

Research Article

Maxim Romanov

Observations of a Medieval Quantitative Historian?

Abstract: The comparison of the results of the computational analysis of the *Ta'rikh al-islām*, al-Dhahabī'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 dhawāt al-āthār*, al-Dhahabī's 4-folio epistle, suggests that al-Dhahabī 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-Dhahabī as one of the earliest quantitative historians. Although we do not have conclusive evidence about how exactly al-Dhahabī worked with his data, the paper argues that all necessary mathematical, visual and 'mechanical' techniques that would facilitate data analysis already existed, and that historians could have used them..

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** (<https://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** (<https://www.r-project.org/>), a free software environment for statistical computing and graphics, and **D3** (<http://d3js.org/>), a JavaScript library for building interactive data-driven documents.

Keywords: Historiography, al-Dhahabī, historical method, quantitative history

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

1 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 professor² 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-Dhahabī's *Ta'rikh 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 *Ta'rikh al-islām* is a 50-volume mammoth of Islamic biographical literature that covers the first seven centuries of Islamic history and includes over 30,000⁶ biographical records arranged chronologically into decades. However, this giant

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 *Ta'rikh al-islām* from *al-Jāmi' al-kabīr*, a collection of about 2,400 Arabic texts (mostly premodern) published by al-Turāth (Jordan; comes on an external hard drive). The text itself is based on (and has been collated by me with): AL-DHAHABĪ, *Ta'rikh al-islām*; on this source, see: SOMOGYI, "The *Ta'rikh 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 *Ta'rikh 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 *Ta'rikh al-islām*, whose volume is approximately 2,9 mln. words, is second only to Ibn 'Asākir's (d. 571/1175 CE) *Ta'rikh madīnat Dimashq* (approximately 8,1 mln. words), whose chronological and biographical coverage, however, is significantly smaller.

book lacks anything that could offer even a preliminary answer to whether al-Dhahabī had a wholistic view of his historical and geographical data. Even though al-Dhahabī frequently comments on specific events and individuals, nowhere in the *Ta'rikh 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-Dhahabī—*al-Amṣār dhawāt al-āthār* (“Cities and ports for hearing the reports”)⁷—may be the missing analytical partner text to al-Dhahabī’s book. The exact opposite of the *Ta'rikh al-islām*, the *Amṣār* is a mere 4-folio epistle where al-Dhahabī briefly characterizes the role of different urban centers and provinces of the Islamic world in Ḥadīth scholarship up to his own time. Unlike most of al-Dhahabī’s works which focus on individuals, the subject of the *Amṣār* is cultural geography; most interestingly, al-Dhahabī 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-Dhahabī’s assessments. However, existing assessments of al-Dhahabī’s *Amṣār* are methodologically problematic, since they are based either on al-Dhahabī’s reputation as a prominent Ḥadīth scholar,⁹ or on evaluations of al-Dhahabī’s own sample of biographical

⁷ This excellent translation was offered by Michael Cooperson. Franz Rosenthal, who translated the title as “Main cities in which traditions were cultivated,” decided to exclude this epistle from his translation of al-Sakhawi’s *al-Iclān bi-l-tawbikh li-man dhamma ahl al-ta'rikh*, 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-DHAHABĪ, *al-Amṣār*, 1985; AL-DHAHABĪ, *al-Amṣār*, 1986. On the *Amṣār* also see: AL-SHAYKH, *al-Hāfiẓ al-Dhahabī*, 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” (*thumma fī athnā'i l-mi'ati l-thalithati tanāqaṣa 'ilmu l-haramayni wa-kathura bi-ghayri-himā*). AL-DHAHABĪ, *al-Amṣār*, 1985, 20.

⁹ Fuat Sezgin accepts that in his *Amṣār* al-Dhahabī “gives us comprehensive information about the centres for *hadīth*-study and their distribution in different centuries throughout the Muslim world.” (Sezgin, Fuat. “Dār al-Ḥadīth,” in *EI2, Brill Online*.) However, the epistle is very short and sketchy to take it as a reliable assessment on its own. See also

data in the *Amsār*.¹⁰ In many ways, scholarly attempts to assess the reliability of al-Dhahabī's statements in the *Amsār* 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.

2 Methodological considerations

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

The dataset formed from the *Ta'rikh al-islām* includes about 29,000 biographies of individuals who died in the period of c. 40–700/661–1300 CE.¹¹ The prevailing majority of individuals were included in the *Ta'rikh al-islām* because they were involved in the transmission of Ḥadīth (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-Dhahabī's descriptions of those regions in the *Amsār*.

Before we turn to the comparison of statements from the *Amsār* and the graphs of relevant data from *Ta'rikh al-islām*, some methodological assumptions must be made explicit. The graphs that will follow are based on “descriptive

the above-mentioned Arab editions of the *Amsār*, and AL-SHAYKH, *al-Hāfiẓ al-Dhahabī*, 447–48.

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

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-rāwā “an fulān bn fulān*, “he transmitted from so-and-so,” and their numerous variations and equivalents.

names” (sing. *nisba*), 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 *nisba*, 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 Qudāma—is quite interesting from the perspective of relying on the face value of *nisbas*. The *nisba* “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-Dhahabī writes that Jerusalem was never a center of learning, and as the data from the *Ta’rīkh 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.

Although the situation with *nisbas* may appear confusing, it in fact can be resolved through collocation analysis—that is by looking into what other *nisbas* are applied to individuals who bear the name “al-Muqaddasī.” In the *Amṣār*, al-Dhahabī writes about the Muqaddasīs (pl. al-Maqādisa) in the context of Damascus, and—if we look at the *Ta’rīkh al-islām*—the most frequent *nisbas* of the Muqaddasīs in the period of 500–700/1107–1301 CE are “al-Dimashqī” and “al-Ṣāliḥī,” with the first referring to the city of Damascus, and the second to the Ḥanbalī quarter of this city. Similar *nisba*-usage cases can be observed with other

13 See, for example, AL-SAMĀNĪ, *al-Ansāb*, 5:235–239.

14 ROMANOV, “Computational Reading,” 28–35.

toponymic names as well. Referring to the town of Suhravard and the region of Jīlān (both in Iran), the *nisbas* “al-Suhrawardī” and “al-Jīl[ān]ī” feature in the *Ta’rīkh al-islām* only when the Suhrawardī and the al-Jīl[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 *nisba* of both the Suhrawardīs and the al-Jīl[ān]īs during this period is “al-Baghdādī.” Such instances of re-appropriation are not frequent and happen only with *nisbas* that are not frequent; more importantly, the way my method is designed, the Muqaddasīs will be counted also as Damascenes, and al-Suhrawardīs and al-Jīl[ān]īs as Baghdadis.

My computer-aided analysis of the 29,000 biographies yields about 700 unique *nisbas* (with over 300 toponymic ones) that identify a group of at least 10 different individuals in the *Ta’rīkh al-islām*; the overall number of these *nisbas* runs into over 70,000 instances, considering that individuals are often described with more than one *nisba*. While 70,000 data points can hardly be called “big data”, this dataset is too big to make exact identification of each and every *nisba* possible. Thus, under these circumstances, considering *nisbas* 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 *nisbas* in biographical collections improves—and this can be achieved only through large-scale analysis—these methodological assumptions can be adjusted.¹⁵

3 The Cultural Geography of the *Amṣār*

Al-Dhahabī includes over 80 urban centers and provinces in the *Amṣār*. Starting with the sacred cities of Islam he moves through the regions of al-Shām, al-‘Irāq, Miṣr, al-Yaman, al-Andalus, the regions of al-Maghrib and Ifriqiyā, al-Jazīra, northwestern Iran, northeastern Iran (Khurāsān), Mā-warā²-al-nahr and Khwārizm, 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, Ḥadramawt, and al-Ḥabasha. 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 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

¹⁵ For the detailed discussion of methodological assumptions see, ROMANOV, “Computational Reading,” 28–40.

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 *Ta’rīkh al-islām*: the more data he had in the *Ta’rīkh al-islām*, the more certain and detailed were his statements in the *Amṣār*.¹⁶

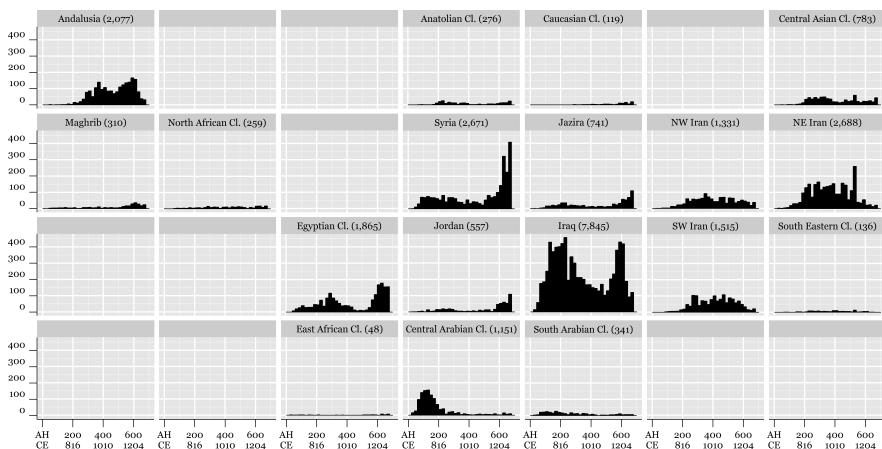


Fig. 1. Graphs of chronological coverage of different regions in the *Ta’rīkh al-islām* through toponymic *nisbas*.

Of particular interest are his chronological statements when he marks periods during which a region contributed most significantly to Ḥadīth sciences. In doing so, he names centuries (e.g., *al-mi³a al-thāniya*, “the second [hijrī] century”) or gives a reference to milestone events (such as foundation, conquest, destruction); in other cases he names most prominent Ḥadīth scholars, from which the period of prosperity can be inferred.

In terms al-Dhahabī’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

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].” (*Makka ... kāna al-‘ilm bīhā yasīrañ fī zamani l-ṣahāba thumma kathura fī awākhiri ‘aṣri l-ṣahāba wa-kadhālikā fī ayyāmi l-tābi‘īn wa-zamani aṣhābi-him ... thumma fī athnā‘i l-mi³ai l-thālithai tanāqasha ‘ilmu l-haramayni wa-kathura bi-ghayri-himā*).

clear chronological statements. Figure 1 shows chrono-geographical coverage of the *Ta'rikh al-islām*, while Figure 2 displays how this coverage compares with al-Dhahabī's statements in the *Amṣār*.

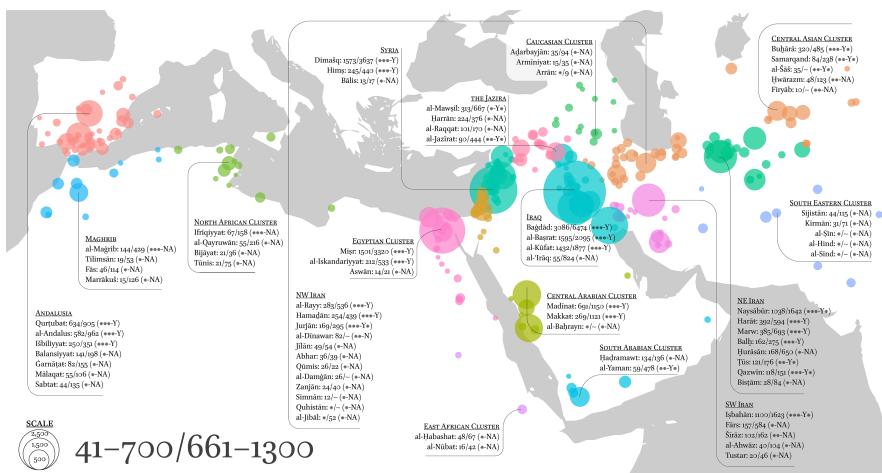


Fig. 2. The cartogram of geographical distribution of individuals in the *Ta'rikh al-islām* with corresponding comments from the *Amṣār*. For example, “Madīna: 691/1150 (***-Y)”, which says that in the *Ta'rikh al-islām* there are 691 individuals that can be associated with Madīna through their *nisbas*, 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 *Ta'rikh 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 *Ta'rikh al-islām*; Y*—can be interpreted as agreeing with the data from the *Ta'rikh al-islām*; N—does not agree; NA—not applicable, since the place is not explicitly mentioned in the *Amṣār*.

Listed places. More than half of places are simply listed by al-Dhahabī without any inferable information on their role and importance in the area of Ḥadīth sciences. Comparison with the *Ta'rikh 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-Dhahabī lists one or two prominent Ḥadīth scholars associated with a place, but refrains from any broader statements. Comparison with data in the *Ta'rikh 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-Dhahabī's “History”.

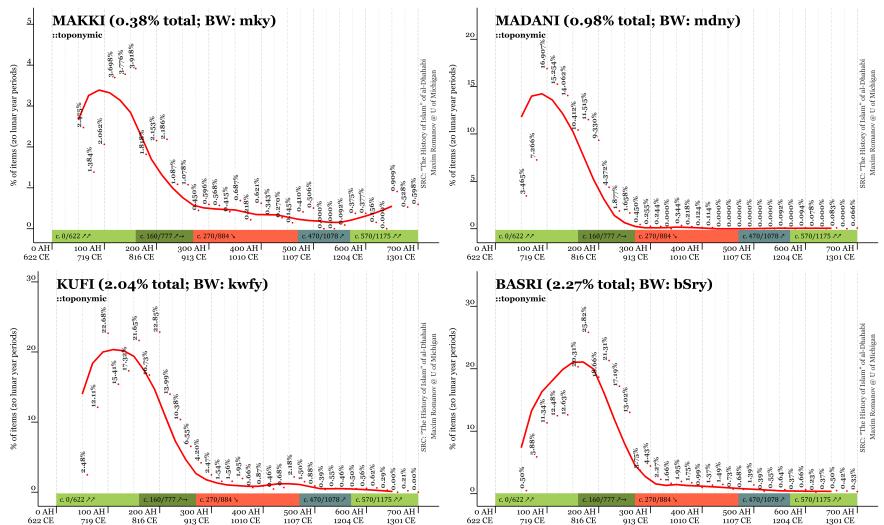


Fig. 3. Chronological distribution of individuals in the *Ta'rikh al-Islām* associated with the early Islamic centers.

Certain statements. al-Dhahabī's most certain statements are about places for which he has most data in the *Ta'rikh 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 *Ta'rikh 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 *Ta'rikh al-Islām* who bear toponymic names associated with these places. In the *Amsār al-Dhahabī* says that Medina (*nisba al-Madani*) and Mecca (*nisba 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 (*nisba al-Kūfi*) and Basra (*nisba al-Baṣrī*), they also began to gain prominence during the time of the Companions; al-Dhahabī marks the end of the Kufan prominence with Ibn 'Uqba 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-Dhahabī's statements correspond perfectly to the curves: his statements of floruit—different forms and variations of *kathura al-*

cilm bi-hā, “the learning was abundant there”—agree with the peaks of curves, while his statements of decline—forms and variations of *tanāqasa al-cilm 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).

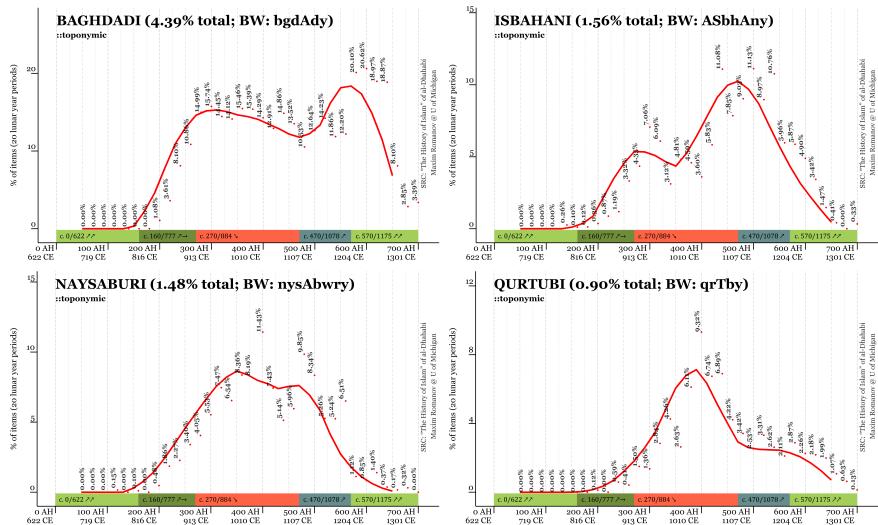


Fig. 4. Chronological distribution of individuals in the *Ta'rikh al-islām* associated with the intermediate Islamic centers.

The graphs of intermediate centers (Figure 4) feature Baghdad (*nisba al-Baghdādī*), Isfahan (*nisba al-Isbahānī*), Nishapur (*nisba al-Naysābūrī*), and Cordova (*nisba al-Qurṭubī*). In the *Amṣār al-Dhahabī* writes that Baghdad remained the key center from its foundation by the caliph al-Mansūr (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

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

¹⁸ al-Dhahabī talks about Andalusia in general, without detailed statements on its cities. See, AL-DHAHABĪ, *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*.

Nishapur¹⁹ starts its history as a center with Ibrāhīm b. Tāhmān who died in 163/779 CE and ends with the coming of the Mongols in 617/1220 CE, after which it disappeared, “as if it never existed.” al-Dhahabī’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

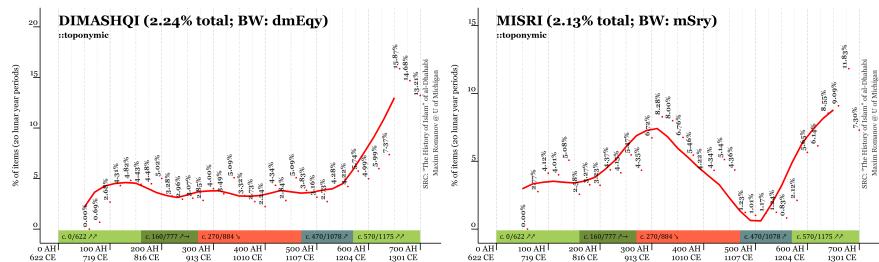


Fig. 5. Chronological distribution of individuals in the *Ta'rikh al-islām* associated with the late Islamic centers.

Unlike al-Dhahabī’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 (*nisba al-Dimashqī*) and Egypt (*nisba 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 Mu‘āwiya (r. 41–60/661–680 CE) and ‘Abd 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 ‘Asākir (d. 571/1175 CE), the Muqaddasī family, Ibn Taymīya (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

¹⁹ AL-DHAHABĪ, *al-Amṣār*, 1986, 205–208.

²⁰ AL-DHAHABĪ, *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-DHAHABĪ, *al-Amṣār*, 1986, 160–166.

Fāṭimids in 358/968 CE, whose Ismā‘īlī/Shī‘īte rule marked drastic decline for Sunnī Ḥadīth learning²³ in the province, until Salāḥ 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-Dhahabī 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. 913–1107 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.

4 The Status Quo of the Islamic Sciences

The comparison of al-Dhahabī’s two texts makes it highly plausible that al-Dhahabī’s statements in the *Amṣār* regarding major regions of the Islamic world are informed by the quantifiable data from his *Ta‘rīkh al-islām*. One, of course, may object, arguing that al-Dhahabī’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 Ḥadīth learning versus other religious sciences. Here al-Dhahabī laments that Ḥadīth learning declined—often to the point of non-existence—in most previously prominent regions, surviving now only in Egypt, Greater Syria,

23 As data from the *Ta‘rīkh al-islām* shows, the Mālikī legal school suffered in a similar way.

24 AL-DHAHABĪ, *al-Amṣār*, 1986, 167–170. NB: Alexandria became prominent during the residence of al-Silafi, 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-DHAHABĪ, *al-Amṣār*, 1986, 170–171.

25 This decline is discernible on the graph with absolute numbers. According to the data from the *Ta‘rīkh 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).

and the immediately adjacent regions. Despite the decline of Ḥadīth 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^c-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-Dhahabī’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 *Ta’rīkh al-islām*. First, the network of geographical connections of individuals from the latest volumes of the *Ta’rīkh al-islām* shows (Figure 6) that the Islamic world [of scholarship?] indeed shrunk to the crescent of Egypt, Greater Syria and northern Iraq (the Jazīra), 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-Dhahabī’s inability to get access to the later historical and biographical writings of his peers from remote regions—a possibility that al-Dhahabī himself considered²⁸—this very inability may be a witness to the fact that the cultural integration of the 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 described as specialists in the “Islamic trivium”—the Qur’ānic, Ḥadīth, 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-Dhahabī). As to Ḥadīth specialists, their prime time—the 3rd/9th century CE—had long passed. The curve of Ḥadīth specialists aggregates all major *ḥadīth*-related “descriptive names” that al-Dhahabī used in the *Ta’rīkh al-islām*.²⁹ The most frequent *nisbas* are *hāfiẓ*,

²⁶ *mukaddar ... bi-‘ulūm al-awā’il wa-ārā’ al-mutakallimīn wa-l-mu‘tazila*, AL-DHAHABĪ, *al-Amṣār*, 1986, 235.

²⁷ Librande quotes a couple of similar laments of Ḥadīth scholars of earlier times. See, LIBRANDE, “al-Dhahabi’s Essay,” 128.

²⁸ AL-DHAHABĪ, *Ta’rīkh al-islām*, 1:11–16.

²⁹ Each individual is counted only once even if he is described with more than one *hadīth*-related *nisba*.

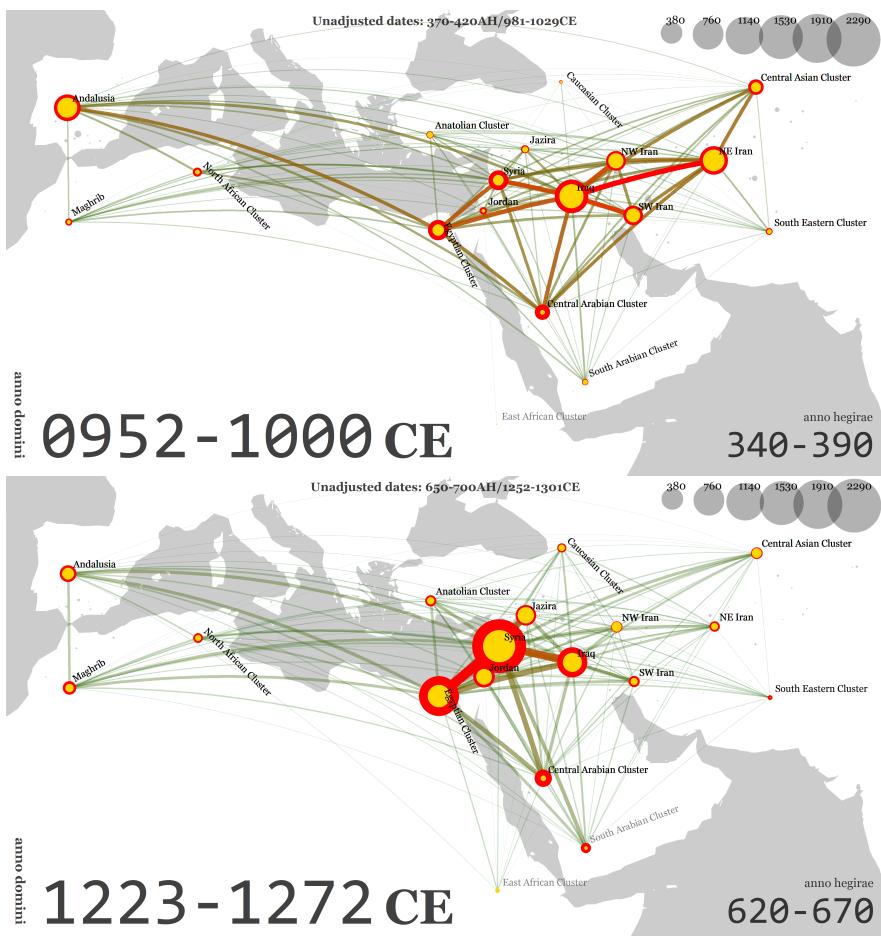


Fig. 6. The cartograms show how, according to al-Dhahabi's *Ta'rikh al-islām*, the Islamic world was connected during two different periods: the cartogram at the top shows a more even representation of major regions and their more comprehensive interconnectedness, while the cartogram at the bottom demonstrates 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.

thīqa, *raḥḥāl/a]*, and *muhaddith*.³⁰ The pattern of distribution of the first three *nisbas* 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 Ḥadīth when the six canonical collections (*al-kutub al-sitta*) were compiled, and during which Ḥadīth scholars travelled most actively (as the name *raḥḥāl/a* and its counterpart *jawwāl* imply).³¹ At the same time, while the number of Ḥadīth specialists declined, Ḥadīth continues playing a central, and perhaps even more important role, in the life of Islamic society, as described by al-Dhahabī. As was stated above, over 90% of all biographees in the *Taṣrīkh al-islām* were involved in the transmission of Ḥadīth, 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 *hadīths* from so-and-so. Additionally, the variety of social backgrounds of those involved in the transmission of Ḥadīth expands to the point that we now even find military commanders (sing. *amīr*) among the transmitters of the words of the Prophet.

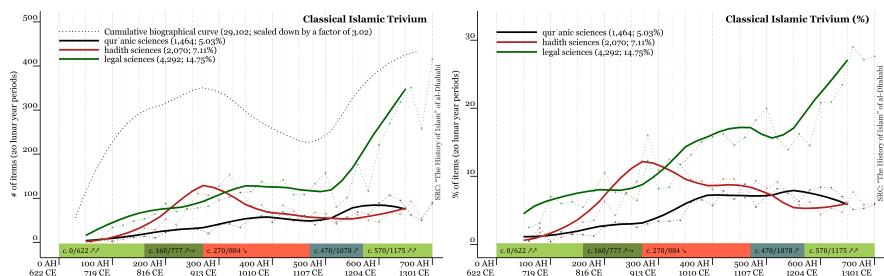


Fig. 7. Fluctuations of individuals involved in Qur'ānic, Ḥadīth and legal sciences—in absolute (left) and relative (right) numbers.

5 In search of al-Dhahabī's method

The correlation between al-Dhahabī's certain statements in the *Amṣār* and visual representations of the data from his *Taṣrīkh al-islām* is rather intriguing

30 Although the growing numbers of the *muhaddithūn* slightly push the declining curve of Ḥadīth specialists upward, this does not affect 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.

(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 *Ta'rikh al-islām*: to reiterate, the more data on a certain location there is in the *Ta'rikh al-islām*, the more certain al-Dhahabī's statement are about that location in the *Amṣār*; and vice versa—there is practically no data in the *Ta'rikh al-islām* about places that are simply mentioned in the *Amṣār* (see, Figure 8).

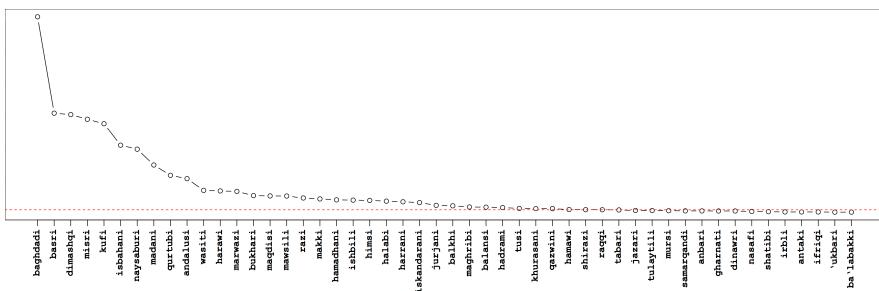


Fig. 8. Al-Dhahabī's certain statements in the *Amṣār* can be found only about places that feature most frequently in his *Ta'rikh al-islām*: on the graph these are places that appear only above the red-dashed line, which demarcates the threshold of 100 individuals per location. (The graph shows the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

These factors lead to questions about al-Dhahabī's method. How exactly did he collect and organize over 30,000 biographies and about 10,000 descriptions of events into what became his *Ta'rikh 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-Dhahabī was one of hundreds of scholars who were engaged in similar activities both before and after him,³² particularly in the

³² Moreover, in organizing his *Ta'rikh al-islām*, he must have followed in the footsteps of Ibn al-Jawzī (597/1201 CE), who was first combine a chronicle with a biographical collection in his *al-Muntażam fi-l-ta'rikh*. See, SOMOGYI, “Ibn al-Jauzi’s School of Historiography.”

fields of lexicography,³³ Ḥadīth,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-Sakhāwī’s *al-I‘lān bi-l-tawbīkh li-man dhamma ahl al-ta‘rīkh* is teeming with references to such activities.³⁷ More importantly al-Sakhāwī offers an insight into the mechanics of how exactly such activities could have occurred: we find that the *Mu‘jam al-safar* of al-Silafi (d. 576/1180 CE) was initially written on separate sheets of paper, with each biography written on its own sheet (*fī jazāzāt kull tarjama fī jazāza*).³⁸ Some autographs of the *Ta‘rīkh al-islām* include such loose sheets with writing in al-Dhahabī’s hand.³⁹ We find a similar example a few centuries later in the draft (*musawwada*) of Ḥājjī Khalīfa’s (d. 1067/1656 CE) biographical collection of poets entitled *Sullam al-wusūl ilā tabaqāt al-fuhūl*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

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, BA‘LABAKKĪ, *The Arabic lexicographical tradition*.

34 Ḥadīth 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-Sakhāwī’s *al-I‘lān bi-l-tawbīkh*, 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 geograficheskaiia 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-SAKHĀWĪ, *al-I‘lān*, 211.

39 Ma‘rūf considers them to be his methodological tool, see: MACRŪF, *al-Dhahabī wa-manhaju-hu*, 395.

40 See, BIRNBAUM, “The Questing Mind”; BIRNBAUM, “Kātib Chelebi (1609–1657) and Alphabetization.”

41 BIRNBAUM, “The Questing Mind,” 148.

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 *Ta’rīkh 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 sources⁴² 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 continue the database metaphor; see Figure 9) could have been what he wanted to write in the first place.

Variable/Source	<i>Ta’rīh al-islām</i>	<i>Ta’rīh Bağdād</i>	<i>Tabaqāt al-ḥanābilat</i>
<i>period from</i>	-52 AH	145 AH	164 AH
<i>period until</i>	700 AH	473 AH	527 AH
<i>personalia</i>	all	all	Ḥanbalis
<i>geography</i>	all	Baghdad	all
<i>arrangement</i>	chronological	alphabetical	generational

Fig. 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-Dhahabi’s *Ta’rīkh 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-Khatib al-Baghdadi’s *Ta’rīkh Baghdađ*—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 Ya‘lā’s *Tabaqāt al-ḥanābilat*, 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”.

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 *Ta’rīkh al-islām*) that can be traced back to the *Ta’rīkh madīnat Dimashq* of Ibn ‘Asākir (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.

The very organization of the *Ta'rikh al-islām* suggests that al-Dhahabī'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-Dhahabī'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 downs of a group in question. Figure 10 offers a visual representation of this point.

Two of al-Dhahabī's own "abridgments" of the *Ta'rikh al-islām* can be seen as such thematic subsets/queries: namely, his books on prominent scholars of Ḥadīth (*Tabaqāt al-huffāz*) and prominent reciters of the Qur'ān (*Ma'rifat al-qurrā' 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 *Tabaqāt al-huffāz* points to the same period of florescence—250–300 AH / 864–912 CE—as on the graph of Ḥadīth specialists based on the *Ta'rikh al-islām*, also showing a similar declining trend of Ḥadīth sciences by the end of the period. The *Ma'rifat al-qurrā' al-kibār* on the other hand clearly shows the rise of the Qur'ān reciters by the end of the period.

Al-Dhahabī's two abridgments, *Tabaqāt al-huffāz* and *Ma'rifat al-qurrā' al-kibār*, and a possible method of working with biographies ("the mechanical histogram"), may explain the certainty of al-Dhahabī's statements regarding the status quo of Ḥadīth and Qur'ān sciences at his time. From what we know, he has not written a comparable summary on jurists, but he did thoroughly work

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

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.

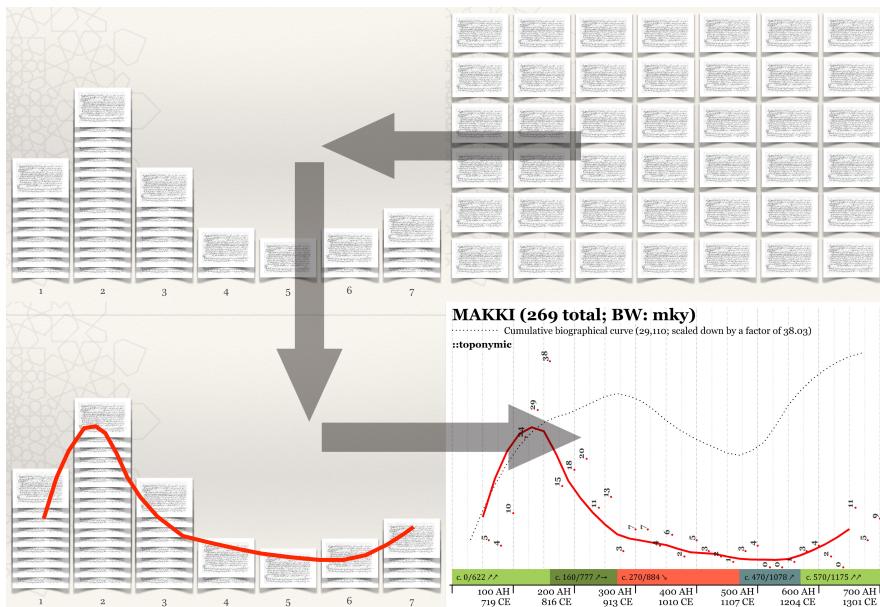


Fig. 10. A possible analytical tool: (top-right) relevant biographies are collected from the “databank” and (top-left) arranged into periods (here, centuries), which (bottom-left) offers an analytical summary similar to a modern graph (bottom-right).

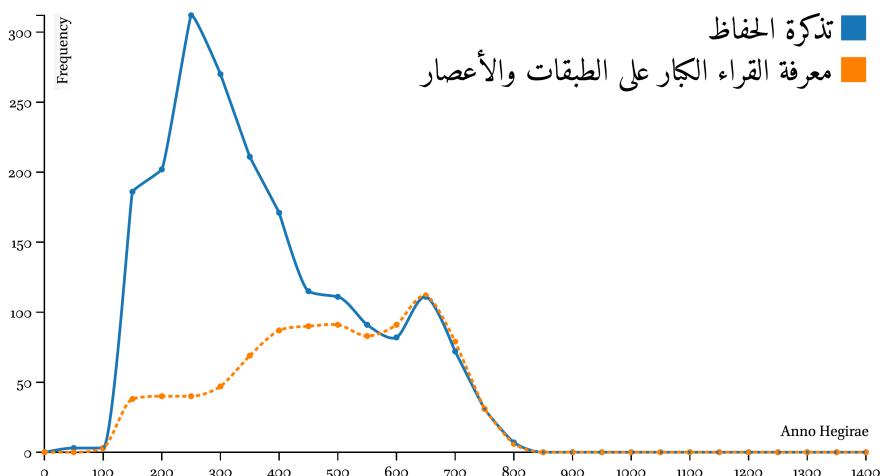


Fig. 11. Date statements from the *Tabaqāt al-huffāz* (blue) and the *Ma'rifāt al-qurrā al-kibār* (orange) display the same chronological patterns for Ḥadīth scholars and Qur'ān reciters as the graphs based on the *Ta'rīkh al-islām*; see Figure above on the status quo of the Islamic sciences for comparison. (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).

with all major *tabaqāt* collections of legal schools (since they are listed as his sources in the introduction to the *Ta'rikh 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 non-destructive alternative could have been counting and graphing. This possibility is not completely far-fetched, since premodern Islamic scholars were not alien to mathematical⁴⁶ and visual methods (see, Figures 12 & 13) when working with information in what can be characterized as a humanistic inquiry.⁴⁷ Us-

45 At the same time, the number of jurist at the late period was so significantly higher than those of Ḥadīth and Qur’ān 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-Khalīl al-Farāhidī (d. c. 170/786 CE) designed his dictionary of the Arabic language, *Kitāb al-‘ayn*, 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 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 BA‘LABAKKĪ, *The Arabic lexicographical tradition*, 292–96. See also, a section on *combinatorial analysis* in R.Rashed’s “al-Riyādiyyāt”, *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-‘ayn and its attribution to Khalil b. Ahmad*, 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

ing 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-hisāb bi-l-takht wa-l-turāb*; “calculation [by means] of dust”, *hisāb al-ghubār*).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:

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-Riyādiyyāt”, *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-Hisāb*” and M. Souissi’s “*Hisāb al-ghubār*” in *EI2-Online*.

50 See, AL-KADI, “Origins of Cryptology”; MRĀYĀTĪ, *Ilm al-ta‘miyya wa-istikhrāj al-mu‘ammā cinda al-‘arab*. Whether this method was actually devised by al-Kindī is not relevant for our argument.

51 MRĀYĀTĪ, *Ilm al-ta‘miyya wa-istikhrāj al-mu‘ammā cinda al-‘arab*, 207, see also AL-KADI, “Origins of Cryptology”, 108.

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

53 AL-MUQADDASĪ, *The Best Divisions for Knowledge of the Regions*.



Fig. 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. *tashjīr*) 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 *Jawāmi‘ al-‘ulūm* (“Connections of the sciences”) of Shāyā b. Farīghūn’s (the 4th/10th century CE). On Shāyā b. Farīghūn and for additional bibliography, see: Bosworth, C. E.. “Ibn Farīghūn,” *EI2-Online*. Brill Online, 2016; a digitized microfilm of the manuscript (El Escorial 950, 84 folios) of this work can be accessed through *al-Jāmi‘ al-makhtūtāt al-islāmiyya* (<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>).

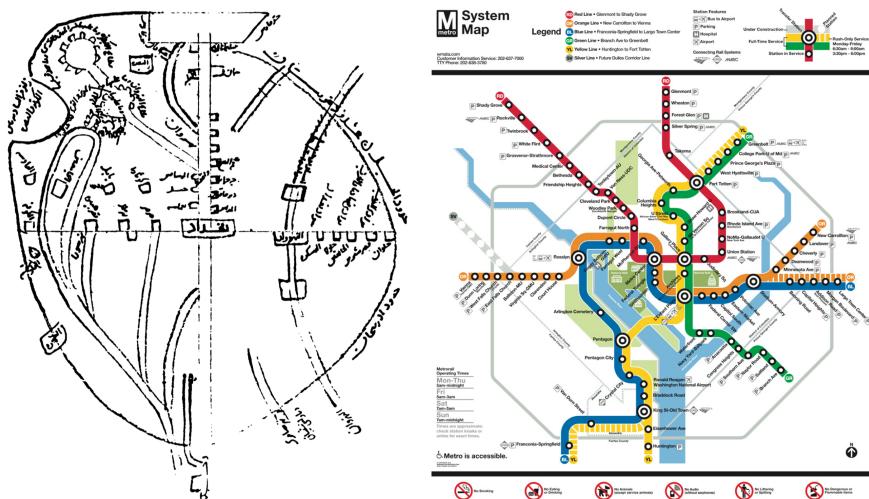


Fig. 13. The most striking examples of the visual representation of complex data are found in geographical maps and cosmographical diagrams.⁵² While early Islamic maps may appear too primitive to our modern eye, since they never represent space correctly, as *cartograms*, i.e. analytical representations 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-ṣIrāq (*left*, from al-Muqaddasī's *Ahsan al-taqāṣīm fī maṣrīfat 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 *ghayn* and *zā'*, 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-Dhahabī 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 *Ta'rikh al-islām* with other biographical texts. In this regard, al-Dhahabī's sampling of Andalusian sources is particularly interesting (Figure 15), showing that al-Dhahabī's

⁵⁴ Translation is from AL-KADI, “Origins of Cryptology,” 107–9; for Arabic, see MRĀYĀTĪ, *‘Ilm al-ta‘miyya wa-istikhraj al-mu‘ammā ‘inda al-‘arab*, 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, *‘Ilm al-ta‘miyya wa-istikhraj al-mu‘ammā ‘inda al-‘arab*, 1:77; cf. AL-KADI, “Origins of Cryptology,” 112.

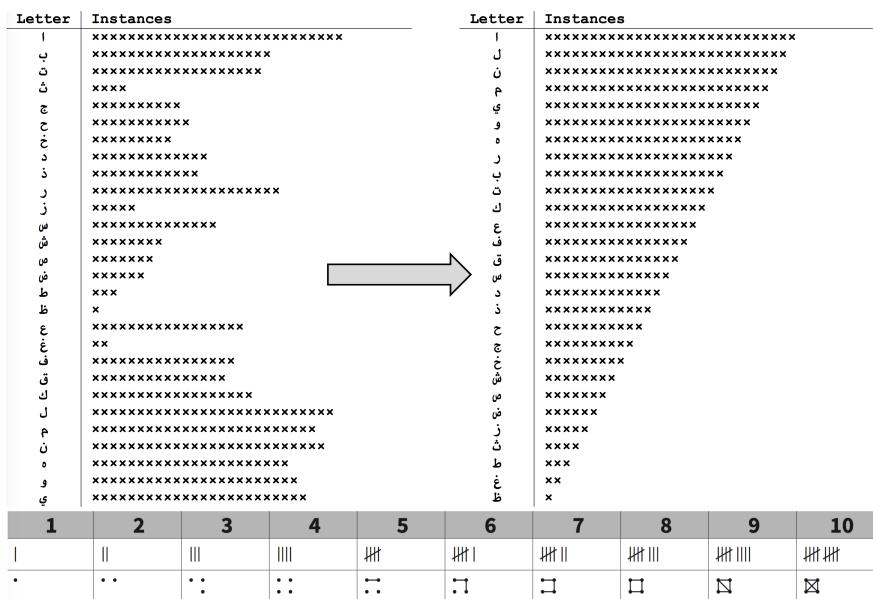


Fig. 14. A possible method for counting letter frequencies: (*left*) the initial stem-and-leaf plot (one can clearly see that *alif* and *lām* are the most frequent letters, while *ghayn* and *zāwī* are the least frequent ones; this ranking is based on the text of the Qur'ān); (*right*) the same plot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

included roughly 40 to 50% of biographies from each decade covered in his Andalusian sources, thus offering a quantitatively representative sample.⁵⁸

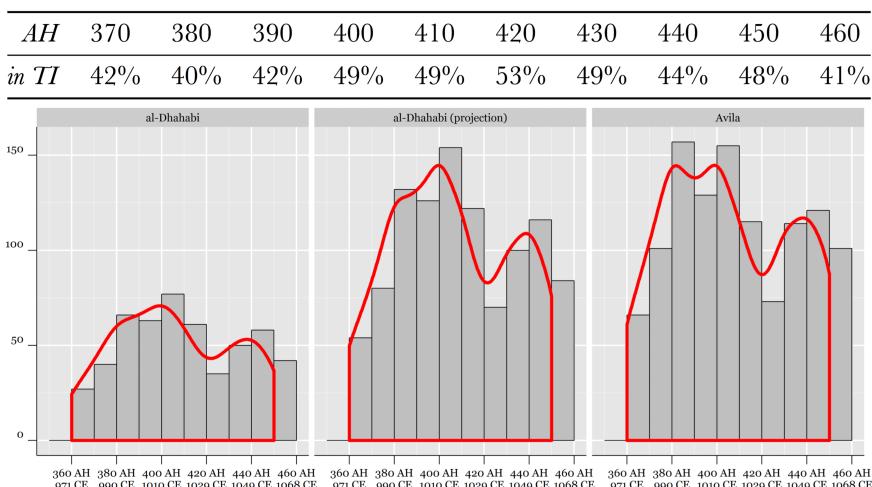


Fig. 15. In her study of the demography of al-Andalus, Avila⁵⁹ collected all relevant data for the period of 360–460 AH / 971–1068 CE from four major Andalusian biographical collections (almost 1,150 individuals), of which al-Dhahabī explicitly names three as sources of his *Ta’rīkh al-islām*. The (center) graph shows a projected representation of Andalusia had al-Dhahabī chosen to write his history in 100 volumes, an opportunity he himself considered. The table (top) shows that al-Dhahabī’s selected 40–50% of individuals from each decade!

Another example suggests that al-Dhahabī was not the only one making quantitatively representative sampling of their sources. Figure 16 of Ibn al-‘Imād’s *Shadharāt al-dhahab* and al-Dhahabī’s *Ta’rīkh 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

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.

59 AVILA, *La sociedad hispanomusulmana*.

60 For more details, see ROMANOV, “Computational Reading,” 97–99. The graph for Ibn al-‘Imād’s *Shadharāt al-dhahab* is from BULLIET, *Conversion to Islam*, 8.

is even more striking in the light of the significant difference in the overall volume of both sources: about 30,000 biographies in the *Ta'rikh al-islām* for the period of 700 lunar years (about 74% individuals with toponymic *nисбат*) versus about 8,500 biographies in the *Shadharāt al-dhahab* for the period of 1000 lunar years (about 72% individuals with identifiable places of origin for the period of 100–1000/719–1592 CE).

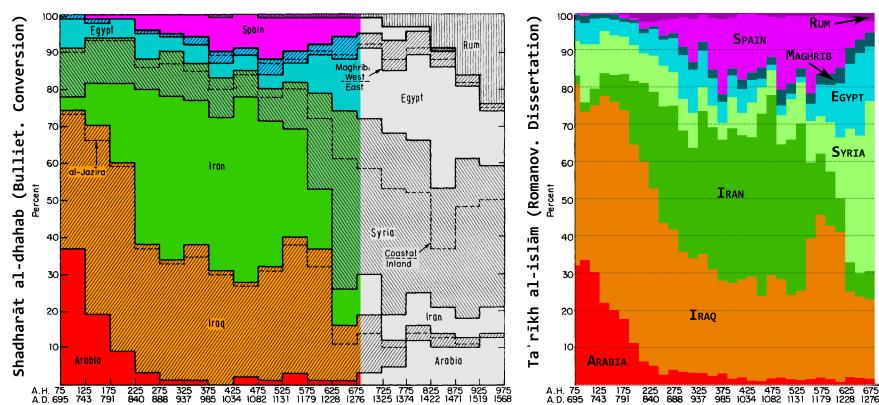


Fig. 16. A comparative graph showing chrono-geographical coverage of Ibn al-‘Imād’s *Shadharāt al-dhahab* and al-Dhahabī’s *Ta’rikh al-islām*. NB: Bulliet’s graph has been modified slightly for readability—the area in bright colors shows the same period as al-Dhahabī’s *Ta’rikh al-islām*.

In conclusion, I still have no explicit evidence that al-Dhahabī—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-Dhahabī, 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 analysis requires they do so. After all, if a hammer in a hand

61 REYNOLDS, “Do Historians Count Anymore?”

makes everything look like nails, would not lots of nails beg for something that works like a hammer?

References

- AL-DHAHABĪ. *al-Amṣār dhawāt al-ātār*. Edited by Qāsim Ḩalī Sa‘dī. 1st ed. Bayrūt: Dār al-Bashār al-Islāmīya, 1986.
- . *al-Amṣār dhawāt al-āthār*. Edited by Ḩabd al-Qādir Arnā’ūt and Maḥmūd Arnā’ūt. 1st ed. Dimashq ; Bayrūt: Dār Ibn Kathīr, 1985.
- AL-DHAHABĪ. *Ta’rīkh al-islām wa-wafayāt al-mashāhīr wa-al-a’lām*. Edited by Ḩumar Tadmurī. 2nd ed. 52 vols. Bayrūt: Dār al-Kitāb al-‘Arabī, 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 ma’rifat al-aqālīm)*. Translated by Basil Anthony Collins. The Great Books of Islamic Civilization. Reading: Garnet, 2001.
- AL-SAKHĀWĪ. *al-I‘lān bi-l-tawbikh li-man dhamma ahl al-ta’rīkh*. Edited by (Franz Rosenthal). Al-Ṭabā’ al-ūlā [1st Edition]. Bayrūt: Mu’assasa al-risāla, 1986.
- AL-SAMĀNĪ. *al-Ansāb*. 5 vols. Bayrūt: Dār al-fikr, 1998.
- AL-SHAYKH, Ḩabd al-Sattār. *al-Hāfiẓ al-Dhahabī: mu’arrikh al-islām, nāqid al-muhaddithīn, imām al-mu’addilīn wa-al-mujarrihīn (673-748 Hijrī)*. Dimashq; Bayrūt: Dār al-Qalam, 1994.
- AVILA, María Luisa. *La sociedad hispanomusulmana al final del califato: (aproximación a un estudio demográfico)*. Madrid: Consejo Superior de Investigaciones Científicas: Instituto de Filología, Departamento de Estudios Árabes, 1985.
- BA’LABAKKĪ, 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-zunūn* 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 geograficheskaiia 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, Hadīth 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.
- MACRŪF, Bashshār ʿAwwād. *al-Dhahabī wa-manhaju-hu fī kitābihi Ta’rīkh al-Islām [al-Dhahabī and his technique in writing “The History of Islam”]*. Al-Ṭabā'a al-ūlā [1st Edition]. al-Qāhira [Cairo]: Matba'a ʿĪsā al-Bābī al-Ḥalabī, 1976.
- MRĀYĀTĪ, Muḥammad. *ʻIlm al-ta‘mīyya wa-istikhrāj al-mu‘ammā ʻinda al-‘arab*. Dimashq: Matbū‘at Mujamma‘ al-lugha al-‘arabiyya bi-Dimashq, 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 *Ta’rīkh 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-‘ayn and its attribution to Khalil b. Ahmad*. 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.

Acknowledgment: ...