

Text Mining Islamic Law

Christian Lange

Faculty of Humanities, Department of Philosophy and Religious Studies,
Utrecht University, Utrecht, The Netherlands
c.r.lange@uu.nl

Maksim Abdul Latif

Independent researcher, Utrecht, The Netherlands
m.a.latif@icloud.com

Yusuf Çelik

Faculty of Humanities, Department of Philosophy and Religious Studies,
Utrecht University, Utrecht, The Netherlands
y.celik@uu.nl

A. Melle Lyklema

Faculty of Humanities, Department of Philosophy and Religious Studies,
Utrecht University, The Netherlands
a.m.lyklema@uu.nl

Dafne E. van Kuppevelt

Research engineer, Netherlands eScience Center, Amsterdam,
The Netherlands
d.vankuppevelt@esciencecenter.nl

Janneke van der Zwaan

Independent researcher, Amsterdam, The Netherlands
jannekevdzwaan@gmail.com

Abstract

Digital humanities has a venerable pedigree, stretching back to the middle of the twentieth century, but despite noteworthy pioneering contributions it has not become a mainstream practice in Islamic Studies. This essay applies humanities computing

to the study of Islamic law. We analyze a representative corpus of works of Islamic substantive law (*furū' al-fiqh*) from the beginnings of Islamic legal jurisprudence to the early modern period (2nd/8th-13th/19th c.) using several computational tools and methods: text-reuse network analysis based on plain-text annotations and HTML tags, clustered frequency-based analysis, word clouds, and topic modeling. Applying machine-guided distant reading to Islamic legal texts over the *longue-durée*, we study (1) the role of the Qurʾān, (2) patterns of normative qualifications (*aḥkām*), and (3) the distribution of topics in our corpus. In certain instances the analysis confirms claims made in the scholarly literature on Islamic law, in other instances it corrects such claims.

Keywords

Islamic law – digital humanities – schools of law in Islam – Qurʾān

1 Digital Humanities and Islamic Law

Digital humanities, the practice of conducting or facilitating humanities research by using technological and computational means, has a venerable pedigree, stretching back to the middle of the twentieth century.¹ Since the turn of the millennium, this practice has undergone a tangible acceleration.² As regards Islamic Studies and its sister disciplines, including Arabic, Persian, Ottoman and Middle Eastern Studies, increasing numbers of Arabic (and to a lesser extent, also Persian and Ottoman) texts are not just scanned and made available as Pdfs, but also converted into full-text, sometimes in annotated digital format, which makes them searchable and amenable to computational analysis. In addition, the expanding digital infrastructure in which these texts are hosted³ and the concomitant increase in websites and software allow

1 See Willard McCarty, “Humanities Computing,” in *Encyclopedia of Library and Information Science*, ed. Miriam A. Drake, 2nd ed. (New York: Marcel Dekker, 2003), 1224–35.

2 Some prominent contemporary proponents tout digital humanities not just as “the next big thing”, but as “the Thing”. See Matthew K. Gold, “The Digital Humanities Moment,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis-London: University of Minnesota Press, 2012), ix-xvi, at ix, referring to two issues of the *Chronicle of Higher Education*, from 2009 and 2011, respectively. For a robust and eloquent defense of the digital humanities, see Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary Analysis* (Urbana-Chicago-Springfield: University of Illinois Press, 2013).

3 See, in particular, the Islamic Open Texts Initiative (<https://iti-corpus.github.io/>).

researchers to sift through this growing corpus of texts with ever-increasing ease.⁴ For more than a decade, scholars have harvested these digital archives and made use of their various functionalities.

It is no exaggeration to say, however, that digital humanities—if understood as not just combing the digital archive but also as tool-building and machine-guided analysis⁵—has only begun to have an impact in Islamic Studies. This is not to deny that there have been noteworthy pioneering efforts,⁶ and developments in the last decade suggest that scholars may in fact be witnessing a new dawn of digital research in Islamic Studies and related disciplines. A series of recent roundtables and conferences, as well as the appearance, in 2016, of the first edited volume dedicated to Islamicate digital humanities (henceforth:

4 For an overview, see Travis Zadeh, “Uncertainty and the Archive,” in *Digital Humanities and Islamic and Middle East Studies*, ed. Elias Muhanna (Berlin: de Gruyter, 2016), 11–64. Updates on the swiftly expanding domain of digital collections in Islamic and Middle East Studies, as well as other useful information, can be found on <https://digitalorientalist.com/>. See also the information about digital resources on the website of the American Oriental Society at <https://www.americanorientalsociety.org/digital-resources/>.

5 In the introduction to the second volume of *Debates in the Digital Humanities* (2016), Lauren F. Klein and Matthew K. Gold draw attention to the fact that over the past ten years, digital humanities has grown to include a wide range of methods and practices. See Lauren F. Klein and Matthew K. Gold, “Digital Humanities: The Expanded Field,” in *Debates in the Digital Humanities 2016*, ed. Lauren F. Klein and Matthew K. Gold (Minneapolis, MN: University of Minnesota Press, 2016), ix–xv.

6 These include María Luisa Ávila, *La sociedad hispano-musulmana al final del califato: aproximación a un estudio demográfico* (Madrid: Consejo de Investigaciones Científicas, 1985); John Nawas and Monique Bernards, “A Preliminary Report of the Netherlands Ulama Project (NUP): The Evolution of the Class of Ulama in Islam with Special Emphasis on the Non-Arab Converts (*Mawālī*) from the First through Fourth Century A.H.,” in *Law, Christianity and Modernism in Islamic Society*, ed. Urbain Vermeulen and J.M.F. Van Reeth (Leuven: Peeters, 1998), 97–107. Scholarship based on digitized prosopographical literature continues to produce impressive results. See, for example, the PUA database at the Escuela de Estudios Árabes in Granada, a database comprising more than 11,000 biographical entries on Andalusī *‘ulamā*, culled from bio-bibliographical dictionaries (<https://www.eea.csic.es/pua/>), and the resulting publications (most of them in Spanish). See also the two Mamluk prosopography projects (<http://www.mms.ugent.be/mpp/>) at Gent University (2009–16, 2016–20), which put a similar database and interface for the Mamluk period at the disposal of researchers. For a recent, computer-driven research monograph, see Cornelis van Lit, *Among Digitized Manuscripts: Philology, Codicology, Paleography in a Digital World* (Leiden-Boston: Brill, 2020).

7 Elias Muhanna (ed.), *Digital Humanities and Middle East Studies* (Berlin: de Gruyter, 2016). We leave aside here contributions to the field of Arabic computational linguistics, for an overview of which see Everhart Ditter, “Issues in Arabic Computational Linguistics,” in *The Oxford Handbook of Arabic Linguistics*, ed. Jonathan Owens (Oxford: Oxford University Press, 2013, updated version 2017: [10.1093/oxfordhb/9780199764136.013.010_update_001](https://doi.org/10.1093/oxfordhb/9780199764136.013.010_update_001)). For the proceedings of a roundtable and two panels held at the 2013 MESA meeting in

IDH), indicate that the field is consolidating.⁷ It should be noted, however, that most efforts in IDH have been directed first and foremost at creating digital corpora and indexing them with metadata, and only secondly to machine-driven analysis. To cite an example, Harvard Law School's SHARIASource, a project of obvious relevance to the computational study of Islamic law, so far has been functioning as a platform and repository first and foremost; the same may be said about other, equally impressive initiatives, such as the two projects based in Germany, *Corpus Coranicum* and *Bibliotheca Arabica*.⁸ While interest in developing novel computational methods and analytical tools for IDH is palpable in a number of current collaborative research projects,⁹ there is at present no sustained output of publications, let alone a journal or book series dedicated to IDH. Studies that foreground data-based computational analysis in Islamic Studies are still a rare phenomenon in traditional publication venues.

Although IDH may be moving in a promising direction, the field is still in its infancy. This situation carries a certain risk as a growing number of researchers, especially those of a younger generation, follow the siren call of IDH and invest time and energy in acquiring coding skills and contributing to the building of a digital infrastructure. However, up to the present, IDH cannot be said to have produced results that would seem to justify such an investment. There is, to this day, no substantial body of achievement in IDH. One may therefore be forgiven for wondering whether the perception of IDH as the vanguard of a scholarly revolution is accurate. Perhaps a recalibration of our expectations is in order?

Truth be told, the digital analysis of textual corpora, in Arabic and in other languages, tends to underwhelm, by simply confirming what we already know. Faced with results that seem intuitive or even trivial, digital humanists hesitate

New Orleans, see "Digital Humanities in Middle East Studies," *IJMES* 50 (2018), 103–39. The "Digital Islamic Humanities Project" at Brown University (<https://islamicdh.org/>) convened workshops and conferences in 2013, 2014, 2015 and 2016. In 2018, the Utrecht-based project, "Bridging the gap: Digital humanities and the Arabic-Islamic corpus" (<https://sensis.sites.uu.nl/digital-humanities/>), in collaboration with the Brown project, organized a conference (Amsterdam, 13–15 December) at the Netherlands Royal Academy of Arts and Sciences, at which an early and shorter version of this article was presented. Mention should also be made of the "Islamicate Digital Humanities Network" (<https://idhn.org/>), founded following the Amsterdam conference.

8 See <https://ilsp.law.harvard.edu/shariasource/>, and the introduction to SHARIASource by Intisar Rabb, "Digital Islamic Law: Purpose and Prospects," *IJMES* 50 (2018), 113–18. For the two German initiatives, see <https://corpuscoranicum.de/> and www.saw-leipzig.de/bibliotheca-arabica.

9 Such as the London-based KITAB project, see <http://kitab-project.org/>.

to publish the results of their computational forays in the archive.¹⁰ For this reason it is instructive to study the history of computational scholarship in other disciplines of the humanities. In the 1990s, a palpable sense of disappointment struck scholars engaged in the computational study of western literature.¹¹ As Anthony Kenny, an accomplished practitioner, noted in 1992, “after thirty-odd years of this kind of [computational] research” there were “embarrassingly few books and articles” that were “both (a) respected as an original scholarly contribution within their own discipline and (b) could clearly not have been done without a computer.”¹² IDH lives under the shadow of these disappointments, and despite all the excitement we must keep looking for “evidence of value”¹³ in IDH—including in the present study, which proposes a computer-guided investigation of Islamic jurisprudential literature.

We still lack a basic digital infrastructure for the computational study of Islamic law.¹⁴ The available digital archives provide an “illusion of totality,”¹⁵ but in fact their unexplicated selection criteria produce eclectic and discontinuous corpora, tilted towards certain periods and currents of Islamic jurisprudence. The search function included in open-access electronic libraries such as *al-Maktaba al-shāmila* or its Shi’i counterpart, *Noorlib*, offers the possibility to run simple keyword searches, at best a combination of several such keyword searches. As a rule, the tools that are embedded in these repositories to recognize Arabic trilateral roots are not explained and the underlying code is not open source. Thus, the reliability of such root prediction tools is far from certain.¹⁶ None of

-
- 10 See Matthew Thomas Miller and Sarah Bowen Savant, “Tell Me Something I Don’t Know!: The Place and Politics of Digital Methods in the (Islamicate) Humanities,” *IJMES* 50 (2018), 135–9.
- 11 See *Computers and the Humanities* 27 (1993), entitled “A New Direction for Literary Studies?”. This special issue was sparked by a paper by Mark Olsen, first delivered at the Modern Language Association Meeting (San Francisco 1991), “What Can and Cannot Be Done with Electronic Text in Historical and Literary Research”. See Olsen, “Signs, Symbols, and Discourses: A New Direction for Computer-Aided Literature Studies,” *Computers and the Humanities* 27 (1993), 309–14.
- 12 Anthony Kenny, *Computers and the Humanities* (Ninth British Library Research Lecture, London: British Library, 1992), 8.
- 13 Willard McCarty, “A Telescope for the Mind?,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis-London: University of Minnesota Press, 2012), 118.
- 14 Elias Muhanna has underscored the need to create such a basic infrastructure, and not just in the area of Islamic law. See Elias Muhanna, “What Does ‘Born Digital’ Mean?,” *IJMES* 50 (2018), 110–12, at 111. As regards Islamic law, in addition to SHARIASource, mention should be made of the CALD (Corpus of Arabic Legal Documents) database, which collects editions of legal documents from the 2nd/8th to the 9th/15th centuries. See <http://cald.irht.cnrs.fr/php/ilm.php>.
- 15 Zadeh, “Uncertainty and the Archive,” 28.
- 16 On the reliability of root prediction tools for Arabic, see Janneke van der Zwaan, Maksim Abdul Latief, A. Melle Lyklema, Dafne van Kuppevelt, and Christian Lange, “Are You Sure

the large digital repositories is equipped to facilitate clustered searches or diachronic searches over the *longue-durée*, let alone network analysis, comparison of texts, topic modeling, or adequate visualizations thereof. There are no self-serve text-analysis tools, such as *Voyant*, that work well for Arabic or other languages of the Islamic world.

Before it will be possible to study Islamic legal literature computationally, two things must be put in place: (1) a controllable, logically organized corpus that uses clear selection criteria; and (2) a computational toolbox that makes it possible to study text reuse, networks of people, places and texts, frequencies of concepts and semantic fields, distribution of topics, and more. The present study is the product of an attempt to create such an infrastructure, that is, to design a representative corpus of Islamic substantive law treatises (*furūʿ al-fiqh*) from the beginnings of Islamic jurisprudence in the 2nd/8th and 3rd/9th centuries to the 13th/19th century, and to analyze this corpus by using several advanced computational tools and methods.

1.1 *Building the Corpus*

The corpus created for this study (see Fig. 1) is comprised of fifty-five books of the *furūʿ al-fiqh* type. All these works are comprehensive in the sense that they cover the three areas of acts of worship (*ʿibādāt*), transactions between human beings (*muʿāmalāt*), and penal law (*jināyāt*). For each century of the Islamic calendar, we select one book from each of the four surviving Sunni schools, as well as one Jaʿfarī book, resulting in five *madhhab* subcorpora. All Sunni texts in the corpus are taken from *al-Maktaba al-shāmila* and the electronic library *islamweb*; the Jaʿfarī ones come from several Shiʿi online repositories.¹⁷ For convenience, we combine the 2nd/8th and 3rd/9th centuries, for which

Your Tool Does What It is Supposed To Do? Validating Arabic Root Extraction," *Digital Scholarship in the Humanities* 2019 (<https://doi.org/10.1093/llc/fqz045>).

¹⁷ See the Appendix. All data relevant to this study, including the text files that make up the corpus, are available on <https://github.com/arabic-digital-humanities/fiqh>. Additional details can be found in the metadata file. For part 2 of this article, we have also relied on existing online readers found on *islamweb.net*, *masaha.org*, and *rafed.net*, largely because these readers have relatively specific and reliable annotations (HTML tags) indicating Qurʾān verses. For example, *masaha.org* labels citations from the Qurʾān with a specific "quran" CSS class. This makes it easier to identify Qurʾān citations than it would be by relying on plain-text annotations (e.g., plain brackets around Qurʾān verses, something that might also be applied, accidentally, to a hadith, and thereby elicit false positives). The software developed for this study, as well as various plots related to sections two and four of this essay (network of Qurʾānic verses, topic modeling), are available as well. See Janneke van der Zwaan and Dafne van Kuppevelt, "arabic-digital-humanities/adhtools: v0.1.0 (Version v0.1.0)," <http://doi.org/10.5281/zenodo.2633820> (9 April 2019).

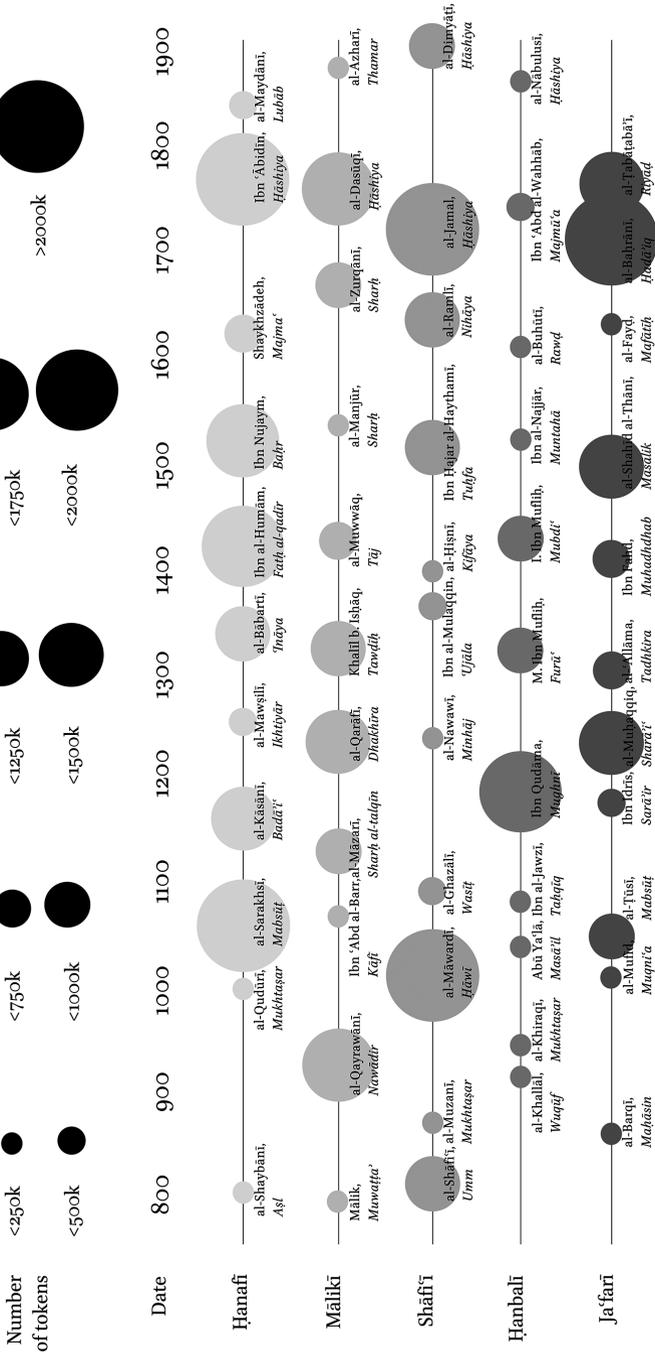


FIGURE 1 The corpus

few such books have survived. The last century to be included in the corpus is the 13th/19th century. In addition to comprehensiveness, our selection criteria are popularity, size and, of course, availability in digitized format. Inevitably, some compromises have been made, as not all of these criteria can be met in every instance.¹⁸

Consider the Ḥanafī subcorpus as an example. For the 4th/10th century, our corpus includes al-Qudūrī's *Mukhtaṣar*, even though al-Qudūrī, who died in 428/1037, is arguably an early 5th/11th-century author. Indeed, instead of al-Qudūrī's *Mukhtaṣar*, it would have been preferable to include the *K. al-Kāfi* of al-Ḥakīm al-Shahīd (d. 334/945), the first known commentary on al-Shaybānī's (d. 189/905) *K. al-Aṣl*. However, this work is not currently available in digitized format, and we were unable to find any other digitized Ḥanafī *furū' al-fiqh* book from the 4th/10th century. At the other end of the chronological spectrum, the 13th/19th century, al-Maydānī's (d. 1363/1886) *al-Lubāb* may seem an odd choice, because, compared to the other books in the corpus (average length: 895,082 tokens),¹⁹ it is short (231,605 tokens). However, a more extensive Ḥanafī *furū' al-fiqh* text from the 13th/19th century is not available digitally at the present time. The only important exception is Ibn 'Abidīn's (d. 1252/1836) celebrated *Ḥāshiya*, but in our Ḥanafī subcorpus—again this is not unproblematic—it represents the 12th/18th century.

There are other problems. By limiting Ḥanafī *furū' al-fiqh* literature to one book per century, the Ḥanafī tradition is stripped of much of its richness and complexity. By cutting a single path through a veritable jungle of texts, the subcorpus of Ḥanafī texts included in our corpus suggests a one-directional, linear development of Ḥanafī jurisprudence. The Ḥanafī textual heritage, however, is not a seamless whole, not even when the scope is restricted to *furū' al-fiqh* texts. There are different genres (from textbooks [*matns*] and summaries [*mukhtaṣars*] to commentaries [*sharḥs*] and super-commentaries or glosses [*ḥāshiyas*]) as well as multiple regional traditions within Ḥanafī *furū' al-fiqh*.²⁰

18 Strictly speaking, the earliest Ja'farī work included in the corpus, al-Barqī's *K. al-Maḥāsin*, is not a work of *fiqh*, but rather a compilation of hadiths. Also, it is transmitted incompletely and lacks important sections. See *EP*², s.v. "Al-Barqī" (Charles Pellat). This special character of the *K. al-Maḥāsin* also explains why textmining it produces results that are different from those produced by textmining the other texts in our corpus. See below, Fig. 2.4 and *passim*.

19 By token—the unit of measurement that is recognizable to a computer—we mean the unbroken string of Arabic characters between two spaces in a text. For example, the chain of characters *f-l-l-m-'-m-n-y-n*, in Arabic consists of four words: the conjunction *fa-*, the preposition *li-*, the definite article *al-*, and the plural noun *mu'minīn*. However, in our corpus these four words form a single token.

20 On the Ḥanafī *furū' al-fiqh* tradition, reconstructed on the basis of Ibn Quṭlūbughā's (d. 879/1474) *al-Taṣḥīḥ wa'l-tarjīḥ*, see Talal Al-Azem, *Rule-Formulation and Binding Precedent in*

Thus, it is risky to rely unconditionally on the subcorpus in order to derive synthetic insights about Ḥanafī legal thought. This risk points to a fundamental tension, perhaps unresolvable, between the approach of humanities computing and that of the traditional humanities. In the computational approach, the goal is to identify basic patterns and structures in large data sets (“devices, themes, tropes—or genres and systems”²¹), at the expense of specifics. By contrast, in the traditional humanities approach, scholars seek to identify possible biases of (usually) single texts and of their own reading practices, especially the danger of glossing over difference and diversity within a tradition. Hence the often-voiced injunction to use distant reading techniques, whether machine-guided or not, in combination with close reading, in an effort to inform and control these techniques—a hermeneutic circle for the digital age. In such circumstances, is it wise to advocate for a computational, data-driven analysis of Islamic legal literature? What questions can digital humanities meaningfully ask about Islamic law? We propose here that we should move forward by trial and error. The balance between gains and risks can only be determined at the end of this exploration, perhaps only after several more studies such as the present one have been conducted.²²

Our corpus includes approximately 49,300,000 tokens²³ and, assuming a multiplier of 1.5,²⁴ around 74,000,000 words. It would be hasty, perhaps, to speak of this corpus in terms of “big data”, but in comparison with, say, the corpus of digitized classical Greek and Roman texts assembled in the well-known Perseus Digital Library (68,925,971 words, according to the library’s website),²⁵ it is not insignificant.

The five subcorpora include: (1) Ḥanafī texts: 13,006,501 tokens; (2) Shāfi‘ī texts: 10,495,516 tokens; (3) Mālīkī texts: 9,224,967 tokens; (4) Ḥanbalī texts:

the Madhhab-Law Tradition: Ibn Quṭlūbughā’s Commentary on The Compendium of Qudūri (Leiden-Boston: Brill, 2016), 50–101, 227–233.

21 Franco Moretti, “Conjectures on World Literature,” in idem, *Distant Reading* (London-New York: Verso, 2013), 49.

22 To echo Willard McCarty, “failure is our most important product, partly for the shock-value, as antidote to the hype of pervasive techno-triumphalism, but also to stress that computing is an ongoing, never ending *experimental* process.” See Willard McCarty, “Getting There from Here: Remembering the Future of Digital Humanities,” *Literary and Linguistic Computing* 29:3 (2014), 284.

23 For the definition of a token, see above, note 19.

24 Based on a sample count of 1,000 tokens in a standard *furū’ al-fiqh* text.

25 <http://www.perseus.tufts.edu/hopper/collections>.

5,051,702 tokens; and (5) Ja'fari texts: 11,450,826 tokens. The five largest texts in the corpus are al-Baḥrānī (Shi'i, d. 1186/1772), *al-Ḥadā'iq al-nāḍira* (3,599,589 tokens); al-Māwardī (Shāfi'i, d. 450/1059), *al-Ḥawī al-kabīr* (2,713,331 tokens); the *Ḥāshiyā* of Ibn 'Ābidīn (Ḥanafī, d. 1836, 2,550,368 tokens); al-Sarakhsī (Ḥanafī, d. ca. 482/1090), *al-Mabsūṭ* (2,494,216 tokens); and the *Ḥāshiyā* of Sulaymān b. 'Umar al-Jamal (Shāfi'i, d. 1204/1789, 2,251,387 tokens). By contrast, the five smallest texts in the corpus are al-Khallāl (Ḥanbalī, d. 311/923), *al-Wuqūf wa'l-tarajjul* (32,349 tokens); al-Khiraqī (Ḥanbalī, d. 334/945–6), *Mukhtaṣar* (36,539 tokens); al-Qudūrī (Ḥanafī, d. 428/1037), *Mukhtaṣar* (43,452 tokens); al-Nawawī (Shāfi'i, d. 676/1277), *Minhāj al-ṭālibīn* (88,670 tokens); and the *Ḥāshiyā* of al-Nābulusī (Ḥanbalī, d. 1319/1901) (118,608 tokens). These numbers suggest that, in terms of words written, the Ḥanafīs are the most prolix of the five schools, and the Ḥanbalīs are the least prolix. This inference is based on the assumption that the corpus is representative with respect to the length of texts, one of the selection criteria of our corpus.

The total number of tokens per century in the corpus is as shown in Table 1:

Again assuming the representativeness of the corpus, we infer from these numbers that the length of *furū' al-fiqh* texts increased substantially in the 5th/11th century. The texts of the following centuries, as a rule, do not exceed in length the 5th/11th-century texts. Only in the 12th/18th century do we find texts of greater length than those written in the 5th/11th century, including the three aforementioned texts, that is, al-Baḥrānī's *al-Ḥadā'iq al-nāḍira* and the two *Ḥāshiyas* of Ibn 'Ābidīn and al-Jamal.

TABLE 1 Number of tokens per century

2nd/8th and 3rd/9th century	1,854,587 tokens
4th/10th century	2,145,490 tokens
5th/11th century	6,418,738 tokens
6th/12th century	3,419,602 tokens
7th/13th century	5,254,132 tokens
8th/14th century	4,304,694 tokens
9th/15th century	5,065,767 tokens
10th/16th century	4,556,013 tokens
11th/17th century	2,960,586 tokens
12th/18th century	10,430,366 tokens
13th/19th century	2,819,537 tokens

1.2 *Analyzing the Corpus*

After eliminating paratext (front and end matter, and footnotes) from the texts in the corpus, and annotating major divisions (parts [*kitābs*], chapters [*bābs*] and sections [*faṣls*]), we adapted BlackLab, a text search engine developed at the Dutch Language Institute (INL),²⁶ to our corpus. The resulting BlackLab Arabic Digital Humanities (henceforth: BlackLab ADH, <http://arabic-dh.hum.uu.nl/corpus-frontend/>) allows for fast, complex searches of the corpus, by word(s), stem(s) or root(s), while also making it possible to filter and group searches by century, geographic region, and law school. These features make it easy, for example, to search for the frequencies of specific terms, ratios, or ontologies of terms over the centuries, or to compare how law schools differ or overlap in the way they use certain concepts.

Counting the frequency with which a specific term, or a string or cluster of terms, appears in a text, of course, does not tell us *how* this term is used, that is, whether it is used positively or negatively, or in a technical, non-technical, argumentative, or regurgitative manner. The fact, for example, that one jurist frequently uses the term *ijtihād* (“personal juristic reasoning”) does not prove that he approves the concept; it is possible that he is arguing polemically *against* the use of *ijtihād*. In the examples discussed below, we reflect on such ambiguities and propose ways to overcome them. Beyond the frequency-based analysis powered by BlackLab ADH, we also employ a number of other techniques, such as text-reuse network analysis based on plain-text annotations and HTML tags, word embedding, parsimonious word clouds, and topic modeling.

The fact that such techniques are largely untested on the Arabic legal corpus means that we should be modest about the results of this study, which is exploratory—a point that bears emphasis. We do not aspire to develop a new data-driven, digital grand theory of Islamic jurisprudence. As Elias Muhanna reminds us, “instead of seeking out the latest digital methodologies and tools in the hope that they will unlock the secrets of our archives, scholars would be better served by asking the same questions they would in an analogue project.”²⁷ With due modesty, therefore, we seek to demonstrate the usefulness of humanities computing for the study of Islamic law by digitally analyzing the corpus to generate quantitative answers to the following three, basic questions. What role is played by the Qur’ān in Islamic jurisprudence, and how do the law schools differ in the way they rely on and refer to it (see below, 2.)? What kind of normative categorizations, or *aḥkām*, are salient in *furū’ al-fiqh*, and how restrictive or permissive is *furū’ al-fiqh*, considered both as a whole and per law

²⁶ <https://inl.github.io/BlackLab>.

²⁷ Muhanna, “What Does ‘Born Digital’ Mean?,” 112.

school (see below, 3.)? What are the fundamental concerns and topics of *furū' al-fiqh*, according to the entire corpus and each of the law schools, over the *longue-durée* (see below, 4.)?

Not in all instances, however, do we ask the “same questions” asked by previous scholars, as Muhanna proposes. The ability of IDH to digest large amounts of texts that no single scholar could ever hope to master, in our view, *does* enable us to think of and to ask questions that previously were unanswerable, and it would be a missed opportunity not to do so.²⁸ Ideally, IDH will help us to move beyond traditional frameworks and to read our corpus, in Anver Emon’s apt phrasing, “openly and freely but with attention to the text and its limits... [without] the hegemonic hoof of philology... stamping one’s back”.²⁹ In this spirit, we seek to provide a corrective to, or to confirm, some of the claims made in the scholarly literature on Islamic law.³⁰

2 The Qur’ān in Muslim Jurisprudence

Scholars usually consider the Qur’ān and the hadith to be the two textual sources of Islamic law, a legal system that Bernard Weiss, one of its most well-known

²⁸ Some critics contend that the *only* way forward for digital humanities is to abandon “traditional objects of study” and to focus on examining “large amounts of simple linguistic features”. See Olsen, “Signs, Symbols, and Discourses,” 309.

²⁹ Anver M. Emon, “On Reading *Fiqh*,” in Anver Emon and Rumea Ahmed (eds.), *The Oxford Handbook of Islamic Law* (Oxford: Oxford University Press, 2018), 56.

³⁰ We base ourselves primarily on the following selection of well-known studies of Islamic law written in the second half of the 20th century: Joseph Schacht, *An Introduction to Islamic Law* (Oxford: Clarendon Press, 1964); Noel J. Coulson, *A History of Islamic Law* (Edinburgh: Edinburgh University Press, 1964); idem, *Conflicts and Tensions in Islamic Jurisprudence* (Chicago-London: The University of Chicago Press, 1969); Bernard G. Weiss, *The Spirit of Islamic Law* (Athens-London: The University of Georgia Press, 1998); Aron Zysow, *The Economy of Certainty: An Introduction to the Typology of Islamic Legal Theory* (Atlanta, Georgia: Lockwood Press, 2013). We also rely on a number of recent state-of-the-art overviews: Wael B. Hallaq, *Shari’a: Theory, Practice, Transformations* (Cambridge: Cambridge University Press, 2009); Rudolph Peters and Peri Bearman (eds.), *The Ashgate Research Companion to Islamic Law* (Farnham: Ashgate, 2014); Knut S. Vikør, *Between God and Sultan: A History of Islamic Law* (London: Hurst & Company, 2005); Anver Emon and Rumea Ahmed (eds.), *The Oxford Handbook of Islamic Law* (Oxford: Oxford University Press, 2018); Khaled Abou El Fadl, Ahmad Atif Ahmad and Said Fares Hassan (eds.), *The Routledge Handbook of Islamic Law* (London: Routledge, 2019). We also refer to various entries dealing with Islamic law topics in several standard encyclopaedias: Stanley N. Katz (gen. ed.), *Oxford International Encyclopaedia of Legal History* (Oxford: Oxford University Press, 2009); *EI²* (Leiden: E. J. Brill, 1960–2005); *EI³* (Leiden-Boston: Brill, 2007-). In addition, we engage the specialized literature as much as possible, in order to enrich distant with close reading.

interpreters, has described as having a “textualist bent”.³¹ Scholarly discussion has largely revolved around the question of the origin and authenticity of these sources, and around the hermeneutical principles that scholars of *uṣūl al-fiqh* developed to interpret them.³² By comparison, little attention has been devoted to analyzing the reliance of Muslim jurists on the Qurʾān and the hadith in quantitative terms. While Western scholars such as Alexander Knysh occasionally highlight that Muslim jurists “[p]eriodically... made attempts to restrict the discretionary power of the judges by inviting them to ‘return’ to the letter of the Qurʾān,”³³ Salafī authors such as Nāṣir al-Dīn al-Albānī (d. 1999) criticize *madhhab* traditionalists for relying on the Qurʾān too little, or only indirectly, allowing the opinions of later *madhhab* authorities to accumulate, layer after layer, on top of, and eventually covering up, Islamic scripture.³⁴ Can we test such global intuitions on the basis of our data-driven, quantitative approach? For example, were some schools of law more likely than others to refer to the Qurʾān?³⁵ Do certain schools of law privilege certain parts or verses of the Qurʾān? Are there significant differences between individual jurists in terms of how they rely on the Qurʾān?

A statistical analysis of our corpus shows that Shāfiʿī authors refer to the Qurʾān most frequently, in 10,619 citations.³⁶ That is, a Qurʾānic verse is cited after an average of every 982 tokens or an average of 4.9 pages in a printed edition of a Shāfiʿī work.³⁷ Shāfiʿīs are followed by Jaʿfarīs (9,133 citations, after

31 See Weiss, *Spirit*, 38–65.

32 For the scholarly concern with the origin and authenticity of the two sources, see, for example, Herbert Berg, “The Divine Sources,” in *The Ashgate Research Companion to Islamic Law*, 27–40. Fine examples of studies on the hermeneutics of the two sources are A. Kevin Reinhart, “Jurisprudence,” in *The Blackwell Companion to the Qurʾān*, ed. Andrew Rippin (New York, NY: Wiley-Blackwell, 2008), 434–449; Amr Osman, “The Qurʾān and the Hadith as Sources of Islamic Law,” in *The Routledge Handbook of Islamic Law*, 127–40.

33 Alexander Knysh, “Multiple Areas of Influence,” in *The Cambridge Companion to the Qurʾān*, ed. Jane Dammen McAuliffe (Cambridge: Cambridge University Press, 2006), 217.

34 See Emad Hamdeh, “Qurʾān and Sunna or the *Madhhabs*?: A Salafī Polemic Against Islamic Legal Tradition,” *Islamic Law and Society* 24 (2017), 211–53.

35 In the following, we limit ourselves to an analysis of the Qurʾān’s footprint in *fiqh*, to the exclusion of hadith. Until the words of the Prophet are stringently annotated in electronic corpora, it will be difficult to disambiguate them from those of the Companions or from other historical sayings and anecdotes. Hence, a direct text re-use detection via conventional search methods (e.g. fuzzy search) will result in numerous false positives. An element of our present method, that is, the detection of text re-use based on HTML tags extracted from popular digital readers, may assist in creating a Deep Learning training set for a more stable entity extraction (in this case hadith).

36 This count is based on the number of verified citations, not on the number of annotations. Not every annotation is necessarily a citation, since a digital scribe may sometimes misplace annotations, erroneously indicating that a passage is a citation from the Qurʾān.

37 For this calculation, we assume an average length of six tokens for each cited Qurʾānic verse, and an average length of 200 tokens per page of a printed edition.

every 1247 tokens or 6.2 pages), Ḥanbalīs (3,888 citations, after every 1,293 tokens or 6.4 pages), Ḥanafīs (8,908 citations, after every 1,454 tokens or 7.2 pages) and Mālikīs (4,726 citations, after every 1,946 tokens or 9.7 pages).³⁸ It is commonly asserted that Ḥanafī jurists are lax in their reliance on the Qurʾān, but in our corpus it is the Mālikī authors who rely the least on Qurʾānic evidence.

The Mālikī tendency not to refer to the Qurʾān may be related to the centrality of the figure of Mālik b. Anas (d. 179/795) and his *Muwattʾā* in the Mālikī school. To illustrate this hypothesis computationally, we draw attention to one of the most sizable Mālikī books in the corpus, Ibn Abī Zayd's (d. 386/966) *al-Nawādir wa'l-ziyādāt 'alā mā fi'l-Mudawwana* (1,715,091 tokens). There are 198 Qurʾān citations in this text (that is, occurring after an average of every 8,657 tokens or 43.3 pages). By comparison, a similarly sized work (1,464,584 tokens), *Badā'ir al-ṣanā'ir* by the Ḥanafī jurist al-Kāsānī (d. 587/1191), contains 2,113 citations (that is, occurring after an average of every 688 tokens or 3.4 pages). However, the expression *qāla Mālik* occurs 4,151 times in Ibn Abī Zayd's text, and only 123 times in that of al-Kāsānī. This reliance on Mālik's opinions recalls an anecdote reported in Muḥammad al-Rāṭī's (d. 853/1449) *Intiṣār al-faqīr*, which treats the merits of Mālik b. Anas and of Mālikism. As al-Rāṭī relates, the 'Abbasid caliph al-Manṣūr (r. 136–58/754–75) intended to hang the text of the *Muwattʾā*, written in gold letters, on the walls of the Ka'ba, an idea obliquely approved by Mālik himself.³⁹ The extraordinary veneration accorded to Mālik by his followers has been studied by Abdelmajid Turki,⁴⁰ while Robert Brunschvig has written about polemical reactions against Mālik by non-Mālikī authors, including the accusation that he elevated the practice (*'amal*) of the people of Medina above Islamic scripture.⁴¹ Our computer-aided analysis provides quantitative support for Turki's and Brunschvig's conclusions.⁴²

38 Each of these citations can be read at <https://quran-in-fiqh.hum.uu.nl/> by using the text re-use search and visualization tools. An export of the dataset containing all the discovered verses and their relevant pages can be accessed on Yoda (Utrecht University's data publication platform). See Yusuf Çelik and Christian Lange, "Computing and Visualizing the Qurʾān's Footprint in Fifty-Five Works of Islamic Substantive Law." 1.0, Utrecht University, 2021, doi:10.24416/UU01-AADM01.

39 See Yasin Dutton, *Original Islam: Malik and the Madhhab of Madina* (London-New York: Routledge, 2006), 73.

40 See Abdelmajid Turki, "La vénération pour Mālik et la physionomie du malikisme andalou," *Studia Islamica* 33 (1971), 41–65.

41 See Robert Brunschvig, "Polémiques médiévales autour du rite de Mālik," *Al-Andalus* 15 (1950), 377–413.

42 We must, however, differentiate carefully between Mālikī authors. Thus, while an author like Ibn Abī Zayd, in his *al-Nawādir wa'l-ziyādāt*, cites the Qurʾān infrequently (as per the above, 198 times, that is, after every 8,657 tokens per average), another author, al-Qarāfi (d. 684/1285) in his *al-Dhakhira*, does so frequently (1,330 citations, that is, after every 1,060 tokens per average).

TABLE 2 Most-frequently cited Qur'anic verses in the corpus, all five schools combined

1	Q 2:196 (701) [42/55]	Rules concerning the pilgrimage: obligatoriness of either pilgrimage or expiation; shaving the head during the pilgrimage and exceptions to this requirement.
2	Q 5:6 (666) [41/55]	Ritual purity required for prayer; <i>tayammum</i> .
3	Q 2:282 (515) [40/55]	Contractual debts should be recorded in writing; the proper ways of recording debt; importance of witnesses (two men or one man and two women).
4	Q 4:11 (456) [38/55]	Rules concerning inheritance: inheritance of sons, daughters, and parents.
5	Q 4:43 (455) [40/55]	Impermissibility of praying intoxicated or in a state of ritual impurity.
6	Q 5:95 (434) [37/55]	Prohibition of killing game in the sacred state of pilgrimage (<i>ihram</i>); expiatory acts by someone who does so intentionally.
7	Q 4:23 (410) [34/55]	Prohibitions regarding incestuous marriages.
8	Q 4:92 (340) [39/55]	Punishment for unintentional homicide: blood-money; freeing of a slave; fasting for two consecutive months.
9	Q 2:237 (300) [36/55]	Divorce before consummation of marriage; injunction to forego divorce and to be generous.
10	Q 2:233 (283) [35/55]	Suckling and weaning; wet-nurses.

In addition to asking ourselves *how often* jurists refer to the Qur'an, we also should ask *which* Qur'anic verses are cited by Muslim jurists and in which areas of Islamic jurisprudence this happens. In Tables 2 and 3, we list the most frequently cited Qur'an verses in our corpus. *Sūra* and verse numbers are followed, in brackets, by the total number of quotations of the verse and, in square brackets, the total number of texts that reference a verse. First, we give the most frequently cited verses in the entire corpus, all five law schools combined.⁴³

43 Calculation of the number of quotations is based on a set of rules. A citation that can

Table 2 sheds light on the topics with regard to which the jurists in our corpus most frequently invoked the Qurʾān, that is, the topics for which they felt Qurʾānic evidence was relevant. Next to the verse in Islamic contract law (Q 2:282), Qurʾānic verses relating to ritual practice (pilgrimage and prayer) and family law (marriage, divorce, inheritance) figure prominently. Given that worship, as noted by Bernard Weiss, “is not a realm in which one expects to find the accent on human freedom”,⁴⁴ it makes sense that the jurists rely on the Qurʾān in this area. It is particularly in the area of the *ʿibādāt*, ritual actions, that Muslim jurists treat God’s logic as inscrutable, as Kevin Reinhart has shown.⁴⁵ More broadly, *tawqīf*, that is, reliance on revelation in formulating the law, is declared essential by Muslim jurists in areas that are “non-rational”, or to use the language of *fiqh*, “divinely imposed” (*muqaddar*). As Christian Lange has demonstrated, the schools of law define the scope of the “divinely imposed norms (*muqaddarāt*)” in Islamic law differently; generally speaking, however, the *muqaddarāt* are understood to refer to the *ʿibādāt* as well as to all “numerical norms” (the so-called *maqādīr*, e.g. fasting *ten* days to make up for not making an offering during *ḥajj*, see Q 2:196; the permission to divorce *twice* by repudiation, see Q 2:229; the ratios according to which inheritance is to be divided, see Q 4:11).⁴⁶ The fact that the verses in Table 2 touch so closely on the area of the *muqaddarāt* and *maqādīr* suggests that Muslim jurists understood that not all areas of the law are amenable to human reasoning to the same degree.

That this holds true not only for the corpus as a whole but also for individual schools is demonstrated by Table 3, which shows the ten most frequently cited verses in each of the five schools.

be linked unequivocally to a single Qurʾānic verse is counted as one instance of a Qurʾān quotation. In cases in which a citation can be linked to multiple verses, each verse is counted separately. For example, the expression “I am better than he” (*anā khayrun minhu*) can be linked to two verses in the Qurʾān: 7:12 and 38:76. Accordingly, if a scholar cites this expression, two verses are counted. Citations shorter than three tokens are ignored because they skew the final numbers.

44 Weiss, *Spirit*, 146.

45 Kevin Reinhart, “Ritual Action and Practical Action: The Incomprehensibility of Muslim Devotional Action,” in *Islamic Law in Theory: Studies in Honor of Bernard Weiss*, ed. Kevin Reinhart and Robert Gleave (Leiden-Boston: Brill, 2014), 76 (al-Nazzām), 80 (the Muʿtazila), 90 (Abū Ḥanifa and al-Shāfiʿī).

46 Christian Lange, “Sins, Expiation, and Non-Rationality in Ḥanafī and Shāfiʿī *Fiqh*,” in *Islamic Law in Theory*, 152–159.

TABLE 3 Most frequently cited Qur'anic verses in the corpus, according to individual schools

#	Ḥanafīs	Shāfi'īs	Mālikīs	Ḥanbalīs	Ja'farīs
1	Q 2:196 (185) [10/11]	Q 2:282 (199) [10/11]	Q 5:6 (109) [10/11]	Q 2:196 (75) [4/11]	Q 2:196 (189) [9/11]
2	Q 5:6 (165) [11/11]	Q 5:6 (187) [10/11]	Q 4:43 (73) [10/11]	Q 5:6 (57) [4/11]	Q 5:6 (148) [8/11]
3	Q 2:233* (129) [9/11]	Q 2:196 (181) [10/11]	Q 2:282 (72) [9/11]	Q 4:11 (54) [4/11]	Q 4:11 (115) [8/11]
4	Q 4:11 (124) [8/11]	Q 5:95 (145) [10/11]	Q 2:196 (71) [10/11]	Q 5:89 (47) [3/11]	Q 4:23 (102) [7/11]
5	Q 4:23 (120) [9/11]	Q 4:43 (143) [10/11]	Q 4:11 (62) [9/11]	Q 5:95 (45) [4/11]	Q 2:282 (99) [9/11]
6	Q 4:43 (106) [9/11]	Q 2:229* (107) [10/11]	Q 4:23 (56) [6/11]	Q 4:92 (44) [4/11]	Q 4:43 (98) [8/11]
7	Q 2:282 (103) [9/11]	Q 65:2* (105) [10/11]	Q 5:95 (47) [8/11]	Q 2:282 (42) [4/11]	Q 5:95 (98) [9/11]
8	Q 5:95 (99) [7/11]	Q 4:23 (105) [10/11]	Q 2:187 (41) [8/11]	Q 2:228 (37) [6/11]	Q 3:97* (91) [9/11]
9	Q 5:89 (97) [10/11]	Q 2:228 (103) [10/11]	Q 4:6* (40) [6/11]	Q 2:187 (36) [4/11]	Q 2:237* (88) [8/11]
10	Q 4:92 (90) [10/11]	Q 4:11 (101) [10/11]	Q 4:25 (40) [8/11]	Q 4:24* (36) [3/11]	Q 4:92 (81) [9/11]

Table 3 points to a large overlap of verses across all five schools, including the Ja'farīs. The few tangible differences are indicated by asterisks, which tag verses that are among the ten most frequently cited verses of one law school but not of the other schools. Ḥanafīs, for example, show a pronounced interest in Q 2:233, a verse with rules relating to suckling and weaning, while Shāfi'īs pay disproportional attention to Q 2:229, a verse regulating divorce.

Next, we visualize the results of our analysis in an interactive network with the help of an open access online tool called "Footprinter" (<https://quran-fiqh.hum.uu.nl/>). In this network, grey dots represent Qur'anic verses; the more often a verse is cited by different authors, the closer to the center of the network is the dot. Colored nodes represent texts in our *furū' al-fiqh* corpus, with Ḥanafī texts appearing in orange, Shāfi'ī texts in light blue, Ḥanbalī texts

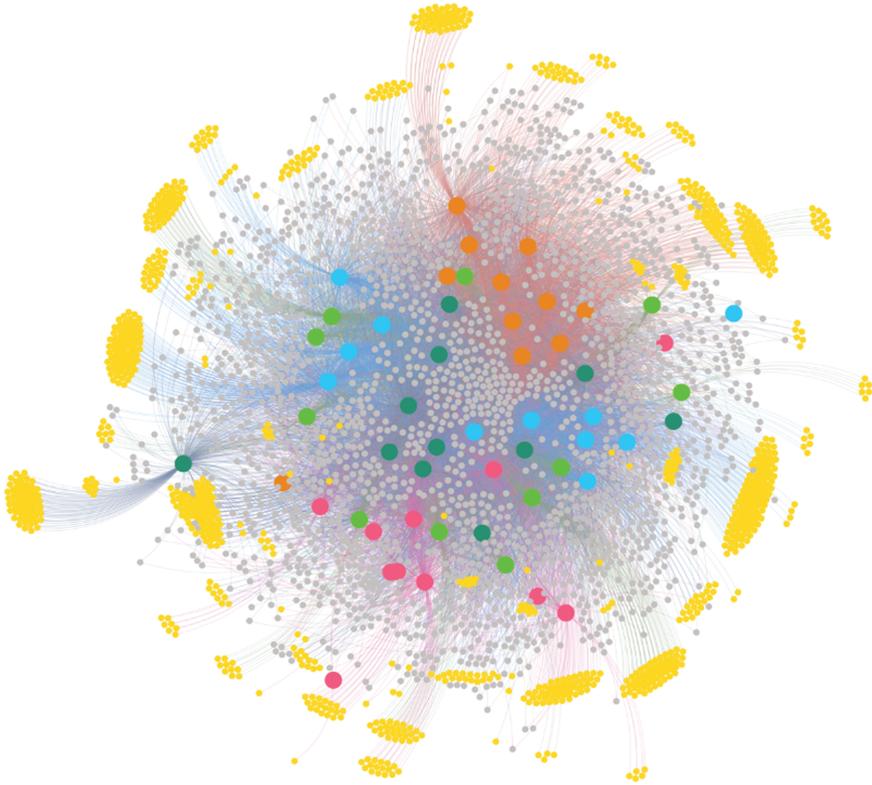


FIGURE 2.1 Overview of interactive network

in purple, Mālikī texts in green, and Jaʿfarī texts in olive (Fig. 2.1). Zooming in allows for identification of individual texts (Fig. 2.2). Qurʾānic verses can be displayed in isolation from other verses (Fig. 2.3). Gold dots represent verses that are only cited by a single author; they cluster, in the form of islands, around the margin of the network (Fig. 2.4). Several other selection criteria and filters can be applied. A built-in reader allows users to follow verses back to the texts in which they are cited.

There are 2,954 dots (verses) in the network, which means that almost half (47%) of *all* Qurʾānic verses (6,236 verses in the standard Egyptian edition) are quoted in our corpus.⁴⁷ This is significant because, in theory, only 350 to

47 Certain citations refer to distinct expressions in the Qurʾān and are unequivocally linked to a single verse. However, some citations found in the corpus contain expressions that are recurrent in the Qurʾān, e.g., *fa-sʾalū ahla ʾl-dhikri*. Such citations can be linked to more than one verse, in this case 16:43 and 21:7, both of which are included in the final total of 2,954 verses. However, there are not too many such instances. Including only citations that have a single source in the Qurʾān, we count 2,461 verses on which the jurists rely.

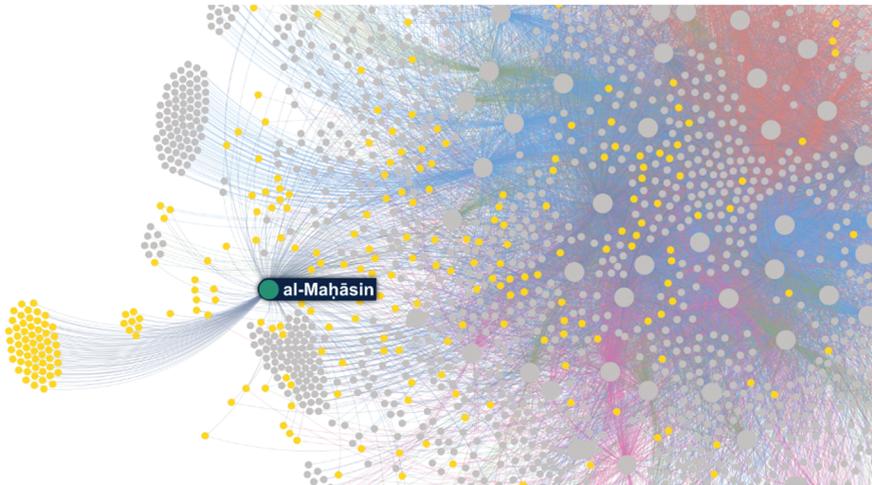


FIGURE 2.4 Al-Barqī's *al-Mahāsin* in the interactive network

of rule formation in Islamic law; they add that only about 80 verses are legal in a narrow sense.⁴⁹ The fact that Muslim jurists quoted such a large array of verses does not mean, of course, that they considered all of them relevant in a legal sense, or relied on them for the purposes of legal reasoning. However, the fact that 2,954 verses are quoted demonstrates that the *furū' al-fiqh* literature interacts with the Qur'ān in ways that are multifunctional and not limited to the immediately "legally relevant" verses. For example, Q 112:1 ("Say: He is God, One"), a verse without direct legal import, is quoted frequently and for different purposes in the corpus (132 times by 37 different authors, see Fig. 2.3). Muḥammad Ibn Mufliḥ (d. 763/1362) and Sulaymān b. 'Umar al-Jamal (d. 1204/1789), for example, discuss whether a person may use a lavatory with a *dīnār* embossed with Q 112:1.⁵⁰ In the introduction to his massive work, al-Baḥrānī mentions that according to some unnamed Akhbārī scholars, the Qur'ān must be interpreted only on the basis of the opinions of the Prophet and the Imams (the *aṣḥāb al-iṣma*), even with regard to such a straightforward

49 We follow here Joseph E. Lowry, *Early Islamic Legal Theory: The Risāla of Muḥammad ibn Idrīs al-Shāfi'i* (Leiden-Boston: Brill, 2007), 207, who refers to Coulson, *History*, 12, and Wael Hallaq, *A History of Islamic Legal Theories: An Introduction to Sunni Uṣūl al-Fiqh* (Cambridge: Cambridge University Press, 1997), 10. See further Reinhart, "Jurisprudence," 436; Knysh, "Multiple Areas of Influence," 215.

50 Muḥammad Ibn Mufliḥ, *K. al-furū'*, 6 vols. (Beirut: Dār al-Kutub al-'Ilmiyya, 1997), 1:84; Sulaymān b. 'Umar al-Jamal, *Hāshiyat al-Jamal 'alā sharḥ al-Manḥaj*, ed. 'Abd al-Razzāq Mahdī, 8 vols. (Beirut: Dār al-Kutub al-'Ilmiyya, 1434/2013), 1:120.

verse as Q 112:1.⁵¹ Al-Qarāfi (d. 684/1285) relates from Ibn Rushd al-Jadd (d. 520/1126) that Mālik b. Anas disliked the three-fold recitation of *sūra* 112 in a single genuflection unit (*rak'a*) during prayer. According to a Prophetic hadith, *sūra* 112 is tantamount to one-third of the Qur'ān, but Mālik wanted to dispel the notion that reciting *sūra* 112 three times would carry the same reward as reciting the entire Qur'ān.⁵²

Not all the verses included in the specialized works of the *Aḥkām al-Qur'ān* genre are included in our network. For example, the Ḥanafī Abū Bakr al-Jaṣṣāṣ (d. 370/981), in his *Aḥkām al-Qur'ān*, refers to 1,045 Qur'ānic verses,⁵³ 436 of which are absent in our Ḥanafī subcorpus. Similarly, the Shāfi'ī scholar al-Kiyā al-Harrāsī (d. 504/1010 or 1011) refers to 569 verses in his *Aḥkām al-Qur'ān*,⁵⁴ 125 of which are missing in our Shāfi'ī subcorpus. Moreover, there are instances in which both al-Jaṣṣāṣ and al-Harrāsī discuss verses that appear in neither the Ḥanafī nor the Shāfi'ī subcorpora, a case in point being Q 2:22 (“[God] made the earth a resting place [*firāsh*] for you...”). This phenomenon prompts the observation that *furū' al-fiqh* texts and *Aḥkām al-Qur'ān* texts are “genealogical”: authors of *Aḥkām al-Qur'ān* rely on earlier works in the genre rather than on *furū' al-fiqh* texts to determine the body of Qur'ān verses that merit attention. Conversely, *furū' al-fiqh* texts omit Qur'ānic verses that are stock-in-trade for authors of *Aḥkām al-Qur'ān* texts.

Returning to our network, texts that share numerous Qur'ānic citations with other texts in the corpus, and thus can be said to rely on a cross-*madhhab* core of Qur'ānic verses, cluster around the center: these include al-Shāfi'ī's (d. 204/820) *K. al-Umm*, al-Māwardī's *al-Ḥāwī*, al-Kāsānī's *Badā'ī' al-ṣanā'ī'*, and Ibn Idrīs' *al-Sarā'ir* (d. 598/1202) (see Fig. 2.2). By contrast, a peripheral position in the network indicates anomalous behavior. Here we find, among others, al-Shaybānī's *K. al-Aṣl*, al-Khallāl's *al-Wuqūf wa'l-tarajjul*, the *Mukhtaṣars* of al-Khiraqī and al-Qudūrī, and al-Nawawī's *Minhāj al-ṭālibīn*. These are texts that, by virtue of being short, do not cite the Qur'ān frequently (there are three citations in *Minhāj al-ṭālibīn*, four in the *Mukhtaṣar* of al-Qudūrī, five in

51 Al-Bahrānī, *al-Ḥadāq'iq al-nādira*, ed. Muḥammad Taqī al-Ayrawānī, 25 vols. (Beirut: Dār al-Aḍwā', 1400/1980) 1:27.

52 Al-Qarāfi, *al-Dhakhira*, ed. Muḥammad Ḥajjī, 14 vols. (Beirut: Dār al-Gharb al-Islāmī, 1994), 2:228.

53 This count is based on the index provided in the digital edition (<https://al-maktaba.org/book/23579>) of Abū Bakr al-Jaṣṣāṣ, *Aḥkām al-Qur'ān*, ed. Muḥammad Ṣādiq Qamḥāwī, 5 vols. (Cairo: Dār al-Muḥḥaf, [1965?]).

54 This count is based on the index in the digital edition (<https://al-maktaba.org/book/23582>) of al-Kiyā al-Harrāsī, *Aḥkām al-Qur'ān*, 2 vols. (Beirut: Dār al-Kutub al-'Ilmiyya, 1403/1983).

al-Wuqūf wa'l-tarajjul, seven in the *Mukhtaṣar* of al-Khiraqī, and fifty-two in the *K. al-Aṣl*).

A different case is the *K. al-Maḥāsin* of al-Barqī (d. 274/887) (Fig. 2.4), also fairly short and also situated at the fringe of the network, but which refers to the Qur'ān 426 times, a relatively large number. The peripheral position of this text results from the fact that it regularly refers to Qur'ānic verses *not* cited by other texts in the network. This is indicated by the large island of yellow dots connected to al-Barqī's text. When one 'visits' al-Barqī's island, by following up on the verses behind the yellow dots and reading the passages in which they are embedded, one observes that al-Barqī's anomalous status in the network results from his repeated references to the *āl al-bayt*. A case in point is al-Barqī's exclusive citation of Q 26:100, in which the unbelievers say on the Day of Judgment that "there are no intercessors [*shāfi'īn*] for us." Al-Barqī here relates the view of al-Ḥusayn, the grandson of the Prophet, that the "intercessors" are "the Imams (*al-āmma*) and the righteous amongst the believers".⁵⁵

Also noteworthy on the periphery of the network are long texts with a surprisingly low number of Qur'ān citations, for example, al-Muḥaqqiq al-Ḥillī's (d. 676/1277) *Sharā'i' al-islām*, "one of the most influential Twelver Shi'ite legal compendia",⁵⁶ which quotes the Qur'ān a mere sixteen times. As al-Ḥillī posits in a work on legal hermeneutics (*uṣūl al-fiqh*), jurists discover laws by relying on "theoretical considerations that are in most cases not derived from the exoteric meanings of the [revealed] texts" (*i'tibārāt nazariyya laysat mustafāda min ḡawāhir al-nuṣūṣ fi 'l-akthar*).⁵⁷ Quantitative analysis shows that al-Ḥillī did not hesitate to follow this hermeneutical maxim in his own legal reasoning.

3 Legal Deontology

3.1 *Ḥalāl and ḥarām*

We begin our machine-supported, quantitative exploration of Islamic legal deontology with the two most basic qualifications of actions in Islamic law, lawful (*ḥalāl*) and unlawful (*ḥarām*). It is no doubt true, as Yūsuf al-Qaraḍāwī asserts, that "[i]n Islam, the sphere of prohibited things is very small, while

55 Al-Barqī, *al-Maḥāsin*, 2 vols. (Qom: Dār al-Kutub al-Islāmiyya, n.d.), 1:184.

56 *Encyclopædia Iranica*, s.v. Ḥelli, Najm al-Dīn (E. Kohlberg), online publication.

57 Al-Muḥaqqiq al-Ḥillī, *Ma'arīj al-uṣūl*, ed. Muḥammad Ḥusayn Raḡawī (Qum: Mu'assasat Āl al-Bayt, 1403/1983), 179. See Wilferd Madelung, "Authority in Twelver Shiism in the Absence of the Imam," in *La notion d'autorité au Moyen Âge: Islam, Byzance, Occident* (Paris: Presses Universitaires de France, 1982), 163–73, at 168.

that of permissible things is extremely vast"⁵⁸—even if the same may be said about most, if not all moral-legal systems in human societies. And yet, certain legal systems are thought to be more “liberal” than others. “Muslim jurists,” writes Bernard Weiss, “acknowledge that there is a large sphere in which human beings must be able to conduct their own affairs so as to achieve maximal advantage for themselves... [b]ut in all human social life, freedom must have its limits, and Islamic law stands in contrast to the liberalism of the West in the drawing of these limits.”⁵⁹ Is Weiss treating Islam and its legal tradition as the “external other” of Western liberalism, to echo Joseph Massad?⁶⁰ As Massad argues, while Western Orientalists have tended to characterize Islamic law as uniquely illiberal, restricting freedom through irrational prohibitions and thereby undermining the ability of Muslims to modernize their societies, Western liberalism has long sought to “transform” Islamic law and bring it in line with Western liberal sensibilities.⁶¹

Mohammad Fadel suggests that scholars of Islamic law should “transcend the limitations of the Islam/liberalism dichotomy”, and that they should study the “moral language” of each of the two formations more closely.⁶² A machine-guided analysis of *furūʿ al-fiqh*, we suggest, has something to offer in this search for the “moral language” of the *fiqh* tradition. Here, we compute the ratio between the two concepts of lawful and unlawful in our *furūʿ al-fiqh* corpus. Such ratios can be determined for each of the subcorpora of the law schools as well as for the entire corpus over time. In Figure 3, we represent the relative frequency of usage of the two terms in each of the five schools as vertical bars in which values are calibrated to a common scale, and in which the percentage of references to *ḥalāl* is stacked on the percentage of references to *ḥarām*.⁶³ This representation is not based on a simple frequency search of the two terms. To achieve more balanced results, we extend our search to include not only a wild card search for *ḥarām*, but also *y-ḥ-r-m* and *t-ḥ-r-y-m*, and we subtract

58 Yusuf al-Qaradawi, *The Lawful and the Prohibited in Islam* (London: Al-Birr Foundation, 1423/2003), 4.

59 Weiss, *Spirit*, 145.

60 Massad, *Islam in Liberalism* (Chicago: University of Chicago Press, 2015), 1.

61 *Ibid.*, 112.

62 Ideally, this approach would result in identifying “compatible elements” of the two traditions. See Mohammad Fadel, “The True, the Good, and the Reasonable: The Theological and Ethical Roots of Public Reason in Islam,” *Canadian Journal of Law and Jurisprudence* 21:1 (2008), 6.

63 Total numbers: Jaʿfarīs: 10,426/3,158, Ḥanbalīs: 3,307/1,164, Shāfiʿīs: 10,722/3,995, Mālikīs: 4,898/2,238, Ḥanafīs: 7,637/3,823.

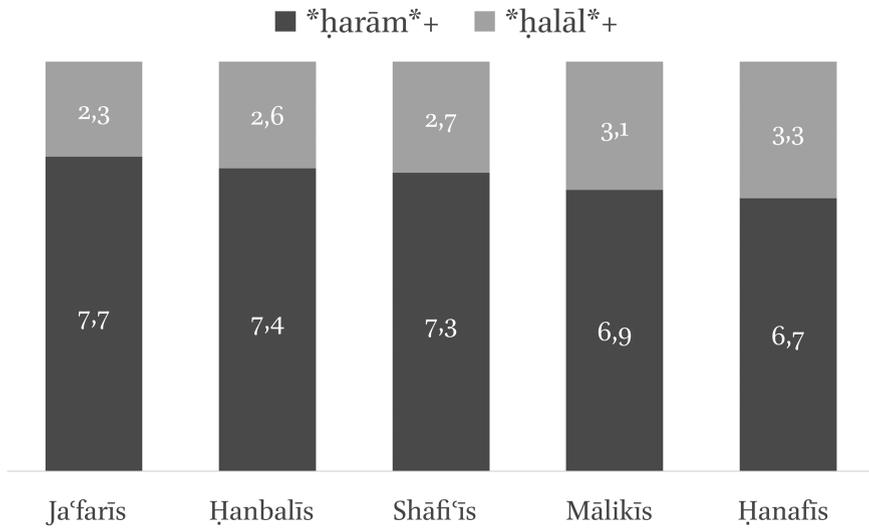


FIGURE 3 Ratios of of *ḥarām*+/*ḥalāl*+ according to law schools

instances of *lā y-ḥ-r-m, *lam y-ḥ-r-m, and *iḥrām*.⁶⁴ Similarly, we search for *ḥalāl* in combination with *t-ḥ-l-y-l* and *y-ḥ-l, subtracting instances of *lā y-ḥ-l and *lam y-ḥ-l. We call the resulting search clusters *ḥarām*+ and *ḥalāl*+.⁶⁵

While the bar diagram does not show how “liberal” or “illiberal” Islamic law is in comparison to other legal systems, it suggests that Ja'farīs and Ḥanbalīs are slightly more likely to use the prohibition *ḥarām*+, and that in this sense they are more “illiberal” than the other three *madhhabs*. 76.8% of the combined number of *ḥarām*+ and *ḥalāl*+ in the Ja'farī subcorpus refer to *ḥarām*+, while 23.2% refer to *ḥalāl*+ (Ḥanbalīs: 74% vs 26%). Then follow, in descending order, Shāfi'īs (72.9% vs 27.1%), Mālikīs (68.3% vs 21.7%), and Ḥanafīs (66.5 vs 33.5%). Paul Powers has observed that scholars of Islamic law have a “careless tendency... to imply that Ḥanafīs are ‘liberal’ and Ḥanbalīs are ‘conservative’”,⁶⁶ and that there is, to date, no “systematic historical study of such

64 There are 178 instances in the entire corpus of *laysa bi-ḥarām* and *ghayr ḥarām*. This is less than 1% of all 26,444 instances of *ḥarām*, and the same holds for *laysa bi-ḥalāl* and *ghayr ḥalāl*, making the search for other forms of negation (e.g., *lā ḥarām*) redundant.

65 That is: *ḥarām*+ = *ḥarām*+ *t-ḥ-r-y-m* + *y-ḥ-r-m - *lā y-ḥ-r-m - *lam y-ḥ-r-m - *iḥrām*; *ḥalāl*+ = *ḥalāl*+ *t-ḥ-l-y-l* and *y-ḥ-l - *lā y-ḥ-l - *lam y-ḥ-l-m.

66 Paul R. Powers, “The Schools of Law,” in *The Ashgate Companion to Islamic Law*, 49. See, for example, Raymond Charles, *Le droit musulman* (Paris: Presses Universitaires de France, 1956), 27, who casually notes that the Ḥanafī school is regarded as “the most liberal” (*le plus libéral*), while the Ḥanbalī school is “the strictest” (*le plus strict*).

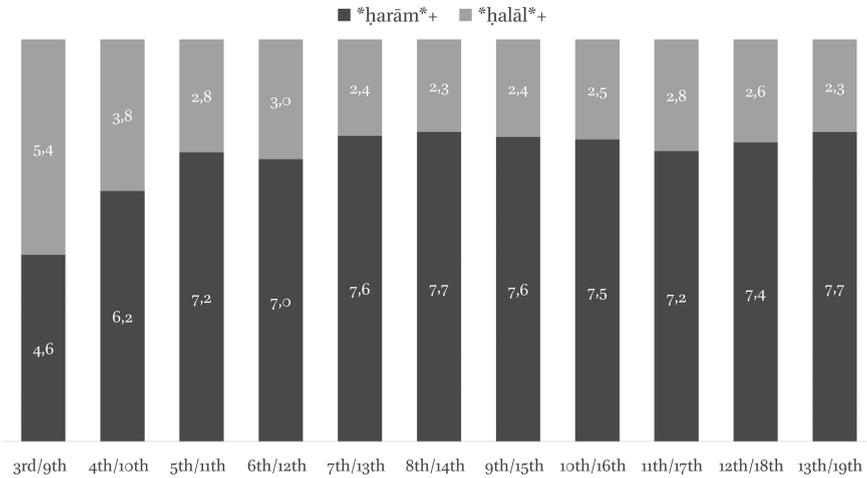


FIGURE 4 Ratios of *ḥarām*+/*ḥalāl*+ in *furū' al-fiqh* over the centuries

differences”, which means that “no sweeping characterizations of the doctrinal tone of individual *madhhabs* are warranted.”⁶⁷ Our digital, distant-reading approach is a first step in the direction of such a systematic historical study. The differences between the schools in our model are not, we acknowledge, significant from a strictly statistical point of view. But they do show tendencies that appear to support the scholarly intuition about Ḥanafī ‘liberalism’ and Ḥanbalī ‘conservatism’.

To respond to the need, highlighted by Powers, to pay attention to the “doctrinal tone” of the schools as it developed over time, in Figure 4 the *ḥarām*+/*ḥalāl*+ ratio is again shown in vertical bars that are calibrated to a common scale, but arranged according to century.⁶⁸

Figure 4 shows an increase in the use of *ḥarām*+ in proportion to *ḥalāl*+ from the 3rd/9th century until the 5th/11th century, at which time many of the major ‘classical’ comprehensive *fiqh* texts appear, such as the *Mabsūṭs* of al-Sarakhsī and al-Ṭūsī, the *Kāfī* of Ibn ‘Abd al-Barr, and the *Ḥawī al-kabīr* of al-Māwardī. A second, smaller increase occurs between the 5th/11th century and the 8th/14th century, the century of al-Bābartī, Khalīl b. Ishāq, al-Nawawī, Muḥammad Ibn Mufliḥ, and al-Muḥaqqiq al-Ḥillī. After the 8th/14th century, the ratio remains stable, approximately 3:1. It is again

67 Ibid., 50.

68 Total numbers: 2nd/8th and 3rd/9th c.: 970/119; 4th/10th c.: 871/527; 5th/11th: 4723/1844; 6th/12th c.: 2427/1030; 7th/13th c.: 3008/948; 8th/14th c.: 4231/1264; 9th/15th c.: 3801/1219; 10th/16th c.: 4288/1421; 11th/17th c.: 2734/1053; 12th/18th c.: 8820/3038; 13th/19th c.: 3052/915.

possible to challenge the statistical significance, or “evidence of value”, of our findings. Be that as it may, Figure 4 illustrates that computer-supported distant reading puts us in a position to think about *longue-durée* dynamics in *furū‘ al-fiqh* that cannot be seen, or even thought about, by the usual close reading of the sources: a “haramization” of the law from the formative to the classical period, and a stability in the discourse on lawful and unlawful in the postclassical period up to the modern period. In the 20th and 21st centuries—a period that does not fall within the scope of this study—the arc of *ḥarām* and *ḥalāl* seems to bend towards “halalization”, as has been argued with regard to certain consumption-driven and affluent Muslim societies of the late 20th and early 21st centuries.⁶⁹

3.2 *The Five Qualifications (al-ahkām al-khamsa)*

In addition to the distinction between *ḥarām* and *ḥalāl*, Islamic legal deontology operates with five basic normative qualifications (*ahkām taklifiyya*): obligatory (*wājib*), recommended (*mandūb*), neutral (*mubāḥ*), disapproved (*makrūh*), and forbidden (*ḥarām*). “The intermediate categories,” writes Mohammad Hashim Kamali, “consist essentially of options that offer scope for personal freedom.” He concludes that the “scope of liberty [in Islamic law] is thus much wider than that of *wājib* and *ḥarām*.”⁷⁰ A machine-supported analysis of our corpus, we argue, can establish how wide this scope is, quantitatively speaking.

A complication results from the fact that over the course of the centuries, Muslim jurists have used a number of synonyms for each of these five qualifications, as well as other subcategories.⁷¹ Thus, Ḥanafis distinguish between two categories of the obligatory: absolutely obligatory (*farḍ*) and binding (*wājib*).⁷²

69 See Johan Fischer, *Proper Islamic Consumption: Shopping among the Malays in Modern Indonesia* (Copenhagen: NIAS Press, 2008), 29, 74–203; Karim Douglas Crow, “Consuming Islam: Branding ‘Wholesome’ as Lifestyle Fetish,” *Islamic Sciences* 13:1 (2015), 3–26.

70 Mohammad Hashim Kamali, *Shari‘ah Law: An Introduction* (Oxford: Oneworld, 2008), 202.

71 Erwin Gräf, “Zur Klassifizierung der menschlichen Handlungen nach Ṭūsī dem Šaiḫ at-Ṭā‘ifa (gest. 460) und seinen Lehrern,” *Zeitschrift der deutschen morgenländischen Gesellschaft*, Suppl. 3, pt. 1 (1977), 388–422. Gräf identifies more than 100 different terms used to classify human acts in a single chapter (on *‘ibādāt*) of al-Ṭūsī’s *al-Mabsūt*. See *ibid.*, 388. See also Weiss, *Spirit*, 18–21; *idem*, *Search*, 92–111; Zysow, *Economy*, 63 n70.

72 Zysow, *Economy*, 52; Hallaq, *Sharī‘a*, 86; Baber Johansen, “Legal and Ethical Qualifications,” in *The Oxford International Encyclopedia of Legal History*, ed. Stanley Katz (Oxford: Oxford University Press, 2009), 3:322–3. See the detailed study by Kevin Reinhart, “Like the Difference Between Heaven and Earth: Ḥanafī and Shāfi‘ī Discussions of *Farḍ* and *Wājib* in Theology and Uṣūl,” in *Studies in Islamic Legal Theory*, ed. Bernard Weiss (Leiden: E.J. Brill, 2002), 205–34.

Mālikīs distinguish between recommended (*mandūb*) and (good) practice (*sunna*), while for the other schools, the two terms are largely synonymous.⁷³ In order to account for this terminological variety, we trained a word embedding tool on our corpus to identify the closest semantic neighbors of the five terms.⁷⁴ Thus, we base the following analysis on the following pairs of terms: *wājib*+*farḍ* (henceforth: *wājib*+); *mandūb*+*mustaḥabb* (henceforth: *mandūb*+); *mubāḥ*+*ḥalāl* (henceforth: *mubāḥ*+); *makrūh*+*qabīḥ* (henceforth: *makrūh*+); *ḥarām*+*maḥzūr* (henceforth: *ḥarām*+). The following pie chart (Fig. 5) shows the ratio between these five pairs in the entire corpus, not counting the most common negations (that is, instances of any of the ten terms preceded by *ghayr* or *laysa bi-*).⁷⁵

More than half (54.3%) of all qualifications in the corpus are *wājib*+, about one-fifth (22%) *ḥarām*+, almost one-eighth (12.3%) *mubāḥ*+, and

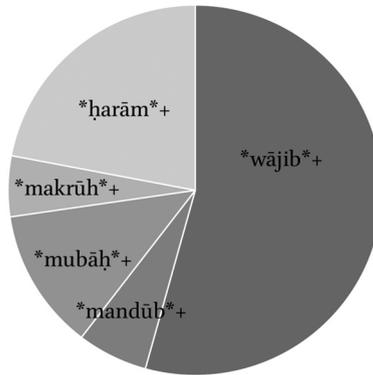


FIGURE 5 Ratio of normative qualifications across all schools

73 Ignazio Guidi, "Sunna e Nadb presso i Giuristi Malechiti," in *Festschrift Eduard Sachau zum siebenzigsten Geburtstag*, ed. Gotthold Weil (Berlin: Riemer, 1915), 333; Zysow, *Economy*, 63 n70; Johansen, "Legal and Ethical Qualifications," 3:322.

74 Word embedding tools identify the words that occur most frequently in similar contexts. This relationship is represented in fractions: the closer the words are numerically (1 = a full match), the more closely related they are semantically. For this essay, we used the Python package Gensim to train a word2vec model with a window size 5 on our corpus. This approach is similar to the one proposed in Abu Bakr Soliman, Kareem Eissa and Samhaa R. El-Beltagy, "Aravec: A Set of Arabic Word Embedding Models for Use in Arabic NLP," *Procedia Computer Science* 117 (2017), 256–65.

75 In fact, negations of the five categorizations tend to cancel each other out. This means that the results of searches that *include* the negations are virtually the same as the results of searches that *exclude* the negations, except for a few decimals: 54.8%, 6%, 12.1%, 5.4%, 21.7% (total corpus, including negations) against 54.3%, 6.1%, 12.3%, 5.3%, 22% (total corpus, excluding negations).

a little more than one-tenth *makrūh*+ and *mandūb*+ combined (5.3% and 6.1%, respectively). The individual law schools follow this pattern, with minor variations. The two ends of the spectrum are occupied by the Ḥanbalī and the Mālikī schools (Fig. 6.1 and Fig. 6.2): Ḥanbalīs have the smallest sum total of middle categories, roughly one-fifth of all qualifications (*mandūb*+: 3.8%; *mubāḥ*+: 14.1%; *makrūh*+: 2.9%; total: 20.8%), while Mālikīs have the largest middle category, more than a quarter of all qualifications (*mandūb*+: 7.2%; *mubāḥ*+: 11.2%; *makrūh*+: 6.8%; total: 26.4%).

We detect here an echo of the notion that Ḥanbalīs, as Noel Coulson put it, embrace a certain “moralist attitude”, that is, they divide human actions starkly into obligatory and forbidden, while the other schools, especially Mālikīs and Ḥanafīs, have a “legally formalist attitude” that favors use of a broader range

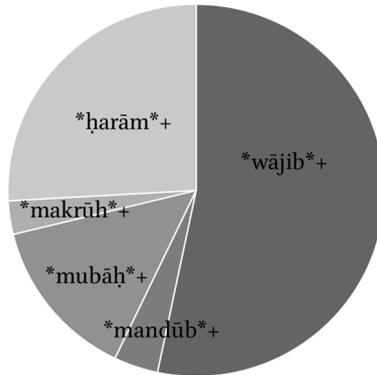


FIGURE 6.1 Ratio of normative qualifications, Ḥanbalīs

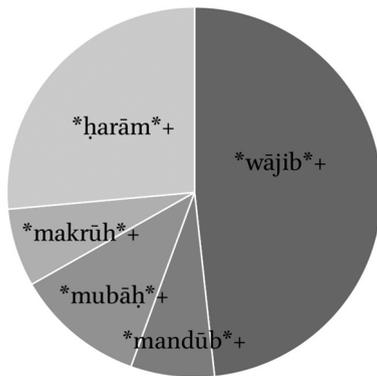


FIGURE 6.2 Ratio of normative qualifications, Mālikīs

of categories.⁷⁶ As for Ḥanbalīs, it should be noted that their perceived strictness is balanced by their greater willingness, comparatively speaking, to argue in terms of “dispensations” or “alleviations” (*rukḥaṣ*, sg. *rukḥṣa*), a category of legal norms that, as Goldziher remarked, is “appended” to the five qualifications.⁷⁷ In the Ḥanbalī subcorpus, the term *r-kh-ṣ* appears once every 6,345 tokens, whereas it is less commonly used by the Shāfi‘īs (once every 7,694 tokens), Mālikīs (once every 7,952 tokens), Ḥanafīs (once every 9,801 tokens) and Ja‘farīs (once every 11,160 tokens).⁷⁸

To return to the question we posed at the beginning of this section, our analysis suggests that, *pace* Kamali, it is not certain that the scope of the intermediate categories is “much wider” in Islamic law than the scope occupied by the two categories of *ḥarām* and *wājib*. Kamali, it should be said, is not alone in his view of the relationship between the five qualifications. Scholars frequently state that the moral sphere (demarcated by the terms *makrūh* and *mandūb*) and the legal sphere (the domain of *wājib* and *ḥarām*) are seamlessly connected in Islamic law, and that both spheres are equally important for Muslim jurists. Bernard Weiss, for example, opines that “[i]t is important always to bear in mind that the Shari‘a is as much concerned with recommending and disapproving as it is with prescribing and forbidding.”⁷⁹ Ahmad Alkhamees states that “Shari‘a pays similar attention to recommended and disapproved acts as to prescribed and prohibited acts.”⁸⁰ Wael Hallaq, finally, finds that the theological and eschatological nature of the intermediate categories “does not relegate [Hallaq’s emphasis] them to a category below, and thus outside, the law,” and that “[m]eshing the moral with the legal, these norms were subject to a great deal of articulation and discussion.”⁸¹ Computational quantification of

76 See Coulson, *Conflicts and Tensions*, 86. In the case of the Mālikīs, the *mandūb*+ domain would likely increase even more if the category of *sunna* were also taken into account, something our machine-driven approach does not enable us to do. It would be interesting to conduct a quantitative study of the differences in the use of the five qualifications in the three fields of the law (*‘ibādāt*, *mu‘āmalāt*, *jināyāt*). See Johansen, “Legal and Ethical Qualifications,” 3:323–5.

77 Ignaz Goldziher, *Die Zāhiriten: ihr Lehrsystem und ihre Geschichte* (Leipzig: Otto Schulze, 1884), 64.

78 The *rukḥaṣ* are included under what Bernard Weiss calls “nonnormative categorizations” (*aḥkām waḍ‘īyya*). See Weiss, *Search*, 109.

79 Weiss, *Spirit*, 19. Rudolph Peters and Peri Bearman write that the middle categories “are integral to Sharia as a moral code”. See Peri Bearman and Rudolph Peters, “Introduction: The Nature of the Sharia,” in *The Ashgate Companion to Islamic Law*, 5. A similar position is adopted by Schacht, *Introduction*, 200–1.

80 Ahmad Alkhamees, *A Critique of Creative Sharī‘a-Compliance in the Islamic Finance Industry* (Leiden-Boston: Brill Nijhoff, 2017), 20–21.

81 Hallaq, *Sharī‘a*, 87–88.

the five legal qualifications provides a corrective to this view, suggesting that the jurists were first and foremost interested in determining legal prescriptions and prohibitions, and only secondarily, and at some distance, in voicing moral approval or disapproval.

4 Topics

4.1 Word Clouds

In the final section of this study we use a computational, distant-reading approach in order to identify the salient topics in our *furūʿ al-fiqh* corpus. In digital humanities, word clouds are a popular means to visualize, in one image, the most-frequently used words (or, as in our case: tokens) in a given corpus. Because ordinary, frequency-based word clouds (FWCs) bring to the fore often-used stopwords and particles, the use of filters is imperative. In Figure 7, we filter out all particles, cardinal and ordinal numbers, and verbs; we only include adjectives, nouns and proper names, in both prefixed and suffixed forms.

The two central terms in the corpus are “prayer” (*al-ṣalāt*) and “property” (*al-māl*).⁸² The layer around this core includes “messenger” (*rasūl*), “the Prophet” (*al-nabī*), “sale” (*al-bayʿ*), “contract” (*al-ʿaqd*), “Muslim individual/

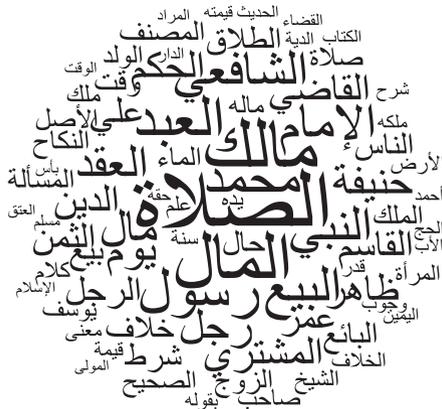


FIGURE 7 FWC entire corpus

82 We do not include the term *mālik* in this “central” group, as it can be read both as a proper name (e.g., Mālik b. Anas) and as a noun, meaning “owner”. Counting the first and the last 100 instances of the term *mālik* in a random sample of major works in the corpus (al-Kasānī, al-Māwardī, Ibn Qudāma, al-Qarāfī, and al-Baḥrānī), we find that the term is used as a noun

For example, the PWC of the Ḥanbalī subcorpus (Fig. 8.1) shows that Ḥanbalī jurists refer to Aḥmad b. Ḥanbal more frequently than jurists of the other schools, and it highlights the central position of al-Khiraqī (d. 334/945–6), author of the first short summary (*mukhtaṣar*) of Ḥanbalī law.⁸⁷ Two other peculiarities of the Ḥanbalī subcorpus are the use of the term *al-ri'āya* and the use of *wa-ʿanhu* (“and/also from him [is related]”). Regarding the first peculiarity, a close reading of the relevant passages in the Ḥanbalī subcorpus reveals that the term *al-ri'āya* refers to the title of a work, *al-Ri'āya al-kubrā* of Ibn Ḥamdān al-Ḥarrānī (d. 695/1296), a major source for Muḥammad Ibn Muflīḥ (d. 763/1362) and Ibrāhīm Ibn Muflīḥ (d. 884/1479), which is also quoted frequently by other post-Mongol Ḥanbalī jurists in our corpus, such as Ibn al-Najjār (d. 972/1564), al-Buhūṭī (d. 1051/1641), and al-Nābulusī (d. 1319/1901).

Ibn Ḥamdān is remembered for teaching that by his time, fully independent, or “absolute” (*mutlaq*) *mujtahids* had disappeared from the lands of Islam.⁸⁸ This connects him to the second peculiarity of the Ḥanbalī PWC, the frequent use of *wa-ʿanhu*, which indicates the propensity of Ḥanbalī jurists, compared to those of the other schools, to transmit the received opinions of earlier authorities, especially those of Aḥmad b. Ḥanbal.

The frequent use the term *wa-ʿanhu* is characteristic of two Ḥanbalī authors in particular, the aforementioned two Ibn Muflīḥ: Muḥammad and his great-grandson, Ibrāhīm. Together, their texts account for more than 90% of all instances of *wa-ʿanhu* in the Ḥanbalī subcorpus, even though, in terms of the size of their texts, they make up only approximately 35%. More is at stake here than a curious stylistic preference shared by two authors hailing from the same Damascene dynasty of Ḥanbalī scholars. Muḥammad Ibn Muflīḥ was a student of Ibn Taymiyya (d. 728/1328) and also studied with the traditionists al-Dhahabī (d. 748/1438 or 753/1352–3) and al-Mizzī (d. 742/1341).⁸⁹ According to George Makdisi, the *K. al-Furūʿ* by Muḥammad Ibn Muflīḥ, “one of the most prolific writers of the Ḥanbalī school of his period,” is “one of the most important Ḥanbalī works for the establishment of the true legal doctrine of Aḥmad b. Ḥanbal”.⁹⁰ His great-grandson Ibrāhīm related that his grandfather was “a virtuous expert,

0.001. Word weights were calculated using the weighwords software. See <https://github.com/larsmans/weighwords>.

87 Christopher Melchert, *The Formation of the Sunni Schools of Law, 9th-10th Centuries C.E.* (Leiden-New York: Brill, 1997), 148–49.

88 See Ahmed Fekry Ibrahim, “Rethinking the *Taqlid* Hegemony: An Institutional, *Longue-Durée* Approach,” *JAOS* 136:4 (2016), 812.

89 *EI*³, s.v. “Ibn Muflīḥ” (Daniella Talmon-Heller).

90 *EI*², s.v. “Ibn Muflīḥ” (George Makdisi). On the place of Shams al-Dīn Ibn Muflīḥ in the history of the Ḥanbalī school, see further Henri Laoust, *Le Ḥanbalisme sous les Mamlouks bahrides* (Paris: Geuthner, 1960), 68–9, and notes 369–70.



FIGURE 8.3 PWC Mālikī subcorpus

place to Mālik’s student Ibn al-Qāsim (d. 191/807) and to Ibn al-Qāsim’s student Saḥnūn (d. 240/855), while also drawing attention to the formative early figures of al-Mājishūn (d. 164/780–1), Ashhab (d. 204/819), Aṣḥab (d. 225/839), Ibn Ḥabīb (d. 238/853) and Ibn al-Mawwāz (d. 269/882), in addition to the arguably lesser well-known 5th/11th-century jurist al-Lakhmī (d. 478/1085), and to the *K. al-Mustakhraja*, commonly called *al-Utbīyya*, of al-Utbī (d. 255/869).

Finally, the three major terms in the Ja’farī PWC (Fig. 8.4) refer to characteristically Ja’farī ways of framing an argument: *wa-ālihi* is part of the *taṣliya* formula used in Shi’i texts (*ṣallā ‘alayhi wa-ālihi*, “God bless him and his family”); *al-akhbār*, used in reference to the Ja’farī hadith corpus, occurs in phrases such as *fī ba’ḍ al-akhbār* (“according to certain traditions”); *al-aṣḥāb* refers to



FIGURE 8.4 PWC Ja’farī subcorpus

“companions” or “adherents” of the Jaʿfarī law school, invoked anonymously in phrases such as *ka-mā qāla baʿḍ al-aṣḥāb* (“as a certain companion/a certain number of companions said”). As revealed by a close-reading check in BlackLab ADH, in earlier Jaʿfarī texts there are almost no instances of authors referring to “a certain tradition” or to “a certain companion” to buttress an argument. Such phrases become commonplace in Jaʿfarī law only in the 10th/16th and the following centuries, starting with al-Shahīd al-Thānī’s (d. 965/1557) *Masālik al-afhām*. The 10th/16th century witnessed the emergence of the Safavid state and its patronage of Jaʿfarī law, and the fact that Jaʿfarī jurists, from this period onwards, invoke their collective tradition of legal scholarship to frame an argument demonstrates their confidence in their school’s institutional strength.

4.2 *Topic Modeling*

Another computational bird’s eye approach to the corpus is provided by topic modeling, a common technique in digital text mining.⁹⁵ The term refers to the application of a statistical model to a corpus, divided into segments, in order to identify for each segment the salient set of words which, together, form a “topic”.⁹⁶ In Arabic corpora, topic modeling does not work well with words because Arabic words frequently appear with many morphological variations. Topic modeling based on roots is promising, but root recognition for Arabic, as mentioned above, remains a challenge.⁹⁷ Here, we choose to model topics on the basis of stems rather than words or roots. For this purpose, we use a stemmer that removes prefixes and suffixes from words (or rather, what it identifies as such, not always successfully), to the exclusion of infixes.⁹⁸ The twenty most salient topics in the corpus are as shown in Table 4:

95 In this essay, we use Latent Dirichlet Allocation (LDA) to arrive at topics. The LDA model arrives at a predefined number of topics. For each topic, it assigns a weight to each possible word, denoting its importance for the topic. Per topic, the table shows the ten words with the highest weights. See David M. Blei, Andrew Y. Ng and Michael I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research* 3 (January, 2003), 993–1022.

96 The segments for which topics were identified in this study are the tagged chapters (*bābs*), subchapters (*faṣls*), and sometimes books (*kitābs*) within a work. *Furūʾ al-fiqh* texts are often divided into several hundred such sections. For lack of resources, we could not comprehensively tag the corpus. Therefore, we exclude from the analysis all works that are divided into less than twenty tagged sections. Here, as elsewhere, we salute Willard McCarty: “[D]eep encoding is very laborious.” See McCarty, “Humanities Computing,” 1230.

97 See above, note 16.

98 See Leah S. Larkey, Lisa Ballesteros and Margaret E. Connell, “Light Stemming for Arabic Information Retrieval,” in *Arabic Computational Morphology: Knowledge-Based and Empirical Methods*, ed. Abdelhadi Soufi, Antal van den Bosch and Günther Neumann (Dordrecht: Springer, 2007), 221–43.

TABLE 4 Color-coded topics in the corpus

#	Color	Topic	Stems
1		legal reasoning	<i>dh-k-r</i> , [<i>alif</i>]- <i>w-l</i> , [<i>alif</i>]- <i>dh</i> , <i>f-l</i> , <i>sh-r-h</i> , <i>z</i> -[<i>alif</i>]- <i>h-r</i> , <i>w-l</i> , <i>k-l</i> -[<i>allif</i>]- <i>m</i> , <i>b-kh-l</i> -[<i>alif</i>]- <i>f</i> , <i>sh-y-kh</i>
2		legal reasoning: general; <i>akhbār</i>	<i>s-l</i> -[<i>alif</i>]- <i>m</i> , <i>h-dh</i> , [<i>alif</i>]- <i>b</i> , [<i>alif</i>]- <i>kh-b</i> -[<i>alif</i>]- <i>r</i> , <i>r-w</i> -[<i>alif</i>], <i>sh-y-kh</i> , <i>dh-k-r</i> , <i>z</i> -[<i>alif</i>]- <i>h-r</i> , <i>k-l</i> -[<i>alif</i>]- <i>m</i> , [<i>alif</i>]- <i>ṣ-h</i> -[<i>alif</i>]- <i>b</i>
3		legal reasoning: general; <i>khilāf</i>	<i>kh-l</i> -[<i>alif</i>]- <i>f</i> , <i>a-w-l</i> , ^ʿ <i>d-m</i> , <i>m-t-l-q</i> , [<i>alif</i>]- <i>j-m</i> -[<i>alif</i>]- ^ʿ , [<i>alif</i>]- <i>ṣ-l</i> , <i>th</i> -[<i>alif</i>]- <i>n</i> , <i>w-l</i> , <i>ṣ-h-y-h</i> , [<i>alif</i>]- <i>m</i>
4		legal reasoning: Prophetic hadith and early authorities	<i>ṣ-l</i> , [<i>alif</i>]- <i>b</i> , <i>r-s-w-l</i> , ^ʿ <i>m-r</i> , <i>m</i> -[<i>alif</i>]- <i>l-k</i> , <i>ḥ-d-y-th</i> , <i>n-b</i> , <i>sh</i> -[<i>alif</i>]- <i>f</i> ^ʿ , <i>r-j-l</i> , [<i>alif</i>]- <i>b</i>
5		legal reasoning: Ḥanafis	[<i>alif</i>]- <i>dh</i> , <i>h-dh</i> , <i>dh-k-r</i> , [<i>alif</i>]- <i>b</i> , <i>b-kh-l</i> -[<i>alif</i>]- <i>f</i> , <i>ḥ-n-y-f</i> , <i>y-w-s-f</i> , <i>f-l</i> , <i>h-q</i> , <i>w-k-dh</i>
6		legal reasoning: Shāfiʿīs	<i>th</i> -[<i>alif</i>]- <i>n</i> , <i>m-s</i> -[<i>alif</i>]- <i>l</i> , <i>f-ṣ-l</i> , [<i>alif</i>]- <i>b-w</i> , <i>sh</i> -[<i>alif</i>]- <i>f</i> ^ʿ , <i>w-j-h</i> , <i>w-l</i> , [<i>alif</i>]- <i>ḥ-d-h-m</i> , <i>m-dh-h-b</i> , <i>dh-k-r</i>
7		legal reasoning: Mālikīs	<i>q</i> -[<i>alif</i>]- <i>s-m</i> , <i>m</i> -[<i>alif</i>]- <i>l-k</i> , [<i>alif</i>]- <i>dh</i> , <i>f-l</i> , <i>h-dh</i> , <i>w-l</i> , <i>m-d-w-n</i> , <i>k-t</i> -[<i>alif</i>]- <i>b</i> , [<i>alif</i>]- <i>sh-h-b</i> , <i>s-h-n</i>
8		ritual law: ritual purity	<i>gh-s-l</i> , <i>w-l</i> , [<i>alif</i>]- <i>dh</i> , <i>w-d</i> , <i>ṣ-l</i> -[<i>alif</i>], <i>n-j</i> -[<i>alif</i>]- <i>s</i> , <i>d-m</i> , <i>ḥ-y-d</i> , <i>m-s-h</i> , <i>n-j-s</i>
9		ritual law: prayer	<i>ṣ-l</i> -[<i>alif</i>], [<i>alif</i>]- <i>m</i> -[<i>alif</i>]- <i>m</i> , [<i>alif</i>]- <i>dh</i> , <i>w-l</i> , <i>ṣ-l</i> , <i>a-w-l</i> , <i>w-q-t</i> , <i>r-k</i> ^ʿ , <i>s-l</i> -[<i>alif</i>]- <i>m</i> , <i>h-dh</i>
10		ritual law: oaths; fasting; expiation	<i>ṣ-w-m</i> , <i>k-f</i> -[<i>alif</i>]- <i>r</i> , <i>y-w-m</i> , <i>ḥ-l-f</i> , <i>w-l</i> , <i>y-m</i> , <i>ḥ-n-th</i> , [<i>alif</i>]- <i>dh</i> , <i>y-ḥ-n-th</i> , <i>n-dh-r</i>
11		ritual law: pilgrimage	<i>ḥ-j</i> , <i>w-l</i> , <i>ṣ-y-d</i> , [<i>alif</i>]- <i>dh</i> , <i>ḥ-r-m</i> , ^ʿ <i>m-r</i> , [<i>alif</i>]- <i>ḥ-r</i> -[<i>alif</i>]- <i>m</i> , <i>m-ḥ-r-m</i> , <i>t-w</i> -[<i>alif</i>]- <i>f</i> , <i>d-m</i>
12		ritual law: alms; taxes	<i>z-k</i> -[<i>alif</i>], <i>w-l</i> , <i>m</i> -[<i>alif</i>]- <i>l</i> , <i>ḥ-w-l</i> , <i>ṣ-d-q</i> , ^ʿ <i>sh-r</i> , <i>kh-m-s</i> , <i>n-ṣ</i> -[<i>alif</i>]- <i>b</i> , [<i>alif</i>]- <i>dh</i> , <i>ḥ-d</i>
13		personal law: marriage	<i>z-w-j</i> , <i>t-l</i> -[<i>alif</i>]- <i>q</i> , <i>w-l</i> , <i>n-k</i> -[<i>alif</i>]- <i>ḥ</i> , [<i>alif</i>]- <i>dh</i> , <i>ḥ-d</i> , [<i>alif</i>]- <i>m</i> , <i>m-h-r</i> , <i>f-l</i> , ^ʿ <i>d</i>
14		personal law: inheritance	<i>th-l-th</i> , <i>w-ṣ</i> , [<i>alif</i>]- <i>b</i> , <i>m</i> -[<i>alif</i>]- <i>l</i> , [<i>alif</i>]- <i>m</i> , ^ʿ <i>sh-r</i> , <i>n-ṣ-f</i> , <i>th-l</i> -[<i>alif</i>]- <i>th</i> , <i>w-r-th</i> , <i>m-w-ṣ</i>
15		private law: con- tracts; sales	<i>b-y</i> ^ʿ , <i>m-sh-t-r</i> , <i>th-m-n</i> , ^ʿ <i>q-d</i> , <i>q-b-d</i> , [<i>alif</i>]- <i>dh</i> , <i>b</i> -[<i>alif</i>]- ^ʿ , <i>w-l</i> , <i>h-dh</i> , <i>kh-y</i> -[<i>alif</i>]- <i>r</i>

TABLE 4 Color-coded topics in the corpus (cont.)

#	Color	Topic	Stems
16		private law: contracts; estates	<i>m</i> -[<i>alif</i>]- <i>l</i> , [<i>alif</i>]- <i>r-d</i> , <i>w-l</i> , [<i>alif</i>]- <i>j-r</i> , [<i>alif</i>]- <i>dh</i> , <i>f-l</i> , <i>ʿ-m-l</i> , <i>m-l-k</i> , [<i>alif</i>]- <i>kh-dh</i> , <i>d</i> -[<i>alif</i>]- <i>r</i>
17		private law: slaves	<i>ʿ-t-q</i> , <i>w-l-d</i> , <i>m</i> -[<i>alif</i>]- <i>l</i> , <i>w-l</i> , [<i>alif</i>]- <i>m</i> , <i>h-r</i> , <i>m-l-k</i> , <i>q-y-m</i> , <i>m-k</i> -[<i>alif</i>]- <i>t-b</i> , <i>m-w-l</i>
18		procedural law: litigation; proof	<i>sh-h</i> -[<i>alif</i>]- <i>d</i> , <i>w-l</i> , [<i>alif</i>]- <i>q-r</i> -[<i>alif</i>]- <i>r</i> , <i>q</i> -[<i>alif</i>]- <i>d</i> , <i>m-d-ʿ</i> , [<i>alif</i>]- <i>dh</i> , <i>y-d</i> , <i>h-d</i> , <i>h-q</i> , <i>h-dh</i>
19		public law: homi- cide; punishments	<i>q-t-l</i> , <i>h-d</i> , <i>w-l</i> , <i>q-t-ʿ</i> , <i>q-s</i> -[<i>alif</i>]- <i>s</i> , [<i>alif</i>]- <i>dh</i> , <i>r-j-l</i> , <i>ʿ-m-d</i> , <i>n-f-s</i> , <i>y-d</i>
20		public law: warfare	<i>m-s-l-m</i> , <i>w-l</i> , [<i>alif</i>]- <i>s-l</i> -[<i>alif</i>]- <i>m</i> , [<i>alif</i>]- <i>h-l</i> , <i>h-r-b</i> , [<i>alif</i>]- <i>m</i> -[<i>alif</i>]- <i>m</i> , <i>q-t-l</i> , <i>ʿ-l-y</i> , <i>d</i> -[<i>alif</i>]- <i>r</i> , [<i>alif</i>]- <i>dh</i>

As Table 4 demonstrates, topic modeling produces several coherent topics, but it also produces topics that are difficult to label because they include stems such as *w-l* (< *wa-lahu*) and *h-dh* (< *hādihā*), which we consider noise. Topic 19 (public law: homicide; punishments), for example, seems robust, as it includes stems related to killing/murder (*q-t-l*), statutory punishment (*h-d*), cutting (*q-t-ʿ*), hand (*y-d*), talion (*q-s*-[*alif*]-*s*), and intentionality (*ʿ-m-d*). This means that, if a certain section of a work has topic 19 as its salient topic, we can state with some confidence that issues of criminal law and punishment are paramount in this section. By contrast, topic 5 (legal reasoning: Ḥanafis) is elusive: it includes several stems referring to stopwords ([*alif*]-*dh*, *h-dh*, *w-k-dh*), as well as stems that refer to two famous early Ḥanafī authorities (Abū Yūsuf and Abū Ḥanīfa), and to the stem *b-kh-l*-[*alif*]-*f* (*bi-khilāf*). This combination of stems *may* suggest that sections that have topic 5 as their salient topic revolve around differences of opinion within the Ḥanafī school, especially those involving its founding figures. All in all, the attribution of labels to the topics generated by the model remains, to a significant degree, an exercise of the imagination. Manually checking the topics against the source texts, as we did in order to create Table 4, can help the process of labeling, but a measure of subjectivity and doubt remains.

First, we plot topics of individual texts. For some topics, there is no major difference between the texts in the corpus. For example, topic 15 (private law: contracts, sales) is stable at around 10% of all sections in the majority of texts in the corpus, with al-Sarakhsī's *al-Mabsūṭ* occupying the top position (112 of 761 tagged sections [13.8% of the number of tokens]). In other areas,

however, specific texts manifest a preference for a certain topic. For example, al-Shaybānī's *K. al-Aṣl*, described by Eric Chaumont as "a collection of opuscula, dealing with different aspects of practical law",⁹⁹ focuses on public law issues. In 94 of the 242 tagged sections in this work (29.1% of tokens), topic 20 is salient (public law: warfare). The work that follows the *K. al-Aṣl* in this respect is al-Sarakhsī's *al-Mabsūṭ*, with 115 of 761 tagged sections (11.6% of tokens). Similarly, in 24 of the 242 tagged sections (11.2% of tokens) in al-Shaybānī's *K. al-Aṣl*, topic 19 is salient (public law: homicide and punishments), followed by al-Khiraqī's *Mukhtaṣar* (9.4% of tokens) and al-Shaykh al-Mufīd's *al-Muqni'a* (8.7% of tokens). Another Ḥanafī, Ibn 'Abidīn, pays considerable attention to topic 12 (ritual law: alms; taxes) (90 of 227 tagged sections [38% of tokens]), followed by two other Ḥanafī authors, al-Bābartī (28 of 167 tagged sections [20.3% of tokens]) and al-Kāsānī (66 of 841 tagged sections [5% of tokens]). These quantitative findings point to the proximity of Ḥanafīs to the state throughout the history of Islamic law.

It is striking that ritual law topics (especially prayer) are salient especially in early texts. Three texts from the 3rd/9th century appear in the top five texts devoted to ritual law topics (topics 8–12): al-Barqī, *Mahāsīn* (#1, 377 of 404 tagged sections [98.5% of tokens]), al-Khallāl, *Wuqūf* (#3, 66 of 85 tagged sections [90.8% of tokens]), and Mālik, *Muwattā'* (#5, 453 of 664 tagged sections [63.5% of tokens]). Also al-Shāfi'ī's *Umm* ranks high (#9, 276 of 790 tagged sections [28.9% of tokens]). The only text after the 4th/10th century that devotes similar attention to ritual law topics (mostly prayer) is Ibn 'Abd al-Wahhāb, *Majmū'* (#2, 99 of 104 tagged sections [96.7% of tokens]).

The following horizontal bars are normalized visualizations of five texts. Reading from left to right, each tagged section appears in the color of the topic that is salient in it.

Topics 4 (legal reasoning: Prophetic hadith and early authorities; dark brown) and 3 (legal reasoning: general; *khilāf*; brown) are salient in the early sections of each work. In these sections, the vocabulary referring to Prophetic tradition and to transmitted knowledge in general forms the dominant topic. As is well known, *furū' al-fiqh* texts begin with sections on ritual law. It is in these sections, as Figures 9.1 to 9.5 demonstrate, that the authority of the Prophet in these matters was especially important to the jurists.

The horizontal bars feature green sections somewhere between the first and the fourth quarter, a phenomenon that points to the importance of private law (topics 15–17) in our texts, as compared to, for example, ritual and public law.

99 *EP*, s.v. "Al-Shaybānī" (Eric Chaumont).

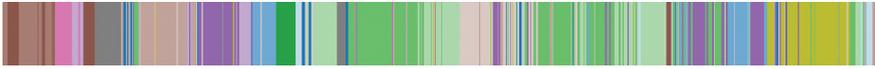


FIGURE 9.1 Sarakhsi, *Mabsūt* [2,494,216 tokens]



FIGURE 9.2 Māwardī, *Ḥāwī* [2,713,331 tokens]



FIGURE 9.3 Qarāfi, *Dhakhīra* [1,417,018 tokens]



FIGURE 9.4 Ibn Muflīh, *Mubdi'* [998,921 tokens]



FIGURE 9.5 Ṭūsī, *Mabsūt* [800,420 tokens]

Also noteworthy is that private law issues appear to be discussed in the same sequence in all five texts, from slaves (topic 17; dark green), to estates (topic 16; grass green), to sales (topic 15; light green). Public law (topics 19–20; light purple and dark purple) unsurprisingly occupies a position near the end of our texts. The intersection of public law topics with ritual law (topic 10: oaths; fasting; expiation; powder blue) is noteworthy, suggesting that jurists regularly thought about punishment in terms of expiation.¹⁰⁰ In al-Sarakhsī's *al-Mabsūt*, public law topics are evenly distributed across the entire text, which confirms our observation made in the previous section that al-Sarakhsī is an author with an above-average interest in public law and its institutions—according to his biographers, he dictated most parts of his *K. al-Mabsūt* to students while in prison (he spent a total of fourteen years in captivity).¹⁰¹ Finally, let us note

100 As noted by Christian Lange, “Public Order,” in *The Ashghate Companion to Islamic Law*, 164.

101 See Osman Taştan, “Al-Sarakhsī (d. 483/1090),” in *Islamic Legal Thought: A Compendium of Muslim Jurists*, 243, 246.

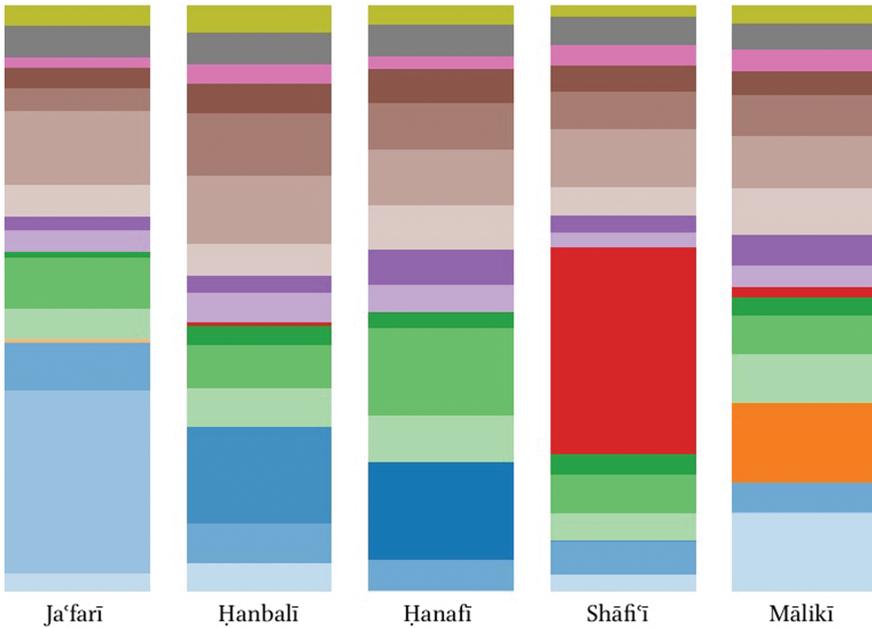


FIGURE 10 Topics per law school

that topic 14 (personal law: inheritance; orange) surfaces only in al-Qarāfi, for reasons that remain to be explored (see below).

In addition to plotting the topics of individual texts, we here present combined plots of topics according to all five schools. In Figure 10, the normalized vertical bars indicate the relative weight of topics in the five subcorpora.

Certain peculiarities of individual law schools are visible here. Shāfiʿī, for example, devote an unusual amount of attention to topic 18 (procedural law; red). If we pursue this matter to the level of individual texts in our digital corpus,¹⁰² we find that this peculiarity appears first in the three texts of Ibn Ḥajar al-Haytamī (Cairo/Mecca, d. 973/1565), *Tuḥfat al-muḥtāj*, Shihāb al-Dīn al-Ramlī (Cairo, d. 1004/1595), *Nihāyat al-muḥtāj*, and Sulaymān b. ʿUmar al-Jamal (d. 1204/1789), *Ḥāshiyat al-Jamal*, a commentary on the *Manhaj al-tullāb* by Zakariyyāʾ b. Muḥammad al-Anṣārī (Cairo, d. 926/1520), a judge who taught *fiqh* to both Ibn Ḥajar al-Haytamī and al-Ramlī.¹⁰³ In other words, if there

¹⁰² For a visualization, see cell 38 of the plot found on <https://github.com/arabic-digital-humanities/adhtools/blob/master/notebooks/TopicModelingVisualization20.ipynb>.

¹⁰³ See Mahmood Kooriadhodi, “Cosmopolis of Law: Islamic Legal Ideas and Texts across the Indian Ocean and Eastern Mediterranean World” (PhD Leiden 2016), 158. On al-Anṣārī, see *EF*³, s.v. “Al-Anṣārī, Zakariyyāʾ” (Richard J. McGregor).

is a predilection for procedure among the Shāfiʿī authors in our corpus, it is especially visible in the texts of Egyptian Shāfiʿī jurists of the early Ottoman era.

In addition to the *Tuhfa* and the *Nihāya*, two highly influential commentaries on the *Minhāj al-ṭālibīn* of al-Nawawī (d. 676/1277),¹⁰⁴ Ibn Ḥajar al-Haytamī and al-Ramlī are famous for attaching their names to vast collections of *fat-wās*. This demonstrates their commitment to the practical application of legal doctrine and thus may explain their interest in procedural questions. On the one hand, the two authors were continuing the interest of al-Nawawī's *Minhāj al-ṭālibīn* in questions of procedure. Al-Nawawī devotes three separate chapters to, respectively, the office of the judge (*k. al-qadāʾ*), witnessing (*k. al-shahādāt*) and legal claims and proofs (*k. al-daʿwā wa'l-bayyināt*), some thirty pages (out of 540 pages) in the printed editions of his work.¹⁰⁵

On the other hand, there is a noticeable surge of topic 18 in Ibn Ḥajar al-Haytamī and al-Ramlī's texts, and later, in that of al-Jamal. We should note that Ibn Ḥajar al-Haytamī and al-Ramlī, as well as their teacher al-Anṣārī, wrote during a period of political and legal insecurity, brought about by the Mamluk-Ottoman war (890–923/1485–1517) and its aftermath.¹⁰⁶ They witnessed, in the words of Leslie Peirce, the “integration of the court[s] into an empire-wide legal system, and a program of legal reform that was being scripted in Istanbul.”¹⁰⁷ As a result of this process, Ḥanafī legal doctrine and Ḥanafī judges were granted precedence over the doctrines and judges of the other schools.¹⁰⁸ One of the more contentious issues dividing Ḥanafīs and Shāfiʿīs was procedural law. Shāfiʿī jurists, unless they ‘converted’ to the Ḥanafī *madhhab*,¹⁰⁹ experienced a loss of control over the judicial process.

104 Joseph Schacht refers to these two commentaries as “the two authoritative textbooks of the Shāfiʿī school”. See *EP*, s.v. “Ibn Ḥadjar al-Haytamī” (Joseph Schacht [C. Van Arendonk]). On Shihāb al-Dīn al-Ramlī and his son Shams al-Dīn, see *EP*, s.v. “Al-Ramlī” (Aron Zysow).

105 Abū Zakariyyāʾ al-Nawawī, *Minhāj al-ṭālibīn wa-ʿumdat al-muftiyīn* (Beirut: Dār al-Minhāj, 1426/2005), 568–90.

106 The social disarray at the time is famously lamented by Aḥmad b. ʿAli al-Maqrīzī, *Kitāb al-mawāʿiẓ wa'l-iʿtibār bi-dhikr al-khiṭaṭ wa'l-āthār*, 2 vols. (Cairo: Maktabat al-Thaqāfa al-Dīniyya, 1987), 2:221.

107 Leslie Peirce, *Morality Tales: Law and Gender in the Ottoman Court of Aintab* (Berkeley: University of California Press, 2003), 10.

108 On this process, see Rudolph Peters, “What Does it Mean to Be an Official *Madhhab*? Ḥanafism and the Ottoman Empire,” in *The Islamic School of Law: Evolution, Devolution, and Progress*, ed. Peri Bearman, Rudolph Peters and Frank Vogel (Cambridge, MA: Harvard University Press, 2005), 151, 155–57; Guy Burak, *The Second Formation of Islamic Law: The Ḥanafī School in the Early Modern Ottoman Empire* (Cambridge: Cambridge University Press, 2015), 10–20, passim. See also the pertinent comments by Kooriadathodi, “Cosmopolis of Law,” 173, 179; Peirce, *Morality Tales*, 6.

109 Guy Burak notes the case of Khayr al-Dīn al-Ramlī (d. 1081/1671). See Burak, *The Second Formation of Islamic Law*, 193. On the biography/hagiography of Khayr al-Dīn al-Ramlī,

Our computational analysis highlights how Shāfiʿī jurists of the period claimed a measure of authority over the courtroom in their written work, to compensate for losing influence in the judiciary.

Another clear irregularity is the emphasis in the Mālikī school on topic 14 (personal law: inheritance; orange), which echoes what we observed in the previous section regarding al-Qarāfi's *Dhakhīra*. In the list of texts most heavily focused on topic 14, five Mālikī texts lead in the ranking: Ibn Abī Zayd's (d. 386/966) *al-Nawādir* (91 of 199 tagged sections [32.4% of tokens]); al-Muwwāq's (d. 897/1491) *al-Tāj wa'l-iklīl* (19 of 64 tagged sections [29.1% of tokens]); al-Qarāfi's *al-Dhakhīra* (11 of 261 tagged sections [10.5% of tokens]); Ibn 'Abd al-Barr's (d. 483/1090) *al-Kāfi* (10 of 351 tagged sections [3.3% of tokens]); and Mālik's *Muwattā'* (12 of 664 tagged sections [1.9% of tokens]). Remarkably, no other work in the corpus includes sections in which topic 14 is salient. "The Mālikī system of inheritance," wrote Noel Coulson, "[has] a distinct character of its own," based on Mālikīs' insistence that the Public Treasury, in the absence of agnate relatives (*aṣaba*), succeeds as residuary heir to an estate.¹¹⁰ Applying computational topic modelling to *furū' al-fiqh* broadly confirms this impression.

Given the constraints of space we cannot discuss other noteworthy differences in detail. Let us note, however, that Ja'farīs emphasize ritual law issues (especially topics 8–10) more than the other schools do, especially Shāfiʿīs. The Ḥanafī subcorpus, by comparison with the other four subcorpora, devotes greater attention to topic 12 (ritual law: alms; taxes; dark blue), topic 15 (private law: contracts; sales; light green) and topic 16 (private law: contracts; estates; grass green). Also, topic 19 (public law: homicide; punishments; light purple) and topic 20 (public law: warfare; dark purple) seem to be a Ḥanafī specialty, with Mālikīs a close second. Ḥanbalī texts show a special affinity for terminology related to legal reasoning (topics 1 through 4; in various shades of brown)—presumably, the result of their emphasis on transmitted, hadith-based opinion. Also topic 11 (ritual law: pilgrimage; light azure) is disproportionately important in Ḥanbalī texts.

5 Conclusions

Our approach to the digital corpus of Islamic jurisprudence from different computational perspectives has yielded a heterogeneous range of insights and

see Judith E. Tucker, "Biography as History: The Exemplary Life of Khayr al-Din al-Ramli," in *Auto/Biography and the Construction of Identity and Community in the Middle East*, ed. Mary Ann Fay (New York: Palgrave, 2001), 9–18, esp. 14.

¹¹⁰ Coulson, *History*, 97–98.

conclusions that defy easy summary. And yet, if we wish to claim, as we do, that “the tools are here”, we cannot avoid the question that perennially plagues digital humanists: “what about results?”¹¹¹

Let us recapitulate. In part 2 of our study, we found that, contrary to common wisdom, Ḥanafīs do not rely on the Qurʾān less than the other law schools. If anything, it is Mālikīs who do so. Shāfiʿīs, by contrast, refer to Qurʾānic evidence most frequently of all the five law schools. We further observed a general preference of Muslim jurists to quote verses from the Qurʾān that relate to ritual worship (*ibāda*) as well as to numerically defined norms, especially in the area of inheritance, marriage, and divorce. We visualized Qurʾān reliance of the law schools in an interactive network; this enabled us to appreciate the multifunctional “footprint” of the Qurʾān in Islamic law, that is, that the jurists in our corpus are concerned with far more than only the verses that are immediately relevant in a strictly legal sense. The network also helped us to identify texts that build upon a central repertoire of Qurʾānic verses shared across the entire *furūʿ al-fiqh* tradition, for example, al-Shāfiʿī’s *K. al-Umm*, al-Kāsānī’s *Badāʾiʿ al-ṣanāʿiʿ* and Ibn Idrīs’ *al-Sarāʾir*. We also identified texts that are conspicuously ‘unorthodox’ in their use of Qurʾānic evidence—whether because they largely ignore the Qurʾān (e.g., al-Ḥillī’s *Sharāʾiʿ al-islām*) or because they rely on a group of verses ignored by other authors (e.g., al-Barqī’s *al-Maḥāsīn*).

In part 3 of our study we examined the distribution of normative qualifications (*aḥkām*) in our *furūʿ al-fiqh* corpus, as well as in the *madhhab* subcorpora. In search of “moral language” (Fadel) and “doctrinal tone” (Powers) we examined the *ḥalāl/ḥarām* ratio across the five law schools. Although margins are small, we found that in quantitative terms, Jaʿfarīs and Ḥanbalīs tend towards the language of *ḥarām*, while Ḥanafīs and Mālikīs tend towards the language of *ḥalāl*. Our machine-guided analysis of the corpus further suggested a gradual process of “haramization”: the *ḥalāl/ḥarām* ratio slowly shifts in favor of *ḥarām*, across all five law schools, up to the 8th/14th century, after which it remains stable. As regards the five normative qualifications (*al-aḥkām al-khamsa*), we found that the Mālikīs give the widest scope to the middle categories (*mandūb/mustaḥabb*, *mubāḥ/ḥalāl* and *makrūh/qabīḥ*). Both in the Mālikī subcorpus and in the corpus as a whole, however, the middle, “moral” categories are outnumbered by the outer categories (*wājib/farḍ* and *ḥarām/maḥẓūr*)—a finding that casts doubt on the repeated assertions in the scholarly literature that Islamic law is legal and moral in equal measure.

Our examination of the topical distribution in our corpus, in part 4 of our study, demonstrated the centrality of prayer and property in *furūʿ al-fiqh*.

¹¹¹ McCarty, “A Telescope for the Mind?,” 117.

Parsimonious word clouds revealed the prominent role played by certain, not always well-known authors and texts in their respective *madhhabs*, for example, Ibn Ḥamdān's *al-Ri'āya* in the Ḥanbalī school, or Qāḍikhān's and al-Bazzāzī's *Fatāwā* in the Ḥanafī school. In the most experimental part of our study, topic modeling allowed us to see that, in diachronic perspective, questions of ritual law dominate the early texts in our corpus. In synchronic perspective, we found that ritual law occupies more space among Ḥanbalīs and Ja'farīs than in the other schools. Ḥanafīs emphasize public law and commerce, Mālikīs display a great interest in inheritance law, and Shāfi'īs (or at least a certain group of Shāfi'īs writing between the 10th/16th and the 12th/18th century) are much concerned with procedural law.

Beyond these findings, our primary aim in this article has been to introduce and to test the promise of a novel methodology, that is, the computational text mining of *furū' al-fiqh*. Reprising part 1 of this study, we conclude with four methodological reflections, and advance some suggestions for further research along computational lines. First, the results of studies such as the one presented here must be replicable, which means that there must be a sustainable and open-to-all environment in which relevant data are stored. Readers are encouraged to check our findings by referring to the metadata and text files of our corpus released on Zenodo, as well as to the codes developed to support our analysis, made available through Github; and then to run their own analyses on BlackLab ADH and the Qur'ān Footprinter, both hosted by the Digital Humanities Lab at Utrecht University.

Second, the digital corpus of *furū' al-fiqh* deserves to be further curated and expanded. In the future, one important solution to the problem of bias in the corpus will be to grow the corpus in several directions: not one text per century, but several (focusing on those texts that were used most frequently, rather than those that happen to be digitally available), and not a mixture of genres, but full coverage of all genres. Likewise, texts from the Ibādī and other law schools should be included in future reiterations of this study. This, however, will have to wait until a greater number of texts, especially for the later, post-classical centuries, become available, a process that will require teamwork. No single scholar can carry out the laborious task of compiling such a corpus and preparing it for computational analysis.

Third, the text mining tools used in this study are far from exhaustive. A number of existing digital text-mining techniques are absent from our analysis. These include tools to detect text reuse that would enable us, for example, to study the hadith footprint in the corpus. The frequency-based analysis of concepts in Arabic texts or textual corpora, as this study has shown, must be based on complex, clustered searches, rather than on searches for simple

words, stems or roots. Other techniques, such as topic modeling, are largely untested in IDH. The present article is a first step to illustrate their usefulness.

Fourth, and finally, text mining the digital corpus in the full sense of the Digital Humanities requires manpower and time. Researchers in IDH must be aware of this fact, ready to work in teams, and willing to do spadework. Collaboration in local teams should be complemented by international collaboration between research institutions and projects. Only then will IDH—including the computational study of Islamic law, but also of other text genres—emerge from its current niche into full light, and move from experimental exploration to sustained analysis and output.

Acknowledgements

The authors gratefully acknowledge support by the ERC Consolidator Grant project “The senses of Islam” (project no. 724951), by the CLARIAH/Netherlands eScience Center project, “Bridging the gap: Digital humanities and the Arabic-Islamic corpus” (project no. ADAH2016.014), and by the Digital Humanities Lab at Utrecht University during the research for, and writing of, this article. We would also like to thank the two anonymous reviewers for their thorough critique of the original version of this article, as well as all colleagues who generously provided advice and resources along the way.

Appendix: The Corpus

1. Ḥanafis:

al-Shaybānī (d. 189/805), *al-Aṣl* [319,286 tokens, <http://shamela.ws/index.php/book/6164>]; al-Qudūrī (d. 428/1037), *al-Mukhtaṣar* [43,452 tokens, <http://shamela.ws/index.php/book/124336>]; al-Sarakhsī (d. ca. 482/1090), *al-Mabsūṭ* [2,494,216 tokens, <http://shamela.ws/index.php/book/5423>]; al-Kāsānī (d. 587/1191), *Badā'ī' al-ṣanā'ī'* [1,464,584 tokens, <http://shamela.ws/index.php/book/8183>]; al-Mawṣilī (d. 683/1284), *al-Ikhtiyār* [322,102 tokens, <http://shamela.ws/index.php/book/1066>]; al-Bābartī (d. 786/1384), *al-Ināya* [1,190,224 tokens, <http://shamela.ws/index.php/book/9403>]; Ibn al-Humām (d. 861/1457), *Fatḥ al-qadīr* [2,064,650 tokens, <http://shamela.ws/index.php/book/21744>]; Ibn Nujaym (d. 970/1563), *al-Baḥr al-rā'iq* [1,730,647 tokens, <http://shamela.ws/index.php/book/12227>]; Shaykhzādeh (d. 1078/1667–8), *Majma' al-anhur*

[595,367 tokens, <http://shamela.ws/index.php/book/21644>]; Ibn ‘Ābidīn (d. 1252/1836), *al-Hāshiyā* [2,550,368 tokens, <http://shamela.ws/index.php/book/21613>]; al-Maydānī (d. 1363/1886), *al-Lubāb* [231,605 tokens, <https://www.shamela.ws/index.php/book/21496>].

2. Mālikis:

Mālik b. Anas (d. 179/795), *al-Muwatṭāʾ* [151,424 tokens, <http://shamela.ws/index.php/book/28107>]; Ibn Abī Zayd al-Qayrawānī (d. 386/966), *al-Nawādir wa'l-zīyādāt ‘alā mā fī ‘l-Mudawwana* [1,715,091 tokens, <http://shamela.ws/index.php/book/96257>]; Ibn ‘Abd al-Barr (d. 483/1090), *al-Kāfi fī fiqh ahl al-Madīna* [226,537 tokens, <http://shamela.ws/index.php/book/21731>]; al-Māzarī (d. 536/1141), *Sharḥ al-talqīn* [818,339 tokens, <https://www.shamela.ws/index.php/book/121376>]; al-Qarāfī (d. 684/1285), *al-Dhakhīra* [1,417,018 tokens, <http://shamela.ws/index.php/book/1717>]; Khalīl b. Ishāq al-Jundī (d. 776/1365), *al-Tawḍīḥ fī sharḥ Mukhtaṣar [= Jāmi‘ al-ummahāt] Ibn al-Hājib* [1,177,391 tokens, <http://shamela.ws/index.php/book/14442>]; al-Muwawāq (d. 897/1491), *al-Tāj wa'l-iklīl li-Mukhtaṣar al-Khalīl* [781,273 tokens, <http://shamela.ws/index.php/book/21611>]; al-Manjūr (d. 995/1586), *Sharḥ al-manhaj al-muntakhab ilā qawā‘id al-madhhab* [126,657 tokens, <http://shamela.ws/index.php/book/18279>]; al-Zurqānī (d. 1122/1710), *Sharḥ al-Zurqānī ‘alā Muwatṭāʾ al-imām Mālik* [972,714 tokens, <http://shamela.ws/index.php/book/551>]; al-Ḍasūqī (d. 1230/1815), *Hāshiyat al-Ḍasūqī ‘alā al-Sharḥ al-kabīr* [1664,518 tokens, <http://shamela.ws/index.php/book/21604>]; al-Azharī (d. 1335/1916), *al-Thamar al-dānī... sharḥ Risālat Ibn Abī Zayd al-Qayrawānī* [174,005 tokens, <http://shamela.ws/index.php/book/7441>].

3. Shāfi‘is:

al-Shāfi‘ī (d. 204/820), *al-Umm* [1,205,588 tokens, <http://shamela.ws/index.php/book/1655>]; al-Muzanī (d. 264/877), *al-Mukhtaṣar* [201,828 tokens, <http://shamela.ws/index.php/book/1661>]; al-Māwardī (d. 450/1059), *al-Hāwī al-kabīr* [2,713,331 tokens, <http://shamela.ws/index.php/book/6157>]; al-Ghazālī (d. 505/1111), *al-Wasīṭ* [404,649 tokens, <http://shamela.ws/index.php/book/6128>]; al-Nawawī (d. 676/1277), *Minhaj al-ṭālibīn* [88,670 tokens, <http://shamela.ws/index.php/book/12096>]; Ibn al-Mulaqqin (d. 804/1401), *Ujālat al-muḥtāj* [377,609 tokens, <http://shamela.ws/index.php/book/20561>]; al-Ḥiṣnī (d. 829/1426), *Kifāyat*

al-akhyār [222,002 tokens, <http://shamela.ws/index.php/book/6140>]; Ibn Ḥajar al-Haytamī (d. 973/1565), *Tuḥfat al-muḥtāj* [1,051,176 tokens, <http://shamela.ws/index.php/book/9059>];¹¹² al-Ramlī (d. 1004/1595), *Nihāyat al-muḥtāj* [1,017,677 tokens, <http://shamela.ws/index.php/book/3565>];¹¹³ Sulaymān b. ‘Umar al-Jamal (d. 1204/1789), *Ḥāshiyat al-Jamal* [2,251,387 tokens, <http://shamela.ws/index.php/book/21598>]; al-Dimyātī (d. 1369/1950), *Ḥāshiyat al-Dimyātī* [961,599 tokens, <http://shamela.ws/index.php/book/963>].

4. Ḥanbalis:

al-Khallāl (d. 311/923), *al-Wuqūf wa’l-tarajjul* [32,349, <http://shamela.ws/index.php/book/26883>]; al-Khiraqī (d. 334/945–6), *al-Mukhtaṣar* [36,539, <http://shamela.ws/index.php/book/2977>]; al-Qāḍī Abū Ya‘lā (d. 458/1066), *al-Masā’il al-fiqhiyya* [184,234, <http://shamela.ws/index.php/book/13246>]; Ibn al-Jawzī (d. 597/1201), *al-Taḥqīq fī aḥādīth al-khilāf* [215,473, <http://shamela.ws/index.php/book/5907>]; Ibn Qudāma (d. 620/1223), *al-Mughnī* [1,961,758, <http://shamela.ws/index.php/book/8463>]; Muḥammad Ibn Muflīḥ (d. 763/1362), *K. al-Furū‘* [845,176, <http://shamela.ws/index.php/book/12052>]; Ibrāhīm Ibn Muflīḥ (d. 884/1479), *al-Mubdi‘ fī sharḥ al-Muqni‘* [998,921, <http://shamela.ws/index.php/book/21619>]; Ibn al-Najjār (d. 972/1564), *Muntahā al-irādāt fī jam‘ al-Muqni‘* [140,806, <http://shamela.ws/index.php/book/13664>]; al-Buhūtī (d. 1051/1641), *al-Rawḍ al-murbi‘* [153,334, <http://shamela.ws/index.php/book/13664>]; Ibn ‘Abd al-Wahhāb (d. 1206/1791), *Majmū‘at al-ḥadīth ‘alā abwāb al-fiqh* [364,504 tokens, <http://shamela.ws/index.php/book/12061>]; al-Nābulusī (d. 1319/1901), *Ḥāshiyat al-Labadī ‘alā Nayl al-mā‘arib* [118,608 tokens, <http://shamela.ws/index.php/book/97809>].

112 The digital edition stored on *al-Maktaba al-shāmila* comes with the two supercommentaries, or *Ḥāshiyas*, of al-‘Ubbādī and al-Shirwānī. We deleted both commentaries from the version we included in our corpus.

113 The digital edition stored on *al-Maktaba al-shāmila* comes with the two supercommentaries, or *Ḥāshiyas*, of al-Shabrāmālīsī and al-Maghribī al-Rashīdī. We deleted both commentaries from the version we included in our corpus.

5. Ja'faris:

al-Barqī (d. 274/887), *al-Maḥāsīn* [145,940 tokens, <http://siratali.org/maktaba/mahasini/>]; al-Shaykh al-Mufīd (d. 413/1022), *al-Muqni'a* [148,580 tokens, <http://shiaonlinelibrary.com/المقنعة-الشيخ-المفيد-الكتب/11>]; al-Ṭūsī (d. 460/1067), *al-Mabsūṭ* [800,420 tokens, <http://www.yasoob.com/books/html/m001/00/n00032.html>]; Ibn Idrīs (d. 598/1202), *al-Sarā'ir* [516,557 tokens, <http://shiaonlinelibrary.com/السرائر-ابن-إدریس-الحلي-49>]; al-Muḥaqqiq al-Ḥillī (d. 676/1277), *Sharā'i' al-islām* [1,464,584 tokens, <http://www.yasoob.com/books/html/m001/00/n00053.html>]; al-'Allāma al-Ḥillī (d. 726/1325), *Tadkhirat al-fuqahā'* [714,294 tokens, <http://shiaonlinelibrary.com/تذکرة-الفقهاء-ط-ح-العلامة-الحلي-75>]; Ibn Fahd (d. 841/1437), *al-Muḥadḥhab al-bārī'* [998,921 tokens, <http://ar.lib.eshia.ir/10053/1/4>]; al-Shahīd al-Thānī (d. 965/1557), *Masālik al-afḥām* [1,506,727 tokens, <http://ar.lib.eshia.ir/10151/0/7>]; al-Fayḍ al-Kāshānī (d. 1091/1680), *Mafātiḥ al-sharā'i'* [221,494 tokens, <https://www.masaha.org/bookview/view.php?bid=1393>]; al-Baḥrānī (d. 1186/1772), *al-Ḥadā'iq al-nāḍira* [3,599,589 tokens, <http://ar.lib.eshia.ir/10013/1/2>]; al-Ṭabāṭabā'ī (d. 1231/1816), *Riyāḍ al-masā'il* [1,333,720 tokens, <http://ar.lib.eshia.ir/10098/11/3>].

