Qeios

Peer Review

Review of: "Quantifying Hot Topic Dynamics in Scientific Literature: An Information-Theoretical Approach"

Jonathan Scaccia[1]

1. American Institutes for Research, Washington, D.C., United States

This article introduces a new way to explore how the meanings of concepts shift over time in scientific literature, using a mix of mathematical and information-theoretic methods. The centerpiece of the approach is the use of normalized Variation of Information (NVI) to build yearly networks of keywords and measure how relationships between concepts change within a specific topic area. The method is tested on a substantial dataset of 10,370 articles on international security from 2010 to 2023, with a special emphasis on cybersecurity.

The article has several strengths. The use of NVI offers clear advantages over more traditional similarity measures, especially in its ability to capture nuanced semantic shifts. The framework also shows strong potential for use across disciplines like policy research, the digital humanities, and network science. The analysis is timely and relevant, showing how research agendas respond to major world events. It would be really interesting to see how this plays out in other domains as well.

That said, there are a few areas where the manuscript could be improved. The method is currently geared toward narrow, predefined topics, and it's not clear how well it would scale to broader areas like biomedical research. The way concepts are selected feels a bit manual and based mostly on frequency thresholds; it would be better if this step were more automated. The paper would also benefit from a direct comparison with other topic modeling approaches, especially around topic coherence or predictive power. The author claims this method is more efficient than dynamic topic models, but no runtime or resource benchmarks are provided to support that.

The analysis is also limited to yearly time windows. It would be helpful to explore whether different time intervals (like twice per year) might affect the detection of semantic shifts. This could impact the sensitivity of the results, especially in fast-moving fields.

doi.org/10.32388/SYT74E

To strengthen the paper, the author should add a short section comparing their method to existing topic modeling techniques and include performance metrics to back up claims about efficiency. It would also help to explain how the method handles evolving vocabulary, ambiguous terms, and new concepts. And expanding the discussion to show how this approach could be applied to other areas like public health or climate policy would make the work even more compelling.

One more important point: the paper doesn't address the lag in scientific publishing. Research is often conducted years before it appears in published literature. That delay could blur the timing of observed conceptual shifts. Trends that seem to emerge at a certain point might actually reflect earlier developments. This lag could also hide emerging ideas that are active in the research community but haven't yet made it to publication, resulting in a more conservative picture of topic dynamics than what's actually happening on the ground.

Finally, I don't use Mathematica myself, but the shared script seems to make sense. It looks like it supports the visualizations used in the paper and could be a helpful tool for other researchers trying to dig into the data.

Declarations

Potential competing interests: No potential competing interests to declare.