Qeios

Peer Review

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

Prajit Das[1]

1. Security Business Group, Cisco Systems, Inc, San Jose, United States

This paper presents a methodological framework for analyzing conceptual evolution within scientific literature. It introduces an approach that warrants consideration for publication in this journal under the field/topic of social sciences.

Arguments for Publication:

1. Methodological Contribution:

* The paper proposes a framework utilizing the Normalized Variation of Information (NVI) to quantify conceptual distances. This addresses a recognized limitation in some existing co-occurrence network methods by introducing metric properties, which can facilitate more rigorous topological and temporal comparisons.

* The concept of "velocity matrices" is introduced to measure the rate of change in conceptual relationships over time, providing a dynamic perspective on how topics evolve.

* The work positions itself as a complementary tool to established methods like Dynamic Topic Models (DTM) and Neural Topic Models (NTM), specifically focusing on the internal dynamics of predefined conceptual vocabularies.

2. Relevance to Social Sciences:

* The application of the framework to topics within International Security and International Relations demonstrates its utility for addressing questions pertinent to social science research.

* The case study on "cybersecurity" illustrates the method's capacity to correlate conceptual shifts with external geopolitical, technological, and regulatory events, which is relevant for understanding the interplay between discourse and real-world developments in social and political contexts.

doi.org/10.32388/DGPFYA

3. Interpretability and Detail:

* The NVI-based approach provides a measure of conceptual similarity that considers the shared and unique information between concept distributions, offering a nuanced view of semantic relationships.

* The method is presented as enabling a fine-grained analysis of conceptual relationships within a defined topic.

* Visualizations, such as bubble-flow graphs, are employed to aid in the interpretation of temporal patterns of conceptual change.

4. Consideration of Methodological Aspects:

* The paper discusses the limitations of current approaches and how the NVI framework aims to provide an alternative.

* The Supporting Information section addresses potential influences of dataset size on distance measures, indicating an awareness of methodological considerations.

Areas for Further Consideration:

To enhance the paper's contribution and provide a more comprehensive understanding of its capabilities and limitations, the following points could be considered:

1. Specific Examples of Unique Insights from the NVI-based Method:

* The paper suggests that NVI captures "contextual information and shared uncertainty between concept distributions" beyond mere co-occurrence frequency. To further illustrate this, providing specific examples of insights uniquely revealed by the NVI-based method would be beneficial. For instance, how might the NVI distance between "cybersecurity" and "privacy" reveal a nuanced semantic shift that a traditional co-occurrence or topic model might not capture with the same clarity or precision? This could involve scenarios where:

* Subtle Semantic Drift: Two concepts might consistently co-occur, but NVI could detect a shift in the *pattern* of their co-occurrence across different document types or sub-contexts, indicating a change in their underlying semantic relationship that simple frequency counts would miss.

* Information Overlap vs. Co-occurrence: NVI quantifies the information shared, which is a richer measure than simple co-occurrence. It could identify concepts that, while not frequently co-occurring, share significant information due to their distributional properties, revealing implicit connections.

* Enabling Velocity Analysis: The ability to calculate "velocity matrices" is a direct consequence of NVI's metric properties. This allows for the identification of "fast-moving" concepts or periods of rapid

conceptual realignment, which might be harder to discern from other methods focused on topic prevalence or word-topic distributions.

2. Impact of Parameter Variations on Study Conclusions:

* The study's methodology for defining a "topic" involves specific parameters: `k` (minimum frequency for the seed concept), `n%` (minimum document occurrence percentage for candidate concepts), and `b` (NVI distance cutoff). Variations in these parameters could influence the study's conclusions in several ways:

* `k` (Seed Concept Frequency Threshold): A higher `k` would narrow the initial document set (`D`) to only those where the seed concept is highly prominent. This could lead to a more specialized, but potentially less comprehensive, conceptual network, potentially missing broader or more diffuse conceptual shifts. Conversely, a lower `k` might introduce more noise into the document set, potentially diluting the core topic and affecting the precision of NVI distance calculations.

* `n%` (Candidate Concept Document Occurrence Threshold): Increasing `n%` would restrict the analysis to only the most frequently co-occurring concepts within the selected documents. This might result in a sparser network, highlighting only dominant relationships but potentially overlooking emerging or less frequent, yet significant, conceptual connections. A lower `n%` would include more peripheral concepts, potentially increasing network density and complexity, which could make the identification of clear patterns more challenging.

* `b` (NVI Distance Cutoff): A stricter (lower) `b` would limit the topic to concepts very closely related to the seed, providing a highly coherent but potentially narrow view of the conceptual landscape. A looser (higher) `b` would expand the topic to include more distantly related concepts, potentially revealing broader thematic realignments but also introducing less meaningful connections.

* The choice of these parameters directly affects the composition and boundaries of the conceptual network, which in turn influences the calculated NVI distances, the velocity matrices, and ultimately the interpretation of conceptual convergence, divergence, and the identification of "localized hubs." A discussion on the sensitivity of the findings to these parameter choices would enhance the robustness of the conclusions.

Overall, I would state that this publication is worth accepting into the journal and will very much benefit scholars of the domain.

Declarations

Potential competing interests: No potential competing interests to declare.