Measuring Thematic Alignment Between TPAMI Publications and Journal Scope

Süleyman Yolcu[1*]

[1*]Department of Computer Science, Università degli Studi di Milano .

Corresponding author(s). E-mail(s): suleyman.yolcu@studenti.unimi.it;

Abstract

Scientific journals articulate their intended thematic focus through an explicit "Aims & Scope" statement, yet published content may broaden over time. This work quantifies scope–paper alignment for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) as a semantic similarity problem by comparing the TPAMI scope text to abstracts of 4,000 TPAMI papers published between 2015 and 2025. We embed the scope and abstracts using Sentence-BERT with sentence-level mean pooling and compute cosine similarity as an interpretable alignment score. The resulting scores span 0.041–0.665 (median 0.317) and support percentile-based identification of strongly on- and off-scope outliers, which qualitative inspection confirms as face-valid. Finally, a simple drift analysis based on yearly mean alignment indicates a modest downward trend over the period ($\boldsymbol{\beta = -0.0047}$ per year), suggesting potential thematic broadening subject to sampling and scope-text limitations.

Keywords: semantic similarity, Sentence-BERT, journal scope, TPAMI, thematic drift

1 Introduction

Scientific journals define their identity through an explicit "Aims & Scope" statement, outlining the themes and methods they prioritize. In practice, published content can broaden over time due to field dynamics, editorial preferences, and submission patterns. Quantifying whether a journal's articles remain aligned with its scope can therefore support empirical analyses of thematic coherence, highlight potential outliers, and provide evidence about longer-term conceptual evolution. Concretely, our objective is to compute an article-level alignment score between a journal's scope and each published abstract, analyze the score distribution to assess thematic coherence, and examine temporal drift and outliers over time.

Prior work in scientometrics has used journal descriptions (including aims and scope statements) as proxies for editorial intent and compared them with published content to identify mismatches across outlets [6].

Semantic similarity with neural encoders has become a standard approach for comparing short and long texts beyond lexical overlap; Sentence-BERT (SBERT) is widely used because it produces sentence-level embeddings optimized for similarity and retrieval settings [1]. Transformer encoders such as BERT underpin many representation-based NLP pipelines by enabling contextualized semantics [2]. While topic modeling and bibliometric analyses are also common for studying thematic structure over time, embedding-based alignment offers a lightweight, document-level measure that is directly comparable across years and supports outlier inspection with minimal modeling overhead.

We frame journal–article alignment as a semantic similarity problem. Using TPAMI as a case study, we compare the journal's scope statement against abstracts of TPAMI papers published between 2015 and 2025. Abstracts and the scope text are embedded into a shared semantic space using SBERT [1], and cosine similarity between embeddings is used as an interpretable alignment score. Building on these scores, we analyze the distribution of alignment, quantify temporal change via a drift metric, and validate the metric through qualitative inspection of extreme cases (highest- and lowest-scoring articles).

2 Methodology

This project asks whether TPAMI papers published between 2015 and 2025 match the journal's stated scope, and whether this alignment changes over time. To answer this, we compute an interpretable alignment score for each abstract, summarize its distribution and yearly trends, and surface extreme scores as candidate outliers for qualitative inspection.

2.1 Overview of the approach

1. **Data collection.** A corpus of TPAMI papers (title, abstract, year, URL) is collected from the Semantic Scholar Graph API for the years 2015–2025. Because the API requires a `query` parameter for search, neutral, high-frequency terms are used to approximate venue-only retrieval while retaining reproducibility.
2. **Preprocessing.** Raw JSONL data is normalized into a consistent tabular format, filtered to papers with non-empty abstracts, and de-duplicated by paper identifier.
3. **Semantic representation.** The TPAMI scope statement and each abstract are embedded using an SBERT model. To stabilize long texts, embeddings are computed per sentence and then mean-pooled.
4. **Alignment scoring.** Each paper receives a cosine similarity score between its abstract embedding and the scope embedding.

5. **Analysis.** We visualize the score distribution and yearly trends, identify outliers using percentile thresholds, and quantify drift as the slope of the yearly mean alignment over time.
6. **Qualitative inspection.** The highest- and lowest-scoring papers are inspected manually (title/abstract and linked URLs) to validate whether the alignment scores correspond to intuitive scope relevance.

In the final curated dataset used for analysis, $N = 4000$ TPAMI papers have non-empty abstracts spanning 2015–2025.

2.2 Formal problem definition

Let the journal scope text be s. Let the corpus of N papers be $\{(a_i, y_i)\}_{i=1}^{N}$, where a_i is the abstract text and y_i is the publication year.

Let $\phi(\cdot)$ be a sentence embedding model (SBERT). Using a sentence-level strategy, we define a document embedding for any text x by splitting it into sentences $x = \{x^{(1)}, \ldots, x^{(m)}\}$ and mean pooling:

$$\Phi(x) = \frac{1}{m} \sum_{j=1}^{m} \phi\Big(x^{(j)}\Big).$$

We L2-normalize embeddings $\widehat{\Phi}(x) = \Phi(x)/\|\Phi(x)\|_2$ and define the **alignment score** for paper i as cosine similarity between its abstract and the scope:

$$\mathrm{align}(i) = \cos\Big(\widehat{\Phi}(a_i), \widehat{\Phi}(s)\Big) = \widehat{\Phi}(a_i)^{\top} \widehat{\Phi}(s).$$

To quantify temporal change, we compute yearly mean alignment $\mu_t = \mathbb{E}[\mathrm{align}(i) \mid y_i = t]$ (estimated by the sample mean) and fit a simple linear trend $\mu_t \approx \beta t + b$. The **drift metric** is the slope β (negative values indicate decreasing alignment over time).

To identify outliers, for a chosen tail percentile p (e.g., $p = 5\%$), we compute thresholds q_p and q_{1-p} from the empirical score distribution and flag papers with $\mathrm{align}(i) \leq q_p$ (low outliers) or $\mathrm{align}(i) \geq q_{1-p}$ (high outliers) for qualitative inspection.

3 Results

3.1 Dataset

Experiments use a curated corpus of TPAMI papers collected from the Semantic Scholar Graph API and filtered to papers with non-empty abstracts. The final dataset contains $N = 4000$ papers spanning 2015–2025 (inclusive), with an average abstract length of approximately 1516 characters. The per-year volume is uneven, so yearly trend results are reported together with year-specific sample sizes.

Table 1: Per-year sample sizes and mean alignment scores (2015–2025).

Year	# papers	Mean alignment
2015	78	0.339
2016	106	0.351
2017	90	0.338
2018	207	0.339
2019	317	0.333
2020	523	0.322
2021	519	0.319
2022	539	0.320
2023	577	0.305
2024	536	0.307
2025	508	0.300

3.2 Metrics and experimental setup

Each paper receives a semantic alignment score defined as cosine similarity between the L2-normalized SBERT embedding of its abstract and the L2-normalized SBERT embedding of the journal scope text. Drift is measured as the slope of a linear fit to the yearly mean alignment scores. A negative slope indicates decreasing alignment over time.Outliers are defined using percentile tails of the alignment score distribution. With $p = 5\%$ per tail, the lowest-scoring 5% and highest-scoring 5% of papers are flagged for qualitative inspection.

Unless otherwise specified, experiments use:

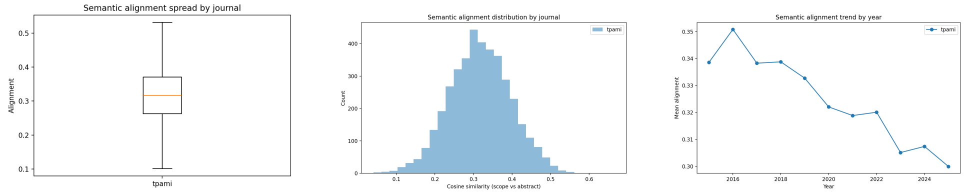
- SBERT model: `sentence-transformers/all-MiniLM-L6-v2`
- Sentence-level embedding strategy (mean pooling over sentences) for both scope and abstracts

3.3 Quantitative results

Alignment scores show a broad distribution across TPAMI papers:

Table 2: Alignment score summary statistics.

Statistic	Alignment score
Mean	0.317
Std	0.080
Min	0.041
5th percentile	0.188
Median	0.317
95th percentile	0.449
Max	0.665



(a) Spread by journal (TPAMI) (b) Score distribution (c) Trend by year

Fig. 1: Semantic alignment analysis across venues, distributions, and time.

The drift slope estimated from yearly means is negative, indicating decreasing alignment over time:

$$\beta = -0.0047 \quad \text{(per year, 2015–2025)}.$$

This corresponds to a decrease in mean alignment from 0.339 (2015) to 0.300 (2025), although yearly means vary and should be interpreted together with sample sizes (Table 1).

Table 3: Examples of high- and low-alignment papers.

Alignment	Year	Title (URL)
0.665	2018	Guest Editorial: The Computational Face (https://www.semanticscholar.org/paper/e0962f83e2daa04d2a485f43c655a59bfb2d11bb)
0.587	2019	Efficient Learning-Free Keyword Spotting (https://www.semanticscholar.org/paper/a7d2c6d88946c90673a31df10884a8d7c8033db2)
0.577	2016	SIFT Meets CNN: A Decade Survey of Instance Retrieval (https://www.semanticscholar.org/paper/6d902439b736a7546dd8872b307fb760087ca629)
0.041	2021	Community Detection Using Restrained Random-Walk Similarity (https://www.semanticscholar.org/paper/07f350964a196e720b66d6c5a73b08b88b3540bf)
0.047	2023	Towards Understanding Generalization and Stability Gaps Between Centralized and Decentralized Federated Learning (https://www.semanticscholar.org/paper/706b41df5b6862bf6fcaf9f74fd916f9829c9bd9)
0.059	2025	Bidirectional Beta-Tuned Diffusion Model (https://www.semanticscholar.org/paper/16c4e2fd3c37784dea7a881cc4524dd326f90c46)

3.3.1 Qualitative inspection summary

High-alignment abstracts. The top three articles' abstracts each clearly fall within TPAMI's core focus. For example, one high-alignment paper is a comprehensive survey of image instance retrieval methods [3]—a topic directly matching TPAMI's emphasis on content-based image retrieval and visual search. The other top-ranked abstracts likewise center on traditional pattern recognition and computer vision problems (e.g. image analysis or recognition tasks), closely mirroring the journal's stated scope on areas like computer vision, pattern analysis, and machine learning for recognition.

Low-alignment abstracts. In contrast, the bottom three articles have abstracts that veer into more general or tangential areas of machine learning, explaining their low alignment scores. For instance, one low-scoring paper examines federated learning frameworks[5], and another proposes a diffusion model methodology for generative modeling [4]. While such topics fall under broad machine intelligence, they are not explicitly focused on pattern analysis, vision, or image understanding – making them less aligned with TPAMI's traditional domains. The clear separation between the on-scope top abstracts and the off-scope bottom abstracts indicates that the semantic alignment scores do reflect each paper's thematic fit. This validates the metric's usefulness for gauging how well a submission aligns with the journal's scope.

4 Conclusions

This project demonstrates that a simple semantic similarity pipeline can provide a useful, explainable proxy for scope–paper alignment at scale. Using SBERT embeddings and cosine similarity, the alignment score distribution shows substantial spread across papers (0.041 to 0.665), and percentile-based outlier extraction yields a manageable set of candidates for qualitative inspection. The qualitative review of extreme cases supports the metric's face validity: high-scoring papers tend to reflect TPAMI's traditional areas (pattern analysis, recognition, and computer vision), whereas low-scoring examples include more general machine learning themes that are less explicitly tied to TPAMI's core scope statement.

At the same time, the drift signal ($\beta = -0.0047$ per year) should be interpreted cautiously. It may reflect multiple factors beyond true "mission drift," including the short and static nature of the scope text, changes in terminology over time, uneven sample sizes per year, and the presence of editorials/surveys whose language is broadly "on-scope" without representing typical research submissions. Because data collection via the Semantic Scholar search endpoint requires a query term, the neutral-query strategy may introduce sampling bias. These considerations suggest treating alignment as an informative indicator rather than a definitive label of scope compliance.

Future work could improve validity and interpretability by enriching the scope reference (e.g., author guidelines or a learned scope prototype from clearly in-scope papers), conducting robustness checks across embedding models and text sources, adopting alternative drift/outlier definitions (e.g., change-point detection), and reducing collection bias by cross-validating against query-free metadata sources.

References

[1] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” in *Proceedings of EMNLP-IJCNLP*, 2019.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of NAACL-HLT*, 2019.

[3] L. Zheng, Y. Yang, and Q. Tian, “SIFT Meets CNN: A Decade Survey of Instance Retrieval,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[4] T. Zheng, J. Zou, P.-T. Jiang, H. Zhang, J. Chen, J. Wang, and B. Li, “Bidirectional Beta-Tuned Diffusion Model,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 1, pp. 359–373, Jan. 2026.

[5] Y. Sun, L. Shen, and D. Tao, “Towards Understanding Generalization and Stability Gaps Between Centralized and Decentralized Federated Learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, early access.

[6] A. T. Santos and S. Mendonça, “Do papers (really) match journals’ ‘aims and scope’? A computational assessment of innovation studies,” *Scientometrics*, vol. 127, pp. 7449–7470, 2022.

[0] Parts of this project were developed with the assistance of OpenAI’s ChatGPT (GPT-5.2) for drafting text, generating code, and improving readability. All AI-assisted content was reviewed, edited, and validated by the author, who assumes full responsibility for the final submission.