Observations of a Medieval Quantitative Historian?

Maxim Romanov
University of Leipzig

September 26, 2016

Abstract

The comparison of the results of the computational analysis of the *Taʾrīḫ al-islām*, al-Ḏahabī's 50-volume biographical collection, with brief statements that describe the rise and decline of cities and provinces of the Islamic world from the *al-Amṣār ḏawāt al-āṯār*, al-Ḏahabī's 4-folio epistle, suggests that al-Ḏahabī had a solid grasp of the tremendous amount of biographical and historical data that he collected, and that his short epistle may be regarded as a missing analytical summary of the most ambitious historical project in the pre-modern Islamic world. In the light of these results, we perhaps may think of al-Ḏahabī as one of the earliest quantitative historians. Although we do not have conclusive evidence about how exactly al-Ḏahabī worked with his data, the paper argues that all necessary mathematical, visual and 'mechanical' techniques that would facilitate data analysis already existed, and that al-Ḏahabī and other premodern Islamic historians could have used them.

The article has been accepted for publication in *Der Islam*.

NB: This PDF file may be a compressed version with low-quality images; the original version with high-quality images can be downloaded from `maximromanov.github.io/files/Romanov-OMQH-2016.pdf`

Technical notes

Note on data and visualizations: All data, graphs and cartograms used in the article were produced by the author. The data was extracted with `Python` (`www.python.org`) scripts from the electronic text of a medieval Arabic biographical collection available online in open access. Graphs and cartograms are based on the extracted data and produced in `R` (`www.r-project.org`), a free software environment for statistical computing and graphics, and `D3` (`d3js.org`), a JavaScript library for building interactive data-driven documents.

Note on transliteration: The article uses a somewhat unconventional transliteration system, which was developed to facilitate computational analysis. Unlike more traditional transliteration schemes the current one uses one-to-one letter representation, with every Arabic letter transcribed distinctively, which allows for an automatic conversion between transliteration and the Arabic script. The overall scheme should be easily recognizable to Arabists (new letters are as follows: *ŧ* for *tā' marbuṭat*; *ã* for dagger *alif*; and *á* for *alif maqṣūrat*). Additionally, all attached conjunctions, prepositions, pronominal suffixes are separated with "-". The final version will adopt the transliteration system of the edition where it it to be published. Whenever applicable, toponyms are given in their current American spelling. Bibliographical references and quotations preserve their original transliteration schemes.

If a hammer in a hand makes everything look like nails, wouldn't lots of nails beg for something that works like a hammer?

Introduction

Analyzing quantitative data from vast biographical collections, one may become puzzled with a question whether the author of a certain biographical collection had a comprehensive view of the data that he assembled. Could his familiarity with the data have led him to observations similar to what we can discover relying on advanced graphing techniques of time-series analysis?[1] A professor[2] at the University of Michigan posed this question to me when I was discussing with him some early results of my computational analysis of al-Ḏahabī's *Taʾrīḫ al-islām*,[3] at the moment being utterly overwhelmed by the avalanche of frequency lists, graphs, cartograms, collocation tables and word clouds.[4]

"[O]ne of the most ambitious histories of the entire world of Islam,"[5] the *Taʾrīḫ al-islām* is a 50-volume mammoth of Islamic biographical literature that covers the first seven centuries of Islamic history and includes over 30,000[6] biographical records arranged chronologically into decades. However, this giant book lacks anything that could offer even a preliminary answer to whether al-Ḏahabī had a wholistic view of his historical and geographical data. Even though al-Ḏahabī frequently comments on specific events and individuals, nowhere in the *Taʾrīḫ al-islām* does he attempt to put his historical data into analytical perspective: the book has no concluding section and features only a brief introduction that consists mostly of the list of over forty sources that he used for its composition.

It seems that an unusually brief work of al-Ḏahabī—*al-Amṣār ḏawāt al-āṯār* ("Cities and ports for hearing the reports")[7]—may be the missing analytical partner text

[1] Time series refers to a chronologically ordered sequence of values of a variable at equally spaced time intervals. Time-series analysis is a set techniques that are used to study patterns in such data. The most frequent of these techniques is a graph of chronological change, which you find in this article.

[2] This article is an accidental response to one of the many intriguing questions posed to me by Andrew Shryock, then a member of my dissertation committee. See, Romanov, "Computational Reading.".

[3] My work is based on the text of the *Taʾrīḫ al-islām* from *al-Ǧāmiʿ al-kabīr*, a collection of about 2,400 Arabic texts (mostly premodern) published by al-Turāṯ (Jordan; comes on an external hard drive). The text itself is based on (and has been collated by me with): al-Ḏahabī, *Taʾrīḫ al-islām*; on this source, see: Somogyi, "The *Taʾrīkh al-islām* of adh-Dhahabī."

[4] Over 800 visualizations ended up being included in my dissertation, which is, however, is but a small part of over 20,000 exploratory visualizations that resulted from my computational analysis of the *Taʾrīḫ al-islām*.

[5] Lucas, *Constructive Critics*, 43.

[6] In terms of chronological scope and biographical coverage it is indeed the most ambitious biographical-cum-annalistic work ever composed in the course of Islamic history; lengthwise the *Taʾrīḫ al-islām*, whose volume is approximately 2,9 mln. words, is second only to Ibn ʿAsākir's (d. 571/1175 CE) *Taʾrīḫ madīnat+ Dimašq* (approximately 8,1 mln. words), whose chronological and biographical coverage, however, is significantly smaller.

[7] This excellent translation was offered by Michael Cooperson. Franz Rosenthal, who translated

to al-Ḏahabī's book. The exact opposite of the *Taʾrīḫ al-islām*, the *Amṣār* is a mere 4-folio epistle where al-Ḏahabī briefly characterizes the role of different urban centers and provinces of the Islamic world in Ḥadīṯ scholarship up to his own time. Unlike most of al-Ḏahabī's works which focus on individuals, the subject of the *Amṣār* is cultural geography; most interestingly, al-Ḏahabī occasionally characterizes periods when these centers thrived by using direct or indirect chronological statements: in the case of direct statements, he explicitly names periods;[8] alternatively, he refers to specific generations or particular individuals.

The *Amṣār* has already attracted its share of scholarly attention, and modern scholars who studied this epistle tend to agree with al-Ḏahabī's assessments. However, existing assessments of al-Ḏahabī's *Amṣār* are methodologically problematic, since they are based either on al-Ḏahabī's reputation as a prominent Ḥadīṯ scholar,[9] or on evaluations of al-Ḏahabī's own sample of biographical data in the *Amṣār*.[10] In many ways, scholarly attempts to assess the reliability of al-Ḏahabī's statements in the *Amṣār* and the main question of the article are related, and a methodological solution that will be offered in what follows should shed light on both issues.

Methodological considerations

My previous work on al-Ḏahabī's *Taʾrīḫ al-islām* allows me to compare the contents of this enormous collection with al-Ḏahabī's statements in the *Amṣār*. The results should give us a better understanding of this short, but arguably crucial epistle, as well as to offer an insight into al-Ḏahabī's historical methodology.

The dataset formed from the *Taʾrīḫ al-islām* includes about 29,000 biographies of

---

the title as "Main cities in which traditions were cultivated," decided to exclude this epistle from his translation of al-Saḫāwī's *al-Iʿlān bi-l-tawbīḫ li-man ḏamma ahl al-taʾrīḫ*, where it was included in full by the author. See, Rosenthal, *A History of Muslim Historiography*, 409. Otherwise, the epistle was published at least three times in the 1980s (because of extensive annotations, which, however, do not add to our understanding of the epistle, some editions exceed a hundred pages): Librande, "al-Dhahabī's Essay"; al-Ḏahabī, *al-Amṣār*, 1985; al-Ḏahabī, *al-Amṣār*, 1986. On the *Amṣār* also see: al-Šayḫ, *al-Ḥāfiẓ al-Ḏahabī*, 447–48.

[8] For example, "then, in the course of the third century, the learning in the sacred cities diminished, but became abundant in other places" (*ṯumma fī aṯnāʾi l-miʾat+i l-ṯāliṯat+i tanāqaṣa ʿilmu l-ḥaramayni wakaṯura bi-ġayri-himā*). al-Ḏahabī, *al-Amṣār*, 1985, 20.

[9] Fuat Sezgin accepts that in his *Amṣār* al-Ḏahabī "gives us comprehensive information about the centres for *ḥadīth*-study and their distribution in different centuries throughout the Muslim world." (Sezgin, Fuat. "Dār al-Ḥadith," in *EI2, Brill Online*.) However, the epistle is very short and sketchy to take it as a reliable assessment on its own. See also the above-mentioned Arab editions of the *Amṣār*, and al-Šayḫ, *al-Ḥāfiẓ al-Ḏahabī*, 447–48.

[10] Librande offered a convincing analysis of this epistle by identifying its place in the larger context of the *ʿilm al-riǰāl*, "the science of the transmitters [of Ḥadīṯ]," and looking into 187 Ḥadīṯ specialists who were listed by al-Ḏahabī in this epistle as exemplar representatives of different settled regions. Occasionally puzzled by al-Ḏahabī's choices, Librande nonetheless found al-Ḏahabī's representation convincing. However, 187 scholars is but a tiny sample (compared to the *Taʾrīḫ al-islām*), which makes Librande's assessment equally problematic. See, Librande, "al-Dhahabī's Essay," 123–29 in particular.

individuals who died in the period of *c.* 40–700/661–1300 CE.[11] The prevailing majority of individuals were included in the *Taʾrīḫ al-islām* because they were involved in the transmission of Ḥadīṯ (over 90%),[12] even though they did not necessarily make any noteworthy contributions to this area. Relying on onomastic data we can compare chronological curves of individuals associated with particular regions with al-Ḏahabī's descriptions of those regions in the *Amṣār*.

Before we turn to the comparison of statements from the *Amṣār* and the graphs of relevant data from *Taʾrīḫ al-islām*, some methodological assumptions must be made explicit. The graphs that will follow are based on "descriptive names" (sing. *nisbaṯ*), and anyone who ever worked with biographical collections is likely to object that not every individual identified as, for example, "al-Madanī" was actually a Medinan, as well as there are Medinans who are not identified as such with this specific toponymic *nisbaṯ*, not to mention that the "descriptive name" al-Madanī (and its variation al-Madīnī) may refer to urban centers other than Medina.[13] The situation with "descriptive names" is indeed complicated, and such objections are not invalid. However, at this point of our understanding of overabundant Islamic onomastic data—as well as biographical data more broadly—both sides of the issue of whether we can or cannot use "descriptive names" at their face value are impossible to prove:[14] we simply do not know to what extent the presence of false positives (i.e., Madanīs who have nothing to do with Medina) and the absence of false negatives (i.e., the Medinans who are not identified as Madanīs) actually affects the overall picture. Until some solid data are provided to convincingly support either side of the issue, historians can operate only on the level of explicit methodological assumptions.

The case of the Muqaddasī family—the famous Ḥanbalī family of the Banū Qudāmaṯ—is quite interesting from the perspective of relying on the face value of *nisbaṯ*s. The *nisbaṯ* "al-Maqdisī/al-Muqaddasī" refers to Jerusalem (Bayt al-Maqdis, or al-Bayt al-Muqaddas), and, technically, the family name al-Muqaddasī does refer to the region of Jerusalem. From the history of this clan we know that they indeed were natives of Palestine, but as a strong scholarly family they appear only after they establish themselves in Damascus. In the *Amṣār* al-Ḏahabī writes that Jerusalem was never a center of learning, and as the data from the *Taʾrīḫ al-islām* shows, indeed there are almost no individuals with the name "al-Muqaddasī" until after 500/1107 CE—the period when Damascus becomes the leading center. It seems that in the scholarly circles the name "Jerusalemite" was not much in use, which allowed this toponymic name to be re-appropriated for as a family one.

[11] The first three volumes of this text, which cover the period up to 40/660 CE, have different structure (biographies are not presented as distinct units), and for this reason I had to exclude from the analysis.

[12] More specifically, these individuals are identified through the presence of transmission statements of various kinds in their biographies, such as, for example, *wa-rawá ʿan fulān bn fulān*, "he transmitted from so-and-so," and their numerous variations and equivalents.

[13] See, for example, al-Samʿānī, *al-Ansāb*, 5:235–239.

[14] Romanov, "Computational Reading," 28–35.

Although the situation with *nisbat*s may appear confusing, it in fact can be resolved through collocation analysis—that is by looking into what other *nisbat*s are applied to individuals who bear the name "al-Muqaddasī." In the *Amṣār*, al-Ḏahabī writes about the Muqaddasīs (pl. al-Maqādisaṯ) in the context of Damascus, and—if we look at the *Taʾrīḫ al-islām*—the most frequent *nisbat*s of the Muqaddasīs in the period of 500–700/1107–1301 CE are "al-Dimašqī" and "al-Ṣāliḥī," with the first referring to the city of Damascus, and the second to the Ḥanbalī quarter of this city. Similar *nisbaṯ*-usage cases can be observed with other toponymic names as well. Referring to the town of Suhravard and the region of Ǧīlān (both in Iran), the *nisbat*s "al-Suhrawardī" and "al-Ǧīl[ān]ī" feature in the *Taʾrīḫ al-islām* only when the Suhrawardī and the al-Ǧīl[ān]ī/Qādirī families are prominent in Baghdad (roughly late 12th–early 13th centuries CE), and—similarly to the case of the Muqaddasīs—the most frequent co-occurring *nisbaṯ* of both the Suhrawardīs and the al-Jīl[ān]īs during this period is "al-Baġdādī." Such instances of re-appropriation are not frequent and happen only with *nisbat*s that are not frequent; more importantly, the way my method is designed, the Muqaddasīs will be counted also as as Damascenes, and al-Suhrawardīs and al-Ǧīl[ān]īs as Baghdadis.

My computer-aided analysis of the 29,000 biographies yields about 700 unique *nisbat*s (with over 300 toponymic ones) that identify a group of at least 10 different individuals in the *Taʾrīḫ al-islām*; the overall number of these *nisbat*s runs into over 70,000 instances, considering that individuals are often described with more than one *nisbaṯ*. While 70,000 data points can hardly be called "big data", this dataset is too big to make exact identification of each and every *nisbaṯ* possible. Thus, under these circumstances, considering *nisbat*s at their face values is simply the most logical way to begin the large-scale analysis of biographical data. As our knowledge about the "behavior" of *nisbat*s in biographical collections improves—and this can be achieved only through large-scale analysis—these methodological assumptions can be adjusted.[15]

The Cultural Geography of the *Amṣār*

Al-Ḏahabī includes over 80 urban centers and provinces in the *Amṣār*. Starting with the sacred cities of Islam he moves through the regions of al-Šām, al-ʿIrāq, Miṣr, al-Yaman, al-Andalus, the regions of al-Maġrib and Ifrīqiyya, al-Jazīraṯ, northwestern Iran, northeastern Iran (Ḫurāsān), Mā-warāʾ-al-nahr and Ḫwārizm, southern Iran (spanning from al-Ahwāz to Sijistān, all lumped together), and, in the very end, he briefly mentions the very fringes: al-Hind, al-Sind, Ḥaḍramawt, and al-Ḥabašaṯ. As his coverage shows, he was very familiar with the geography of the Islamic world, but his chronological statements are more or less certain—i.e., he names the periods of prosperity one way or

[15] For the detailed discussion of methodological assumptions see, Romanov, "Computational Reading," 28–40.

another—only for about two dozen places, most of which feature in the first part of the epistle. Furthermore, not all of his descriptions are equally thorough and detailed, and it seems that the certainty of his statements and the level of details of his assessments in the *Amṣār* correspond to the amount of relevant data in the *Taʾrīḫ al-islām*: the more data he had in the *Taʾrīḫ al-islām*, the more certain and detailed were his statements in the *Amṣār*.[16]

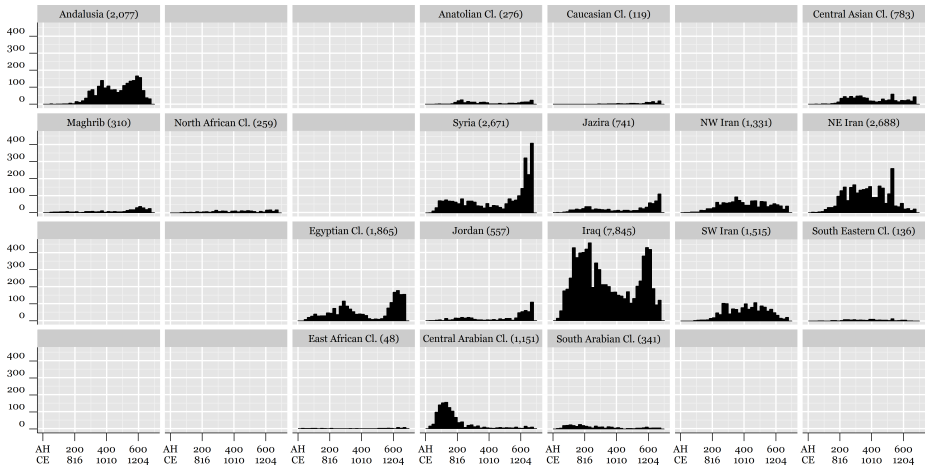

Figure 1: Graphs of chronological coverage of different regions in the *Taʾrīḫ al-islām* through toponymic *nisbat*s.

Of particular interest are his chronological statements when he marks periods during which a region contributed most significantly to Ḥadīṯ sciences. In doing so, he names centuries (e.g., *al-miʾat al-ṯānīyat*, "the second [*hijrī*] century") or gives a reference to milestone events (such as foundation, conquest, destruction); in other cases he names most prominent Ḥadīṯ scholars, from which the period of prosperity can be inferred.

In terms al-Ḏahabī's evaluations, urban centers and provinces in the *Amṣār* can be divided into three major groups: those that are merely listed, those that are characterized with some uncertainty, and those that are characterized with clear chronological statements. Figure 1 shows chrono-geographical coverage of the *Taʾrīḫ al-islām*, while Figure 2 displays how this coverage compares with al-Ḏahabī's statements in the *Amṣār*.

Listed places. More than half of places are simply listed by al-Ḏahabī without any inferable information on their role and importance in the area of Ḥadīṯ sciences.

[16]Here is an example of his certain and detailed statement: "[In] Mecca, the learning was sparse at the time of the Companions. Then, it became abundant at the end of their time, and then at the time of the Followers and their companions. Then, in the course of the third century (816–913 CE), the learning diminished in the two sacred cities (i.e., Mecca and Medina), but became abundant in other [cities of Islam]." (*Makkat … kāna al-ʿilm bi-hā yasīran fī zamani l-ṣaḥābat ṯumma kaṯura fī awāḫiri ʿaṣri l-ṣaḥābat wa-kaḏālika fī ayyāmi l-tābiʿīn wa-zamani aṣḥābi-him … ṯumma fī aṯnāʾi l-miʾati l-ṯāliṯati tanāqaṣa ʿilmu l-ḥaramayni wa-kaṯura bi-ġayri-himā*).

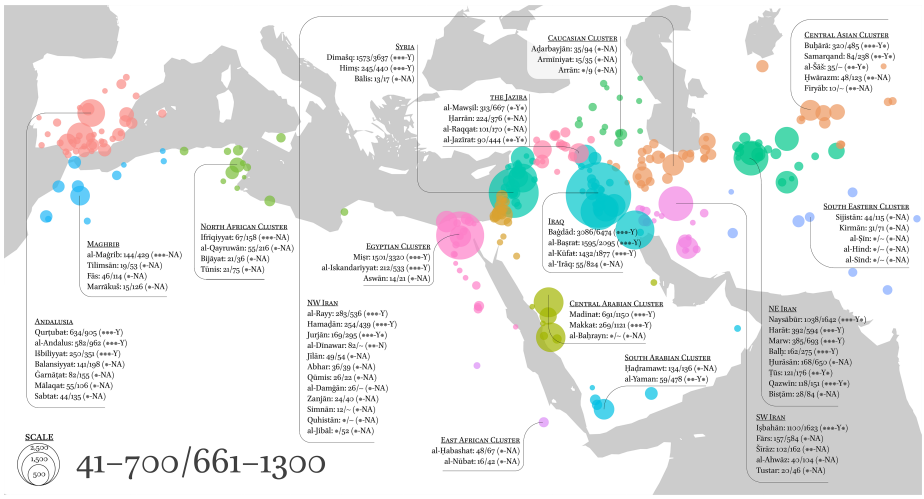

Figure 2: The cartogram of geographical distribution of individuals in the *Taʾrīḫ al-islām* with corresponding comments from the *Amṣār*. For example, "Madīnat: 691/1150 (***-Y)", which says that in the *Taʾrīḫ al-islām* there are 691 individuals that can be associated with Madīnat through their *nisbat*s, while the city itself is mentioned in 1,150 biographies; in the *Amṣār* the prominence of the city is described with a certain statement (***), which agrees (Y) with the graph based on the data from the *Taʾrīḫ al-islām*. The legend for the part of comments in parenthesis is as follows: ***—certain statement; **— uncertain statement; *—mere mention; Y—agrees with the data from the *Taʾrīḫ al-islām*; Y*—can be interpreted as agreeing with the data from the *Taʾrīḫ al-islām*; N—does not agree; NA—not applicable, since the place is not explicitly mentioned in the *Amṣār*.

Comparison with the *Taʾrīḫ al-islām* shows that these are the least represented locations both through onomastic data and toponymic frequencies (see Figure 8).

Uncertain statements. In such cases (less than two dozen), al-Ḏahabī lists one or two prominent Ḥadīṯ scholars associated with a place, but refrains from any broader statements. Comparison with data in the *Taʾrīḫ al-islām* shows these are places that are not sufficiently represented, and more often than not individuals associated with the place are spread thinly across the entire period of almost seven Islamic centuries covered in al-Ḏahabī's "History".

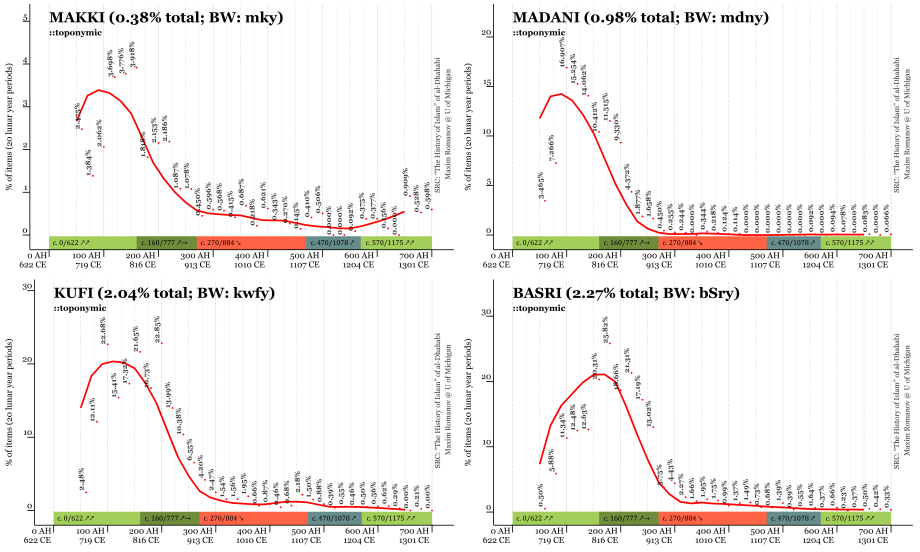

Figure 3: Chronological distribution of individuals in the *Taʾrīḫ al-islām* associated with the early Islamic centers.

Certain statements. al-Ḏahabī's most certain statements are about places for which he has most data in the *Taʾrīḫ al-islām*. Such statements are not only certain—i.e., he defines the period rather specifically—but they also closely correspond to the graphs based on the *Taʾrīḫ al-islām*. Let's take a look at the most vivid examples of centers that flourish in the beginning, the middle and the end of the covered period (early, intermediate and late centers, respectively). On Figure 3, the graphs of the most prominent early centers show curves of individuals from the *Taʾrīḫ al-islām* who bear toponymic names associated with these places. In the *Amṣār* al-Ḏahabī says that Medina (*nisbaṯ* al-Madanī) and Mecca (*nisbaṯ* al-Makkī) were prominent centers of knowledge since the time of the Companions, although Mecca started as a center under the last of the Companions and never became home to as many learned Muslims as did Medina; both cities lost their prominence as centers of knowledge in the course of the 3rd Islamic century (*c.* 815–912 CE). As to Kufa (*nisbaṯ* al-Kūfī) and Basra (*nisbaṯ* al-Baṣrī), they also began to gain prominence during the time of the Companions; al-Ḏahabī marks the end of the Kufan prominence with Ibn ʿUqbaṯ who died in 332/943 CE; Basra prospered

until the beginning of the 3rd Islamic century (*c.* 815 CE), after which it started to decline rapidly. The graph of the early Islamic centers (Figure 3) shows that al-Ḏahabī's statements correspond perfectly to the curves: his statements of floruit—different forms and variations of *kaṯura al-ʿilm bi-hā*, "the learning was abundant there"—agree with the peaks of curves, while his statements of decline —forms and variations of *tanāqaṣa al-ʿilm bi-hā*, "the learning declined there"—to the low points of the curves, with all four centers practically disappearing from the cultural map of the Islamic world by the beginning of the 4th Islamic century (*c.* 912 CE).

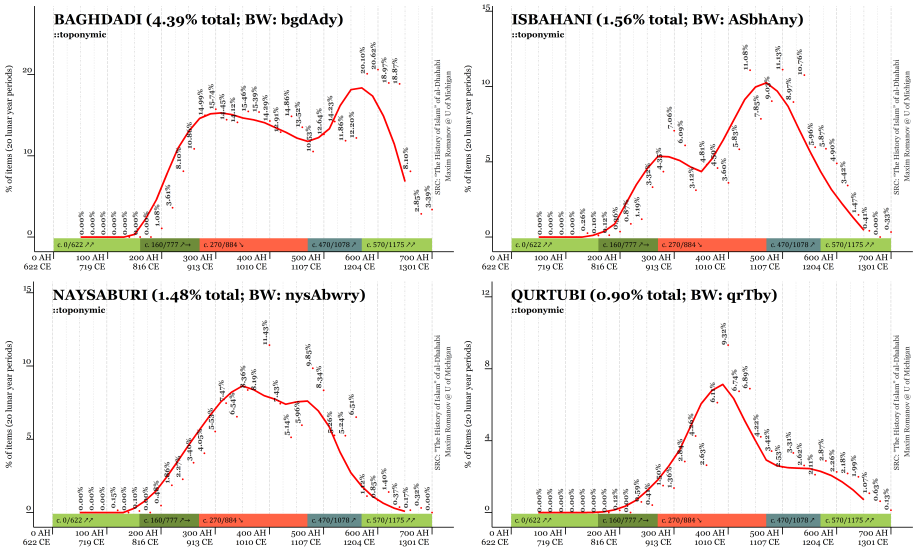

Figure 4: *Chronological distribution of individuals in the* Taʾrīḫ al-islām *associated with the intermediate Islamic centers.*

The graphs of intermediate centers (Figure 4) feature Baghdad (*nisbaṯ* al-Baġdādī), Isfahan (*nisbaṯ* al-Iṣbahānī), Nishapur (*nisbaṯ* al-Naysābūrī), and Cordova (*nisbaṯ* al-Qurṭubī). In the *Amṣār* al-Ḏahabī writes that Baghdad remained the key center from its foundation by the caliph al-Manṣūr (r. 136–158/754–775 CE) until it was sacked by the Mongols in 656/1258 CE.[17] Andalusia prospered from the 3rd Islamic century (c. 815–912 CE) until Cordova and Seville fell into the Christian hands (633/1235 CE and 646/1248 CE, respectively).[18] Nishapur[19] starts its history as a center with Ibrāhīm b. Ṭahmān who died in 163/779 CE and ends with the coming of the Mongols in 617/1220 CE, after which it disappeared,

[17] Here, however, I should add that the curve of Baghdad actually starts plummeting two decades before the Mongol invasion.

[18] al-Ḏahabī talks about Andalusia in general, without detailed statements on its cities. See, al-Ḏahabī, *al-Amṣār*, 1986, 184–188. This "generality" may come from his perspective as an easterner; a similar eastern perspective be seen in al-Muqaddasī, *The Best Divisions for Knowledge of the Regions*.

[19] Al-Ḏahabī, *al-Amṣār*, 1986, 205–208.

"as if it never existed." al-Ḏahabī's statement regarding Isfahan is rather vague though: he simply writes that it used to be a center that vied with Baghdad in prominence.[20] Here again, both graphs and statements closely correspond.[21]

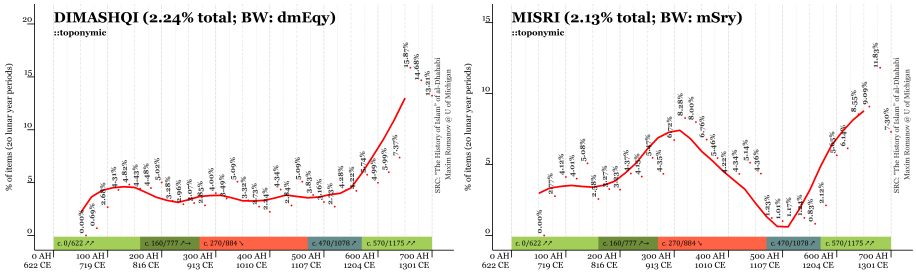

Figure 5: Chronological distribution of individuals in the *Taʾrīḫ al-islām* associated with the late Islamic centers.

Unlike al-Ḏahabī's statements about intermediate centers, where he is often using references to conquests and invasions as turning points, his statements about the late centers are much more interesting. The graphs of the late centers (Figure 5) feature Damascus (*nisbat* al-Dimašqī) and Egypt (*nisbat* al-Miṣrī). In the *Amṣār* one finds that the history of Damascus as a center of learning begins during the time of the Companions; it flourishes during the time of the Umayyad caliphs Muʿāwiyat (r. 41–60/661–680 CE) and ʿAbd al-Malik (r. 65–86/685–705 CE), declines in the course of the 4th and 5th Islamic centuries (c. 912–1106 CE), and comes back to prominence after that, especially during the time of the Zangid *amīr* Nūr al-Dīn (r. 541–569/1146–1174 CE), Ibn ʿAsākir (d. 571/1175 CE), the Muqaddasī family, Ibn Taymīyat (d. 728/1327 CE), al-Mizzī (d. 742/1341 CE), and their followers.[22] Egypt began to gain prominence during the time of the Followers and continued on that course until the coming of the Fāṭimids in 358/968 CE, whose Ismāʿīlī/Shīʿīte rule marked drastic decline for Sunnī Ḥadīṯ learning[23] in the province, until Ṣalāḥ al-Dīn put an end to their rule in 567/1171 CE,[24] after which Egypt starts regaining its position as a center of learning.

These statements of al-Ḏahabī are particularly interesting since he also describes temporal fluctuations. Although the relative graph does not allow us to discern the decline of Damascus during the 4th and 5th Islamic centuries (c. 913–

[20] Al-Ḏahabī, *al-Amṣār*, 1986, 232–233.

[21] It should be added, however, that most statements regarding the intermediate centers are punctuated by milestone dates, often for both the beginning and the end of periods, such as the foundation or the Muslim conquest of a city—to mark its beginning, and the [re]conquest, destruction, or invasion of a city—to mark the end of its period.

[22] Al-Ḏahabī, *al-Amṣār*, 1986, 160–166.

[23] As data from the *Taʾrīḫ al-islām* shows, the Mālikī legal school suffered in a similar way.

[24] al-Ḏahabī, *al-Amṣār*, 1986, 167–170. NB: Alexandria became prominent during the residence of al-Silafī, who moved there from Isfahan in 511/1117 CE and resided there until his death in 576/1180 CE; the prominence of Alexandria started to decline soon after that, which agrees with the onomastic graph of this city, see, al-Ḏahabī, *al-Amṣār*, 1986, 170–171.

1107 CE),[25] one can clearly see how the curve of the city soars up in the 6th Islamic century (after 1107 CE). The decline of the Egyptian curve during the reign of the Ismā'īlī dynasty, on the other hand, is as clear as its rapid recovery after their reign.

The Status Quo of the Islamic Sciences

The comparison of al-Ḏahabī's two texts makes it highly plausible that al-Ḏahabī's statements in the *Amṣār* regarding major regions of the Islamic world are informed by the quantifiable data from his *Taʾrīḫ al-islām*. One, of course, may object, arguing that al-Ḏahabī's statements are informed by the general flow of Islamic history—after all he does often use important historical events, such as conquests, as chronological markers of change (he does this, however, only for intermediate centers whose "life cycles" are marked by such events). Yet, in the concluding part of the *Amṣār* one also finds an interesting discussion of the fate of Ḥadīṯ learning versus other religious sciences. Here al-Ḏahabī laments that Ḥadīṯ learning declined—often to the point of non-existence—in most previously prominent regions, surviving now only in Egypt, Greater Syria, and the immediately adjacent regions. Despite the decline of Ḥadīṯ learning, he continues, Qur'ānic sciences and Islamic law prosper both in the west and in the east of the Islamic world, even though they are "contaminated … with pre-Islamic sciences, and the opinions of speculative theologians and the Mu'tazilites."[26] Such lamentations about the good olden days are so common among Muslim scholars that one may be tempted to dismiss them as a literary trope.[27] However, my analysis strongly suggests that al-Ḏahabī's statements are more than just the grumblings of an old man who idealizes the past, and that they also closely correspond to the data that he collected in the *Taʾrīḫ al-islām*. First, the network of geographical connections of individuals from the latest volumes of the *Taʾrīḫ al-islām* shows (Figure 6) that the Islamic world [of scholarship?] indeed shrunk to the crescent of Egypt, Greater Syria and northern Iraq (the Jazīraṯ), with other regions neither significantly represented, nor strongly integrated into what became the core by the end of the 7th/13th century. Although it can be argued that the "shrinking" of the Islamic world reflects nothing but al-Ḏahabī's inability to get access to the later historical and biographical writings of his peers from remote regions—a possibility that al-Ḏahabī himself considered[28]—this very inability may be a witness to the fact that the cultural integration of the

[25] This decline is discernible on the graph with absolute numbers. According to the data from the *Taʾrīḫ al-islām*, the 4th and 5th centuries were the period of decline for the provinces of Egypt, Syria and Iraq; in fact, the entire cumulative biographical curve is affected by this decline during c. 270–470/884–1078 CE (the period is marked with the red block at the bottom of the graph); the decline is clearly visible on the curve of Baghdad (on the graph of intermediate centers).

[26] *mukaddar … bi-ʿulūm al-awāʾil wa-ārāʾ al-mutakallimīn wa-l-muʿtazilat*, al-Ḏahabī, *al-Amṣār*, 1986, 235.

[27] Librande quotes a couple of similar laments of Ḥadīṯ scholars of earlier times. See, Librande, "al-Dhahabī's Essay," 128.

[28] Al-Ḏahabī, *Taʾrīḫ al-islām*, 1:11–16.

Islamic world has been shattered. Yet, no matter how we interpret this, his statement still correlates with his data.

Second, we can take a look at the graph that aggregates all individuals who can be describes as specialists in the "Islamic trivium"—the Qur'ānic, Ḥadīṯ, and legal sciences. And, as the graph shows, time indeed had changed, and the jurists—whose curve clearly goes up, skyrocketing in the 6th/12th century CE—are now the dominant group of religious scholars. The curve of the Qur'ān specialists (most prominently, Qur'ān reciters, sing. *muqri'*) may be interpreted as slowly moving upward (Figure 7, *left*, with absolute numbers), and their peak is likely to be after the period covered by al-Ḏahabī). As to Ḥadīṯ specialists, their prime time—the 3rd/9th century CE—had long passed. The curve of Ḥadīṯ specialists aggregates all major *ḥadīṯ*-related "descriptive names" that al-Ḏahabī used in the *Taʾrīḫ al-islām*.[29] The most frequent *nisbat*s are *ḥāfiẓ*, *ṯiqat*, *raḥḥāl[at]*, and *muḥaddiṯ*.[30] The pattern of distribution of the first three *nisbat*s points to the period of 200–300/815–912 CE as the peak of florescence (and even more so to 250–300/864–912 CE), which was an important period in the development of Ḥadīṯ when the six canonical collections (*al-kutub al-sittat*) were compiled, and during which Ḥadīṯ scholars travelled most actively (as the name *raḥḥāl[at]* and its counterpart *jawwāl* imply).[31] At the same time, while the number of Ḥadīṯ specialists declined, Ḥadīṯ continues playing a central, and perhaps even more important role, in the life of Islamic society, as described by al-Ḏahabī. As was stated above, over 90% of all biographees in the *Taʾrīḫ al-islām* were involved in the transmission of Ḥadīṯ, and during the period of 500–700/1106–1300 CE we find more and more individuals whose often very brief biographies tell us nothing but that they transmitted some *ḥadīṯ*s from so-and-so. Additionally, the variety of social backgrounds of those involved in the transmission of Ḥadīṯ expands to the point that we now even find military commanders (sing. *amīr*) among the transmitters of the words of the Prophet.

In search of al-Ḏahabī's method

The correlation between al-Ḏahabī's certain statements in the *Amṣār* and visual representations of the data from his *Taʾrīḫ al-islām* is rather intriguing (particularly about the state of the Islamic trivium), but even more so is the correlation between the level of certainty of his statements and the amount of data he had collected in the *Taʾrīḫ al-islām*: to reiterate, the more data on a certain location there is in the *Taʾrīḫ al-islām*, the more certain al-Ḏahabī's statement are about that location in the *Amṣār*; and vise versa—there is practically no data in the *Taʾrīḫ al-islām* about

[29] Each individual is counted only once even if he is described with more than one *ḥadīṯ*-related *nisbat*.

[30] Although the growing numbers of the *muḥaddiṯūn* slightly push the declining curve of Ḥadīṯ specialists upward, this does not affects the overall situation.

[31] For more details, see "Phase 3: The age of 'six books'" (*c.* 200–400/912–1009) in: Lucas, *Constructive Critics*, 73–86.

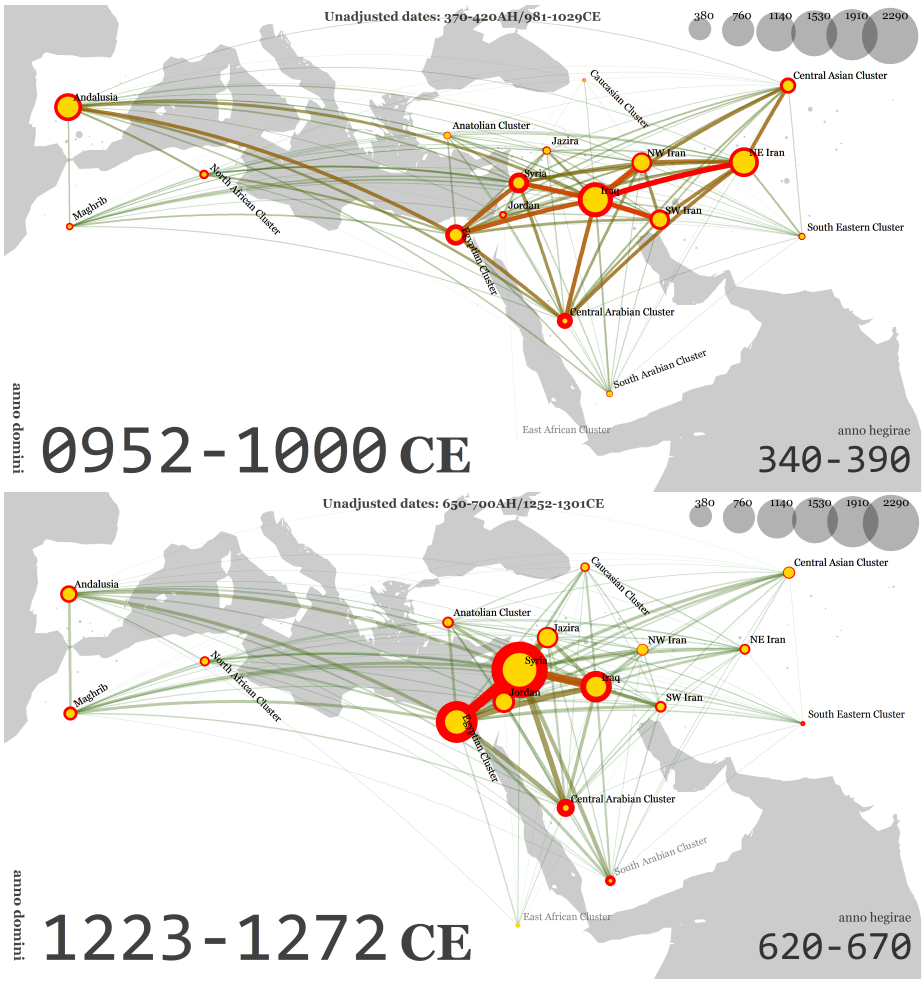

Figure 6: The cartograms show how, according to al-Ḏahabī's *Taʾrīḫ al-islām*, the Islamic world was connected during two different periods: the cartogram at the top shows a more even representation of major regions and their more comprehensive interconnectedness, while the cartogram at the bottom demonstrates that the Islamic world "shrunk" to the fertile crescent region, with other regions neither strongly represented nor integrated. *NB:* Redder and thicker lines mean more connections; greener and thinner lines mean less connections.

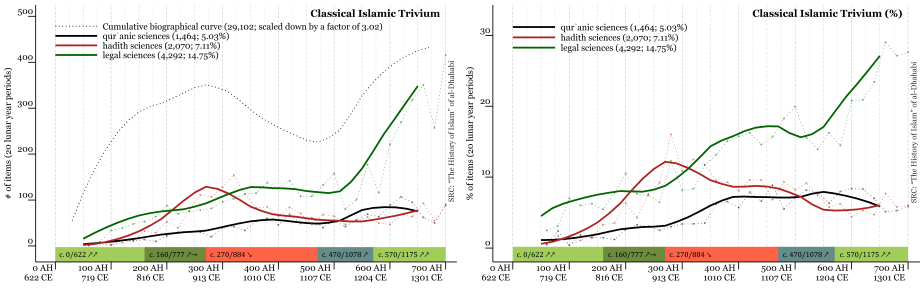

Figure 7: Fluctuations of individuals involved in Qurʾānic, Ḥadīṯ and legal sciences—in absolute (*left*) and relative (*right*) numbers.

places that are simply mentioned in the *Amṣār* (see, Figure 8).

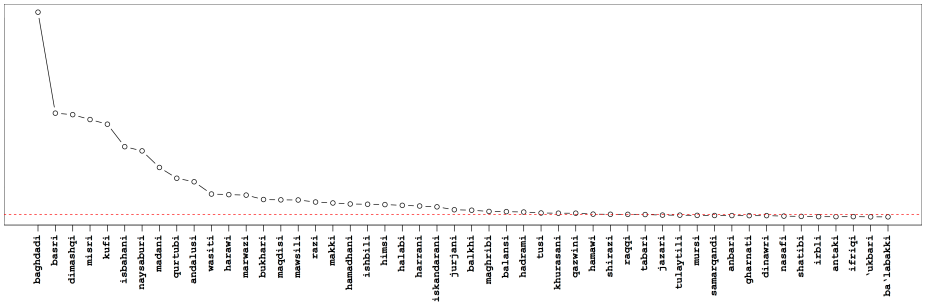

Figure 8: Al-Ḏahabī's certain statements in the *Amṣār* can be found only about places that feature most frequently in his *Taʾrīḫ al-islām*: on the graph these are places that appear only above the red-dashed line, which demarcates the threshold of 100 individuals per location. (The graph shows the top 50 most frequent toponymic *nisbat*s)

These factors lead to questions about al-Ḏahabī's method. How exactly did he collect and organize over 30,000 biographies and about 10,000 descriptions of events into what became his *Taʾrīḫ al-islām*, and, later, reorganized it into a number of his other books? Could his observations have resulted from the use of some quantitative and, perhaps, simple visualization techniques?

The enterprise of collecting and organizing knowledge across all fields of learning is one of the most salient features of scholarship in the premodern Islamic world. In this regard, al-Ḏahabī was one of hundreds of scholars who were engaged in similar activities both before and after him,[32] particularly in the fields of lexi-

[32]Moreover, in organizing his *Taʾrīḫ al-islām*, he must have followed in the footsteps of Ibn al-Ǧawzī (597/1201 CE), who was first combine a chronicle with a biographical collection in his *al-Muntaẓam fī-l-taʾrīḫ*. See, Somogyi, "Ibn al-Jauzī's School of Historiography.".

cography,[33] Ḥadīṯ,[34] genealogy, biography/prosopography, history (or, perhaps better, "chronography"),[35] bibliography, and geography.[36]

In these and other fields of learning, scholars were repeatedly producing continuations and abridgments of the writings of their predecessors. They were updating, expanding, combining, and rearranging them. They were alphabetizing them and creating indices for them. Al-Saḫāwī's *al-Iʿlān bi-l-tawbīḫ li-man ḏamma ahl al-taʾrīḫ* is teeming with references to such activities.[37] More importantly al-Saḫāwī offers an insight into the mechanics of how exactly such activities could have occurred: we find that the *Muʃam al-safar* of al-Silafī (d. 576/1180 CE) was initially written on separate sheets of paper, with each biography written on its own sheet (*fī ǧazāzāt kull tarǧamat fī ǧazāzat*).[38]. Some autographs of the *Taʾrīḫ al-islām* include such loose sheets with writing in al-Ḏahabī's hand.[39] We find a similar example a few centuries later in the draft (*musawwadat*) of Ḥāǧǧī Ḫalīfaṭ's (d. 1067/1656 CE) biographical collection of poets entitled *Sullam al-wuṣūl ilá ṭabaqāt al-fuḥūl*[40], whose "pages [often] consist of small slips of paper arranged in alphabetical order of authors, all neatly stuck together and mounted to form folio-size pages", representing "his flexible, expandable information retrieval system, a forerunner of the 3×5 inch library-card catalogue, centuries before such cards were invented."[41]

These examples suggest that collecting and keeping biographical information must have been a common approach (as well as probably any other kind of "serialized data", to borrow a computer-science term). If the initial version of the *Taʾrīḫ al-islām* was indeed stored in such a format, we may think of it as a premodern analog database of historical and biographical information, which he "stitched" together from earlier sources[42] to serve as his main research tool for writing his other books. Although usually considered "abridgments"—in a sense, books of secondary importance—these shorter books (or, "thematic queries", if we are to

[33] On the Arabic lexicographical tradition, the interdependence of its specimens and various themes and principles of organization, see Rybalkin, *Klassicheskoe arabskoe iazykoznanie*, 259–337, in particular; and, most recently, Baʿlabakkī, *The Arabic lexicographical tradition*.

[34] Ḥadīṯ collections, their interdependence and various organizational principles are nicely overviewed in Brown, *Hadith: Muhammad's Legacy in the Medieval and Modern World*, 15–66.

[35] The interdependence of historical and biographical works is discussed in Rosenthal, *A History of Muslim Historiography*, passim. (al-Saḫāwī's *al-Iʿlān bi-l-tawbīḫ*, translated in Rosenthal's book, is particularly rich on notes about who updated and reorganized whose work).

[36] For a similar discussion of the "classical school of Arabic geography," see: Krachkovskii, *Arabskaia geograficheskaia literatura*, 194–218.

[37] On alphabetization, for example, see Rosenthal, *A History of Muslim Historiography*, 233, 234, 346, 347, 355, 360, 363, 373, 381, etc.

[38] Rosenthal, *A History of Muslim Historiography*, 366; for Arabic: al-Saḫāwī, *al-Iʿlān*, 211.

[39] Maʿrūf considers them to be his methodological tool, see: Maʿrūf, *al-Ḏahabī wa-manhaju-hu*, 395.

[40] See, Birnbaum, "The Questing Mind"; Birnbaum, "Kātib Chelebi (1609-1657) and Alphabetization."

[41] Birnbaum, "The Questing Mind," 148.

[42] For example, using computational methods for identifying text reuse, we were able to establish that there are at least 800 pages worth of text (over 245,000 words, 7.5% of the entire volume of the *Taʾrīḫ al-islām*) that can be traced back to the *Taʾrīḫ madīnat Dimašq* of Ibn ʿAsākir (571/1175 CE), with 50% of quotations in the range of 22-48 words. "Text reuse", or who quotes whom and to what extent, will be discussed in details in the next article.

continue the database metaphor; see Figure 9) could have been what he wanted to write in the first place.

Variable/Source	*Taʾrīḫ al-islām*	*Taʾrīḫ Baġdād*	*Ṭabaqāt al-ḥanābilaṯ*
period from	-52 AH	145 AH	164 AH
period until	700 AH	473 AH	527 AH
personalia	all	all	Ḥanbalīs
geography	all	Baghdad	all
arrangement	chronological	alphabetical	generational

Figure 9: *Biographical collections as queries.* If we imagine a pan-Islamic biographical database, each and every individual biographical collection may be viewed as a specific query into that database. For example, al-Ḏahabī's *Taʾrīḫ al-islām* itself can be viewed as a very broad query that selects all biographical records from all available regions of the Muslim world for the period from the Prophet's lifetime until 700/1301 CE, and arranges them chronologically by decades; al-Ḫaṭīb al-Baġdādī's *Taʾrīḫ Baġdād*—as a query that selects only biographies of those affiliated with the city of Baghdad (the type of affiliation does not matter), considers the period from the foundation of Baghdad (or, actually, including the entire lifetime of caliph al-Manṣūr, the founder of Baghdad) until the author's death, and arranges them alphabetically by first name (sing. *ism*); Ibn Abī Yaʿlá's *Ṭabaqāt al-ḥanābilaṯ*, as a query that limits biographical records to people affiliated with the Ḥanbalī community, considers the period from Ibn Ḥanbal's lifetime until the author's death, applies no geographical limitations, and arranges records by "generational cohorts".

The very organization of the *Taʾrīḫ al-islām* suggests that al-Ḏahabī's "mechanical" system also grouped biographies into decades,[43] and could have had other enhancements that made his workflow more efficient. Yet, even without any additional bells and whistles, an organizational system that uses "movable media" becomes an efficient tool: it allows one to insert new records where they belong, retrieve existing ones so that they can be updated, but most importantly, it allows one to subset ("query") records and to rearrange them according to the purposes of specific projects. This last feature—to subset and to rearrange—also turns this system into a tool for visual time-series analysis. If al-Ḏahabī's certain statements are indeed data-driven, he could have obtained his insights by collecting sheets of relevant biographies from his databank and then arranging them chronologically (or, in fact, just maintaining the chronological order of his databank). The very "mechanical" arrangement of these extracted sheets would be an equivalent of a *histogram*—the most common method for studying data distribution in modern statistics—which would offer a visual point of entry into the historical ups and

[43] Maʿrūf's comments also suggest that al-Ḏahabī might have kept historical events separate from biographical material, which makes a lot of sense methodologically, allowing him for more efficient information retrieval. See, Maʿrūf, *al-Ḏahabī wa-manhaju-hu*.

downs of a group in question. Figure 10 offers a visual representation of this point.

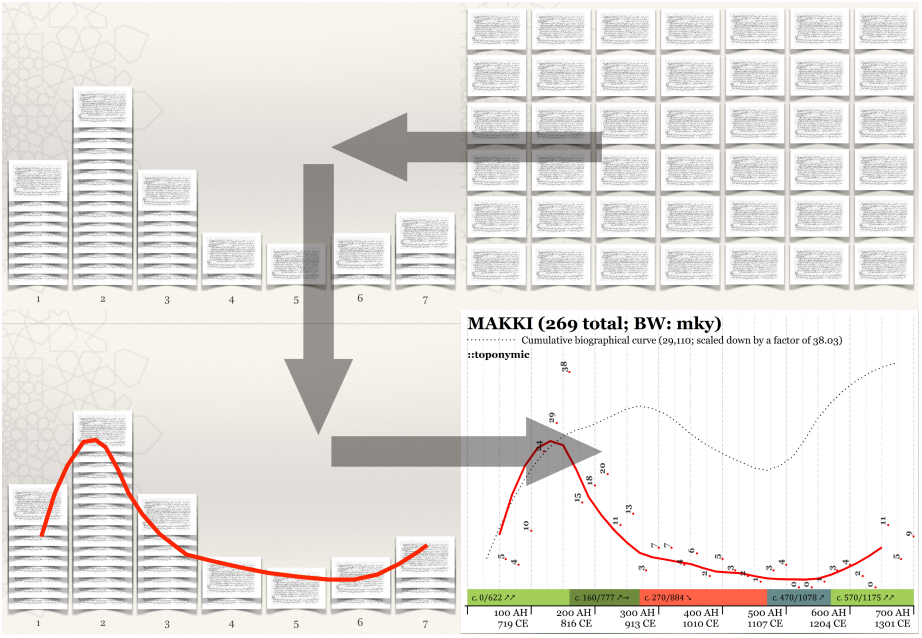

Figure 10: A possible analytical tool: (*top-right*) relevant biographies are collected from the "databank" and (*top-left*) arranged into periods (here, centuries), which (*botom-left*) offers an analytical summary similar to a modern graph (*bottom-right*).

Two of al-Ḏahabī's own "abridgments" of the *Taʾrīḫ al-islām* can be seen as such thematic subsets/queries: namely, his books on prominent scholars of Ḥadīṯ (*Ṭabaqāt al-ḥuffāẓ*) and prominent reciters of the Qur'ān (*Maʿrifat+ al-qurrāʾ al-kibār*). If we look at the chronological coverage of these two books (distribution of date statements in these titles on Figure 11),[44] we see that the *Ṭabaqāt al-ḥuffāẓ* points to the same period of florescence—250–300 AH / 864–912 CE—as on the graph of Ḥadīṯ specialists based on the *Taʾrīḫ al-islām*, also showing a similar declining trend of Ḥadīṯ sciences by the end of the period. The *Maʿrifat+ al-qurrāʾ al-kibār* on the other hand clearly shows the rise of the Qur'ān reciters by the end of the period.

Al-Ḏahabī's two abridgments, *Ṭaḏkirat al-ḥuffāẓ* and *Maʿrifat+ al-qurrāʾ al-kibār*, and a possible method of working with biographies ("the mechanical histogram"), may explain the certainty of al-Ḏahabī's statements regarding the status quo of Ḥadīṯ and Qur'ān sciences at his time. From what we know, he has not written a

[44] Arguably, we can treat date statements (here, references to years, grouped into 50-year periods) as indicators of the chronological focus of a chronicle or a biographical collection: the more dates there are for a certain period, the stronger the focus of a book on that period.

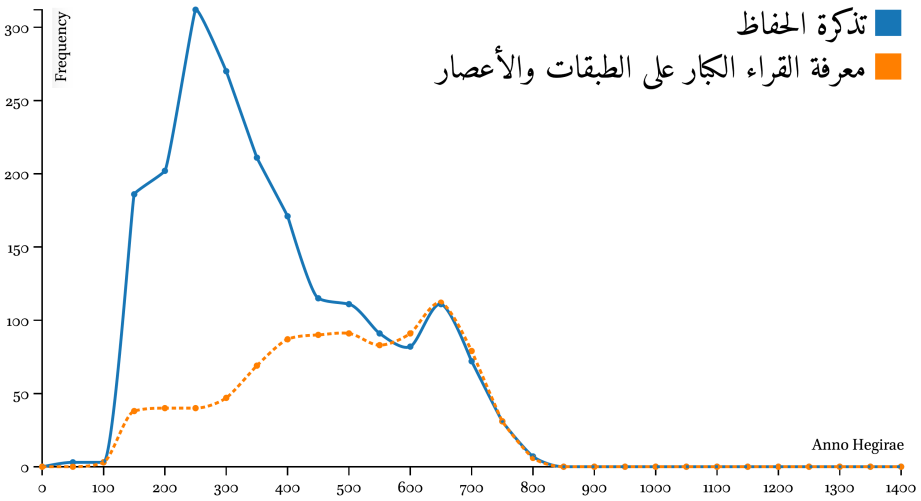

Figure 11: Date statements from the *Ṭaḏkirat al-ḥuffāẓ* (*blue*) and the *Maʿrifat al-qurrāʾ al-kibār* (*orange*) display the same chronological patterns for Ḥadīṯ scholars and Qurʾān reciters as the graphs based on the *Taʾrīḫ al-islām;* see Figure above on the status quo of the Islamic sciences for comparison. (*NB*: the decline of both curves during 650–700 AH / 1252–1300 CE is likely to indicate the lag in the availability of information for the latest period; all chronological datasets—premodern as well as modern—show a similar lagging pattern).

comparable summary on jurists, but he did thoroughly work with all major *ṭabaqāt* collections of legal schools (since they are listed as his sources in the introduction to the *Taʾrīḫ al-islām*) and could have created a similar query.[45]

Explaining his statements about geographical regions, however, is more difficult. He has not written any geographically-focused collections and creating "mechanical histograms" even for the top dozen locations would have been a very time-consuming process, not to mention that the last thing one would want to do is to break the arrangement of 40,000 units of information. A nondestructive alternative could have been counting and graphing. This possibility is not completely far-fetched, since premodern Islamic scholars were not alien to mathematical[46]

[45] At the same time, the number of jurist at the late period was so significantly higher than those of Ḥadīṯ and Qurʾān experts that it could have been unnecessary to research this issue.

[46] A prominent Arab philologist who, however, was not particularly known to be a mathematician, al-Ḫalīl al-Farāhidī (d. *c.* 170/786 CE) designed his dictionary of the Arabic language, *Kitāb al-ʿayn*, relying on what is now referred to as combinatorics: the approach allowed him establish all possible Arabic words mathematically, considering all combinations of letters with and without vowels. Here is a quote to illustrate the method of perhaps the earliest computational linguist: "If you want to exhaustively know all of the Arabic language double letter words, either meaningful or not, which the Arabs either used or rejected, such as *qd*, *km*, *an* … etc., take the [Arabic] alphabet letters which are 28, then multiply them with each other to get 784 [= 28^2]. A single letter is not a word. If you take two letters [without reversal], you get 392 [= 784/2] such as *dm* and the like. If you reverse [the two letter positions] it comes back to 784, 28 of which have identical letters | like *hh* which do not change when

and visual methods (see, Figures 12 & 13) when working with information in what can be characterized as a humanistic inquiry.[47] Using the style that was algorithmic and demonstrative, mathematicians offered solutions to algebraic equations through geometric constructions—perhaps, the most vivid visual approach in mathematics.[48] One of the most common methods of performing calculations in general was with the use of a dust board ("board and dust calculation", *al-ḥisāb bi-l-taḫt wa-l-turāb*; "calculation [by means] of dust", *ḥisāb al-ġubār*).[49] This method allowed one to split complex calculations into smaller steps and keep track of them visually on a dust board, rubbing out and displacing numbers with the final result replacing one of the given numbers. In the works of al-Kindī (d. 260/873 CE) we find an approach to textual data which is both statistical and, potentially, visual. To "the Philosopher of the Arabs" is attributed a method for decrypting a cipher in which each letter in the alphabet is substituted with a randomly selected character from the same or different alphabet (*polyalphabetic substitution cipher*).[50] Unlike Caesar's code in which each letter in the plain text is "shifted" a certain number of places down the alphabet, this type of cipher had been considered unbreakable because of too many possible combinations. Acting on the premise that each language has its most and least frequent letters, al-Kindī describes how one can use letter frequencies to break this code—his method is now considered one of the basic approaches for solving such problems. What is particularly interesting is that al-Kindī stresses that both ciphered and normal texts should be long enough, otherwise distribution of letter frequencies will be incorrect, which clearly demonstrates statistical awareness. Here is the gist of it:

---

reversed. 600 of these [784 − 28 = 28 × 27 = 756 words] are perfect words [i.e., consonants only] with no *Wāw*, *Yā* or *Hamzah* [these are the three basic vowels in Arabic], which come to 300 before reversal [(28 − 3)(27 − 3) / 2 = 300]. 150 words [of the 756] contain one of these [vowels]: *Wāw*, *Yā* or *Hamzah*, with 75 before reversal [25 × 3]. 6 words [of the 756] contain two [different] vowels [3 × 2], with three before reversal. 3 double-letter words [of the 784] contain the same vowel, 25 [double-letter words], contain identical consonants. You should understand what I just explained to you of the double-letter word counts which the Arabs spoke or rejected." Translation is from al-Kadi, "Origins of Cryptology," 122–23; see also al-Kadi, "Origins of Cryptology," 104, 121–24; for a more detailed description of his permutation system, see Baʿlabakkī, *The Arabic lexicographical tradition*, 292–96. See also, a section on *combinatorial analysis* in R.Rashed's "al-Riyāḍiyyāt", *EI2-Online*. Some biographical reports highlight his interest in practical arithmetics, but the episode seems to be comical more than anything. See, Rybalkin, *Klassicheskoe arabskoe iazykoznanie*, 148–49; Talmon, *Arabic grammar in its formative age: Kitāb al-ʿayn and its attribution to Ḫalīl b. Aḥmad*, 49.

[47] Particularly within the framework of the field of the digital humanities, or *humanities computing*, which will be a more appropriate term in the context of the premodern Islamic world. The digital humanities is a very broad umbrella term that includes any kind of humanistic engagement with the digital, while *humanities computing* is an area of computationally-driven text analysis. Humanities computing are often seen as the precursor of the digital humanities (, See Schreibman, Siemens, and Unsworth, *A companion to digital humanities*, 3–19.).

[48] On these mathematicians and the wider use of geometrical methods, see, R.Rashed's "al-Riyāḍiyyāt", *EI2-Online*.

[49] Gandz, "Did the Arabs Know the Abacus?"; Gandz, "The Origin of the Ghubār Numerals, or the Arabian Abacus and the Articuli.". See also, A.I. Sabra's "'Ilm al-Ḥisāb" and M. Souissi's "Ḥisāb al-ghubār" in *EI2-Online*.

[50] See, al-Kadi, "Origins of Cryptology"; Mrāyātī, *ʿIlm al-taʿmiyyat wa-istiḫrāǧ al-muʿammá ʿinda al-ʿarab*. Whether this method was actually devised by al-Kindī is not relevant for our argument.

[51] Mrāyātī, *ʿIlm al-taʿmiyyat wa-istiḫrāǧ al-muʿammá ʿinda al-ʿarab*, 207, see also al-Kadi, "Origins of Cryp-

Figure 12: It seems that Islamic scholars of different backgrounds fully appreciated the value of the visual in the representation of complex ideas, since we find a great number of examples of visual representations of things that are difficult to describe efficiently with words. Tree diagrams (Ar. *taš̆ǧīr*) were used especially frequently to convey complex relationships among multiple objects—the diagram above shows al-Kindī's (d. 260/873 CE) classification of ciphers.[51] One even finds an entire book consisting exclusively of such conceptual diagrams—the *Ǧawāmiʿ al-ʿulūm* ("Connections of the sciences") of Šaʿyā b. Farīġūn's (the 4th/10th century CE). On Šaʿyā b. Farīġūn and for additional bibliography, see: Bosworth, C. E.. "Ibn Farīg̲hūn," *EI2-Online*. Brill Online, 2016; a digitized microfilm of the manuscript (El Escorial 950, 84 folios) of this work can be accessed through *al-Ǧāmiʿ al-maḫṭūṭāt al-islāmiyyat* (`http://wqf.me/?p=16138`, Record *no0950*. On a more popular level, see The Guardian's "How 1,000 years of Arabic scholarship advanced scientific debate—in pictures" (`http://gu.com/p/42y46/sbl`).

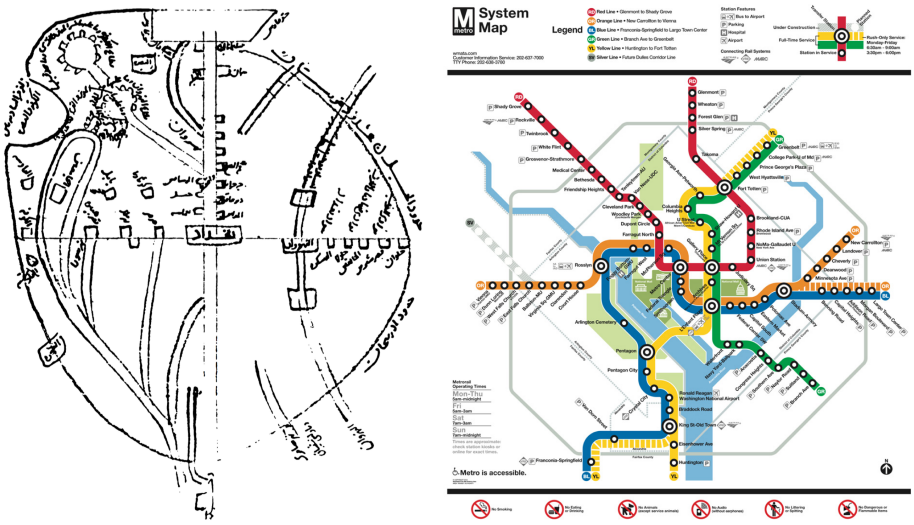

Figure 13: The most striking examples of the visual representation of complex data are found in geographical maps and cosmographical diagrams.[52] While early Islamic maps may appear too primitive to our modern eye, since they never represent space correctly, as *cartograms*, i.e. analytical representations of space, they are an incredibly powerful tool for the purpose of showing relative positions of settlements and connections among them. The cartogram of al-ʿIrāq (*left*, from al-Muqaddasī's *Aḥsan al-taqāsīm fī maʿrifat+ al-aqālīm*[53]) provides its user with all possible navigation options (also most likely serving as a visual mnemonic device), relying on the same data-visualization principle as one finds in modern schemes of public transportation around the world. Deforming geographical space in a very similar manner, the cartogram of the DC Metro (*right*) nonetheless effectively shows how to get from one place to another. (For comparison, the georectified version of the the same scheme can be found at `http://benschmidt.org/dcmetro/`).

> One way to solve an encrypted message, if we know its [original] language, is to find a [different clear] text of the same language *long enough to fill one sheet or so* (italics mine) and then we count [the occurrences of] each letter of it. We call the most frequently occurring letter the "first", the next most occurring the "second", the following most occurring the "third" and so on, until we finish all different letters in the cleartext [sample]. Then we look at the cryptogram we want to solve and we also classify its symbols. We find the most occurring symbol and change it to the form of the "first" letter [of the cleartext sample], the next most common symbol is changed to the form of the "second" letter, and the following most common | symbol is changed to the form of the "third" letter and so on, until we account for all symbols of the cryptogram we want to solve.[54]

We do not have descriptions (or examples) of how one practically does the calculation of letter frequencies, but the simplest way would be to produce something similar to a stem-and-leaf plot—the basic but powerful visualization technique from the pencil-and-paper days of exploratory data analysis.[55] Constructing such a plot (Figure 14, *left*), one begins with writing all letters in some order vertically on either side of a sheet of paper, then goes through the text and adds some counting symbol (say, ×) for every instance of every letter into a relevant raw. As a result, one ends up with a visual representation of the distribution of letter frequencies, which clearly shows the most frequent and the least frequent letters in the text (*alif* and *lām*, and *ġayn* and *ẓā'*, respectively); when a very large number of items is to be counted, tallying marks (Figure 14, *bottom*) could have been used.[56] The plot then can be then resorted—Figure 14, *right*—for a more efficient representation of letter frequencies. Al-Kindī's own calculations of letter frequencies in Arabic are impressively close to modern calculations which are based on much larger samples of text (al-Kindī used a sample of 3,667 letters).[57]

Whether al-Ḏahabī used any of the suggested methods is ultimately hard to say, at least at the moment. Yet, that he did use some kind of quantification is further suggested by the results of the comparison of the *Taʾrīḫ al-islām* with other biographical texts. In this regard, al-Ḏahabī's sampling of Andalusian sources is particularly interesting (Figure 15), showing that al-Ḏahabī's included roughly 40 to 50% of biographies from each decade covered in his Andalusian sources, thus offering a quantitatively representative sample.[58]

---

tology.", 108.

[52] *Cartography in the Traditional Islamic and South Asian Societies*, 1–89.

[53] Al-Muqaddasī, *The Best Divisions for Knowledge of the Regions*.

[54] Translation is from al-Kadi, "Origins of Cryptology," 107–9; for Arabic, see Mrāyātī, *ʿIlm al-taʿmiyyat wa-istiḫrāǧ al-muʿammá ʿinda al-ʿarab*, 2:216.

[55] For the classical description of the method, see Tukey, *Exploratory Data Analysis*, 1–25.

[56] For an explanation of tallying techniques, see Tukey, *Exploratory Data Analysis*, 16–18.

[57] Mrāyātī, *ʿIlm al-taʿmiyyat wa-istiḫrāǧ al-muʿammá ʿinda al-ʿarab*, 1:77; cf. al-Kadi, "Origins of Cryptology," 112.

[58] Data for the graph is from Avila, *La sociedad hispanomusulmana*. For additional details, see Romanov, "Computational Reading," 276–77. The quality of his selection is the subject for the study to follow.

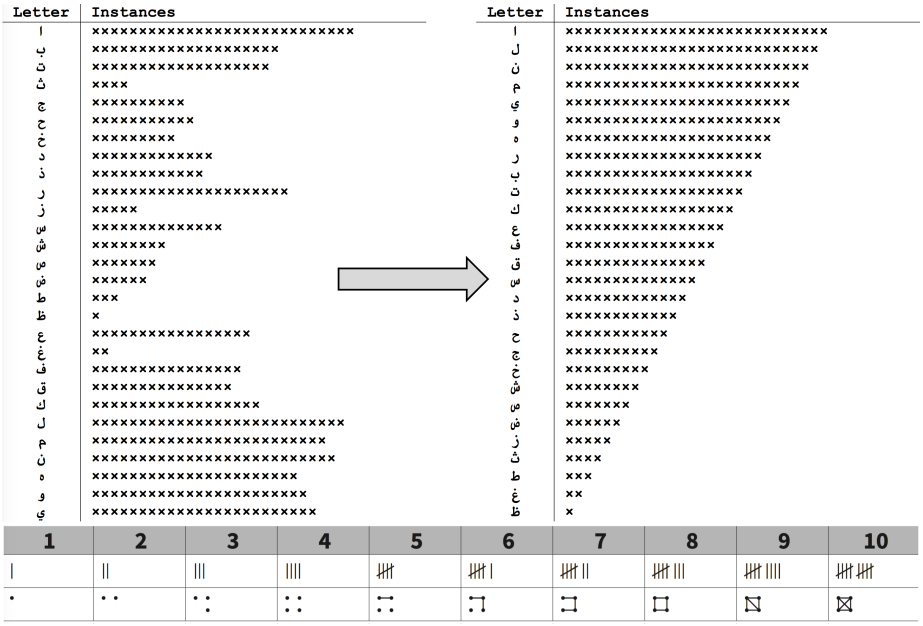

Figure 14: A possible method for counting letter frequencies: (*left*) the initial stem-and-leaf plot (one can clearly see that *alif* and *lām* are the most frequent letters, while *ġayn* and *ẓāʾ* are the least frequent ones; this ranking is based on the text of the Qurʾān); (*right*) the same plot rearranged by frequencies, convenient for the task of deciphering; (*bottom*) examples of tallying marks that could have been used for large scale calculations.

AH	370	380	390	400	410	420	430	440	450	460
in TI	42%	40%	42%	49%	49%	53%	49%	44%	48%	41%

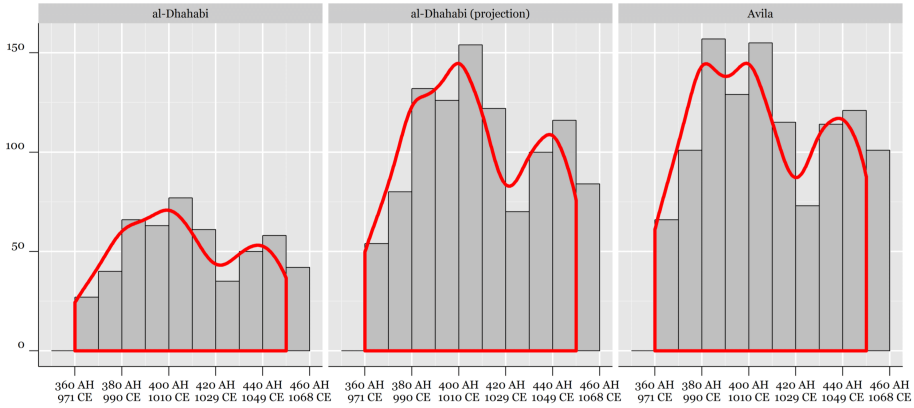

Figure 15: In her study of the demography of al-Andalus, Avila[59] collected all relevant data for the period of 360–460 AH / 971–1068 CE from four major Andalusian biographical collections (almost 1,150 individuals), of which al-Ḏahabī explicitly names three as sources of his *Taʾrīḫ al-islām*. The (center) graph shows a projected representation of Andalusia had al-Ḏahabī chosen to write his history in 100 volumes, an opportunity he himself considered. The table (top) shows that al-Ḏahabī's selected 40-50% of individuals from each decade!

Another example suggests that al-Ḏahabī was not the only one making quantitatively representative sampling of their sources. Figure 16 of Ibn al-ʿImād's *Šaḏarāt al-ḏahab* and al-Ḏahabī's *Taʾrīḫ al-islām* shows the chrono-geographical coverage of both sources.[60] Similarities between them are striking, to say the least, especially if one considers that the authors belonged to different ideological camps (at least in terms of legal affiliations), lived in different provinces and were chronologically separated by almost three centuries. The similarity in the proportional representation of Islamic provinces is even more striking in the light of the significant difference in the overall volume of both sources: about 30,000 biographies in the *Taʾrīḫ al-islām* for the period of 700 lunar years (about 74% individuals with toponymic *nisbat*s) versus about 8,500 biographies in the *Šaḏarāt al-ḏahab* for the period of 1000 lunar years (about 72% individuals with identifiable places of origin for the period of 100–1000/719–1592 CE).

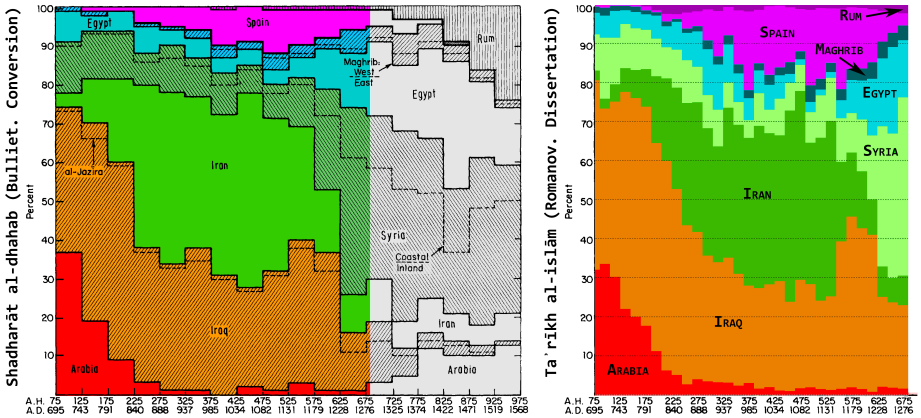

Figure 16: A comparative graph showing chrono-geographical coverage of Ibn al-ʿImād's *Šaḏarāt al-ḏahab* and al-Ḏahabī's *Taʾrīḫ al-islām.* NB: Bulliet's graph has been modified slightly for readability—the area in bright colors shows the same period as al-Ḏahabī's *Taʾrīḫ al-islām*.

In conclusion, I still have no explicit evidence that al-Ḏahabī—or any other Muslim historian—used any of the methods I am theorizing above. Yet, the modularity of their data, format in which this data was most likely collected and stored, the existence of relevant methods, and, most importantly, discoverable statistically meaningful patterns suggest that there was a quantitative methodology behind the work of al-Ḏahabī, and by extension of other scholars who worked with massive amounts of textual data. Even with quantitative methods out of vogue after the "cultural turn,"[61] modern historians still employ them when historical anal-

[59] Avila, *La sociedad hispanomusulmana*.

[60] For more details, see Romanov, "Computational Reading," 97–99. The graph for Ibn al-ʿImād's *Šaḏarāt al-ḏahab* is from Bulliet, *Conversion to Islam*, 8.

[61] Reynolds, "Do Historians Count Anymore?"

ysis requires they do so. After all, if a hammer in a hand makes everything look like nails, would not lots of nails beg for something that works like a hammer?

References

al-Ḏahabī. *al-Amṣār ḏawāt al-āṯār*. Edited by Qāsim ʿAlī Saʿd. 1st ed. Bayrūt: Dār al-Bašāʾir al-Islāmīyaṯ, 1986.

———. *al-Amṣār ḏawāt al-āṯār*. Edited by ʿAbd al-Qādir Arnāʾūṭ and Maḥmūd Arnāʾūṭ. 1st ed. Dimašq ; Bayrūt: Dār Ibn Kaṯīr, 1985.

al-Ḏahabī. *Taʾrīḫ al-islām wa-wafayāt al-mašāhīr wa-al-aʿlām*. Edited by ʿUmar Tadmurī. 2nd ed. 52 vols. Bayrūt: Dār al-Kitāb al-ʿArabī, 1990.

al-Kadi, Ibrahim A. "Origins of Cryptology: The Arab Contributions." *Cryptologia* 16, no. 2 (April 1992): 97–126.

al-Muqaddasī, Muḥammad b. Aḥmad. *The best divisions for knowledge of the regions (Aḥsan al-taqāsīm fī maʿrifaṯ al-aqālīm)*. Translated by Basil Anthony Collins. The Great Books of Islamic Civilization. Reading: Garnet, 2001.

al-Saḫāwī. *al-Iʿlān bi-l-tawbīḫ li-man ḏamma ahl al-taʾrīḫ*. Edited by (Franz Rosenthal). Al-Ṭabʿaṯ al-ūlá [1st Edition]. Bayrūt: Muʾassasaṯ al-risālaṯ, 1986.

al-Samʿānī. *al-Ansāb*. 5 vols. Bayrūt: Dār al-fikr, 1998.

al-Šayḫ, ʿAbd al-Sattār. *al-Ḥāfiẓ al-Ḏahabī: muʾarriḫ al-islām, nāqid al-muḥaddiṯīn, imām al-muʿaddilīn wa-al-mujarriḥīn (673-748 Hijrī)*. Dimašq; Bayrūt: Dār al-Qalam, 1994.

Avila, María Luisa. *La sociedad hispanomusulmana al final del califato: (aproximación a un estudio demográfico)*. Madrid: Consejo Superior de Ivestigaciones Cietíficas: Instituto de Filología, Departamento de Estudios Arabes, 1985.

Baʿlabakkī, Ramzī. *The Arabic lexicographical tradition: from the 2nd/8th to the 12th/18th century*. Handbook of Oriental Studies. Section 1, the Near and Middle East, volume 107. Leiden ; Boston: Brill, 2014.

Bearman, P., Th. Bianquis, C.E. Bosworth, E. van Donzel, and W.P. Heinrichs, eds. *Encyclopaedia of Islam (EI2-Online)*. Second edition. Malden, MA: Brill Online, 2016. http://referenceworks.brillonline.com/.

Birnbaum, E. "Kātib Chelebi (1609-1657) and Alphabetization: A Methodological Investigation of the Autographs of His Kashf al-ẓunūn and Sullam al-wuṣūl." In *Scribes et Manuscrits Du Moyen-Orient. Sous La Dir. de F.Déroche & F.Richard*, 235–63. Bibliothèque Nationale de France, 1997.

———. "The Questing Mind: Kātib Chelebi, 1609-57: A Chapter in Ottoman Intellectual History." In *Corolla Torontonensis: Studies in Honour of Ronald Morton Smith.*

Ed. E.Robbins, S.Sandahl, 133–58. TSAR, 1994.

Brown, Jonathan. *Hadith: Muhammad's Legacy in the Medieval and Modern World*. Oxford: Oneworld, 2009.

Bulliet, Richard W. *Conversion to Islam in the Medieval Period: An Essay in Quantitative History*. Cambridge, Mass.: Harvard University Press, 1979.

Cartography in the Traditional Islamic and South Asian Societies. The History of Cartography, v. 2, bk. 1. Chicago: University of Chicago Press, 1992.

Gandz, Solomon. "Did the Arabs Know the Abacus?" *The American Mathematical Monthly* 34, no. 6 (1927): 308–16.

———. "The Origin of the Ghubār Numerals, or the Arabian Abacus and the Articuli." *Isis* 16, no. 2 (1931): 393–424.

Krachkovskii, I. Iu. *Arabskaia geograficheskaia literatura*. Moskva: Izdatelskaia firma "Vostochnaia literatura" RAN, 1957–.

Librande, Leonard. "al-Dhahabī's Essay 'Amṣār Dhawāt al-āthār'." *Bulletin d'études Orientales* 32/33 (1980): 113–60.

Lucas, Scott C. *Constructive Critics, Ḥadīth Literature, and the Articulation of Sunnī Islam: The Legacy of the Generation of Ibn Sa'd, Ibn Ma'īn, and Ibn Ḥanbal*. Leiden ; Boston: Brill, 2004.

Maʿrūf, Baššār ʿAwwād. *al-Ḏahabī wa-manhaju-hu fī kitābi-hi Taʾrīḫ al-Islām [al-Ḏahabī and his technique in writing "The History of Islam"]*. Al-Ṭabʿaṯ al-ūlá [1st Edition]. al-Qāhiraṯ [Cairo]: Maṭbaʿaṯ ʿĪsá al-Bābī al-Ḥalabī, 1976.

Mrāyātī, Muḥammad. *ʿIlm al-taʿmiyyaṯ wa-istiḫrāǧ al-muʿammá ʿinda al-ʿarab*. Dimašq: Maṭbūʿāt Muǧammaʿ al-luġaṯ al-ʿarabiyyaṯ bi-Dimašq, 1987.

Reynolds, John F. "Do Historians Count Anymore? The Status of Quantitative Methods in History, 1975-1995." *Historical Methods* 31 (1998): 141–48.

Romanov, Maxim G. "Computational Reading of Arabic Biographical Collections with Special Reference to Preaching in the Sunnī World (661-1300 CE)." PhD thesis, University of Michigan, 2013.

Rosenthal, Franz. *A History of Muslim Historiography*. Leiden: E. J. Brill, 1952.

Rybalkin, Valeriy. *Klassicheskoe arabskoe iazykoznanie*. Kiev: "Stylos", 2003.

Schreibman, Susan, Raymond George Siemens, and John Unsworth, eds. *A companion to digital humanities*. Blackwell Companions to Literature and Culture 26. Malden, MA: Blackwell Pub, 2004. http://www.digitalhumanities.org/companion/.

Somogyi, Joseph de. "Ibn al-Jauzī's School of Historiography." *Acta Orientalia (Academiae Scientiarum Hungaricae)* 6 (1956): 207–14.

———. "The *Taʾrīkh al-islām* of adh-Dhahabī." *Journal of the Royal Asiatic Society of*

Great Britain and Ireland, no. 4 (October 1, 1932): 815–55.

Talmon, Rafael. *Arabic grammar in its formative age: Kitāb al-ʿayn and its attribution to Ḫalīl b. Aḥmad*. Studies in Semitic Languages and Linguistics. Leiden: Brill Academic Publishers, 1997.

Tukey, John W. *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science. Reading, Mass.: Addison-Wesley Pub. Co., 1977.