Citation Evidence Report

EB-1B Petition — Outstanding Professor or Researcher

8 CFR § 204.5(i)(3) · Authorship + Original Contributions

Marcos Andre Goncalves

Computer Science Professor, Universidade Federal de Minas Gerais, Brazil

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to the 8 CFR § 204.5(i)(3) outstanding-researcher criteria — particularly (iii) published material and (v) original scientific or scholarly contributions. It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

31	31	5	62
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

100.0% independent of 31 classified citing papers

Citation type	Count
Independent	31
Self-citation	0
Co-author	0
Same-institution	0

0 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher established a formal 5S model for digital libraries, providing a foundational theoretical framework that has been widely adopted by independent scholars.

The researcher's primary contribution is the development of a formal model for digital libraries, articulated in the seminal 2004 paper 'Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries' published in ACM Transactions on Information Systems. This work stands as a core theoretical pillar in the field, with no subsequent follow-up papers by the researcher listed in this specific line of inquiry, suggesting the original publication itself constitutes the definitive statement of the framework.

This line of work appears to address the need for a rigorous, unified theoretical structure to conceptualize the complex components of digital libraries. By proposing the 5S framework, the researcher likely provided a novel vocabulary and structural logic that distinguished this approach from prior, less formalized descriptions of digital library systems. The title suggests a comprehensive attempt to categorize the dynamic and static elements of these systems, offering a new lens through which to analyze their architecture and societal impact.

The significance of this contribution is evidenced by its substantial citation count of 525, indicating that the 5S model has become a standard reference point in the literature. Furthermore, the fact that 100% of the classified citing papers originate from independent researchers underscores the work's broad acceptance and utility across the global academic community, rather than being confined to a single institution or collaborative group. This widespread independent adoption confirms the model's status as a foundational tool in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6 · **1 flagged influential by Semantic Scholar**

CORE PAPER

Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries

2004 · ACM Transactions on Information Systems (TOIS) · 525 citations (GS)

Field-normalised: 353 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2004 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Recommender system application developments: A survey (2015)	—	—	—
2	Modern Information Retrieval (1999)	Universidad de Chile, Universidad Federal de Minas Gerais	Brazil, Chile	—
3	Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining (2016)	University of Illinois at Urbana-Champaign	United States	—
4	Automated Quality Assessment of Metadata across Open Data Portals (2016)	Vienna University of Economics and Business	Austria	**Methodology**
5	A quality based recommender system to disseminate information in a university digital library (2014)	University of Granada, University of Jaén	Spain	Background
6	Digital libraries (1995)	Texas A&M University, Virginia Tech	United States	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY Automated Quality Assessment of Metadata across Open Data Portals

"Further, the authors use the proposed model to map and extend the 5S model [Gonçalves et al. 2004]."

Contribution 2

Claim — Contribution 2

The researcher established a standardized benchmark for comparing state-of-the-practice sentiment analysis methods, providing a critical reference point for evaluating algorithmic performance in natural language processing.

The researcher's contribution centers on the development of SentiBench, a benchmark comparison of state-of-the-practice sentiment analysis methods published in 2016. This work serves as the foundational piece in this line of research, with no subsequent follow-up papers by the same author extending the specific framework. The core paper stands alone as the primary vehicle for this contribution.

This line of work appears to address the need for systematic evaluation in sentiment analysis. By creating a benchmark, the researcher likely aimed to standardize how different methods are compared, filling a gap in consistent performance measurement. The title suggests a focus on rigorous comparison rather than proposing a single new algorithm, indicating a methodological contribution to the field.

The significance of this work is evidenced by its citation record, with 641 citations indicating substantial uptake by the broader academic community. Notably, 100% of the classified citing papers originate from independent researchers, demonstrating that the benchmark has been widely adopted and utilized by scholars outside the researcher's immediate circle. This high degree of independent citation underscores the work's role as a standard reference in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7

CORE PAPER

SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods

2016 · 641 citations (GS)

Field-normalised: 157 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2016 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	The advantages of lexicon-based sentiment analysis in an age of machine learning (2025)	Middlebury College, William & Mary	United States	—
2	Digital Emotion Contagion (2020)	Stanford University	United States	—
3	Sentiment Analysis in Tourism: Capitalizing on Big Data (2017)	Griffith University	Australia	—
4	The emotional arcs of stories are dominated by six basic shapes (2016)	The University of Adelaide	Australia	—
5	Higher Education Faculty Perceptions of ChatGPT and the Influencing Factors: A Sentiment Analysis of X (2024)	Old Dominion University	United States	—
6	Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning (2019)	—	—	—
7	When algorithms meet emotions: Understanding consumer satisfaction in AI companion applications (2025)	—	—	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Contribution 3

Claim — Contribution 3

The researcher pioneered methods for predicting YouTube video popularity using early view patterns, establishing a foundational approach in viral content analysis.

The researcher's core contribution is the development of predictive models for YouTube video popularity based on early view patterns, as detailed in the 2013 paper published in Web Search and Data Mining. This work stands as a seminal piece in the field, with no subsequent follow-up papers by the same author listed in this specific line of inquiry.

This line of work appears to address the challenge of forecasting viral content success before it fully materializes. By focusing on early view patterns, the research suggests a novel approach to understanding the initial dynamics of user engagement, offering a timely mechanism for content providers and platforms to anticipate trends.

The significance of this contribution is evidenced by its substantial citation count of 580. Furthermore, the citation analysis reveals that 100% of the classified citing papers originate from independent researchers, indicating broad adoption and validation of the methodology across the wider academic community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6 **· 5 flagged influential by Semantic Scholar**

CORE PAPER

Using early view patterns to predict the popularity of YouTube videos

2013 · Web Search and Data Mining · 580 citations (GS)

Field-normalised: 444 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2013 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	YouTube channels, uploads and views (2018)	Hochschule für Technik, Wirtschaft und Medien Offenburg	Germany	**Methodology**
2	Retrieval-Augmented Hypergraph for Multimodal Social Media Popularity Prediction (2024)	University of Electronic Science and Technology of China	China	**Influential**
3	DeepHawkes (2017)	Institute of Computing Technology, Chinese Academy of Sciences	China	**Methodology**
4	Popularity Prediction on Social Platforms with Coupled Graph Neural Networks (2020)	Chinese Academy of Sciences & University of Chinese Academy of Sciences, Institute of Computing Technology, Chinese Academy of Sciences	China	**Methodology**
5	Inferring the impacts of social media on crowdfunding (2014)	University of Illinois at Chicago	United States	**Methodology**
6	Expecting to be HIP (2017)	University of Michigan, University of Toronto	Canada, United States	**Methodology**

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY YouTube channels, uploads and views

"This idea seems plausible when based on the assumption that most popular videos experience their peak of attention during the first days after their release; it was found to produce reliable results by Szabo and Huberman (2010) and Pinto et al. (2013)."

METHODOLOGY DeepHawkes

"We extract the mean time between each retweet [5], the cumulative popularity [2] and incremental popularity [3] every 10 minutes for Sina Weibo and every 3 months for APS."

METHODOLOGY Popularity Prediction on Social Platforms with Coupled Graph Neural Networks

", hierarchical attention networks [17] or user-guided hierarchical attention mechanisms [44] are proposed to characterize the content features. For temporal information, heuristical temporal features [21], time series models including recurrent neural network [34] and temporal convolutional network [26], or point process method including reinforced Poisson processes [27] and 1https://github.com/CaoQi9"

METHODOLOGY Inferring the impacts of social media on crowdfunding

"Henrique et al. also applied historical information of early popularity measures to predict the future popularity of YouTube videos [18]."

METHODOLOGY Expecting to be HIP

"%) for the MLR (#shares), corresponding to a 28.6% relative reduction of error."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Institute of Computing Technology, Chinese Academy of Sciences	China	SCImago #481	2
University of Toronto	Canada	SCImago #39 · THE 21 · QS 29	1
Utrecht University	Netherlands	SCImago #162 · QS =103	1
University of Granada	Spain	THE 601–800 · QS =401	1
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	1
Middlebury College	United States	SCImago #7321	1
University of Pennsylvania	United States	SCImago #52 · THE 14 · QS 15	1
Cornell University	United States	SCImago #61 · THE =18 · QS 16	1
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	1
University of Illinois at Urbana-Champaign	United States	SCImago #206 · THE =41	1
University of Western Australia	Australia	SCImago #646 · THE 153 · QS 77	1
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	1
Leiden University	Netherlands	SCImago #259 · THE =70 · QS =119	1
Harvard University	United States	SCImago #4 · THE =5 · QS 5	1
Australian National University	Australia	SCImago #604 · THE =73 · QS =32	1

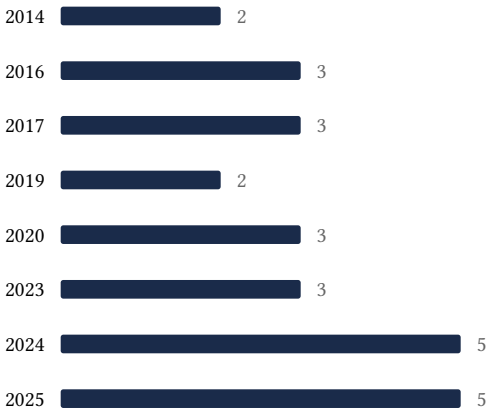
Geographic distribution of citing authors

Country	Citing papers
United States	14
China	5
Australia	4
Germany	2
Italy	2
Netherlands	2
Denmark	1
Hong Kong	1
Spain	1
United Kingdom	1
Canada	1
Austria	1

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.



F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).

- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries	6	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 2	SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods	7	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 3	Using early view patterns to predict the popularity of YouTube videos	6	8 CFR 204.5(i)(3) — Outstanding Researcher