Medieval Citation Networks as Digital Hyperlinks: Transformer-Based Authorship Attribution in Historical Text Collections

Jonathan Schler*
Holon Institute of Technology (HIT)
Holon, Israel

Nati Ben-Gigi*
Binyamin Katzoff*
Maayan Geffet-Tamir*
Bar-Ilan Univerity
Ramat-Gan, Israel

Abstract

Digital libraries containing historical manuscripts face persistent challenges in authorship attribution, particularly for anonymous or misattributed texts where traditional bibliographic metadata is incomplete or disputed. We demonstrate that citation networks in medieval texts function as primitive hyperlink structures, creating navigable knowledge graphs that encode stable authorial signatures across centuries-old document collections. Our transformer-based framework leverages three complementary components: (i) a BERT-CRF deep learning pipeline achieving accuracy of F1 ≈ 0.90 in automatically extracting references from medieval Hebrew and Aramaic texts, (ii) cosine similarity analysis of citation frequency vectors that capture each author's unique "citation fingerprint," and (iii) network-based indicators quantifying cross-community influence patterns in historical corpora. Applied to a corpus of 62.5 million tokens spanning the rabbinic literature of the 10th-15th century, our system successfully extracted more than 230,000 references and constructed comprehensive citation networks. We validate the approach through a contested attribution case: commentary on Tractate Bava Metzia attributed to the medieval scholar "Ritva." Our analysis reveals distinct citation profiles between the attributed text and verified Ritva works (cosine similarity: 0.32), confirming scholarly suspicions of multiple authorship. The methodology identifies Rabbi Shem Tov ibn Gaon as the likely author of disputed sections (similarity: 0.959), corroborated by historical evidence. This work positions medieval citation practices as precursors to modern web hyperlink structures, demonstrating how transformer-based NLP can unlock authorship information embedded in historical reference networks. The language-agnostic methodology offers digital libraries scalable tools for automated manuscript attribution, applicable beyond medieval texts to any citation-rich historical corpus. Our approach bridges computational hypertext analysis with traditional humanities scholarship, providing new pathways for AI-enhanced organization and discovery in digital manuscript collections.

Keywords

digital hyperlinks, network analsyis, authorship attribution, citation network, machine learning, text processing,historical networks

ACM Reference Format:
Jonathan Schler, Nati Ben-Gigi, Binyamin Katzoff, and Maayan Geffet-Tamir. 2025. Medieval Citation Networks as Digital Hyperlinks: Transformer-Based Authorship Attribution in Historical Text Collections. In *Proceedings of Web Archiving and Digital Libraries (WADL) Workshop 2025 (HT).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Long before Berners-Lee conceived the World Wide Web in 1989, medieval scholars had already established sophisticated hyperlinked knowledge networks through systematic citation practices [9]. These historical reference patterns, embedded in manuscripts spanning centuries, represent early examples of linked information structures—networks that connected ideas, authorities, and texts across vast temporal and geographical distances, prefiguring modern hypertext systems.

Digital libraries today confront significant challenges in organizing and attributing historical manuscript collections, particularly in cases involving anonymous or disputed authorship where traditional bibliographic metadata proves insufficient [3]. The problem becomes especially acute in specialized corpora such as medieval religious literature, where centuries of copying, compilation, and mis-attribution have obscured original authorship patterns. Recent scholarship has demonstrated that computational approaches can address these attribution challenges through novel methodologies that leverage citation patterns as stable authorial signatures [16, 23].

Contemporary advances in natural language processing and network analysis offer new pathways for addressing these attribution challenges. Citation networks in historical texts function as computational hyperlink structures, encoding stable authorial signatures that persist across an author's corpus [2]. Unlike traditional stylometric approaches that rely on linguistic features [11, 21], citation-based methods leverage the intellectual fingerprint embedded in how authors reference earlier authorities—patterns that remain remarkably consistent across different works and topics.

The methodological foundation for this approach builds upon established research in authorship attribution. Machine learning algorithms have been successfully applied to create writing profiles based on stylistic features such as word frequency, grammatical characteristics, and syntactic patterns [1, 10]. However, these methods face limitations when applied to historical texts, especially

*All authors contributed equally to this research.

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

in morphologically rich languages like Hebrew, where standard NLP resources are limited [22]. Citation-based approaches offer a complementary methodology that can overcome these linguistic barriers while providing domain-agnostic solutions.

Building on three complementary studies from our research program, this paper presents a transformer-based framework that treats medieval citation practices as precursors to modern web hyperlinks. Our approach integrates: (i) automatic reference extraction using a BERT-CRF pipeline trained on Rabbinic Hebrew achieving a F1 accuracy of 0.90 [3], (ii) cosine similarity analysis of citation frequency vectors for authorship verification, and (iii) network-based diversity indicators that quantify influence patterns across historical scholarly communities [2].

The methodology addresses a fundamental challenge in digital humanities: developing scalable computational approaches for authorship analysis in large historical text collections where traditional attribution methods fail. Applied to contested medieval commentaries, our citation fingerprint approach successfully identifies distinct authorial signatures, confirming scholarly hypotheses about disputed texts and proposing specific alternative attributions with quantifiable confidence measures.

This work contributes to the intersection of hypertext research and digital libraries by demonstrating how historical citation networks prefigure modern web structures. The language-agnostic methodology offers scalable tools for automated manuscript attribution, extending beyond medieval texts to any citation-rich historical corpus. By positioning citation analysis as a form of ancient hypertext navigation, we bridge computational approaches with traditional humanities scholarship, providing new frameworks for AI-enhanced organization and discovery in digital manuscript collections.

2 Related Work

Citation network analysis has emerged as a prominent methodology in digital humanities, enabling researchers to map intellectual relationships and trace knowledge transmission across historical corpora. Early work by [15] demonstrated the application of citation analysis to classical texts, utilizing natural language processing techniques to extract canonical citations and study intertextuality, while [4] constructed citation networks among recent monographs on Venetian history, revealing disciplinary clusters and identifying influential works. Recent advances by [8] survey comprehensive approaches to citation network analysis in digital humanities, highlighting the growing sophistication of computational methods for historical text analysis. In the context of Rabbinic literature specifically, [23] developed semi-automatic approaches to generate networks mapping relationships of Jewish sages across generations, utilizing lexical and syntactic patterns to identify names and relationships within Halachic debates. [16] applied quantitative social network analysis to the Babylonian Talmud, revealing dense core networks formed through relationships among influential rabbis and uncovering insights about historical connections and transmission pathways. These studies demonstrate the particular richness of citation networks in religious scholarly traditions, where systematic referencing practices create dense intellectual webs spanning centuries.

Computational authorship attribution represents a well-established field with methodologies ranging from traditional stylometric analysis to modern machine learning approaches. [11] provide a comprehensive survey of computational methods, highlighting the effectiveness of features such as word frequency, character n-grams, and syntactic patterns in distinguishing authorial signatures, while [6] extends this analysis to modern attribution methods, emphasizing the importance of function words and stylistic markers. However, traditional approaches face significant limitations when applied to historical texts, especially in morphologically rich languages like Hebrew, where standard NLP resources often prove insufficient [22]. Citation-based approaches offer complementary methodologies that can overcome these linguistic barriers. [7] proposed methods for analyzing citation patterns in academic papers to determine authorship, achieving success rates between 30-50% through citation profile matching and self-citation pattern identification, while [19] combined citation analysis with traditional stylometric methods, achieving approximately 60% success rates. In Rabbinic literature specifically, [5] developed supervised machine learning models to identify citations in Hebrew-Aramaic documents, and [12] leveraged these systems to analyze influence networks in Jewish Responsa literature. Recent transformer-based advances [18, 20] have enhanced performance for historical document processing, with [17] introducing BEREL for Rabbinic Hebrew analysis, demonstrating that citation patterns remain remarkably stable across an author's corpus [2].

3 Methodology

Our approach to citation-based authorship attribution consists of three integrated components: (i) automated reference extraction from historical texts using transformer-based deep learning, (ii) construction of citation fingerprint vectors that capture authorial reference patterns, and (iii) similarity analysis using network-based measures to identify authorial signatures. This section details each component and demonstrates how they combine to create a scalable framework for manuscript attribution in digital libraries.

3.1 Automated Citation Extraction Pipeline

The foundation of our methodology relies on accurate extraction of citations from medieval Hebrew and Aramaic texts, a task complicated by the morphological richness of these languages and the non-standardized citation practices of historical authors [3]. We developed a multi-layered system that decomposes the complex reference extraction task into manageable subtasks.

Our corpus consists of medieval Rabbinic literature spanning the 10th-15th centuries, comprising over 62.5 million tokens from approximately 120 authors across six geographic regions, with preprocessing addressing orthographic inconsistencies, abbreviations, and morphological variations through a comprehensive thesaurus containing 240 authors' names and 280 book titles [2].

3.1.1 BERT-CRF Architecture for Reference Identification. Following recent advances in transformer-based language models for historical texts, we employ a BERT-CRF architecture specifically adapted for Rabbinic Hebrew [17]. The model performs two sequential classification tasks that address the structural complexity of references in medieval texts.

The first task identifies reference boundaries within continuous text, addressing the challenge that references often appear in sequences without clear punctuation markers. This boundary detection model tags words that separate consecutive or recursive references, utilizing dedicated labels for different separation types. For recursive references—where one citation contains another citation—the model employs specialized tags to distinguish nested citation structures.

The second task performs component identification within detected reference boundaries, classifying words as author names, book names, reference terms like "chapter" or "page", and other citation elements. Training data consists of manually annotated references from a representative subset of the corpus: 3,301 references with 20,477 named entities for component identification, and 4,744 references with 23,184 entities for boundary detection [3].

The BERT-CRF models achieve strong performance with component identification attaining F1 score of 0.896 and boundary detection achieving F1 score of 0.856. When integrated into the complete pipeline including name normalization and validation steps, the system achieves overall precision of 0.896, recall of 0.905, and F1 score of 0.901 [3].

3.2 Citation Fingerprint Construction

The core innovation of our approach lies in treating citation patterns as stable authorial signatures that can be quantitatively compared across texts and authors. We formalize this concept through citation fingerprint vectors that capture the frequency distribution of an author's references to earlier authorities.

For each author in the corpus, we construct a citation vector $\mathbf{v}_a$ where each dimension corresponds to a cited authority from the medieval period. Formally, if $A = \{a_1, a_2, \ldots, a_n\}$ represents the set of all cited authorities in the corpus, then:

$$\mathbf{v}_a = [c_{a,1}, c_{a,2}, \ldots, c_{a,n}]$$

where $c_{a,i}$ represents the number of times author a cites authority a_i. This representation captures both the diversity of sources an author draws upon and the relative frequency with which they reference different authorities.

A critical challenge involves accounting for significant variations in corpus size across authors. To address this issue, we employ cosine similarity for vector comparison, which measures the angle between vectors rather than their magnitude:

$$\text{similarity}(\mathbf{v}_a, \mathbf{v}_b) = \frac{\mathbf{v}_a \cdot \mathbf{v}_b}{|\mathbf{v}_a| \cdot |\mathbf{v}_b|}$$

This metric ranges from 0 to 1, where values close to 1 indicate similar citation patterns regardless of the absolute number of citations.

3.3 Network-Based Similarity Analysis

Building upon established methods in bibliometric analysis, we extend citation fingerprint comparison to network-level analysis that considers individual authorial patterns and community-wide citation structures [2].

Before applying citation fingerprints to attribution problems, we validated the stability of citation patterns within known authors' corpora. For established authors with multiple works, we divided their complete works into random sections and computed similarity between the resulting citation profiles. Results demonstrate remarkable consistency: the Ramban's commentary sections achieve 0.99 similarity, while other major authors (Rashba, Ritva, Maharam Chalava) consistently score above 0.98 [2]. This validation confirms that citation patterns represent stable authorial characteristics that persist across different works and topics.

3.4 Authorship Attribution Protocol

The complete attribution methodology integrates citation extraction, fingerprint construction, and similarity analysis into a systematic protocol for evaluating disputed attributions. For a text of unknown or disputed authorship, we: (1) extract all citations using the BERT-CRF pipeline, (2) construct a citation fingerprint vector for the disputed text, (3) compare this vector against established citation profiles of candidate authors using cosine similarity, (4) rank potential attributions by similarity scores, and (5) validate results through content analysis and historical plausibility assessment. We require a minimum of 530 citations in disputed texts to ensure statistical reliability (10 time as many data-points as vector dimensions), with similarity scores above 0.85 indicating possible authorial correspondence and scores below 0.4 suggesting different authorship.

4 Case Study: The Ritva Attribution Problem

To demonstrate the effectiveness of our citation-based authorship attribution methodology, we apply it to a well-known disputed attribution in medieval Rabbinic literature: the commentary on Tractate Bava Metzia attributed to Rabbi Yom Tov of Seville (known by the acronym "Ritva"). This case study illustrates how computational analysis can resolve centuries-old scholarly debates while providing quantifiable evidence for attribution decisions.

4.1 Historical Context and Scholarly Debate

The commentary on Bava Metzia attributed to the Ritva presents a classic authorship attribution problem. Two distinct commentaries circulate under the Ritva's name: the *Hiddushei HaRitva* (accepted as authentic) and a commentary printed in Amsterdam in 1729 whose attribution has been questioned by scholars since the 18th century.

The disputed commentary was first partially printed in the responsa of Maharam Galanti (Venice, 1608) covering folios up to 12b, then published in its entirety in Amsterdam in 1729. However, prominent 18th-century scholars including Maharit Algazi and the Hida challenged this attribution based on: (i) discrepancies with Ritva quotations in the 16th-century compilation *Shitah Mekubetzet*, and (ii) differences in writing style and citation patterns compared to authenticated Ritva works.

Rabbi Halpern's analysis [14] proposed that the Amsterdam commentary comprises two distinct parts written by different authors: the first part (folios 1-11) by a student of the Rashba, possibly an earlier version of the Ritva's own commentary, and the second part (folios 12-end) by unknown author(s), possibly scholars from Provence. Rabbi Lichtenstein [13] subsequently argued that the

first part was authored by Rabbi Kreskes Vidal, also a student of the Rashba.

4.2 Computational Analysis

We applied our citation fingerprint methodology to test these scholarly hypotheses and propose alternative attributions based on quantitative evidence.

4.2.1 Baseline Citation Profile Construction. We first established a baseline citation profile for the authentic Ritva by analyzing all accepted works in our corpus: his commentaries on multiple Talmudic tractates, *Hilkhot Berakhot*, and responsa. This comprehensive profile, constructed from 847 distinct citations, represents the Ritva's characteristic pattern of referencing earlier authorities. To validate the stability of this profile, we compared citation patterns across different authentic Ritva works. The average similarity between individual tractate commentaries and the complete Ritva profile was $\mu = 0.82$ ($\sigma = 0.11$), confirming consistent citation behavior across his authenticated corpus.

4.2.2 Attribution Testing of the Disputed Commentary. Initial analysis of the complete disputed commentary revealed a striking divergence from the authentic Ritva profile. The similarity score of 0.32 falls well below our threshold for positive attribution (0.85), strongly suggesting different authorship. This quantitative result supports the scholarly consensus that the Amsterdam commentary was not written by the Ritva.

4.2.3 Two-Part Analysis. Following Halpern's hypothesis of composite authorship, we divided the disputed commentary at folio 12 and analyzed each section independently.

First Part Analysis (Folios 1-12): The citation profile of the first section yielded a similarity score of 0.87 when compared to the authentic Ritva works, indicating substantial correspondence. However, comparison with other contemporary authors revealed an even stronger match: Rabbi Kreskes Vidal achieved a similarity score of 0.91. While both scores fall within the range of positive attribution, the slightly stronger correspondence with Vidal supports Lichtenstein's attribution hypothesis.

Second Part Analysis (Folios 12-end): The second section showed minimal similarity to the authentic Ritva (0.17), confirming that this portion was definitely not authored by him. Systematic comparison against all 120 authors in our corpus identified the most likely alternative attribution.

4.3 Novel Attribution Discovery

Our computational analysis identified Rabbi Shem Tov ibn Gaon as the most probable author of the second part of the disputed commentary, with a remarkable similarity score of 0.959. Rabbi Shem Tov ben Abraham ibn Gaon (c. 1250-1330) was born in Soria, northern Castile, studied under the Rashba, and authored *Migdal Oz*, one of the earliest commentaries on Maimonides' *Mishneh Torah*. Significantly, he explicitly references his own lost Talmudic commentary called "*Shita*" in *Migdal Oz*, writing: "Many commentators have written on this, the first chapter of Bava Metzia, and we expanded upon it in its place in our *Shita*." This attribution had not been previously proposed in traditional scholarship and represents a novel contribution enabled by large-scale computational analysis.

Table 1: Citation Similarity Analysis for Disputed Commentary Section

Author	Similarity Score
R. Shem Tov ibn Gaon	0.959
Ra'avad the third	0.835
Rabbi Shmuel Hasardi	0.728
Authenticated Ritva works	0.170

o validate the computational attribution, we conducted comparative content analysis between the disputed commentary section and *Migdal Oz*, identifying several instances of parallel reasoning and similar interpretative approaches. One striking example appears in both works' treatment of *Hilkhot Gezeila Va'Aveida* 17:11, where both texts present identical arguments, citing the same Talmudic sources and reaching identical conclusions about the given dispute. The attributed section contains approximately 800 citations, well above our minimum threshold of 530 for statistical reliability, with the extremely high similarity score (0.959) representing the strongest match in our entire corpus analysis.

5 Results and Discussion

5.1 Methodological Validation and Performance Assessment

Our citation-based authorship attribution framework demonstrates strong performance across multiple evaluation criteria, establishing its viability as a scalable tool for digital library applications. The technical components achieve consistently high accuracy with reference extraction attaining F1 = 0.901, while citation fingerprint analysis successfully resolves the contested Ritva attribution with quantifiable confidence measures. The stability of citation patterns across known authors' corpora provides crucial validation, with consistency scores exceeding 0.98 for established authors (Ramban, Rashba, Ritva, Maharam Chalava) confirming that citation fingerprints represent stable authorial characteristics rather than random textual variation [2]. Cross-validation through content analysis strengthens confidence in computational results, as demonstrated by the identification of parallel reasoning between the attributed commentary section and Rabbi Shem Tov's *Migdal Oz*, showing convergence between quantitative analysis and qualitative examination.

5.2 Comparison with Traditional Attribution Methods

Traditional approaches to manuscript attribution rely primarily on stylistic analysis, content examination, and historical documentation, but face significant scalability limitations when applied to large digital collections. Our computational approach offers several key advantages: citation fingerprint analysis provides quantifiable similarity measures enabling systematic comparison across large author sets, maintains consistency in evaluation criteria avoiding subjective variations, and enables analysis at scales that reveal subtle authorial signatures invisible to manual examination. However, computational methods complement rather than replace traditional scholarship, as domain expertise remains essential for interpreting

results, assessing historical plausibility, and understanding intellectual contexts that shape citation practices. The most effective attribution analysis combines computational efficiency with scholarly interpretation, as demonstrated in our validation of the Rabbi Shem Tov attribution through content analysis.

5.3 Applications to Digital Libraries and Hypertext Research

The methodology's language-agnostic design and reliance on citation patterns rather than linguistic features enable application beyond medieval Hebrew literature to any citation-rich historical corpus, with potential applications including academic paper attribution for resolving authorship disputes, automated manuscript cataloging, plagiarism detection through unusual citation patterns, and web archive analysis to track intellectual influence patterns in digital scholarship. Our positioning of medieval citation networks as ancient hyperlink structures opens new avenues for hypertext research by demonstrating how both systems create navigable knowledge structures where references serve as pathways between related information sources—medieval scholars relied on citation networks to discover relevant authorities and trace intellectual lineages just as modern web users navigate hyperlinked content. Digital libraries can leverage these insights by treating citation networks as primitive knowledge graphs that encode semantic relationships between texts, authors, and ideas, with visualization tools mapping citation networks geographically and temporally to reveal intellectual transmission patterns and scholarly influence networks previously hidden in traditional library catalogs, thus bridging historical scholarship practices with modern information organization methods.

6 Conclusion

This paper demonstrates how medieval citation networks function as ancient hyperlink structures, encoding stable authorial signatures that enable computational authorship attribution in historical digital libraries. Our transformer-based methodology successfully resolves a centuries-old attribution debate, identifying Rabbi Shem Tov ibn Gaon as the previously unknown author of disputed commentary sections with 95.9% similarity confidence, illustrating how AI and NLP techniques can unlock authorship information embedded in historical reference networks.

The technical contributions establish a scalable framework for citation-based authorship analysis, with the BERT-CRF pipeline achieving F1 $\approx$ 0.90 accuracy in medieval Hebrew reference extraction and citation fingerprint methodology offering digital libraries language-agnostic tools for manuscript attribution. The methodology's applicability extends beyond medieval texts to any citation-rich corpus, including modern academic literature, web archives, and collaborative scholarly environments, with the proposed attribution satisfying historical plausibility criteria including temporal alignment, intellectual tradition, and geographic feasibility.

By positioning citation networks as fundamental hypertext structures, this work reveals how scholarly communities created navigable knowledge graphs long before digital technologies emerged, suggesting that linked information systems represent persistent human approaches to knowledge organization. The convergence of historical scholarship and computational analysis exemplified here demonstrates the value of interdisciplinary collaboration in addressing complex problems in digital cultural heritage, providing new pathways for AI-enhanced organization and discovery in digital manuscript collections.

References

[1] Shlomo Argamon, Moshe Koppel, James W Pennebaker, and Jonathan Schler. 2009. Automatically profiling the author of an anonymous text. *Commun. ACM* 52, 2 (2009), 119–123.

[2] Nati Ben-Gigi, Maayan Zhitomirsky-Geffet, Binyamin Katzoff, and Jonathan Schler. 2024. Citation network analysis for viewpoint plurality assessment of historical corpora: The case of the medieval rabbinic literature. *PLOS ONE* 19, 7 (2024), e0307115.

[3] Nati Ben-Gigi, Maayan Zhitomirsky-Geffet, Jonathan Schler, and Binyamin Katzoff. 2024. Automatic Construction of the Citation Network from the Medieval Jewish Responsa Literature. *ACM Journal on Computing and Cultural Heritage* 18, 2 (2024), 1–18.

[4] Giovanni Colavizza. 2017. The core literature of the historians of Venice. *Frontiers in Digital Humanities* 4 (2017), 14.

[5] Yaakov HaCohen-Kerner, Nadav Schweitzer, and Dror Mughaz. 2011. Automatically identifying citations in Hebrew-Aramaic documents. *Cybernetics and Systems: An International Journal* 42, 3 (2011), 180–197.

[6] Xie He, Arash Habibi Lashkari, Nikhill Vombatkere, and Dilli Prasad Sharma. 2024. Authorship attribution methods, challenges, and future research directions: A comprehensive survey. *Information* 15, 3 (2024), 131.

[7] Shawndra Hill and Foster Provost. 2003. The myth of the double-blind review? Author identification using only citations. *ACM Sigkdd Explorations Newsletter* 5, 2 (2003), 179–184.

[8] Sehrish Iqbal, Saeed-Ul Hassan, Naif Radi Aljohani, Salem Alelyani, Raheel Nawaz, and Lutz Bornmann. 2021. A decade of in-text citation analysis based on natural language processing and machine learning techniques: an overview of empirical studies. *Scientometrics* 126, 8 (2021), 6551–6599.

[9] Michael Kelly and K Patrick Fazioli. 2023. *social and intellectual networking in the early middle ages.* punctum books.

[10] Moshe Koppel and Jonathan Schler. 2004. Authorship verification as a one-class classification problem. In *Proceedings of the twenty-first international conference on Machine learning*. ACM, 489–495.

[11] Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2009. Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology* 60, 1 (2009), 9–26.

[12] Moshe Koppel and Nadav Schweitzer. 2014. Measuring direct and indirect authorial influence in historical corpora. *Journal of the Association for Information Science and Technology* 65, 10 (2014), 2138–2144.

[13] Eliyahu Lichtenstein. 1980. *Mavo le-idushei HaRitva al Masechet Gittin.* Mossad Harav Kook, Jerusalem. Introduction to the Ritva's Novellae on Tractate Gittin.

[14] Ritva. 1962. *Hiddushei Ritba, Baba Metzia.* Private publication, London. Edited from manuscripts and published for the first time.

[15] Matteo Romanello. 2016. Exploring citation networks to study intertextuality in classics. *DHQ: Digital Humanities Quarterly* 10, 2 (2016).

[16] Michael L Satlow and Michael Sperling. 2022. The Rabbinic citation network. *AJS Review: The Journal of the Association for Jewish Studies* 46, 2 (2022), 291–319.

[17] Avi Shmidman, Joshua Guedalia, Shaltiel Shmidman, Cheyn Shmuel Shmidman, Eli Handel, and Moshe Koppel. 2022. Introducing BEREL: BERT embeddings for Rabbinic-encoded language. *arXiv preprint arXiv:2208.01875* (2022).

[18] Pranaydeep Singh, Gorik Rutten, and Els Lefever. 2021. A pilot study for BERT language modelling and morphological analysis for ancient and medieval Greek. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature.* 128–137.

[19] Richard Snodgrass. 2006. Single-versus double-blind reviewing: An analysis of the literature. *ACM Sigmod Record* 35, 3 (2006), 8–21.

[20] Thea Sommerschield, Yannis Assael, John Pavlopoulos, Vanessa Stefanak, Andrew Senior, Chris Dyer, John Bodel, Jonathan Prag, Ion Androutsopoulos, and Nando De Freitas. 2023. Machine learning for ancient languages: A survey. *Computational Linguistics* 49, 3 (2023), 703–747.

[21] Efstathios Stamatatos. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology* 60, 3 (2009), 538–556.

[22] Omri Suissa, Avshalom Elmalech, and Maayan Zhitomirsky-Geffet. 2022. Text analysis using deep neural networks in digital humanities and information science. *Journal of the Association for Information Science and Technology* 73, 2 (2022), 268–287.

[23] Maayan Zhitomirsky-Geffet and Gila Prebor. 2019. SageBook: Toward a cross-generational social network for the Jewish sages' prosopography. *Digital Scholarship in the Humanities* 34, 3 (2019), 676–695.