

Identifying and Analyzing Researchers on Twitter

Asmelash Teka Hadgu and Robert Jäschke

L3S Research Center, Appelstraße 4, 30167 Hannover, Germany
teka@L3S.de, jaeschke@L3S.de

Abstract Twitter is a communication platform, a social network, and a system for resource sharing. For scientists, it offers an opportunity to connect with other researchers, announce calls for papers and the like, communicate and discuss – basically: stay up-to-date. However, the exponential growth of information in society does not exclude social media like Twitter: an abundant number of users court on one’s attention which leads to the question of how (young) researchers can focus on the essential users and tweets?

The classical approach in science to filter information is peer review: only information that is considered to be novel, sound, and significant by experts in the respective field is published. Currently, such a process is at most implemented manually: researchers can subscribe individually to other researcher’s feeds by following them. However, there is no ‘directory’ of scientists on Twitter and finding feeds of experts in a specific discipline or area of interest is cumbersome.

Furthermore, the trend to consider visibility of scientific articles in the social web as a possible (and immediate) alternative or complement to citation counts (with services like Altmetric¹ that provide counts for how often a scientific article has been mentioned on Twitter and other social networks) necessitates the need for peer-review-like mechanisms for the social web. Simple approaches purely based on the popularity of users, tweets, or URLs do not work as a tool for scientists to discover relevant research(ers), since popularity on the social web is fundamentally a matter of the crowd of non-scientists. Articles that are popularized by the media – often independent of their scientific significance – get superior attention compared to other, more important works. Consider the Ig Nobel Prizes,² whose winning (scientific) publications get quite some attention on the social web, e.g., the URL³ of the winner of the 2012 physics prize has been mentioned in more than 230 tweets.⁴ Enabling

Copyright © 2014 by the paper’s authors. Copying permitted only for private and academic purposes. In: T. Seidl, M. Hassani, C. Beecks (Eds.): Proceedings of the LWA 2014 Workshops: KDML, IR, FGWM, Aachen, Germany, 8-10 September 2014, published at <http://ceur-ws.org/>

¹ <http://www.altmetric.com/>

² <http://www.improbable.com/ig/>

³ <http://prl.aps.org/abstract/PRL/v108/i7/e078101>

⁴ <http://topsy.com/trackback?url=http%3A%2F%2Fprl.aps.org%2Fabstract%2FPRL%2Fv108%2Fi7%2Fe078101>

users (and in particular researchers) to access the scientists' perspective in the social web and considering only tweets from physicists would provide a different and likely better picture.

Existing Twitter directories like Wefollow⁵ rely on users' initiative to register and reveal their interests. This clearly limits the set of available profiles, since professionals have limited time and there is no immediate benefit for registration. Therefore, providing an automatically curated directory of scientists would simplify expert finding and the provision of topic-relevant feeds authored by peers. This approach requires to first identify scientists on Twitter and then classify their discipline, topics of interest, and expertise. Since only little is known about scientists on Twitter, such an endeavor should be accompanied by further steps to understand how Twitter is used by them.

In this work we present an approach for the identification and classification of scientists on Twitter together with an empirical analysis of researchers from computer science found on Twitter. We take a pragmatic approach on which users we regard as 'scientists': users being interested in the topics of the target discipline and having similar, Twitter-based features like users that have published scientific papers. We start with a list of seeds that are highly-relevant for the discipline of interest and use it to build and augment a set of candidate users that are likely scientists. For a subset of the candidates that we can match to ground-truth data from a digital library, we build a model for the classification of scientists. We can show that the model is very accurate and use it to classify all of our candidates. Both sets of users (matched and classified) allow us to perform an empirical analysis of scientists on Twitter.

The main contributions of this work are

- a complete framework for discipline-specific researcher classification on Twitter using a small set of seeds only,
- an automatic approach for the generation of ground-truth data by combining different data sources,
- an empirical analysis of computer scientists that are using Twitter, and
- the provision of the used datasets.⁶

The results were published in A.T. Hadgu and R. Jäschke: Identifying and analyzing researchers on Twitter. In *Proceedings of the 2014 ACM Conference on Web Science*, WebSci '14, pages 23–30, New York, NY, USA, 2014. ACM. DOI:10.1145/2615569.2615676

⁵ <http://wefollow.com/>

⁶ <https://github.com/L3S/twitter-researcher>