ture more context around the diffusion process and help identify the *types* of discussions associated with high information retention. However, topic modeling at scale also poses challenges such as the need for manual inspection to label topics, especially with so many scientific fields. Such methods also tend to be noisy when applied to short documents like social media posts. For this study, we focus—as an initial step—on overall cross-platform information retention independent of the details of the surrounding context in which papers are mentioned. A natural next step for future work building on our contributions includes examining such context and its relationship with information retention. Finally, we do not make empirical claims to causality. While we identify clear differences across platforms and trends across them, we cannot say, for example, that an increase in the number of platforms involved in a sequence *leads to* higher information retention. In part, this is because our sequences are not necessarily information cascades transferring information from one to another, but simply temporally ordered. However, this study does not set out to disentangle particular effects, such as the "telephone effect" (**?**), but instead provides an overview of patterns over time in a multi-platform information landscape. Exciting future research directions include parsing out the mechanisms by which platforms affect information retention. One potential such project includes a systematic analysis of the kinds of keywords found in mentions at various points in a sequence, which may not only articulate particular platform effects but also elucidate why the presence of multiple platforms is linked with higher information fidelity overall.

Ethics Statement

A serious consideration for the presentation of any quantitative measure is how it may incorporate biases or black-boxes of complex human concepts. As a difficult construct to quantify, information retention is a par excellence example of this, as noted in our discussion of the validation survey. In order to prevent our proposed quantification from becoming misapplied, we selected the most scrutable version of it, such that interpretation and limitations of what the measure suggests are clear. This simplicity also helps undermine unreasonable extrapolations in arguments about the *quality of information retention* of content when applied beyond this work. Additionally, research utilizing posts from sites containing content from individuals, such as social media and/or blog users, who may not be aware that their content is used for research also encounter important concerns about the integrity of users' privacy. In our work, we consider only public-facing content accessible by any arbitrary individual. In addition, we only use posts that were still available and not deleted by the users at the time of our study. Finally, our results describe user behavior only in aggregate and at a scale that leaves individual posts unidentifiable. While minimizing potential risks, the expected benefits of our contributions to the understanding of information retention in the diffusion of scientific findings across platforms are substantial and foundational for a better appreciation of intentional and unintentional information distortion online. Our findings not only point to general patterns of information retention that might inform media strategies, such as deliberate dissemination across multiple channels, but also are generative in raising potential mechanisms for improving the fidelity of information in online discussions to be tested and examined for causality in future studies.

Conclusion

As scientific findings spread on the web, they are discussed across multiple platforms, shaping what information is retained. Accurately communicating science online has critical implications for policymakers, researchers, and the public alike, but the difficulty in multi-platform data collection has made it extremely challenging to unpack how crucial information is retained in a multi-platform information landscape. In this study, we utilized a large-scale observational dataset that leverages unique identifiers of scientific work (DOIs) to track content across different platforms, and examined information retention in bursts of attention to scientific articles over time. Our study offers three main contributions. First, we provide a view of how online discussions of scientific findings lose information "in the wild", showing a strong propensity for low information retention. This underscores an important need to devise strategies to mitigate such loss. Second, to this end, we show that scientific articles discussed on more platforms tend to have higher information retention. This suggests that *multi*-platform discussions may help improve information retention and highlights future directions to untangle the mechanisms driving this trend. More broadly, this dynamic also highlights that multi-platform work is critical to understanding how online activity inherently shapes societally-relevant information. Finally, we provide a simple, scrutable measure that can reasonably evaluate information retention at scale and a burst-based framework for applying it to study diffusion in science and beyond. Along with our findings, the measure and framework lay the foundations for further work evaluating the quality and fidelity of information for various types of online content. In a time with ongoing debates about what is factual, understanding how information communicated on the web changes as it spreads over time and across platforms is a pressing societal challenge.

Acknowledgments

The authors would like to thank students Amanda Hardy and Joshua Jacobs for their invaluable assistance in data work, Hao Peng for generously sharing code, the Community Data Science Collective for earlier feedback, and reviewers for their thoughtful comments that strengthened this manuscript. This work was supported by the NSF (DGE1842165, IIS-2133963, IIS-2133964, and CAREER Award IIS-1943506) and by the Air Force Office of Scientific Research under award number FA9550-19-1-0029.

Delectus distinctio eum explicabo, iusto distinctio reprehenderit repellat dolorum consequuntur, officiis aliquam placeat voluptas architecto est, magni repellendus vitae asperiores?Quidem excepturi tenetur enim vitae maiores cumque culpa, corporis omnis neque assumenda inventore labore adipisci eligendi, magni sunt architecto optio vero er-

ror, dignissimos esse eius officia debitis quaerat et officiis distinctio, iste ipsam quae