Signature Trajectories in the Multiplatform Diffusion of Science

Keywords: Science of Science, Science on the Web, Multiplatform Diffusion, Science Communication, Science Diffusion.

Extended Abstract

Science communication has undergone dramatic changes over the past decades. According to large-scale surveys, 75 to 80% of researchers use social media to disseminate or promote their work (Sugimoto et al., 2017). Despite the broad adoption of online dissemination platforms by scholars, empirical literature on how science is shared online is nascent (Brossard & Scheufele, 2022). In particular, the complementary role of multiple platforms in determining the visibility of scientific content is understudied. We contribute to this literature by examining the sharing trajectories of academic articles in multiple platforms (social media, news, blogs, and knowledge repositories like Wikipedia). We are interested in understanding how online attention to academic papers differs on these platforms and what common trends emerge in the multiplatform visibility of papers.

We analyze online attention to scientific articles using a subset of Altmetric's dataset, which contains records of 23,450 of the most popular publications mentioned in roughly 20 million posts. Our analysis requires that the papers have online attention, therefore we don't include papers with less than 50 mentions. The dataset includes information relating to the article: title, author, journal, research area, publication date, and timestamped mentions on various platforms including news media, blogs, social media (e.g., Twitter, Facebook, and Reddit), and knowledge repositories. Figure 1 shows how this data can be represented as a time series of mentions on different platforms. The series start at the time of the first mention of an article in any of the platforms and contain the number of mentions every 12 hours. The figure presents the average number of mentions for all the papers in the dataset. We augmented this large-scale dataset with information about scientists from the Web of Science and OpenAlex (Priem et al., 2022). Specifically, we collect yearly citation numbers for both the authors and the papers and fine-grained scientific fields.

Our initial goal is to cluster the time series of multiplatform mentions to identify signature trajectories of attention. To do so, we preprocess the time series of mentions for each paper with minmax normalization and apply k-means clustering (MacQueen, 1967). After selecting the number of clusters using the elbow method, we find shared trends in the mention trajectories of papers. Furthermore, these trends are consistent for subsets of papers attracting different levels of attention (e.g., between 50 and 100 mentions in total, between 200 and 300 mentions in total, or over 600 mentions in total). Figure 2 showcases 4 signature trajectories for the first 5 days after publication for a subset of 17,165 popular papers with over 200 mentions in total. An important difference between these trajectories is the number of platforms involved, clusters 2 and 4 are papers with mainly social media attention whilst papers in cluster 1 also received notable attention in the news and papers in cluster 3 also received substantial attention in news and blogs. We also see that most papers get limited attention on knowledge repositories. These signatures also show that attention concentrates in short intervals of time rather than being evenly distributed. This observation is in line with previous research showing burstiness in the attention to academic papers (Zakhlebin & Horvát, 2020). Furthermore, the peaks of attention on different platforms occur concurrently. Finally, the key differentiator

amongst the papers with negligible coverage in news and blogs (clusters 2 and 4) is the timing of their peak of attention. The barycenters show that papers in cluster 2 exhaust their attention quickly after their first mention, on the other hand attention to papers in cluster 4 peaks considerably later.

To understand what characteristics of a paper determine which cluster their trajectories belong to, we formulated a multiclass classification problem. The characteristics included in the classification model are the field of research, number of institutions involved, number of countries/continents represented by the authors, open access status of the paper, and citation metrics of the authors. Preliminary results show that the open access status and the number of regions represented are significant indicators of whether the paper will belong to a cluster with only social media attention. Papers with open access and a larger number of continents represented are more likely to attract attention from multiple platforms. Previous research has discussed an open access advantage: papers with open access have more downloads, citations, and social media mentions (Wang et al., 2015). Additionally, we now find that open access papers receive attention from multiple platforms, which has recently also been found to indicate higher information retention (Hwang et al., 2022)

Our work advances a better understanding of the reception of academic papers online. Next, we will expand our analysis to different time scales and types of academic papers to better establish the universality of the identified mentioned trajectories. A deeper understanding of shared patterns in these trajectories is an essential building block of understanding which research breakthroughs reach broad audiences and shape tomorrow's scientific advances.

References

Brossard, D., & Scheufele, D. A. (2022). The chronic growing pains of communicating science online. *Science*, *375*(6581), 613–614. https://doi.org/10.1126/science.abo0668

Hwang, S., Horvát, E.-Á., & Romero, D. M. (2022). *Information Retention in the Multi-platform Sharing of Science* (arXiv:2207.13815). arXiv. http://arxiv.org/abs/2207.13815

MacQueen, J. (1967, June). Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability* (pp. 281-297). Los Angeles LA USA: University of California.

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (arXiv:2205.01833). arXiv. http://arxiv.org/abs/2205.01833

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062. https://doi.org/10.1002/asi.23833

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics*, *103*(2), 555–564. https://doi.org/10.1007/s11192-015-1547-0

Zakhlebin, I., & Horvát, E.-Á. (2020). Diffusion of Scientific Articles across Online Platforms. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 762–773.

Figures

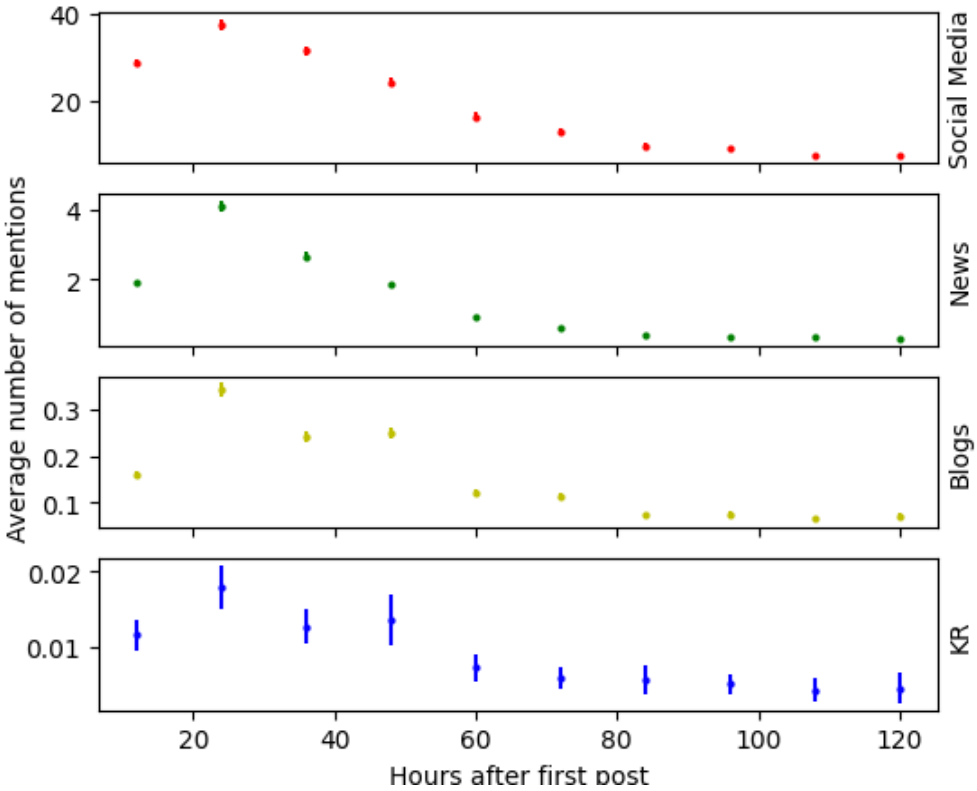

Figure 1 Time series of the average number of mentions in different platforms (Social Media, News, Blogs, and Knowledge Repositories) for the analyzed papers. Error bars denote a confidence interval of 95%.

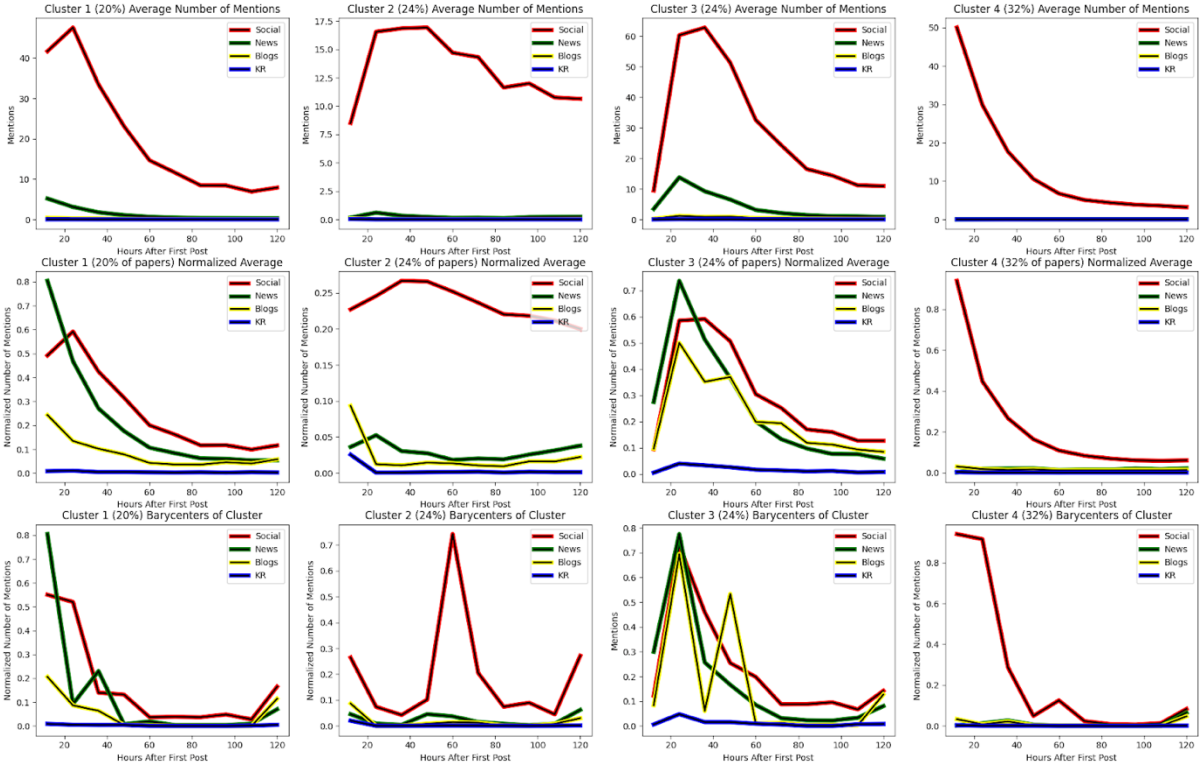

Figure 2 Time series of the number of mentions of papers according to the clustering classification given by the K-means algorithm, the number in parenthesis represents the percentage of papers assigned to that cluster. Top row shows the average number of mentions, the middle row shows the average of the normalized number of mentions of the papers in each cluster and the bottom row shows the barycenters obtained for each cluster.