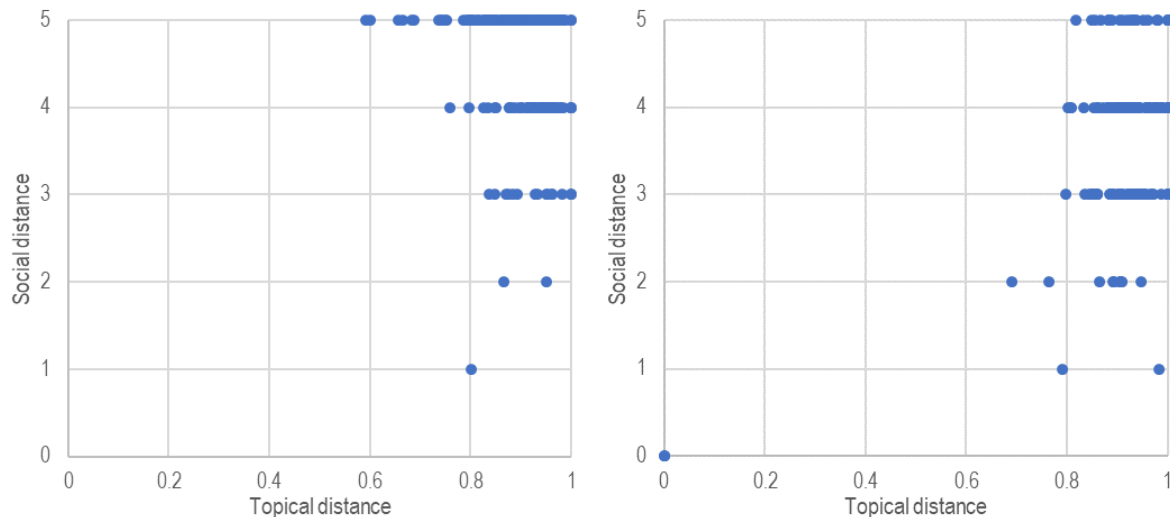


Figure 2. Topical and social distance of the tweets of most prolific tweeter (left) and the most tweeted papers (right)

To show how concepts of topical and social distance can be used in combination to characterise the Twitter activity of scholars and/or papers, figure 3 shows the average topical distance and social distance for the tweets of the 3,770 scholars (left) and the 3,478 LIS papers (right). We see that both for tweeters and papers, topical and social distance appear to be correlated. This makes sense considering that the closer two scholars are in a collaboration network, the more they are likely to be working on similar research topics. Beyond this, we can see that some scholars only tweet their own papers or papers of their close collaborators, while other tweet about papers that are far away from their collaboration network and research topics. We also see that some scholars also tend to tweet the work of their colleagues even though it doesn't necessarily relate to their own work, while others tend to tweet the papers that relate to their research topic even if they do not have collaboration ties with the authors. The same interpretation could be made for papers (right panel of figure 2).

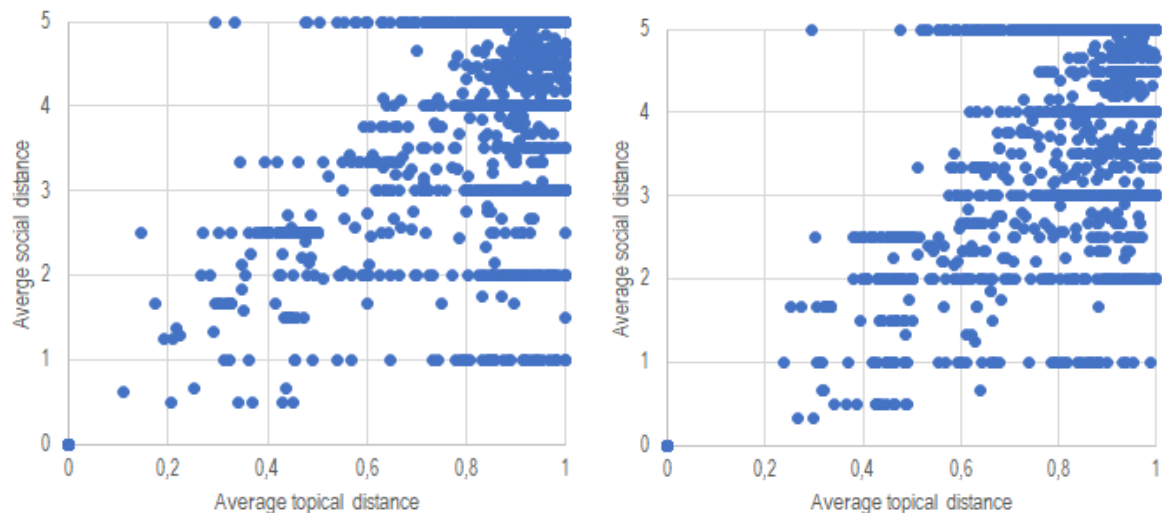


Figure 3. Average topical and social distance of tweeters (left) and the tweeted papers (right)

Finally, we show in table 1 the Spearman correlation between the different variables to see which one correlates the most with citations. Here we separated the topical distance measured by the cosine distance of the noun phrases in the title and abstracts (TDnp), the topical distance calculated with the cosine distance of the cited references (TDref), and their combination (TD). as expected, citations are inversely correlated with all measures of distance. However, it is more strongly correlated with topical distance than with social distance (the stronger correlation overall being with TDref).

Table 1. Spearman correlations between variables

	TDnp	TDref	TD	SD	Cited
TDnp	1.000	0.57	0.987	0.466	-0.383
TDref	0.57	1.000	0.611	0.543	-0.517
TD	0.987	0.611	1.000	0.475	-0.397
SD	0.466	0.543	0.475	1.000	-0.289
Cited	-0.383	-0.517	-0.397	-0.289	1.000

Discussion and Conclusion

The aim of this paper was to demonstrate how topical distance and social distance can provide meaningful results when analysing scholars' tweets linking to scholarly publications. By considering the publication profile of the tweeting authors, we can distinguish between potentially different tweeting reasons and motivations. Furthermore, we have shown that whether a tweet will eventually lead to a citation by the tweeter partly depends on the extent to which the tweeted paper is related to the tweeter's social network and research topic.

From a conceptual point of view, this research can be framed in the more general perspective of the *social media studies of science* (Costas, 2017) in which the interactions between scholarly entities (in this case, individual LIS scholars) with scientific publications becomes the main focus, thus moving away from the more simplistic consideration of social media metrics as mere counts of social media acts around scholarly outputs. In further research we will further improve the operationalization of the concepts of social and topical distance, use the model to analyse larger datasets and investigate disciplinary differences. Further research

may also use topical and social distance to create new typologies of social media use by academics or expand existing ones.

References

- Bowman, T. D. (2015). Differences in personal and professional tweets of scholars. *Aslib Journal of Information Management*, 67(3), 356–371.
- Chretien, K., Azar, J., & Kind, T. (2011). Physicians on Twitter. *Journal of the American Medical Association*, 305(6), 566.
- Costas, R. (2017). Towards the social media studies of science: social media metrics, present and future. *Bibliotecas. Anales de Investigación*, 13(1), 1–5.
- Costas, R., Zahedi, Z., & Wouters, P. (2015a). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective: Do “Altmetrics“ Correlate With Citations? *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019.
- Das, A. K., & Mishra, S. (2014). Genesis of Altmetrics or Article-Level Metrics for Measuring Efficacy of Scholarly Communications: Current Perspectives. *SSRN Electronic Journal*.
- Dinsmore, A., Allen, L., & Dolby, K. (2014). Alternative Perspectives on Impact: The Potential of ALMs and Altmetrics to Inform Funders about Research Impact. *PLoS Biology*, 12(11), e1002003.
- 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), e123.
- Friedrich, N., Bowman, T. D., Stock, W. G., & Haustein, S. (2015). Adapting sentiment analysis for tweets linking to scientific papers. (arXiv e-print No.1507.01967). Retrieved March 30th, 2018 from <https://arxiv.org/abs/1507.01967>.
- Hadgu, A. T., & Jäschke, R. (2014, June). Identifying and analyzing researchers on twitter. *Proceedings of the 2014 ACM conference on Web science* (pp. 23-32). ACM.
- Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter: Tweets as Impact Indicators: Examining the Implications of Automated “bot” Accounts on Twitter. *Journal of the Association for Information Science and Technology*, 67(1), 232–238.
- Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature: Tweeting Biomedicine: An Analysis of Tweets and Citations in the Biomedical Literature. *Journal of the Association for Information Science and Technology*, 65(4), 656–669.
- Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, 101(2), 1027–1042.

Letierce, J., Passant, A., Breslin, J. G., & Decker, S. (2010, May). Using Twitter During an Academic Conference: The# iswc2009 Use-Case. In ICWSM.

Luis Ortega, J. (2016). To be or not to be on Twitter, and its relationship with the tweeting and citation of research papers. *Scientometrics*, 109(2), 1353–1364.

Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the Association for Information Science and Technology*, 47(1), 1-4.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact (arXiv e-print No. 1203.4745). Retrieved March 30, 2018 from <http://arxiv.org/abs/1203.4745>.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto.

Vainio, J., & Holmberg, K. (2017). Highly tweeted science articles: who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics*, 112(1), 345–366.

Robinson-Garcia, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLoS ONE*, 12(8). <http://doi.org/10.1371/journal.pone.0183551>

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

Wang, X., Cui, Y., Li, Q., & Guo, X. (2017). Social media attention increases article visits: An investigation on article-level referral data of PeerJ. *Frontiers in Research Metrics and Analytics*, 2, 11.