BRILL

DEAD SEA DISCOVERIES (2026) 1–27

Dead Sea Discoveries

brill.com/dsd

Deep Learning Methods for Analyzing the Composition of the Hodayot Collection 1QH[a]

Jonathan Ben-Dov | ORCID: 0000-0002-5346-6950
Department of Bible, Tel Aviv University, Tel Aviv, Israel
jonbendov@tauex.tau.ac.il

Yonatan Lourie | ORCID: 0009-0009-0092-9390
Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel
lourie@mail.tau.ac.il

Roded Sharan | ORCID: 0000-0001-8363-4882
School of Computer Science and AI, Tel Aviv University, Tel Aviv, Israel
roded@tauex.tau.ac.il

Received 4 August 2025 | Accepted 5 October 2026 |
Published online 9 November 2026

Abstract

The partition of the 1QH[a] into distinct blocks of hymns—Teacher Hymns and Community Hymns—has been a long-standing consensus, supported by the 4Q copies. Many scholars accepted a block of Teacher hymns in 1QH[a] columns 10–16, with enveloping, transition material in cols. 9, 17–20. Yet the uniformity of the TH block has been doubted due to the presence of CH material and a possible Maskil redaction across the various blocks. In addition, the uniformity of the Community Hymns was also doubted. In this study we implement cutting-edge computerized methods from the field of Natural Language Processing (NLP) to examine the partition of 1QH[a]. The clustering method we have developed does retain the partition despite the suggested interruptions, yet it points out to "hybrid" sections in either of the accepted blocks. In addition, it provides scholars with a new tool for examining the Hodayot beyond the traditional philological tools known so far, prompting a discussion about the value of Deep Learning and AI for the study of ancient texts.

Published with license by Koninklijke Brill BV | DOI:10.1163/15685179-bja10088
© JONATHAN BEN-DOV ET AL., 2026 | ISSN: 0929-0761 (print) 1568-5179 (online)
This is an open access article distributed under the terms of the CC BY 4.0 license.

Keywords

Hodayot – 1QH^a – authorship recognition – Large Language Models – AI – NLP – Qumran poetry – Teacher of Righteousness

1 Introduction

A central question in Hodayot research was the classification of the hymns from the main copy 1QH^a into blocks of material by different authors or belonging to different genres. The discussion particularly involved the distinction of "Teacher Hymns" (= TH) and "Community Hymns" (= CH), in particular their placement within 1QH^a. The evidence from the Cave 4 copies was a significant addition to the discussion and illuminated new insights. The present article is an attempt to add a new kind of evidence to that distinction, produced by computerized Deep Learning. By that we hope to enrich the discussion on the composition of 1QH^a.

A block of hymns can be isolated in the middle of the collection 1QH^a. These hymns open with the incipit אודכה; they describe the personal experience of a person, including one or more of the following motifs: a person enlightened by divine revelation (e.g. 9:21; 12:23–29; 13:11, 14:12–18; 15:10, 26–27, 16:16, 35–37; 17:23–27), who leads a community through acts of enmity and persecution by outsiders (e.g. 10:21–29, 31–34; 12:7–22; 13:6–17), and was even exiled from his place (12:8–9) due to these persecutions. This material is enveloped in 1QH^a by hymns of a different type, whose contents point to a more collective generic experience without the personal motifs noted above. The central block of hymns is known as the Teacher Hymns (TH) while the rest are commonly known as Community Hymns (CH). The distinction began with studies of 1QH^a in the 1960s and became established in numerous newer studies, although it is not devoid of problems as we shall see.[1] It has even become customary to speak of "A Book of the Teacher" spanning the central part of 1QH^a, columns 10–17.[2] The copies 4Q429 and 4Q432 contain exclusively or almost exclusively Teacher Hymns.

The distinction between Teacher Hymns and Community Hymns raises two separate questions that are sometimes combined, mistakenly in our mind. The first question involves the identity of the authors. A long-lasting hypothesis in

1 The classical studies are Morawe, *Aufbau und Abgrenzung*; Jeremias, *Der Lehrer*; Kuhn, *Enderwartung*; Holm-Nielsen, *Psalms from Qumran*. More recent studies consulted here are Douglas, "Teacher Hymn Hypothesis"; Hasselbalch, *Meaning and Context*; Tanzer, *Sages at Qumran*; Newsom, "Farewell"; Johnson, "Look Who's Talking"; Harkins, "New Proposal."

2 Douglas, "Teacher Hymn Hypothesis."

the field connects the Teacher hymns with the agency of the Teacher of Righteousness, a personality known from CD and the Pesharim. This hypothesis was first raised by Eliezer Sukenik in the very early days of scrolls research,[3] and maintained by many since then. More recently however, this idea has been contested, both because the writings that became known with the publication of the DSS show little interest in the Teacher, and because the Teacher is now mostly seen as a legendary figure rather than a historical writer.[4] In general, scholars today are wary of an overly positivistic conception of personages in community history. The question of historical attribution will not concern us here. Rather, we focus on the more fundamental question of distinguishing the various hymns.

The second question, i.e., the actual distinction between the two types of hymns, evoked a lively discussion that has intensified in the recent decade. Tanzer, followed somewhat differently by Harkins, contested the unity of the CH category, which is admittedly less cohesive than the TH.[5] On the other hand, features of unity arise that cut through the entire scroll. Thus, quite a few of the hymns carry Maskil superscriptions, both in 1QH[a], more so in the 4Q copies, and yet other hymns can be subsumed under the Maskil material. Newsom accordingly bid farewell to the Community Hymns and labeled them instead "Maskil Hymns."[6] Johnson went even further by claiming that the entire collection of 1QH[a] is in fact a Maskil collection that encompasses the TH into an overall depiction of a leader.[7]

Even more profoundly, the exact delimitation of the types remains unsettled, and accordingly it is still unclear whether they stand as contiguous blocks in 1QH[a] or rather the two stand mixed. The original proposal by Sukenik for defining the TH included 1QH[a] columns 10–12 and 18:4–14 (converted to DJD 40 numbering).[8] While Newsom follows Douglas and takes for granted the existence of a "Book of the Teacher" in 1QH[a] columns 10–17, with framing material in columns 9 and 18–20,[9] earlier scholars like Jeremias and Kuhn—and Newsom herself—point to sections within that book that are Community Hymns in essence. Douglas himself posited three categories of Teacher Hymns based

3 Sukenik, *Second Survey*, 32–33. This opinion was maintained, for example by Douglas, "Teacher Hymn Hypothesis."

4 Jokiranta, *Social Identity and Sectarianism*; Kratz, "Teacher of Righteousness." But contrast Newsom, "Farewell," 5 n. 22.

5 Tanzer, *Sages at Qumran*; Harkins, "New Proposal."

6 Newsom, "Farewell."

7 Johnson, "Look Who's Talking."

8 Sukenik, *Second Survey*, 32–33.

9 Newsom, "Farewell."

on his level of certainty in their classification. Levels A and B can be considered Teacher Hymns, while the two hymns that fall under his level C (15:29–36, 15:37–16:4) are in fact Community Hymns for all purposes that stand amid the "Book of the Teacher." This fact joins the existence of hybrid hymns that belong to both categories, or of a redaction that integrated the block of TH into the larger scroll 1QH^a by means of unifying elements.[10] Harkins claimed that the hymns in 1QH^a cols. 20–27 (which she calls CH2) are closer to the TH than the CH1 hymns, which in her view had not been part of the original scroll.[11] These possibilities are elaborated below.

The delimitation of the TH block and its distinction from the CH block(s) raise the question how to treat the various hybrid materials, or the various levels of belonging to the strict dichotomy TH-CH. One could ask where these outliers stand on the CH-TH continuum: the distinct Community Hymns within the TH block; the transition material in columns 9, 18–20; The distinction between CH1 and CH2. These fine distinctions call for refined methods of evaluation.

While philological discussions continue, it is worthwhile checking how the new tools of machine learning and AI can tackle this question. Traditional discussions trace exemplary phrases or ideas that mark the most obvious Teacher Hymns and continue by tracing a chain of hymns that relate to them.[12] For example, Jeremias compiled lists of typical vocabulary in the two distinct types.[13] Douglas found the typical phrase הגבירכה בי to be indicative of the Teacher Hymns, and proceeded to include in the group other hymns, which do not contain this phrase but show other similarities with those hymns that do contain it. Thus, distinction is achieved mainly by vocabulary and ideological content. In contrast, we offer to evaluate the material by means of a variety of computational language processing techniques. Some of these techniques are transparent to the user while others are opaque.[14] Can the machine offer new insights, or would it substantiate the traditional distinctions? Would it prioritize one or another of the scholarly opinions? If the answer is positive, what would be the status of the machine's statements? Should we accept them as the basis for further theorizing? These questions are relevant for all studies achieved by computerized learning, the first of which are beginning to appear in DSS research.

Using advanced computer-assisted methods is part of the discipline of Natural Language Processing (NLP). In this paper, we propose a step beyond what

10 Tanzer, *Sages at Qumran*; Johnson, "Look Who's Talking."
11 Harkins, "New Proposal."
12 Thus Douglas, "The Teacher Hymn Hypothesis"; Newsom, "Farewell."
13 Jeremias, *Der Lehrer*, 172–173.
14 Thanks to Dr. Avi Shmidman for his advice on this point.

has previously been achieved with NLP in biblical and DSS studies. Earlier existing classifiers operate mainly by means of word-frequency, while other studies traced sequences of letters (character trigrams) with some success, as surveyed below. We offer a wide set of cutting-edge computerized tools for embedding and clustering texts. The set of tools proposed here integrates various aspects of the language: semantic-lexical methods next to advanced syntactic methods that are employed by Large Language Models (LLMs). Recent studies demonstrated that characterizing texts by means of LLMs outperforms the older methods that were based on word frequency.[15] More specifically, we employ an innovative method for aggregating the various embeddings and propagating them into a more robust result. Our study thus presents an innovation not only for DSS research but also for the discipline of NLP.

While preliminary results and cursory discussion of this research have already been published in a Computer Science venue,[16] the present article presents the research questions in Qumran studies in a less mathematical way and stresses the textual features and their implications for the composition of 1QH[a]. Full information about the algorithms employed in this research can be found in the preliminary article. The code and further information are available on our Github page.[17]

2 Cracks in the Block

The original impetus for this study arose with the sense that some cracks arise in the view of 1QH[a] as an array of distinct blocks of Community and Teacher hymns. The following table collected the assignments of individual hymns by the various authors and shows the complexity of the distinction. Data is collected from the authors who first classified the Teacher Hymns in the early years of research, as well as from three later authors who traced distinctions within the TH block.[18]

15 Rockmore et al., "Literary Landscapes"; Lourie et al., "Semantic and Syntactic Features."

16 Lourie et al., "Semantic and Syntactic Features." That paper discusses other topics in addition to the partition of 1QH[a]. We shall address these topics in separate articles.

17 https://github.com/yonatanlou/QumranNLP.

18 The data for Holm-Nielsen, Morawe, Jeremias, Becker and Kuhn are taken from the table in Douglas, "Teacher Hymn Hypothesis," 245. Schuller's data are taken from Schuller, "The Hodayot Scroll," 269; this Hebrew article is the only place I could locate in Schuller's corpus of writings that counts inconsistencies within the TH block. Tanzer's data are taken from Tanzer, *Sages at Qumran*, 139.

Column and line numbers cited in the table follow the numbering in DJD 40. Hodayot delimitations were normalized according to the list in Schuller and Newsom, *The Hodayot*,

	Holm Nielsen	Morawe	Jeremias	Becker	Kuhn	Douglas	Schuller	Tanzer
10:5–21	TH	TH	TH	TH	TH	TH Level A	TH	Hybrid
10:22–32	TH	TH	CH	TH	CH	TH Level A		
10:33–11:5	TH	TH	TH	CH	CH	TH Level A	TH	
11:6–19	TH	TH	TH	TH	CH	TH Level B	TH	
11:20–37	TH	TH	CH	CH	CH	TH Level B	CH?	Hybrid
11:38–12:5	Unclass.	TH	CH	TH	CH			
12:6–13:6[a]	TH	TH	TH	TH	TH	TH Level A (12:6–31a)	TH	Hybrid
13:7–21	TH	TH	TH	TH	TH	TH Level A	TH	
13:22–15:8	TH	TH	TH	TH	TH	TH Level A	TH	Hybrid
15:9–28	TH	TH	TH	TH	TH	TH Level A	TH	
15:29–36	CH	CH	CH	CH	CH	TH Level C = CH	CH?	
15:37–16:4	CH	CH	CH	CH	CH	TH Level C = CH		Hybrid
16:5–17:36[b]	TH	TH	TH	TH	TH	TH Level B	TH	Hybrid

a Becker and Kuhn consider 12:31b–13:6 to be a secondary addition. Douglas, "The Teacher Hymn Hypothesis," 249, tends to agree.

b Jeremias, Becker, and Kuhn leave 17:1–36 unclassified.

The table highlights notable outliers of Community Hymns even within the block of TH, colored grey. The main examples are the short hymns in 15:29–36 and 15:37–16:4, which are unanimously declared as CH. These two pieces do not invoke the themes of exile and persecution, and do not link to the chain of expressions and motifs that Douglas and others constructed for the TH. Sarah Tanzer considered these hymns to be "hybrids," i.e. community hymns that incorporate TH features or vice versa. For example, she considers 15:29–36 to be a CH appended by a short TH introduction.[19] She counts the following hybrids, which are Teacher Hymns built on themes from the Community Hymns: 10 beginning–19, 12:7–13:6, 13:20–15:5, 11:21–36, 15:37–16:4, 16:4–17:36. Another short hymn, 11:7–20 is also hybrid but not built on a CH core.[20] The main hybrid element according to Tanzer is the *Niedrigkeitsdoxologie*, self-depreciating poems, which the Teacher Hymns adopted from the Community Hymns. Newsom suggested that 12:30b–34a is another self-depreciating piece added by a later

8–9, which in turn follows Stegemann's count. We use the categories CH and TH despite the somewhat different labels in older literature. For Holm-Nielsen's categories, a psalm of thanksgiving = TH and a hymn = CH. For Morawe's categories, *Danklied* = TH and *Hymnus* = CH.

19 Tanzer, *Sages at Qumran*, 62; on hybrids in general see pp. 148–150.

20 Tanzer, *Sages at Qumran*, 128 n. 49.

editor.[21] Such proposals mark the way for a finer division of the blocks into smaller units, further obfuscating the partition.

Yet additional confusion arose with regard to the "transition material," for example the hymn known as "the Hymn of Creation" in column 9. Douglas counted it as transition material based on its content, especially at the end of the hymn in lines 31 onwards.[22] Schuller buttressed this idea by noting that the creation hymn also opened the scroll 4QpapH^{f} (4Q429), and serves as an introduction to the Teacher Hymns collected in it.[23] A similar setting pertains in 1QH^{a}, where the Hymn of Creation opens a new sheet (columns IX–XII).[24] On the other hand, Johnson noted how the content of this hymn, especially the self-depreciating lines 23–29 rather resembles the CH hymn in column 5, again obfuscating the accepted distinctions.[25]

Harkins on her part effected an additional blow to the block division of 1QH^{a}. She does not undermine the uniformity of the TH block, not even 15:29–16:4 that others see as clear CH.[26] However, she distinguishes the CH material that precedes TH block (CH1) from the material that stands after it at the end of the scroll (CH2). The CH2 material is closer to the TH inasmuch as both develop the theme of communion with the angels alongside other Yahad motifs, which are not found in CH1. She even undermined the material constitution of 1QH^{a}, claiming that the CH1 material was written on a separate scroll and had not originally been part of 1QH^{a}.[27]

Objections against the distinctions of TH and CH were expressed on even more fundamental grounds, as scholars began underscoring the role of the Maskil in the composition and performance of the Hodayot.[28] A central hinge for this claim came from the discovery of Maskil superscriptions in several of the CH previously known from the large copy 1QH^{a}: 5:12–14; 7:21; 20:7 (attested in 4QH^{a}, 4QH^{b}); 20:14 (attested in 4QH^{a});[29] 25:34 (attested also in 4QH^{a}), and the psalm in 1QH^{a} frag. 10, whose placement in the scroll is debated.[30] Thus,

21 Newsom, oral communication mentioned by Johnson, "Look Who's Talking," 329 n. 58.

22 Douglas, "Teacher Hymn Hypothesis," 256–257 and n. 31.

23 Stegemann et al, DJD 40:120; Schuller, "Recet Scholarship," 140.

24 See DJD 40:120; Tigchelaar, "Further Look, 210." Harkins, "Another Look," suggested that the scroll 1QH^{a} had in fact started in column IX, containing only the TH block followed by CH2 hymns.

25 Johnson, "Look Who's Talking," 326–327.

26 Harkins, "New Proposal," 104.

27 Harkins, "Another Look." But see the objections to this claim by Johnson, "Scroll Divided"; Tigchelaar, "Further Look."

28 See already Licht, *Thanksgiving Scroll*, 22–26.

29 This one is a mention of the Maskil in the middle of the hymn, not a superscription.

30 Schuller (in DJD 40:37, 99) places it in column 7 (also Harkins, "Another Look," 193), while

Qimron restated Licht's old position that all the hymns are in fact hymns of the leader, which he sees as the Maskil.[31] The Maskil is gradually becoming the main if not the sole character of the Hodayot, both the CH and TH. The distinction was first blurred by Hasselbalch based on socio-linguistic theory.[32] She claimed that while separate authorship of the CH and TH is possible, the bifurcation was blurred due to procedures of hybridization on the two ends. Newsom went forward by claiming that the CH are in fact not by the community but rather by the Maskil.[33] Johnson elaborated on the same idea, adding new arguments. According to him:

> The speaker in 1QH[a] is explicitly identified as a Maskil in the psalms before and after the Teacher's Book, and the introduction and conclusions to the Teacher's book embed the Teacher Hymns into the Maskil's discourse. *There is also evidence that the psalms in the Teacher's Book have been edited or supplemented to bring the material into greater alignment with the rest of the collection*. Even if some or all of the psalms in the Teacher's Book were not originally written with a Maskil in mind as the intended speaker, the sequencing and editing of the material suggests that they have been enfolded, updated, and subsumed within a Maskil's discourse.[34]

Johnson then takes pains to explain how the two main motifs identified with the teacher, i.e., persecution and personal revelation, can in fact be read also as referring to the personality of a Maskil, a generic leader of the community, rather than to a one-time historical figure.

To conclude this section, the traditional distinction between CH and TH has been undermined in various directions. On the one hand, fine divisions of the TH block revealed deviations and redactional pieces that interrupt the flow. On the other hand, the very idea of the division has been called into question, when scholars advance the figure of a Maskil as the inspiration for the entire collection. With these doubts in mind, it is intriguing to see how a non-human algorithm will cluster the corpus.

Qimron (*Dead Sea Scrolls*, 1:61, 66, 105) places it at an unknown place in the first part of 1QH[a]. See the discussion in Newsom, "Farewell," 10–11.

31 Qimron, *Dead Sea Scrolls*, I, 61.

32 Hasselbalch, *Meaning and Context*.

33 Newsom, "Farewell."

34 Johnson, "Look Who's Talking," 337. Italics by the present author.

3 Computational Text Analysis in Biblical Studies

Much if not most of the philological work is based on distinguishing between texts: sources, layers of redaction, multiple authors etc. Philologists carry out this task by means of lists of prominent markers such as characteristic words or syntactic features. It therefore makes sense to assign these tasks to a computer alongside human judgment, whether for tracking authorship or, as in the present case, for tracking literary layers and distinguishing genres. The main challenge for this task is modelling, i.e., finding a way to convert the text into a set of numbers or vectors, which would be available for computerized analysis. Computational language studies developed a variety of methods for achieving this task. Each modelling method highlights a component of the text, either orthography, vocabulary, morphology or syntax or any other distinctive feature that can be presented to the machine as a set of numbers.

Stylometry of biblical Hebrew began already in the 1970s.[35] In this discipline, scholars define a set of grammatical features which they consider distinctive of the style in question. This work is supervised, i.e., a human being instructs the computer what to look for, and at least in its early stages it was operated manually or with statistical software. Stylometric work continued with improved methods.[36] More recently such methods were applied by John Starr to the Aramaic texts from Qumran, with a long and informative appendix about stylometry of Hebrew Bible texts.[37]

Pierre Van Hecke carried out works of linguistic classification in Qumran studies using newer methods.[38] Instead of defining in advance the desired grammatical features, Van Hecke and Johan de Joode modelled the text using an automatic division into character trigrams, that is, sequences of three characters taken with a running window along the text in question. To wit, this is not an N-gram study of word frequency but rather a sophisticated way to take in account various layers of grammar when vectorizing a text. For example, the words בראשית ברא are analyzed as a sequence of the tokens ,ברא ,ראש ,אשי ,שית יתב ,תבר ,ברא etc. A statistical program (R) collects the trigrams from all the studied texts and aggregates them, creating a numerical linguistic profile for

35 Radday, "Computerized Statistical-Linguistic Tests"; Radday and Shore, *Genesis*.

36 Krispenz, *Literarkritik und Stilstatistik*.

37 Starr, *Aramaic Texts from Qumran*, 311–340. Daniel Machiela (*Review of Biblical Literature* 06/21) criticized Starr's selection of features for analyzing the Aramaic texts, but did not object to the section on the Hebrew text.

38 Van Hecke, "Computational Stylometric Approach"; Van Hecke and de Joode, "Promises and Challenges."

each of the texts using programs designed for this task.[39] The method produced good results, allowing Van Hecke to automatically distinguish the biblical scrolls from the non-biblical scrolls, and in addition to group together copies of the same composition. It is hard to say exactly what aspect of language is captured by the character trigrams: orthography? morphology? syntax? style? Yet the procedure creates successful results in distinguishing types of scrolls.

Modelling is often achieved by evaluating the vocabulary and its frequency. For example, the method known as TF-IDF (term frequency—inverse document frequency) traces rare words in a given corpus and gives a larger weight to those words that are frequent in the present corpus but rare elsewhere,[40] thus improving on the mere recording of lexemes. Various classifiers were operated on the biblical text mainly by studying word frequency.[41] These studies show very good results for example in isolating passages of the Priestly source in the Pentateuch or deuteronomistic layers in the Former Prophets. More recently, Feigenbaum-Golovin et al. used distinctive lemmas for authorship recognition of passages in the Pentateuch and Former Prophets, with a more robust statistical method for analyzing the results.[42]

The problems raised by the Hodayot are different from those raised by the Hebrew Bible in several respects. The partition in 1QH^a does not depend on preconceptions about other sections of the DSS corpus, and the Hodayot did not go through as many editorial vicissitudes as D or P. These differences increase the chances for a clear-cut partition. On the other hand, the corpus is smaller and more fragmentary than the Hebrew Bible, which creates a different kind of problems. However, the methods introduced here address these deficiencies by leveraging the available data in a more efficient way. The amount of data as described below, both in 1QH^a and in the entire DSS corpus used for training, is enough for the task.

We attempt to magnify the power of morphology and syntax rather than vocabulary in the modelling process. Thus, in contrast to other methods of computational modelling we do not need to rely on acquiring lemmas and morphology from manually curated repositories, which are by definition subjective

39 See Van Hecke, "Computational Stylometric Approach," 62. The tools used are "Java Graphical Authorship Attribution" and the "stylo" package for textual research in R programming language.

40 Bafna et al., "Document Clustering."

41 Dershowitz et al., "Computerized Source Criticism"; Koppel et al., "Unsupervised Decomposition"; Bühler et al., "Exploring the Stylistic Uniqueness." These studies improve the results achieved in the TIBERIAS classifier: https://tiberias.dicta.org.il; see Berman, "Measuring Style in Isaiah."

42 Faigenbaum-Golovin et al., "Critical Biblical Studies."

and may raise problems. Large Language Models can analyze a text without any prior prompting. A model trained on a given corpus will create a linguistic profile of chunks of text from that corpus and will offer a way to compare them to other chunks of text. This is exactly what models do: create a vector, i.e., a long and complicated number that represents the linguistic profile of a text. Some of this modelling is transparent while some is opaque, as the models use a variety of linguistic features with an unknown weight given to each of them. This opacity is a disadvantage, as the results are not interpretable. Importantly, the present task is a clearly defined procedure on an existing text and does not involve text generation, hence not affected by the phenomenon of text hallucinations.

In the present study we set it as our aim to use several different methods of language representation:

1. Character trigrams (based on the work by Van Hecke)
2. TF-IDF (term frequency—inverse document frequency)
3. BERT, a set of Large Language Models trained on Hebrew, including BEREL, trained on rabbinic Hebrew.[43]

Each of these models will analyze the text of 1QH[a] by capturing a different aspect of its language. For example, while character trigrams capture mainly orthography and morphology, the BERT models give much weight to syntax and style. We survey below the results achieved by each of the methods. Additionally, we aggregate the entire set of results by means of a Graph Neural Network (GNN). This method accepts various data and converts them into a network. Chunks of text are the nodes that constitute the network, with edges connecting the nodes based on their computed similarity to each other. Similarity is defined according to various statistical or semantic features of the chunks, as described in our other, mathematical publication.[44] As we show below, where one modelling method does not provide conclusive results, The GNN is able to provide a more robust score that takes all available data into account. It is thus a way to propagate data that would otherwise be insufficient, by making use of various branches of language: orthography, vocabulary and syntax. This hybrid approach allows us to integrate statistical and semantic features, addressing both the fragmentary nature of the DSS and the inherent challenges of processing ancient Hebrew. Moreover, we show below the results of each method separately. When these results differ from the GNN final result, one may infer which linguistic feature can explain the difference.

43 Seker et al., "AlephBERT"; Shmidman et al., "Introducing BEREL."

44 Lourie, Ben-Dov and Sharan, "Semantic and Syntactic Features."

Graph Neural Networks have emerged as powerful tools for representing textual data, particularly for tasks involving relationships between textual entities. Models such as TextGCN and BertGCN have demonstrated success in text classification by leveraging graph structures.[45] We now apply them to the text of the DSS.

4 Methodology of the Present Study

We used transcriptions of the Qumran scrolls (excluding the scrolls from other Judean Desert sites) based on the data files prepared by Martin Abegg, as represented in the Text-Fabric package.[46] Studies in Natural Language Processing, most of them in English texts, apply various procedures to the texts before the actual processing, for example the removal of stopwords (conjunctions, definite articles or the verb "to be") or of very frequent words. Such pre-processed texts enhance the contrast between sections and allow for more efficient clustering. However, with regard to the Hebrew text of the DSS we found that these methods either reduced performance or significantly diminished dataset size. Consequently, we decided to proceed without pre-processing to maintain data integrity and comparability.

Preparing the Hebrew (excluding Aramaic) scrolls included removing metadata, numbers (fragment numbers, lines, etc.) and non-Hebrew text. We removed all reconstructed text, both hypothetical restorations and the reconstructed text based on parallel copies.[47] We retained letters marked as probable or possible but excluded textual gaps, doubt marks, and redundant (double) spaces. Vacats were also left out; the suggested algorithm does not consider the structuring of the text through spaced paragraphs but rather only the text itself.

45 Yao et al, "Graph Convolutional Networks"; Lin et al., "BertGCN: Transductive Text Classification"; Yang et al., "HGAT"; Huang et al., "Text Level Graph Neural."

46 Roorda, "Text-Fabric." The data is found in: https://github.com/ETCBC/dss/blob/master/docs/about.md#provenance. Note that there are some differences between Abegg's text as contained in Brill's DSSEL and the longer text contained in Text-Fabric. For example, 1QH^a 8:29–37 are not contained in DSSEL but do appear in Text-Fabric and Accordance.

47 We had to remove the latter too, because the two types of restoration are hard to distinguish electronically. We verified that the amount of text restored in 1QH^a from 4Q parallels is meager. While the 4Q copies contain important cues for traditional philological study, like additional Maskil titles, the present algorithm is oriented to cluster the psalms based on their syntax and style, with only little weight given to vocabulary, let alone to specific words. Thus, in this type of study the occurrence of Maskil superscriptions is not a significant factor.

Lacunae are ignored, which means that the words on the two sides of the lacuna are treated as a continuous text. In a well-preserved scroll like 1QH^{a}, the effect of this problem is minimal.[48]

The text of 1QH^{a} was divided into fixed-size chunks. After evaluating various options, we chose a chunk size of 100 contiguous words with an overlap of 15 words to ensure balanced granularity and representation, especially important for shorter scrolls.[49] The automatic, hands-on definition of chunks is essential for the unsupervised method employed here. This division does not exactly correspond to the delineation of hymns in 1QH^{a}, yet in most cases the chunks are small enough to fit within the borders of a thematic unit or protrude from it only to a minimal degree. Note also that there is a 15-word overlap between the machine-divided chunks, which reduces the effect of a potential mismatch between chunks and literary units.

We represented each text chunk using both semantic and statistical features. For the semantic component, we used the Hebrew BERT-based model, BEREL, which is specifically trained on Rabbinic Hebrew texts—closer than modern Hebrew though not identical to the DSS corpus. For the statistical features, we used two approaches: TF-IDF and character trigrams. Next, we represented the data using a graph structure. Each node in the graph corresponds to a text chunk, represented by its BEREL embedding (a vector of 768 numbers). Edges between nodes were constructed based on the TF-IDF and trigram features. Specifically, edges were generated by calculating the cosine similarity separately for each statistical feature type (TF-IDF and trigrams) and combining these similarity matrices. To maintain a meaningful and sparse structure, we included only edges with cosine similarities above the 99th percentile. This resulted in a graph with 978 nodes and 9,391 edges.

Another feature of our use of "black-box" algorithms must be clarified for readers who are not accustomed to such methods. Before applying the classifying algorithm suggested here we carried out a series of tests to verify its reliability.[50] As described in our previous article, we checked the level of success of various classifiers in taking an input of randomly ordered chunks and placing them next to each other in their original order. Another test examined the algorithm's ability to group together chunks from the same composition. The ar-

48 Admittedly, some of the chunks contain text with larger lacunae, especially in the columns 2–4 and 21–23. Nevertheless, the filtering process employed in this study aimed to balance between the accuracy of the input text and a wide coverage of 1QH^{a}.

49 For the efficient use of machine-defined chunks in vectorizing long texts see Rockmore et el., "On the Literary Landscapes."

50 Lourie, Ben-Dov and Sharan, "Semantic and Syntactic Features," Section 2.5.

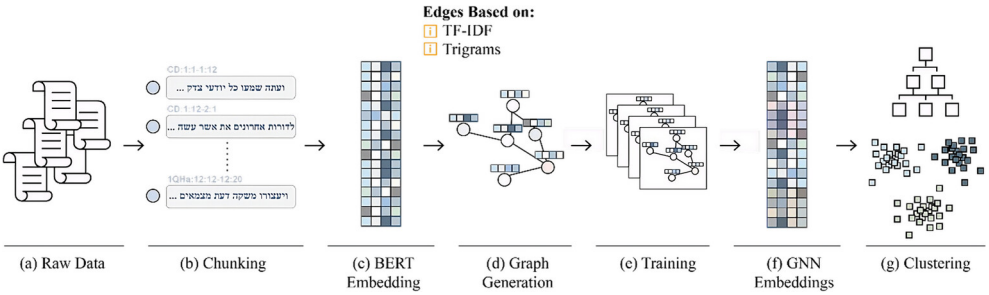


FIGURE 1 A sketch of the research outline

chitecture chosen for this study is thus not random, nor is it an attempt to prove what we already know (circular reasoning). Rather, it is an evidenced-based and amply tested operation of a fully unsupervised method. If the algorithm can successfully join randomly chosen chunks, it should be able to provide a valid clustering of CH and TH.

We then trained a GNN model to generate refined text embeddings. Specifically, we used a Graph Autoencoder, an unsupervised approach designed to learn meaningful representations without relying on potentially subjective or debated labels. The loss function is a classical reconstruction loss (the model task is to reconstruct the initial edge structure). To evaluate and tune the model, we performed cross-validation. After training on each random 80% subsample, we cluster the resulting composition embeddings and repeat this many times, reporting the mean and variance of the clustering scores across runs. This resampling-based evaluation tests stability: if performance is consistently high (and variance is low), the clustering signal is robust rather than a lucky split, serving as a practical alternative to k-fold cross validation given the small sample sizes. Naturally, we did not tune the hyperparameters on the DSS labels, but only on intrinsic clustering scores. Once hyperparameters were finalized, we trained the model on the entire Qumran dataset excluding 1QH^a. Finally, we embedded the text of 1QH^a using the trained GNN model, allowing us to evaluate our methodology on the content of this distinct and significant scroll.

In order to cluster the chunks in an unsupervised manner we used Ward's method of hierarchical clustering which results in a dendrogram depicting the hierarchical relations. Figure 1 describes the pipeline constructed for this algorithm.

5 The Present Analysis

We present here clustering of 1QH^{a} as performed by various embeddings while noting the linguistic features used in each of them. We shall then present the overall product aggregated by the GNN.

A factor that may affect the clustering is that 1QH^{a} was written by several different hands, and, importantly for our purposes, displays different habits of orthography and morphology. 1QH^{a} was penned by at least two, probably three scribes: scribe A penned most of the scroll until 19:25, scribe B penned lines until 19:29, and scribe C penned the rest of the scroll.[51] Several systems of morphology and orthography are reflected in the scroll, displayed in blocks of text: columns 2–8, column 9 beginning–20:6, and column 20:7–the end.[52] As a generalization (inconsistencies notwithstanding): the first block uses short pronouns and the defective spelling לא, בי, the middle section uses long pronouns and defective spellings, while the rest of the scroll generally prefers long forms throughout. It would be interesting to see whether and how the embeddings may take this variance in account, especially the trigram criterion, which presumably gives much weight to orthography.

The following diagrams are dendrograms, presenting the clustering of 1QH^{a} like branches of a two-dimensional tree. The units in the diagram are not individual hymns but rather chunks of 100 words, as delineated above. The diagram should be read from the bottom upwards: the closest pair of chunks is placed next to each other and then placed next to the most similar pair, and so on. Below we shall also present three-dimensional presentations of the clustering, where more complexity can be viewed.

The clustering is unsupervised, which means that the algorithm is oblivious about the scholarly label of the chunks, whether TH or CH. The colors at the bottom part of the diagram are our addition, meant to enhance the reader's understanding. CH are colored black; TH are colored orange; Douglas' transition material in columns 9 and 17 is colored brown. We can now check what kind of clustering is determined by the various embeddings.

5.1 *Clustering by Character Trigrams*

Figure 2 shows the clustering by character trigrams.

51 A convenient discussion of the role of each scribe in the writing and correction of 1QH^{a} is found in Tigchelaar, "Further Look," 205–209. In contrast, Harkins, "Another Look," 194 counts two main scribes, with a third editorial hand that added corrections towards a fuller orthography.

52 Schuller, "Recent Scholarship," 123.

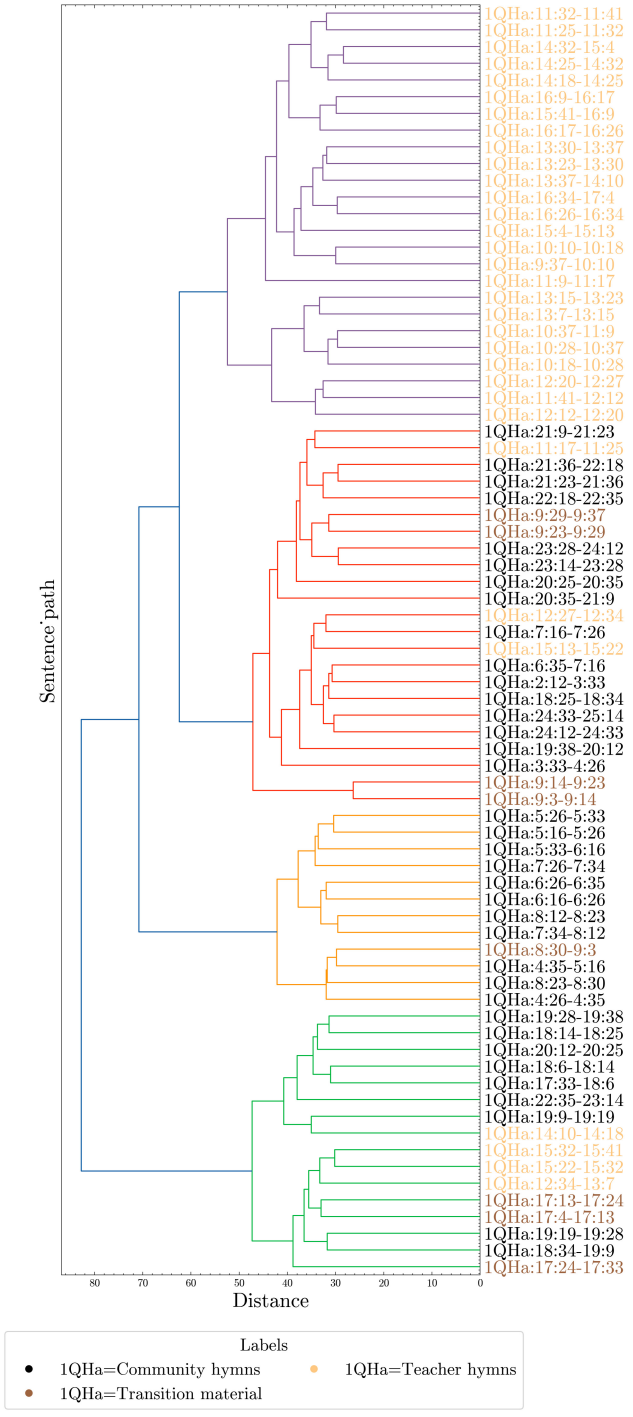


FIGURE 2
Clustering of 1QH[a] according to character trigrams
PRODUCED BY YONATAN LOURIE

The first impression from this diagram is the overall threefold division of the material.

1) Most TH material (yellow) stands at the top end and is subsumed under one cluster (purple)
2) CH material (black) stands at the bottom under three clusters.
3) A block of the transitional column 17 and some hybrid material from column 15 stands further down, clustered (green) with the CH material. The unit 14:10–18 is an additional outlier.
4) CH 2 material from columns 20–23 stands closer to the TH. This is transition material that frames the TH block, while it is also part of Harkins's CH2 block. The proximity thus supports Harkins's hypothesis about these four columns.[53]

The TH block stands apart. The CH block engulfs one chunk of TH from column 14:10–18 and the hybrid material from column 15. Both transitional columns 9 and 17 fall within the CH group, but column 9 is closer to the TH than column 17.

The Trigram clustering generally confirms the CH-TH division. The transition material is scattered at the two ends of the CH block. Some CH 2 material shows similarity with the TH.

5.2 *Clustering by TF-IDF*

Figure 3 presents the clustering by TF-IDF. This system counts frequent and rare words in the chunk while giving extra weight to words that are rare or not represented in other chunks.

Here too, we see separate blocks of TH and CH but the dispersion is wider.

1. A clear block of TH (yellow) stands at the bottom, but not all TH material is contained in it.
2. The CH material appears in two blocks plus some scattered material. Interestingly, these blocks largely correspond with Harkins's groups CH1 and CH2, the former taking the uppermost position while the latter forms a block further down, closer to the TH material.
3. Some material from the TH block stands apart from the rest. This is mostly hybrid material from column 15, with some additional chunks (12:12–27, 12:34–13:7). Note that 14:10–18 stands within the CH in this diagram too.

53 Harkins, "Another Look." The trigrams test is especially sensitive to orthography, but the proximity between TH and CH2 in this test cannot be imbued to orthography, since the method of orthography changed in 19:25 and we would thus expect a separation of cols 20–23 from the TH material that stands before 19:25.

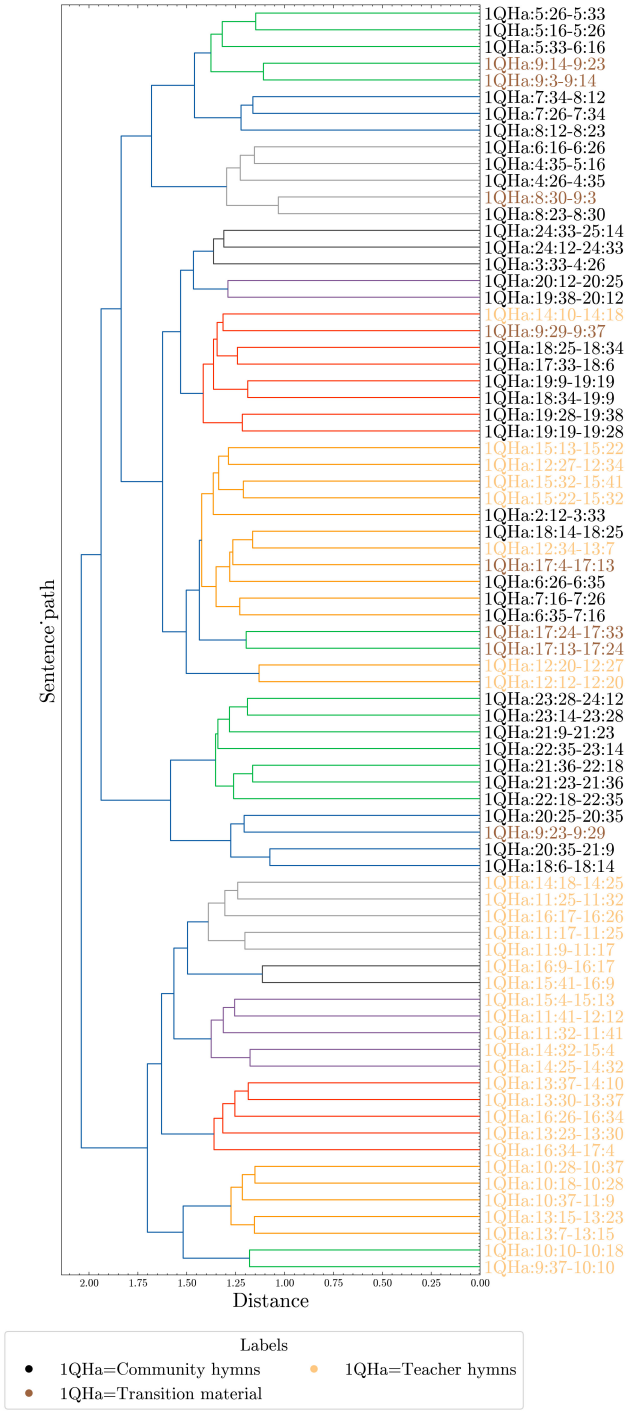


FIGURE 3
Clustering of 1QH[a] according to TF-IDF PRODUCED BY YONATAN LOURIE

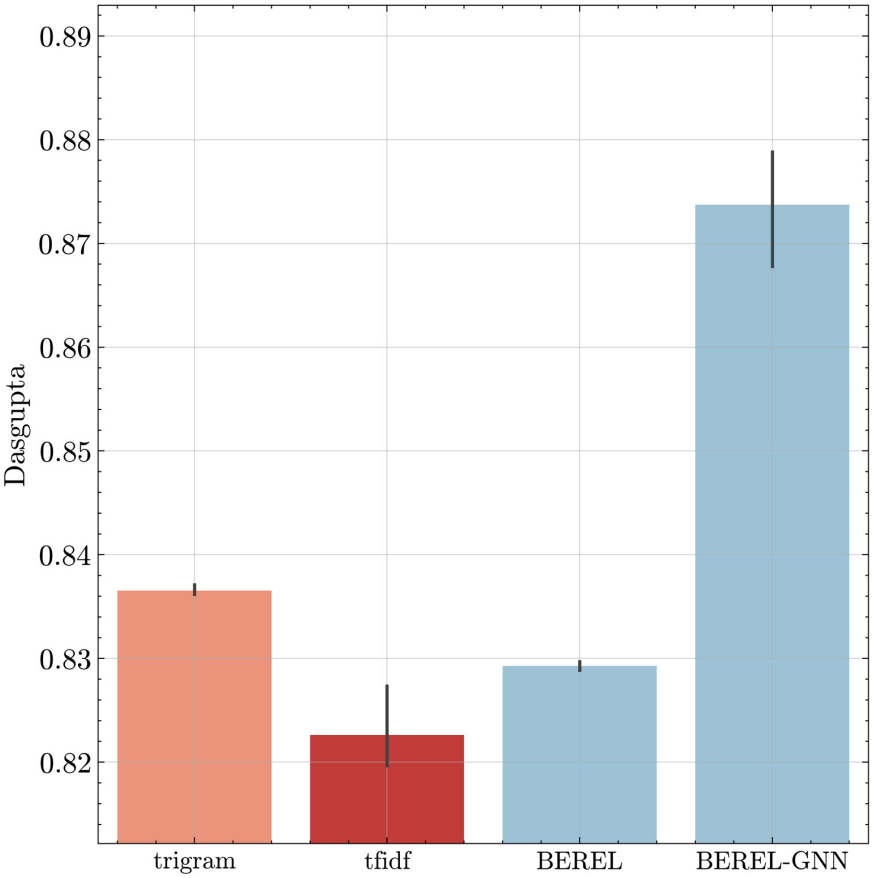


FIGURE 4 Improvement of various embeddings (left) when propagated by GNN
PREPARED BY YONATAN LOURIE

4. Transition material from columns 9 and 17 is scattered within the CH blocks. The only part that is closer to the TH is 9:23–29.

5.3 *The Aggregated Clustering Using GNN*

The Graph Neural Network aggregates all previous embeddings into a network and recalculates the distances, as described above. The nodes of the network are chunks of the 1QH[a] text, as defined by BEREL, a Hebrew BERT model especially trained on rabbinic Hebrew. The GNN propagates the base results and produces a more robust clustering. Figure 4 demonstrates the improvement achieved from the individual embeddings (on the left) to those achieved by the graph (on the right), when tested by on a wide selection of the Qumran corpus. The Dasgupta distance measures the level of success as the ability of

the algorithm to place together chunks that stand next to each other in a given scroll, for example the ability to join two adjacent chunks from CD.[54]

Another way to demonstrate the power of the GNN is to examine scatter plots obtained by various methods. The plots are shown in this set of interactive graphs, parts of which are presented here: https://yonatanlou.github.io/QumranNLP/reports/plots/clustering_within_scroll_scatter_plots.html

This is an interactive graph. Hovering on a node will show its identity and place in 1QH^a. In addition, the box to the left allows choosing one chunk and highlighting it.

Fig. 5 shows the (non-interactive) plot for the TF-IDF (top) and for Character Trigrams (bottom).

A quick exploratory look shows that the groups are more clearly separated in the upper diagram, where there are less outliers. The separation will be even clearer in the plot achieved by the GNN (Fig. 6; see next page).

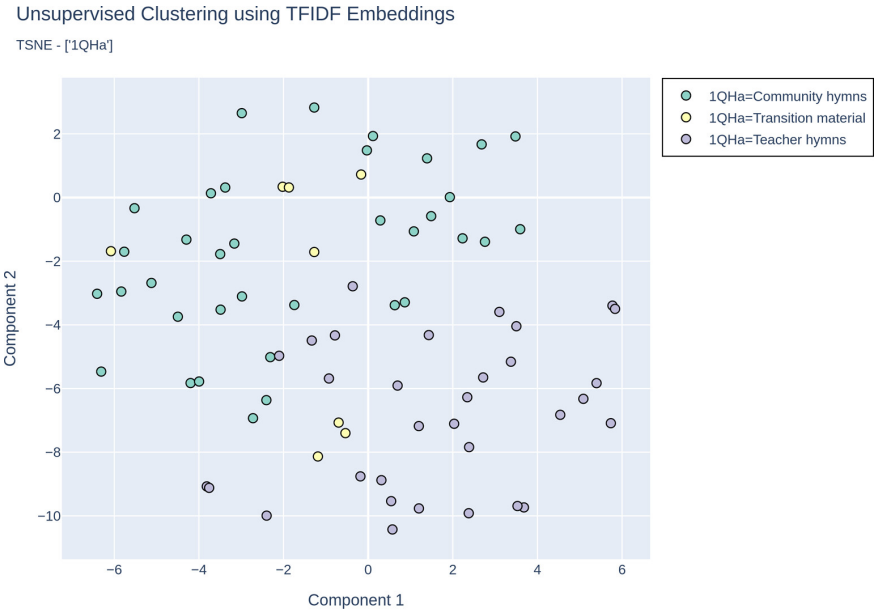
This time the picture is significantly clearer.

1. TH (yellow) and CH (black) stand in clear separation at the two ends of the dendrogram.
2. Within the black block, CH1 material (cols. 3–8) shows some tendency to cluster separately from CH2 (cols. 18–23).
3. Outliers of TH within the "black" block are 14:10–18 and the hybrid 15:32–41. Another hybrid chunk, 15:41–16:9 stands very close to the CH block and is clustered with it in the orange cluster. Clustered with it are also 11:17–25 and 12:27–34.
4. Outliers of CH within the "yellow" material are 18:25–34 and 19:38–20:12
5. The hymns of the transitional column 9 stand clearly within the CH block, while parts of column 17 appear within the TH block.

Interestingly, the "black" outliers 18:25–34 and 19:38–20:12 that stand in the TH block were also diagnosed by the trigram modelling (Fig. 1) as being close to the TH. This proximity was not noted by the TF-IDF modelling (Fig. 2) but was reinvigorated in the combined results.

Different visualizations of the data can offer a more nuanced view of the clustering. Using other scatter plots with the link above for the chunk 15:32–41, for example, will demonstrate its place within the CH rather than in the cluster of Teacher Hymns. Users may gain further insights from examining the chunks that stand close to this one and figuring out the reasons for the similarity.

54 For further details see the above noted Lourie, et al., "Semantic and Statistical Features." For Dasgupta see Dasgupta, "A Cost Function."



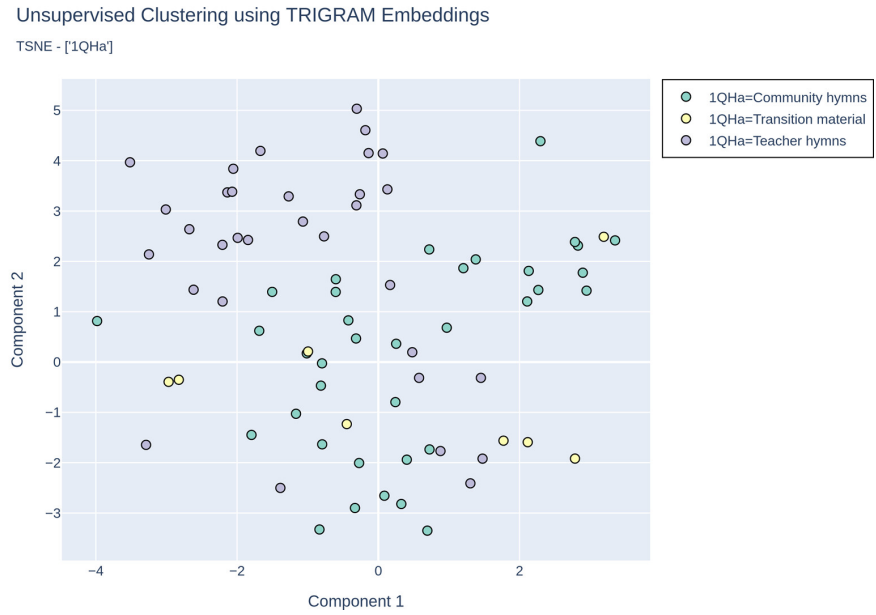


FIGURE 5 Scatter plot of character trigrams and TF-IDF embeddings

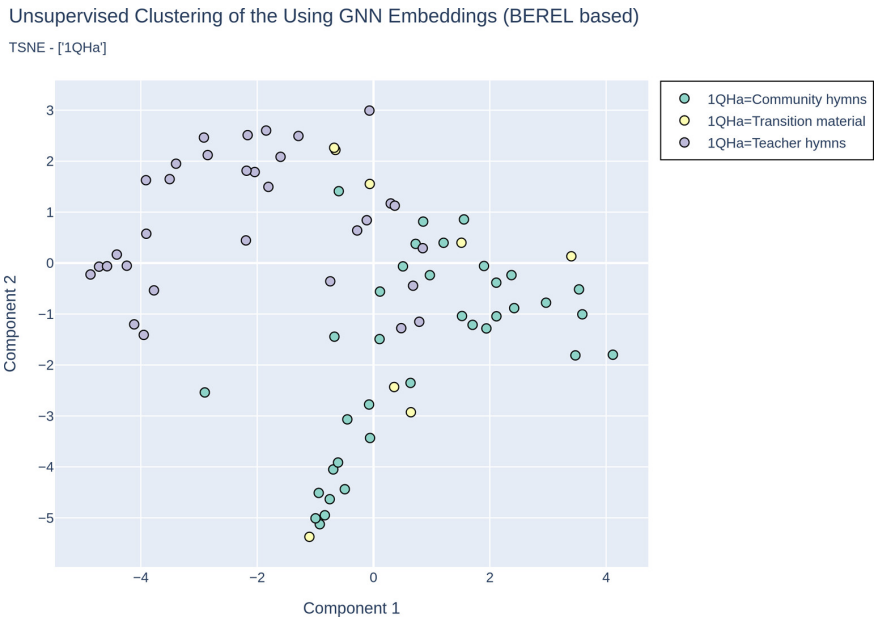


FIGURE 6 Scatter plot of the GNN embeddings

Returning to the dendrogram format, Figure 7 (see next page) presents the final clustering by the GNN.

6 Results and Discussion

The various embeddings generally confirm the distinction between Teacher (or Leader) material in 1QH^a columns 10–16, and the Community material that envelopes it. The division is marked by the TF-IDF and Trigram embeddings and receives final confirmation with the addition of the BEREL model and the GCN.

However, the TH material cannot be taken as one uniform block. The analysis was able to discern the higher resolution adopted by Tanzer, who showed some variety and hybridity within the Teacher's material. Thus the "hybrid" hymns of column 15 stand as a consistent outlier within the TH block. The chunk 14:10–18 is another consistent outlier. Other occasional outliers are: 18:25–34 and 19:38–20:12, 12:12–27, 12:34–13:7. Some of this inconsistency may be due to Maskil passages inserted throughout 1QH^a. However, even if such a maskilic redaction existed, it did not achieve far reaching unification throughout the entire scroll.

There is some evidence (Figs. 2, 7) for the distinction between CH1 and CH2 material. This distinction may be derived from the different orthography and

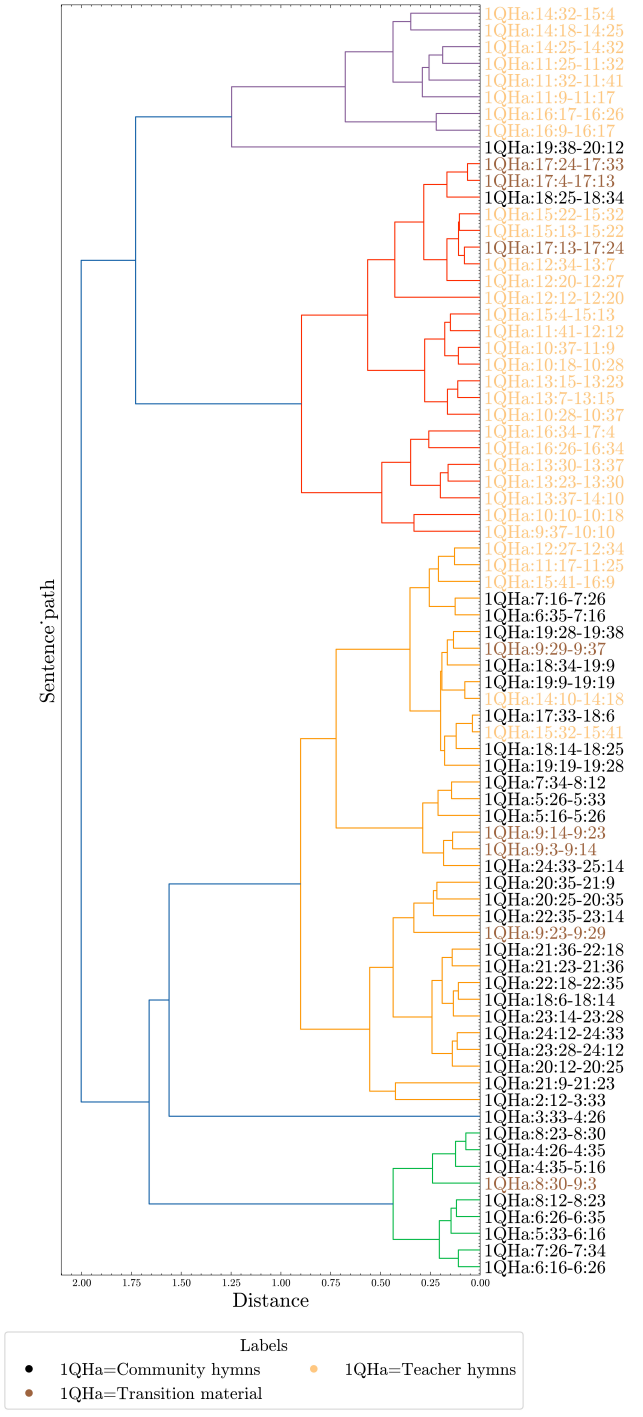


FIGURE 7
Clustering of 1QH[a] by GNN
PRODUCED BY YONATAN LOURIE

morphology of these blocks, amplified by the trigrams test, rather than from their content. However, since this distinction is retained not only in the trigrams results but also at the end product of the pipeline after other clustering systems were applied, there is good grounds to retain the CH1–CH2 distinction.

The transitional material as defined by Douglas in columns 9 and 17 is generally scattered through the dendrograms, the former more closely to the CH while the latter stands closer to TH.

Why does the algorithm place 14:10–18 apart from the TH block despite its rather clear TH traits? It mentions the communion with the angels in line 13 and the phrase "eternal planting" in line 15, in a passage that reminds of the arboreal Teacher Hymn in column 16. Examining this chunk in the scatter plot, however, will show that its separation from the TH is not as decisive as in the dendrogram.

The algorithm is found to be a good philologist. When fed with the entire text of 1QH^a and ordered to cluster it without any further instructions, the algorithm performed well. Significant for us, it did note some fine tuning of the diversity, namely the anomalous hybrid sections in column 15 and the chunk 14:10–18. The computer algorithms and LLMs generally agree with human judgment and add nuance to the bipartite division of 1QH^a.

Acknowledgments

This article was written with the support of the Israel Ministry of Science, Grant number 1001577565.

Bibliography

Bafna, Prafulla, Dhanya Pramod, and Anagha Vaidya. "Document Clustering: TF-IDF Approach." Pp. 61–66 in *International Conference on Electrical, Electronics, and Optimization Techniques* (*ICEEOT 2016*).

Berman, Joshua. "Measuring Style in Isaiah: Isaiah 34–35 and the Tiberias Stylistic Classifier for the Hebrew Bible." *VT* 71.3 (2021): 303–316.

Bühler, Axel, Gideon Yoffe, Nachum Dershowitz, Eli Piasetzky, Israel Finkelstein, Thomas Römer, and Barak Sober. "Exploring the Stylistic Uniqueness of the Priestly Source in Genesis and Exodus Through a Statistical/Computational Lens." *ZAW* 136.2 (2024): 165–190.

Dasgupta, Sanjoy. "A Cost Function for Similarity-Based Hierarchical Clustering." Pp. 118–127 in *Proceedings of the Annual ACM Symposium on Theory of Computing*. ACM, 2016.

Dershowitz, Idan, Moshe Koppel, Navot Akiva, and Nachum Dershowitz. "Computerized Source Criticism of Biblical Texts." *JBL* 134.2 (2015): 253–271.

Douglas, Michael C. "The Teacher Hymn Hypothesis Revisited: New Data for an Old Crux." *DSD* 6.3 (1999): 239–266.

Faigenbaum-Golovin, Shira, Alon Kipnis, Axel Bühler, Eli Piasetzky, Thomas Römer, and Israel Finkelstein. "Critical Biblical Studies via Word Frequency Analysis: Unveiling Text Authorship." *PLOS ONE* 20.6 (2025): e0322905.

Harkins, Angela Kim. "A New Proposal for Thinking about 1QH^{a} Sixty Years after Its Discovery." Pp. 101–134 in *Qumran Cave 1 Revisited: Texts from Cave 1 Sixty Years after Their Discovery: Proceedings of the Sixth Meeting of the IOQS in Ljubljana*, edited by Daniel K. Falk, Sarianna Metso, Donald W. Parry, and Eibert J.C. Tigchelaar. Leiden: Brill, 2010.

Harkins, Angela Kim. "Another Look at the Cave 1 Hodayot: Was CH 1 Materially Part of the Scroll 1QHodayot^{a}?" *DSD* 25.2 (2018): 185–216.

Hasselbalch, Trine. *Meaning and Context in the Thanksgiving Hymns*. Atlanta: SBL, 2015.

Holm-Nielsen, Svend. *Hodayot. Psalms from Qumran*. Aarhus: Universitetsforlaget, 1960

Huang, Lianzhe, Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. "Text Level Graph Neural Network for Text Classification." Pp. 3444–3450 in *EMNLP-IJCNLP 2019–2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*. ACL, 2019.

Jeremias, Gert. *Der Lehrer der Gerechtigkeit*. Göttingen: Vandenhoeck & Ruprecht, 1963.

Johnson, Michael B. "A Scroll Divided?: An Examination of the Wadded Bundle of 1QHodayot^{a}." *DSD* 30.2 (2022): 89–128.

Johnson, Michael B. "Look Who's Talking: Reconsidering the Speaker in the 'Teacher Hymns' (1QH^{a})." Pp. 313–341 in *Emerging Sectarianism in the Dead Sea Scrolls: Continuity, Separation, and Conflict*, edited by John J. Collins and Ananda Geyser-Fouché. Leiden: Brill, 2022.

Jokiranta, Jutta. *Social Identity and Sectarianism in the Qumran Movement*. Leiden: Brill, 2012.

Koppel, Moshe, Navot Akiva, Idan Dershowitz, and Nachum Dershowitz. "Unsupervised Decomposition of a Document into Authorial Components." Pp. 1356–1364 in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, 2011.

Kratz, Reinhard G. "The Teacher of Righteousness and His Enemies." Pp. 515–532 in *Is There a Text in This Cave? Studies in the Textuality of the Dead Sea Scrolls in Honour of George J. Brooke*, edited by Ariel Feldman, Maria Cioată, and Charlotte Hempel. Leiden: Brill, 2017.

Krispenz, Jutta. *Literarkritik und Stilstatistik im Alten Testament*. Berlin: De Gruyter, 2001.

Kuhn, Heinz-Wolfgang. *Enderwartung und gegenwärtiges Heil: Untersuchungen zu den Gemeindeliedern von Qumran mit einem Anhang über Eschatologie und Gegenwart in der Verkündigung Jesu*. Göttingen: Vandenhoeck & Ruprecht, 1966.

Licht, Jacob. *The Thanksgiving Scroll*. Jerusalem: Bialik, 1965. (Hebrew)

Lin, Yuxiao, Yuxian Meng, Xiaofei Sun, Qinghong Han, Kun Kuang, Jiwei Li, and Fei Wu. "BertGCN: Transductive Text Classification by Combining GNN and BERT." Pp. 1456–1462 in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. ACL, 2021.

Lourie, Yonatan, Jonathan Ben-Dov, and Roded Sharan. "Integrating Semantic and Statistical Features for Authorial Clustering of Qumran Scrolls," Pp. 12–21 in *Proceedings of the Second Workshop on Ancient Language Processing. The Albuquerque Convention Center, Laguna*. Ed. Adam Anderson et al.

Morawe, Günter. *Aufbau und Abgrenzung der Loblieder von Qumrân: Studien zur gattungsgeschichtlichen Einordnung der Hodajôth*. Berlin: Evangelische Verlagsanstalt, 1961.

Newsom, Carol. *The Self as Symbolic Space: Constructing Identity and Community at Qumran*. Leiden: Brill, 2004.

Newsom, Carol. "A Farewell to the Hodayot of the Community." *DSD* 28 (2021): 1–19.

Popović, Mladen, Maruf A. Dhali, Lambert Schomaker, Johannes van der Plicht, Kaare Lund Rasmussen, Jacopo La Nasa, Ilaria Degano, Maria Perla Colombini, and Eibert Tigchelaar. "Dating Ancient Manuscripts Using Radiocarbon and AI-Based Writing Style Analysis." *PLOS ONE* 20.6 (2025): e0323185.

Qimron, Elisha. *The Dead Sea Scrolls. The Hebrew Writings*. Volume 1. Jerusalem: Yad Ben-Zvi, 2010.

Radday, Yehuda T. "Two Computerized Statistical-Linguistic Tests Concerning the Unity of Isaiah." *JBL* 89.3 (1970): 319–324.

Radday, Yehuda Thomas, and Haim Shore. *Genesis: An Authorship Study in Computer-Assisted Statistical Linguistics*. Rome: Biblical Institute Press, 1985.

Rockmore Daniel, Chen Jiayi, Jebelli Mohammad J.L., Riddell Allen, Stropkay Harrison. "On the Literary Landscapes of Vector Embeddings." *Computational Humanities Research*. 2025;1.

Roorda, Dirk. "Text-Fabric: Handling Biblical Data with IKEA Logistics." *HIPHIL Novum* 5.2 (2019): 126–135.

Schuller, Eileen M. "The Hodayot Scroll." Pp. 261–273 in *The Qumran Scrolls and their World*, edited by Menahem Kister. Jerusalem: Yad Ben Zvi, 2009 (Hebrew).

Schuller, Eileen M. "Recent Scholarship on the Hodayot 1993–2010." *Currents in Biblical Research* 10 (2011): 119–162

Seker, Amit, Elron Bandel, Dan Bareket, Idan Brusilovsky, Refael Shaked Greenfeld, and Reut Tsarfaty. "AlephBERT: Language Model Pre-Training and Evaluation from

Sub-Word to Sentence Level." Pp. 46–56 in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. ACL, 2022.

Shmidman, Avi, Joshua Guedalia, Shaltiel Shmidman, Cheyn Shmuel Shmidman, Eli Handel, and Moshe Koppel. "Introducing BEREL: BERT Embeddings for Rabbinic-Encoded Language." *arXiv*: 2208.01875 (2022).

Starr, John. *Classifying the Aramaic Texts from Qumran: A Statistical Analysis of Linguistic Features*. London: Bloomsbury T&T Clark, 2017.

Stegemann, Hartmut, Eileen M. Schuller, and Carol Ann Newsom. *1QHodayot^{a}: With Incorporation of 1QHodayot^{b} and 4QHodayot^{a-f}*. DJD 40. Oxford: Clarendon Press, 2009.

Sukenik, Eliezer L. *Megillot Genuzot. Scrolls Stored in an Ancient Geniza Found in the Judean Desert. Second Survey*. Jerusalem: Bialik, 1950. (Hebrew)

Tanzer, Sarah. *The Sages at Qumran: Wisdom in the Hodayot*. Dissertation, Harvard University, 1986.

Tigchelaar, Eibert. "A Further Look at the Cave 1 Hodayot Scroll." *RevQ* 32.2 (2020): 195–211

Van Hecke, Pierre. "Computational Stylometric Approach to the Dead Sea Scrolls: Towards a New Research Agenda." *DSD* 25.1 (2018): 57–82.

Van Hecke, Pierre, and Johan de Joode. "Promises and Challenges in Designing Stylometric Analyses for Classical Hebrew." Pp. 349–374 in *Hebrew Texts and Language of the Second Temple Period*, edited by Steven Fassberg. Leiden: Brill, 2021.

Yang, Tianchi, Linmei Hu, Chuan Shi, Houye Ji, Xiaoli Li, and Liqiang Nie. "HGAT: Heterogeneous Graph Attention Networks for Semi-Supervised Short Text Classification." *ACM Transactions on Information Systems* (*TOIS*) 39.3 (2021).

Yao, Liang, Chengsheng Mao, and Yuan Luo. "Graph Convolutional Networks for Text Classification." Pp. 7370–7377 in *Proceedings of the AAAI Conference on Artificial Intelligence, 2019*.