MAIRAJ SYED, DANNY HALAWI, BEHNAM SADEGHI, NAZMUS SAQUIB

University of California, Davis; University of California, Berkeley; University of Oxford;
Massachussets Institute of Technology

Emails: msyed@ucdavis.edu; dhalawi@berkeley.edu; bsadeghi3@gmail.com; saquib@mit.edu

Verifying Source Citations in the Hadith Literature

ABSTRACT   Historians rely on hadiths (narratives about Muhammad) as a source for writing the history of early Islam. Each hadith is preceded by an *isnād*, which is a list of names purporting to give the sequence of individuals who transmitted it. Scholars apply a variety of methods to the *isnād*s of hadiths in order to determine their dates and geographic origins. These methods presuppose the absence of inadvertent mistakes in the names cited. Because the hadith literature is so voluminous and unwieldy, the systematic discovery and correction of these mistakes using manual methods is unfeasible. We introduce new computational and statistical methods for automatically detecting a subset of these errors and correcting them. We do so by investigating the citations of sources in the *isnād*s documented in the digital hadith repository, *Gawāmiʿ al-Kalim*.

KEYWORDS: hadith, isnād, digital humanities, source criticism, early Islam, computer analysis

INTRODUCTION[1]

Historians have long used a variety of methods to determine the dates and the geographical origins of the reports about the important events and persons of early Islam that constitute our main source for writing history—namely the *āthār* and the subset of the *āthār* about the Prophet Muhammad, called hadiths. Most of these methods rely on analyzing the chain of names, *isnād*, found at the beginning of each hadith that purports to give the sequence of the people who transmitted it until it reached the premodern collector in whose book the report is found. These analytical methods do not take the *isnād*s at face-value as accurate records of a report's transmission history; they are often able to distinguish back-projections from plausible attributions to sources.[2] However, they are at least partly dependent on the absence of egregious inadvertent errors—mistakes such as the

2. As examples, three methods may be mentioned: for source-criticism applied recursively to the corpora of early figures, see Harald Motzki, *The Origins of Islamic Jurisprudence: Meccan Fiqh before the Classical Schools*, trans. Marion H. Katz (Leiden: Brill, 2002). For *isnād*-cum-*matn* analysis, see Harald Motzki, Nicolet Boekhoff-van der Voort, and Sean W. Anthony, *Analysing Muslim Traditions: Studies in Legal, Exegetical and Maghazi Hadith*, vol. 78, Islamic history and civilization (Leiden: Brill, 2010). For the traveling tradition test, see Behnam Sadeghi, "The traveling tradition test: a method for dating traditions," *Der Islam* 85, no. 1 (2010).

unintentional replacement of one name with another. Yet, given the vastness and complexity of the hadith literature, it is no surprise that such mistakes were made frequently, not only by the creators of the digital corpora on which historians increasingly depend and by editors of printed editions, but also by premodern hadith collectors and transmitters.

Thus, the large volume of hadiths, recorded in hundreds of edited published collections, with tens of thousands of persons cited hundreds of thousands of times, makes the production of errors in *isnād*s inevitable. The errors were generated in a multitude of ways: a transmitter in an *isnād* may inadvertently cite a wrong source for a given hadith; a hadith's final collector or a later copyist may incorrectly transcribe an *isnād*; an editor may write the wrong name; and a digitization team may tag a name incorrectly or introduce errors through the process of making a text machine readable.

Scholars have either been unaware of these errors or have corrected them in an *ad hoc* manner, and only for the specific hadiths they were studying, by relying on their expertise in *isnād*s. Given the sheer volume of the hadith literature, systematic identification and correction of such errors through traditional methods is not possible. In this article, we rely on a very large digitized source of hadiths and we introduce systematic computational methods of detecting and correcting some of these errors. Before we engage in a more detailed description of the methods we used and the results we obtained, we briefly describe the nature of information found in the hadith literature and give an overview of the digital repository of hadiths we relied on for our dataset.

BACKGROUND INFORMATION ON THE HADITH LITERATURE

Scholars of hadith rely on information found in two different types of sources to date and locate the transmission of hadiths. They often begin their analysis by collecting all of the versions of a given hadith. Most are found in premodern compendia devoted specifically to the collection of hadiths. Yet, hadiths can be found in other types of sources as well, such as books of law or theology. In either case, the hadiths contain *isnād*s. Scholars diagram the *isnād*s and study their structure and variation along with the variation of the corresponding texts of the hadiths on a given topic. Moreover, determining the date and provenance of a hadith requires identifying the names in the *isnād*s. To do this, they rely on a second source of information: the premodern biographical dictionaries. We introduce both type of sources through the following example:

The idea of the five pillars of Islamic religious practice is attested in hadiths and is found in several different sources. We provide three versions found in three different collections and have highlighted the names that make up the *isnād*s:

1. Bukhārī (d. 256/870)[3] writes:

    ʿ**Ubaydallāh b. Mūsā** narrated to us, saying: **Ḥanẓala b. Abī Sufyān** reported to us about ʿ**Ikrima b. Khālid:** about **Ibn ʿUmar**, may God be pleased with them both, that he said that the Prophet, peace and blessings be upon him,

---

3. The first number in 256/870 is the year in the Islamic hijri calendar; the second is the year according to the Gregorian one.

said: "Islam is built on five things: testifying that there is no god but God and that Muhammad is the messenger of God, establishing daily prayer, giving the charity tax, the pilgrimage, and fasting in Ramaḍān."[4]

2. Muslim (d. 261/875) writes:

> **Sahl b. ʿUthmān al-ʿAskarī** narrated to us: **Yaḥyā b. Zakariyyāʾ** narrated to us: **Saʿd b. Ṭāriq** narrated to us, saying: **Saʿd b. ʿUbayda al-Sulamī** narrated to me about **Ibn ʿUmar**: about the Prophet, peace and blessings be upon him, that he said: "Islam is built on five things: that God be worshipped and others beside him be rejected, establishing prayer, giving the charity tax, making the pilgrimage to the holy house, and fasting in Ramaḍān."[5]

3. Al-Ṭabarānī (d. 360/971) writes:

> **Muḥammad b. Aḥmad b. Ḥammād, Abū Bishr al-Dūlābī,** narrated to us in Egypt: **my father** narrated to us: **Ashʿath b. ʿAṭṭāf** narrated to us about **ʿAbdullāh b. Ḥabīb b. Abī Thābit:** about **al-Shaʿbī:** about **Jarīr b. ʿAbdullāh al-Bajalī:** about the Prophet, peace and blessings be upon him, that he said: "Islam is built on five things: testifying that there is no god but God, establishing prayer, giving the charity tax, [making] the pilgrimage to the holy house, and fasting in Ramaḍān." Only Ashʿath and Sawra b. al-Ḥakam al-Qāḍī have narrated this hadith from ʿAbdullāh b. Ḥabīb.[6]

As seen above, this hadith was recorded by ninth- and tenth-century hadith collectors. Each collector gives the names of the transmitters the hadith passed through before him. In this case, the *isnād*s vary largely from one collector to the next, though texts 1 and 2 both document the hadith as having been transmitted by Ibn ʿUmar. The wording of the hadith (what the Prophet said) also differs slightly from one text to the next.

The three *isnād*s of the hadiths above contain the names of 14 different transmitters. We can find more information about them by consulting the biographical dictionaries on hadith transmitters. Below is a snippet of a biographical entry providing an example of the type of information found in these sources. The tenth-century hadith scholar Ibn Ḥibbān records the following about Ḥanẓala, one of the transmitters cited in the first five-pillars hadith above:

Ḥanẓala b. Abī Sufyān al-Jumaḥī al-Qurashī was an inhabitant of Mecca. [His father] Abū Sufyān's name was al-Aswad. [Ḥanẓala] transmitted from Sālim, Qāsim, Nāfiʿ,

---

4. The translations of all hadith texts are our own. See the following link for an on-line version of the hadith: http://library.islamweb.net/hadith/display_hbook.php?bk_no=146&hid=7&pid=97669 (accessed 2/5/19). For the version found in *GK*'s desktop software, see hadith number 7 in Bukhārī's *Ṣaḥīḥ al-Bukhārī*.

5. See the following link for an on-line version of the hadith: http://library.islamweb.net/hadith/display_hbook.php?bk_no=158&hid=23&pid=105837 (accessed 2/5/19). For the version found in *GK*'s desktop software, see hadith number 23 in Muslim's *Ṣaḥīḥ Muslim*.

6. See the following link for an on-line version of the hadith: http://library.islamweb.net/hadith/display_hbook.php?bk_no=476&hid=784&pid=281037 (accessed 2/5/19). For the version found in *GK*'s desktop software, see hadith number 784 in al-Ṭabarānī's *Muʿjam al-Ṣaghīr li-l-Ṭabarānī*.

Mujāhid, and Ṭāwus. Al-Thawrī, Wakīʿ, and others transmitted from him. He died in the year 151 [AD 768].[7]

Ibn Ḥibbān provides a number of facts about Ḥanẓala, including his tribal genealogy, where he lived, when he died, his mother, and the names of some of his students and teachers. Similar types of facts, to a lesser or greater extent, can be found on thousands of individuals who participated in the transmission of hadiths over many centuries.

DESCRIPTION OF THE DIGITIZED HADITH DATASET

The source for our data is the *Gawāmiʿ al-Kalim* (*GK*) hadith software,[8] created by a team that was supported by the Qatar's General Directorate of Endowments (*al-Idārat al-ʿĀmma liʾl-Awqāf*) and the Islamweb.net website. *GK* has digitized 1,400 sources, including published books and manuscripts. 900 of these are strictly hadith collections, forming *GK*'s core dataset.[9] The rest of the sources are devoted primarily to Islamic law, theology, belles-lettres, or biographies that include hadiths with full *isnād*s. There are 828,841 hadiths found in these 1,400 sources. Importantly, *GK* has largely only parsed the *isnād*s of Prophetic hadiths, meaning they did not parse the *isnād*s that document the statements or actions of authorities other than Muhammad. *GK*'s monumental advance is distinguishing the names of individual transmitters in the *isnād*s, disambiguating them, and linking them to a separate table containing biographical data gleaned from the biographical dictionaries.[10] Of the 828,841 hadiths in *GK*, the hadiths of 447,205 have been parsed, for a total of 638,237 *isnād*s.[11] The 638,237 parsed *isnād*s contain 4,918,709 citations of names of transmitters. This does not mean that there were close to five million unique transmitters, because transmitters were often cited more than once, and some were quite prolific. A total of 49,819 unique persons are cited, some of them once or twice, and some many times. Since a single individual may be known by multiple names, a unique individual narrator can be identified differently

---

7. See https://al-maktaba.org/book/5816/2341 (accessed 2/5/19), which corresponds to the following published source: Muḥammad b. Ḥibbān b. Abī Ḥātim, *Kitāb al-Thiqāt*, ed. Muḥammad ʿAbd al-Muʿīd Khān, 10 vols. (Haidarabad: Dāʾira al-Māʿārif al-ʿUthmānī, 1973), 6:225.

8. The desktop version of the software can be downloaded for free here: https://archive.org/download/G_Kalim/G_Kalim.zip (accessed 8/20/2019) For our analysis we relied on the data found in the desktop version. A less complete version of the data can be found on-line here: http://library.islamweb.net/hadith/index.php (accessed 2/5/19).

9. For a description of the criteria *GK* used to determine the sources they digitized and their method in doing so, see the first section, *jamʿ al-māddat al-ʿilmiyya* (Collection of Scholarly Materials), of the document in which they outline their methodology, *manhaj al-ʿamal* (Methodology). This document is found in the desktop version of the software and may be accessed through the *taʿrīfāt* tab.

10. Other software programs have much more extensive collections of Islamic sources. Those devoted to Sunni texts include the *maktabat al-shāmila*, which can be found here: https://islaamiclibrary.wordpress.com/2009/03/01/thecomprehensivelibrary/ (accessed 2/5/19). For software with digitized Shīʿī texts, see the programs published by Noorsoft: https://www.noorsoft.org (accessed 2/5/19).

11. As we will see in more detail below, the discrepancy between the number of hadith texts and the number of *isnād*s is because many times a hadith's collector will document multiple *isnād*s through which he received the hadith text. According to our analysis, of the 447,205 hadith texts with fully parsed *isnād*s, 104,950 document more than one *isnād*.

in different *isnād*s. *GK* attempted to identify individuals, despite the variation in names, by giving each an unique identification number. This example shows how the names in the *isnād*s are tagged in the first hadith cited above:

> ʿ**Ubaydallāh b. Mūsā** (**5437**) narrated to us saying: that **Ḥanẓala b. Abī Sufyān** (**2567**) reported to us about ʿ**Ikrima b. Khālid** (**5699**): about **Ibn ʿUmar** (**4967**), may God be pleased with them both, that he said that the Prophet, peace and blessings be upon him, said: "Islam is built on five [pillars]: testifying that there is no god but God and that Muhammad is the messenger of God, establishing daily prayer, giving the charity tax, the pilgrimage, and fasting in Ramaḍān."

> 1. 5437 = ʿ**Ubaydallāh b. Mūsā**
> 2. 2567 = **Ḥanẓala b. Abī Sufyān**
> 3. 5699 = ʿ**Ikrima b. Khālid**
> 4. 4967 = **Ibn ʿUmar**

A table in *GK* links each unique personal ID to biographical data about that person gleaned from different biographical sources. The table contains information such as where they lived and died, sects to which they belonged, and a list of their teachers and students.[12]

In addition to disambiguating the names in the *isnād*s by assigning them unique IDs, tagging them in the texts of the hadiths in which they occur, and linking them to biographical information, *GK* has documented, in a separate file, every one of the 628,237 unique *isnād*s by listing the unique IDs of the narrators occurring in them in proper transmission order.[13] For example, the *isnād* of the above hadith is represented in the following manner:

> 4967, 5699, 2567, 5437, 6817

The earliest narrator (4967, Ibn ʿUmar, in the above example), cited in the last position in the text of the *isnād*, occurs first. In most cases, *GK* has also added, in the final position, the unique ID of the collector in whose book the hadith appears. *GK* relies on these lists of narrator IDs to create diagrams of the *isnād*s. The relationships between transmitters, which is a core part of our dataset, come from this file.

GK's digitization efforts make the study of hadiths immeasurably easier. While *GK* has parsed and arranged the *isnād* data in a structured fashion, it has not used the data to create a unified social network. We constructed a unified social network of all of the transmitters based on the transmission relationships among them in all of the *isnād*s *GK* parsed. This network documents hadith transmission activity involving 49,819 unique transmitters, 4,280,472 individual instances of hadith transmission between them, and 320,393

---

12. For a description of the sources *GK* consulted for the biographical information, see section 5, *khidamāt ruwāt al-ḥadīth*, especially its subsection *manhaj al-ʿamal fī tarājim al-ruwāt*. This document is found in the desktop version of their software and may be accessed through the *taʿrīfāt* tab. We created our own table of biographical information by scraping the data found in the "Rawy" file, which is created when one installs the *GK* software on the desktop.

13. The sequence of all narrator IDs for each parsed *isnād* in the *GK* dataset can be found in a file named "SanadRowah," which is created when one installs the *GK* software on the desktop.

unique transmitter relationships. We also created a database and associated programs to allow for a robust analysis of *GK* data. This enables devising methods for automatic detection of errors in *isnād*s, a desideratum given the heavy dependence of historians on them. Moreover, it makes possible the application of a variety of computational methods, including social network analysis and stylometric comparisons of corpora.

ERROR DETECTION THROUGH THE ANALYSIS OF ALL BIDIRECTIONAL TRANSMISSION RELATIONSHIPS IN THE HADITH NETWORK

One way to programmatically detect errors within the dataset is to identify anomalies. We thus identify departures from what we would normally expect in the transmission activity of narrators as documented in the *isnād*s. We expect that hadiths should in most cases flow from older transmitters to younger ones. In the vast majority of cases, transmitters got their texts from older transmitters. This was so for two reasons: first, an older informant would have received texts from deceased transmitters whom the younger ones could not access; and second, transmitters generally preferred having fewer intermediaries between themselves and earlier authorities.[14] Our analysis of the *GK* data confirms this expectation. We calculated the differences in death dates between a transmitter of a text and its receiver when we had death dates for both individuals. Of the 320,393 unique transmission relationships between two narrators in the *GK* dataset, we had such information for 162,520. Figure 1 shows the distribution of the differences in death dates for those 162,520 relationships. Typically, when someone transmitted a text from another person, he/she was decades younger. In 94.8% of the cases, he/she ended up outliving the informant.

Thus, information transfer in the hadith network was broadly unidirectional. This general preference for older authorities makes instances where X transmitted something to Y in one *isnād* and Y transmitted something to X in another anomalous. Such cases raise the suspicion that transmission did not take place in one of those directions: either X→Y or Y→X may be wrong. The anomalous nature of such two-way transmissions is confirmed by the data. Out of the 320,393 unique pairs of transmitters where one or both of them transmitted to the other, in only 3,206, or 1%, do we find transmissions in both directions.

If X has transmitted something to Y or/and vice versa, we call that a "transmission relationship." When each of them has transmitted something to the other, we call their relationship "bidirectional." There are 1,603 bidirectional transmission relationships that can be used to identify errors in *isnād*s, determine error types, and formulate automatic ways of correcting them.

Below is one example of a bidirectional transmission relationship, between two very prominent mid-second century transmitters of hadith, Ibn Jurayj (d. 150/767)[15] and

14. Many people have written on this. See, for example, Pavel Pavlovitch, "Hadith," in *Encyclopedia of Islam, Third Edition*, ed. Kate Fleet, et al. (Leiden: Brill, 2018), and Eerik Dickinson, "Ibn al-Ṣalāḥ al-Shahrazūrī and the *isnād*," *Journal of the American Oriental Society* 122, no. 3 (2002): 491-2.

15. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=5223 for biographical information on the narrator (accessed 2/5/19).
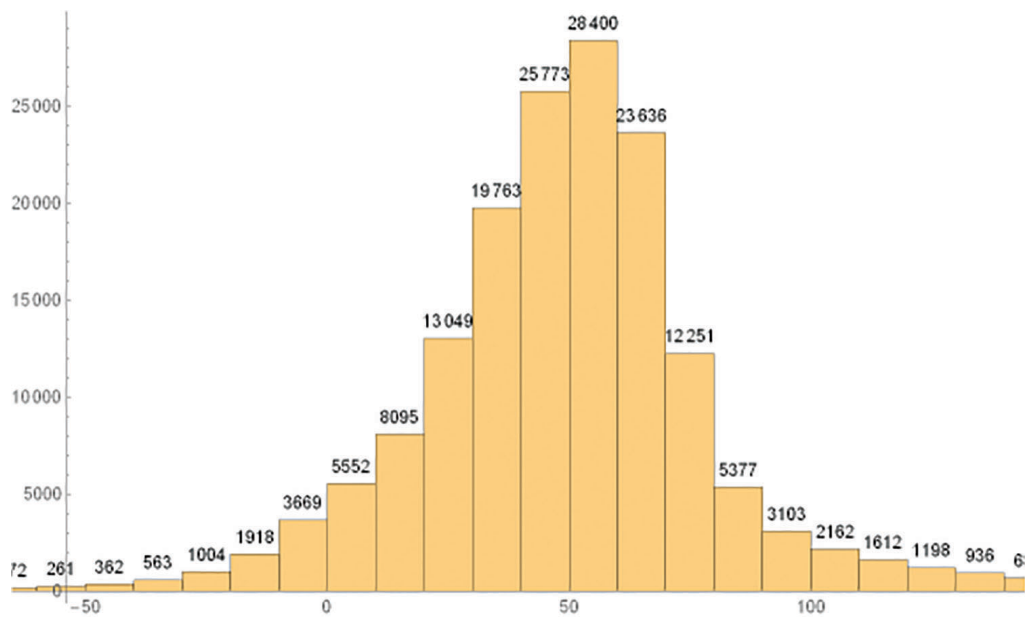
FIGURE 1. Distribution of Differences in Death Dates of Consecutive Transmitters.

Sufyān b. ʿUyayna (d. 198/814).[16] Ibn Jurayj → Sufyān b. ʿUyayna transmissions are cited in the *isnād*s of 291 different hadith texts, which were collected in 110 different books. Transmission in the reverse direction (Sufyān b. ʿUyayna → Ibn Jurayj) occurs in only one hadith. To give a sense of how this transmitter relationship is documented in the *isnād*s of hadiths, we have translated two of them below. The first hadith is one of the 291 with the Ibn Jurayj → Sufyān b. ʿUyayna transmission:

> He [Ibn ʿAbd al-Barr] said: ʿAbd al-Wārith narrated to me, saying: Qāsim narrated to me, saying: al-Khushanī narrated to me, saying: Muḥammad b. Yaḥyā b. Abī ʿUmar narrated to me, saying: **Sufyān b. ʿUyayna** narrated to me about **Ibn Jurayj**: about ʿAṭāʾ that he said: I heard Jābir b. ʿAbdullāh say: "God's messenger . . ."[17]

The transmitter relationships documented in this *isnād* consist of the following:

1. ʿAbd al-Wārith → Ibn ʿAbd al-Barr
2. Qāsim → ʿAbd al-Wārith
3. Al-Khushanī → Qāsim
4. Muḥammad b. Yaḥyā b. Abī ʿUmar → al-Khushanī
5. Sufyān b. ʿUyayna → Muḥammad b. Yaḥyā b. Abī ʿUmar

16. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=3443 for biographical information on the narrator (accessed 2/5/19).

17. Since we are primarily concerned with the *isnād*s cited in a hadith, we do not cite its content (*matn*). For the online version of this hadith, see http://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=&pid=533573&bk_no=1052&startno=68 (accessed 2/4/19). For the version found in *GK*'s desktop software, see hadith number 736 in Ibn ʿAbd al-Barr's *al-Istidhkār*.

6. **Ibn Jurayj → Sufyān b. ʿUyayna**

7. ʿAṭāʾ → Ibn Jurayj

8. Jābir b. ʿAbdullāh → ʿAṭāʾ

In this *isnād*, the sixth transmitter relationship is the one we are interested in: Ibn Jurayj → Sufyān b. ʿUyayna. In the hadith below, the reverse of this relationship, namely Sufyān b. ʿUyayna → Ibn Jurayj, is documented in the *isnād*:

> About **Ibn Jurayj**: about **Ibn ʿUyayna:** about Yaḥyā b. Saʿīd: about someone who heard ʿAmra narrating about ʿĀʾisha that she said: "God's messenger . . ."[18]

The transmitter relationships documented in this *isnād* consist of the following:

1. Ibn Jurayj → ʿAbd al-Razzāq

2. **Ibn ʿUyana → Ibn Jurayj**

3. Yaḥyā b. Saʿīd → Ibn ʿUyayna

4. Someone → Yaḥya b. Saʿīd

5. ʿAmra → Someone

6. ʿĀʾisha → ʿAmra

The fact that the frequent relationship, in this case Ibn Jurayj → Sufyān b. ʿUyayna, occurs in 291 hadiths found in 110 different books, makes the occurrence of its reverse, Sufyān b. ʿUyayna → Ibn Jurayj, which occurs only once, anomalous, leading one to suspect that it may represent an error. The existence of the anomaly raises interesting questions: to what extent are such relationships errors, and to what extent do they represent the historical phenomenon of transmitters trading hadiths? In this case, given the sheer preponderance of the frequent transmitter relationship as opposed to the infrequent one, we surmise that the infrequent one is a mistake. But who made the mistake and at what level? There are two possible scenarios: either *GK* introduced the error,[19] or the mistake is found in *GK*'s source material. In this case, the mistake is found in the published source.[20] It is highly likely that either ʿAbd al-Razzāq, the hadith's collector, the scribes that copied and/or transmitted his work, or the collector's modern editor simply reversed the names that make up the transmitter relationship. Finding such an error in the published text itself illustrates the power of this technique.

The difference between the number of times a transmitter relationship and its reverse are cited in hadiths is often a powerful indicator of an errant relationship. But

18. For the on-line version see: http://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=&pid=26833&bk_no=60&startno=3338 (accessed 2/4/19). For the version found in *GK*'s desktop software, see hadith number 4648 in ʿAbd al-Razzāq's *Muṣannaf ʿAbd al-Razzāq*.

19. *GK* would have made the mistake when constructing the list of historical transmitter relationships represented in the *isnād*, sequentially ordered as simply a list of unique narrator IDs, when *GK* transcribed the hadith from its published or manuscript form into its digitized version.

20. See Abū Bakr ʿAbd al-Razzāq b. Hammām al-Ṣanʿānī, *al-Muṣannaf*, ed. Ḥabīb al-Raḥmān al-Aʿẓamī, 2nd ed., 11 vols., Manshūrāt al-Majlis al-ʿIlmī (Beirut: al-Majlis al-ʿIlmī, 1983), 3:60, which is the specific edition that *GK* relied on for transcription.

unfortunately, many transmissions and their reverses occur in exactly the same number of hadiths. We thus found it useful to divide up the 3,206 bidirectional transmitter relationships into two fundamental types. The first and most populous category, of which the above is an example, denotes transmitter relationships in which one of the two is cited in more hadiths than its reverse. As a measure of the likelihood of error, we calculated the ratio of the transmissions in the more frequent direction to transmissions in either direction. A high ratio means a high likelihood that the infrequent direction is erroneous, while a low ratio suggests both directions are correct. On one end of the spectrum, there is Maʿmar (d. 154/771)[21] → ʿAbd al-Razzāq (d. 211/826),[22] which occurs in 9,695 different hadiths, and its reverse, which occurs just once, with 99.99% of the cases in the frequent direction. On the other end of the spectrum is Muḥammad b. Isḥāq (d. 150/767) → Shuʿba b. al-Ḥajjāj (d. 160/777), for which the frequent and infrequent directions occur respectively in 31 and 33 hadiths, yielding a ratio of 0.516. Figure 2 shows the distribution of the frequency ratios for cases in which the two directions do not occur with equal frequency. It reveals a large number of suspect transmissions. Figure 3 shows the distribution of the death-date differences in the infrequent direction of a bidirectional relationship (where the transmitter's death date is subtracted from the recipient's). In the majority (819) of them, the source/transmitter died after the recipient. The contrast with the differences in death dates represented in Figure 1 above confirms the anomalous nature of the infrequent directions.

The second category consists of relationships in which both directions occur in exactly the same number of *isnād*s. In this category, one cannot rely on the frequency ratio as a diagnostic tool. This category can be divided into two further sub-types: one type, with the majority of the cases, consists of relationships that are cited in exactly one hadith each (454 such cases); the second sub-type consists of relationships in which the two directions occur in an equal number of multiple hadiths (43 such cases).

In order to test the utility of using bidirectional transmissions to discover errors, we have read the hadith texts involved in 958 bidirectional relationships (out of a total of 3,206) to check for errors. Of the 958, we determined that 220 were the result of some type of mistaken reading. Because automatic error detection requires the identification of error types, we classified these 220 mistaken transmitter relationships into four categories: those caused by a mistaken reading of parallel *isnād*s by *GK*; those caused by a suspected misidentification of transmitters in an *isnād* by *GK*; those caused by suspected errors in the sources upon which *GK* relies; and those caused by a misreading of *isnād*s on *GK*'s part not reducible to the first two types. The first two categories seem to be caused by somewhat systematic mistakes made by *GK* in creating the *isnād*s, and can therefore be rectified through automatic correction. The third type of error is especially significant for the identification of errors in published sources upon which many scholars continue to rely.

21. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=7633 for biographical information on the narrator (accessed 2/5/19).

22. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=4533 for biographical information on the narrator (accessed 2/5/19).
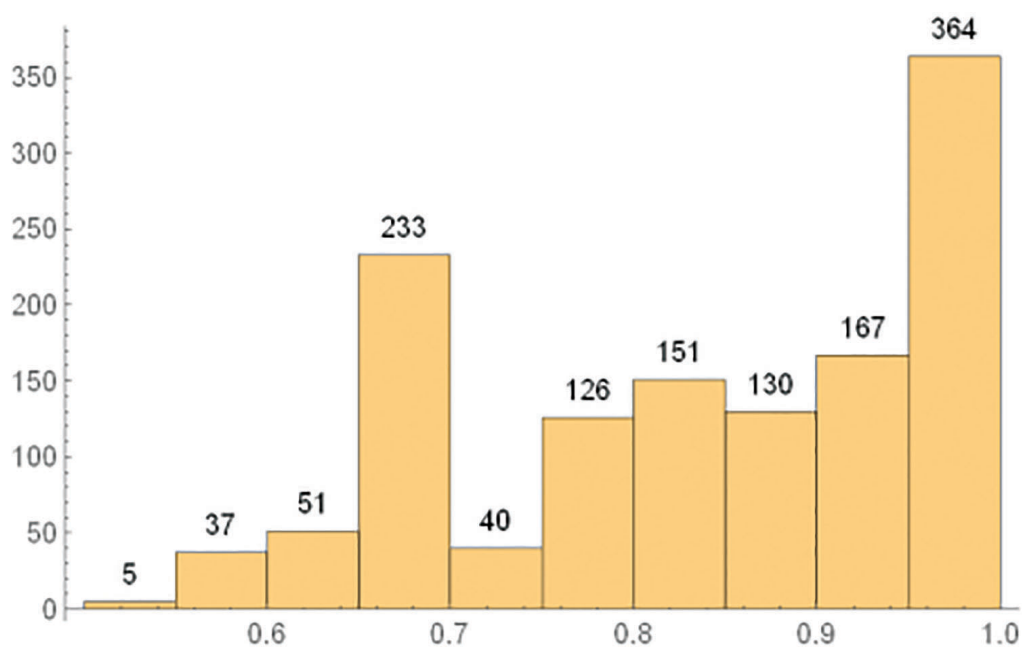
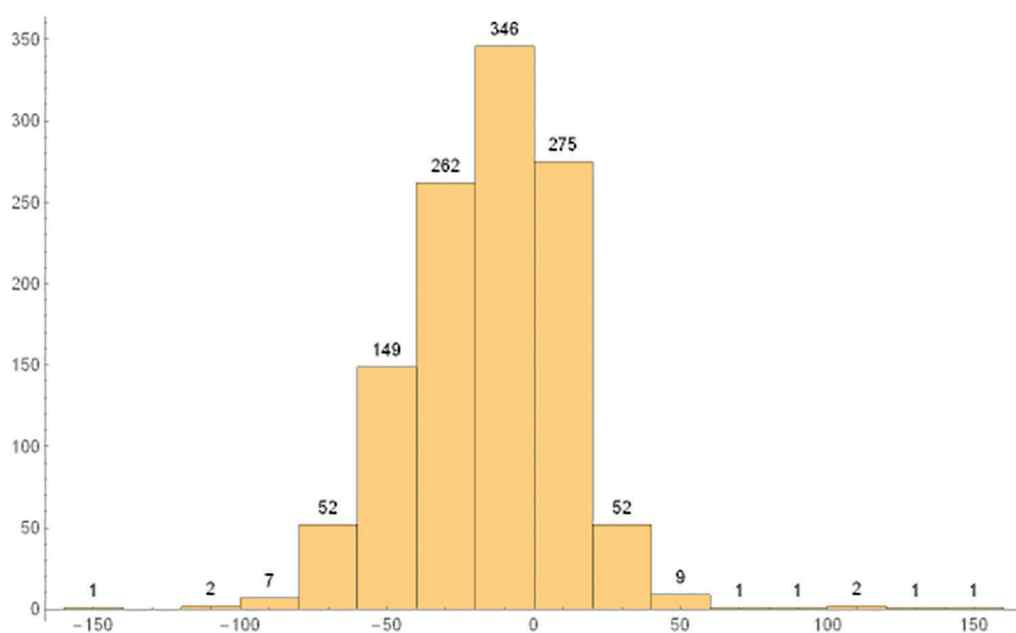FIGURE 2. Distribution of the Frequency Ratios of Bidirectional Transmissions.

FIGURE 3. Distribution of Differences in Death Dates in Infrequent Directions.

TYPES OF MISTAKES FOUND IN *GK* DATA

Parallel *Isnād* Misreading

One type of error has to do with reading a single hadith text that contains multiple *isnād*s, a phenomenon which we label "parallel *isnād*s." The example below concerns this bidirectional

transmitter relationship: ʿAmr b. Dīnār (d. 126/744)[23] → Ibn Jurayj (d. 150/767)[24] and its reverse. The latter direction is cited in only two hadiths, whereas the former is cited in 491. The huge discrepancy signals a possible error. Let us look at the part of the *isnād* that generated *GK*'s erroneous relationship, Ibn Jurayj → ʿAmr b. Dīnār:

> Yūsuf al-Qāḍī *and* Abū Khalīfa al-Faḍl al-Ḥubāb al-Jumaḥī narrated to us saying: Ibrāhīm b. Bashshār al-Ramādī narrated to us: Sufyān b. ʿUyayna narrated to us: ʿ**Amr b. Dīnār** *and* **Ibn Jurayj** narrated to us about ʿAṭāʾ b. Abī Rabāḥ: about Ṣafwān b. Yaʿlā b. Umayya: about his father that he said . . .[25]

This *isnād* indicates four channels of transmission, as noted correctly in *GK*. However, *GK* mistakenly creates the relationship, Ibn Jurayj → ʿAmr b. Dīnār, probably because they simply followed the order of the names as they appeared in the *isnād* without noticing that they were connected with "and." Of the 220 errors, 48 were caused by a misreading of a parallel *isnād*.

Misidentification of Transmitters

35 errors involve mistaken identification of transmitters. We give two examples in which a transmitter is mistakenly assigned the transmitter ID number of someone with a similar name. One example involves the Muḥammad b. Isḥāq al-Ṣāghānī (d. 270/883)[26] → Yazīd b. Hārūn (d. 206/821)[27] relationship and its reverse. The former is cited in one hadith, and the latter in 66. Here is the text of the *isnād* with the sole instance of Muḥammad b. Isḥāq al-Ṣāghānī → Yazīd b. Hārūn:

> Muḥammad b. Maslama narrated to us: **Yazīd b. Hārūn** narrated to us, saying: **Muḥammad b. Isḥāq** said: ʿAbdullāh b. Abī Najīḥ narrated to me about Mujāhid: about Ibn ʿAbbās that he said . . .[28]

The *isnād* simply names Muḥammad b. Isḥāq. This refers to Muḥammad b. Isḥāq al-Yasār (d. 150/767, *GK* narrator ID 6811),[29] but *GK* mistakes him for Muḥammad b. Isḥāq al-Ṣāghānī (*GK* narrator ID 6807). First, *GK* rightly identified the correct

23. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=6123 for biographical information on the narrator (accessed 2/5/19).

24. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=5223 for biographical information on the narrator (accessed 2/5/19).

25. The hadith can be found here in the online version: http://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=&pid=289237&bk_no=477&startno=54 (Accessed 2/4/19). In the desktop version, it is hadith number 18147 in al-Ṭabarānī's *al-Muʿjam al-Kabīr liʾl-Ṭabarānī*.

26. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=6807 for biographical information on the narrator (accessed 2/5/19).

27. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=8488 for biographical information on the narrator (accessed 2/5/19).

28. The hadith can be found here in the online version: http://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=&pid=670037&bk_no=4145&startno=52 (Accessed 2/4/19). In the desktop version, it is hadith number 51 in Muḥammad b. al-ʿAbbās b. Najīḥ al-Bazzāz's *al-Thānī min Ḥadīth Abī Bakr Muḥammad b. al-ʿAbbās b. Najīḥ al-Bazzāz*.

29. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=6811 for biographical information on the narrator (accessed 2/5/19).

Muḥammad b. Isḥāq (6811) in 741 other hadiths in which he is cited as Yazīd's source. Second, given that Yazīd died in 206/821, it is much more likely that Muḥammad b. Isḥāq (6811) who died in 150/767 is Yazīd's source, as opposed to Muḥammad b. Isḥāq al-Ṣāghānī (6807) who died 61 years after Yazīd.

The second example concerns ʿAbdullāh b. Aḥmad b. Ḥanbal (d. 290/903)[30] → Ḥanbal b. Isḥāq (d. 273/886)[31] and its reverse. *GK* has mis-read Abū ʿAbdullāh (GK narrator ID 488) as ʿAbdullāh (GK narrator ID 4657), confusing father and son, in the following *isnād*:

> Abū al-Ḥusayn b. Bishrān informed us, saying: Abū ʿAmr b. Sammāk informed us, saying: **Ḥanbal b. Isḥāq** narrated to us, saying: **Abū ʿAbdullāh** informed us, saying: ʿAbd al-Razzāq narrated to us, saying: Ibn Jurayj informed us, saying: I was informed that . . .[32]

While the transmitter relationship ʿAbdullāh b. Aḥmad b. Ḥanbal → Ḥanbal b. Isḥāq does occur in two other hadiths, the relationship Aḥmad b. Ḥanbal [or Abū ʿAbdullāh] → Ḥanbal b. Isḥāq occurs in 75 other hadiths, indicating that *GK* has misidentified the transmitter in this *isnād*.

Of the 220 errors we detected, 35 were cases of confirmed or suspected misidentified narrators. The 928 transmitter relationships for which we manually read the relevant hadith texts contained 535 unique transmitters, 30 of them misidentified. That is an error rate of 5.6%. Since bidirectional relationships are inherently suspect, the misidentification rate for the entirety of the *GK* dataset, and not for bidirectional transmissions alone, will be much lower.

Misreadings Originating from Mistakes in GK Sources

There are mistakes that originate not from *GK*'s processing of the *isnād*s, but from the sources that *GK* relied on. We found, by checking our results against printed editions on which *GK* relies, that the following procedure allowed programmatic identification of such errors. We first considered a few statistical measures. The most important one was the frequency ratio of frequent and infrequent directions of each bidirectional relationship. A high ratio indicates that the infrequent direction is erroneous. This suspicion is strengthened if the same narrative has come down to us through different *isnād*s, one with the frequent relationship and the other through its infrequent reverse. In addition, the death dates of transmitters, when available, were used to check the plausibility of the infrequent direction. We give two examples of confirmed source mistakes below.

---

30. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=4657 for *GK*'s biographical information on the narrator (accessed 2/5/19).

31. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=15973 for *GK*'s biographical information on the narrator (accessed 2/5/19).

32. This hadith can be found in the following link in the on-line version: http://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=&pid=333775&bk_no=681&startno=3317 (accessed 2/5/19). In the desktop version, this is hadith number 3230 in al-Bayhaqī's *Dalāʾil al-Nubuwwa li-l-Bayhaqī*.

The transmission Sufyān al-Thawrī (d. 161/778)[33] → ʿAbd al-Raḥmān b. Mahdī al-ʿAnbarī (d. 198/814)[34] is recorded in 1,875 different hadiths and its reverse in just one. Here is the *isnād* with the sole reverse direction:

> Al-Ḥasan b. al-Ḥusayn b. al-ʿAbbās al-Niʿālī informed us, saying: Abū Ḥafṣ ʿUmar b. Muḥammad b. ʿAbdullāh b. Aḥmad, known as Ibn Qayyūm al-Muʿaddal al-Nahrawānī informed us of it in the year 362, saying: Abū Bakr Muḥammad b. Ḥamdān b. Baghdād al-Ṣaydalānī narrated to us in Baghdad, saying: Isḥāq b. Muḥammad b. al-Muthannā narrated to us, saying: my father narrated to us, saying: Yaḥyā b. Saʿīd narrated to me about **Sufyān al-Thawrī**: about ʿ**Abd al-Raḥmān b. Mahdī**: about Sufyān b. ʿUyayna: about ʿAmr b. Dīnār: about Jābir, that he said that . . .[35]

GK transcribed and parsed this *isnād* correctly; it is faithful to the printed edition that *GK* relied on.[36] But the purported transmission, ʿAbd al-Raḥmān b. Mahdī → Sufyān al-Thawrī, is suspect, not only because it occurs dramatically less frequently than its reverse, but also because the reverse (more frequent) direction is attested in eight *isnāds* carrying the same narrative that the above *isnād* does. Here is one of the eight:

> Abū al-Fatḥ b. al-Ikhshīd informed us: Muḥammad b. Aḥmad b. Muḥammad informed us: ʿAlī b. ʿUmar al-Ḥāfiẓ informed us: Muḥammad b. Dāwūd b. Sulaymān al-Nīsabūrī narrated to me: Aḥmad b. Maḥmūd b. Muqātil al-Harawī and Muḥammad b. ʿUmayr al-Rāzī both said: Muḥammad b. Ḥammād al-Maṣṣīṣī narrated to us in Ramla: Ibrāhīm b. ʿUthmān b. Ziyād al-Maṣṣīṣī narrated to us: Ibrāhīm b. Saʿīd al-Jawharī narrated to us: Yaḥyā b. Ḥassān narrated to us about ʿ**Abd al-Raḥmān b. Mahdī**: about **Sufyān b. Saʿīd al-Thawrī**: about Yaḥyā b. Saʿīd al-Qaṭṭān [that he said]: Sufyān b. ʿUyayna narrated to us about ʿAmr b. Dīnār: about Jābir b. ʿAbdullāh, may God be pleased with the two of them, that he said . . .[37]

The second example not only identifies an error in a published source, but also has the virtue of confirming the accuracy of one manuscript over another source on which the modern editor relied. The bidirectional transmission relationship is Masrūq (d. 62/682)[38] → Shaʿbī (d. 105/724),[39] which is recorded in 925 hadiths, and its reverse, which occurs

33. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=3436 for biographical information on the narrator (accessed 2/5/19).

34. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=4493 for biographical information on the narrator (accessed 2/5/19).

35. The hadith may be found on-line: http://library.islamweb.net/hadith/display_hbook.php?bk_no=717&pid=363453&hid=3774 (accessed 2/4/19). For the desktop version, see hadith number 3774 in al-Khaṭīb al-Baghdādī's *Taʾrīkh Baghdād*.

36. For the published version of the source see Abū Bakr Aḥmad b. ʿAlī al-Khaṭīb al-Baghdādī, *Taʾrīkh Madīnat al-Salām*, ed. Bashshār ʿAwwād Maʿrūf, 1st ed., 17 vols. (Beirut: Dār al-gharb al-Islāmī, 2001), 13: 112-13.

37. For the on-line version see the following link: http://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=&pid=540287&bk_no=1377&startno=267 (accessed 5/18/19). For the desktop version see hadith number 214 in Muḥammad b. Abī Bakr b. Abī ʿĪsā al-Madīnī's *Kitāb al-Laṭāʾif min ʿUlūm al-Maʿārif*.

38. See https://library.islamweb.net/hadith/RawyDetails.php?RawyID=7430 for biographical information on the narrator (accessed 5/18/19).

39. See https://library.islamweb.net/hadith/RawyDetails.php?RawyID=4099 for biographical information on the narrator (accessed 5/18/19).

only once. Moreover, the hadith narrative to which the anomalous *isnād* is attached has also come down to us via two *isnāds* that instantiate the frequent direction. Finally, the 43-year difference in the two transmitters' death dates greatly favors the frequent direction. Therefore, the infrequent direction is clearly mistaken. The mistake is found in the published edition of Ibn ʿAsākir's (d. 571/1176) *Taʾrīkh Madīnat Dimashq* that *GK* relied on. Here is the text of the *isnād*:

> Abū al-Qāsim b. al-Ḥusayn informed us: Abū Ṭālib b. Ghaylān informed us: Abū Bakr al-Shāfiʿī narrated to us: Muḥammad b. Ghālib narrated to us: ʿAbd al-Ṣamad b. al-Nuʿmān narrated to me: Shaybān narrated to us: about ʿĀṣim: about **Masrūq:** about **Shaʿbī:** ʿAbdullāh b. Jaʿfar narrated to me, saying: . . .[40]

In the footnote on Masrūq in the published edition, the editor, al-ʿUmrawī, notes that two manuscripts have Shaʿbī as receiving the hadith from Masrūq, but an earlier partial edition of the same work has Masrūq transmitting it from Shaʿbī. Our analysis confirms the reading of the earlier partial edition over that of the two manuscripts on which the editor relied. Of the 220 mistakes we have found, we have determined that 22 originate in a source that *GK* relied on.

Other Misreadings

The final category includes misreadings of hadiths that are not caused by identifiable systematic errors, such as those related to reading parallel *isnād*s or the misidentification of transmitters. We give one example of such a misreading.

The transmission, ʿUrwa b. al-Zubayr (d. 94/713)[41] → Hishām b. ʿUrwa (d. 145/762),[42] occurs in 8,570 hadiths, and its reverse only in one. The lone version represents a *GK* misreading of the relevant part of the *isnād*:

> . . . Ḥammād b. Salama narrated to us: **Hishām b. ʿUrwa** narrated to us: about his father [**ʿUrwa b. al-Zubayr**]: about ʿUthmān b. Ṭalḥa . . .[43]

There is nothing wrong with the text of the *isnād* itself, nor is this a case of misidentification. It conforms to the frequent version of the transmitter relationship, ʿUrwa b. al-Zubayr → Hishām b. ʿUrwa. However, in parsing this *isnād*, *GK* mistakenly reversed the above relationship, listing the two transmitter IDs in the wrong order. This category of mistakes represents 150 of the 220 errors we discovered.

40. See ʿAlī b. al-Ḥasan Ibn ʿAsākir, *Tārīkh Madīnat Dimashq*, ed. Muḥibb al-Dīn Abī Saʿīd ʿUmar b. Gharāma al-ʿUmrawī, 80 vols. (Beirut: Dār al-fikr, 1996), 27:258, for the hadith, and footnote 2 for the editor's comments. For the on-line version, see the following link: https://library.islamweb.net/hadith/display_hbook.php?indexstartno=0&hflag=1&pid=383485&bk_no=798&startno=27 (accessed 5/18/19). For the desktop version see hadith number 27566 in Ibn ʿAsākir's *Madīnat Dimashq*.

41. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=5594 for biographical information on the narrator (accessed 2/5/19).

42. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=8055 for biographical information on the narrator (accessed 2/5/19).

43. The link to the hadith in the on-line version may be found here: http://library.islamweb.net/hadith/display_hbook.php?bk_no=636&pid=325259&hid=4508 (accessed 2/5/19). In the desktop version, see hadith number 4508 in Abū Nuʿaym al-Iṣfahānī's *Maʿrifat al-Ṣaḥāba*.

True Bidirectional Transmission

We used the fact that bidirectional transmission is anomalous to explore a portion of the dataset that we thought would be especially prone to errors. Of the nearly 1,000 transmitter relationships we investigated, we suspect that 220 were mistakenly created based on some type of erroneous reading of the *isnād*s. However, we also found several cases where hadith narrators probably did trade hadiths with each other. This scenario seems especially plausible when the number of hadiths that cite a given transmitter relationship are roughly equal to those that cite its reverse. For example, the relationship Muḥammad b. Isḥāq (d. 150/767)[44] → Shuʿba b. al-Ḥajjāj (d. 160/777)[45] occurs in 31 hadiths, and its reverse occurs in 33. This, combined with the fact that their death dates are close, makes it plausible that Muḥammad b. Isḥāq would cite Shuʿba as his source in some hadiths, whereas Shuʿba would cite Muḥammad b. Isḥāq as his in others, a fact that is confirmed in the biographical literature.[46]

PROGRAMMATIC RECOMMENDATIONS FOR DEALING WITH MISTAKEN TRANSMITTER RELATIONSHIPS

The sorts of mistake we have encountered can be identified and corrected automatically by means of programs applying a set of rules based on different metrics, but any such program will have an error rate. For research projects in which no error is acceptable, the scholar may use programs to flag potentially problematic cases; but he/she will still have to research each case manually to confirm an error and correct it. But for many digital humanities applications, those involving analyses of aggregates of information using the methods of statistics or social network analysis, an automatic error-detection-and-correction procedure that itself suffers from an error rate is acceptable as long as it leaves the dataset in an overall better shape than before. For example, a procedure that correctly removes fifty errors from the database but introduces five new ones of its own may be a valuable tool. Moreover, some types of errors may be more acceptable than others.

The Case of Preponderant Transmitter Relationships

Bidirectional relationships in which transmission in one direction occurs in more *isnād*s than in the other are especially ripe for programmatic solutions. One solution would be to reverse the relationship that is less preponderant. This seems to be a safe solution in cases where the discrepancy between the infrequent and frequent relationships is especially large. One may, for example, reverse those relationships that have a ratio larger than 0.8: out of 1,361 bidirectional transmitter relationships that fall into this type, 812

---

44. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=6811 for biographical information on the narrator (accessed 2/5/19).

45. See http://library.islamweb.net/hadith/RawyDetails.php?RawyID=3795 for biographical information on the narrator (accessed 2/5/19).

46. Moreover, al-Mizzī, a premodern hadith scholar, records both scholars as receiving hadiths from each other in both of their biographical notices. In Shuʿba's entry, in the section on the people who narrated from him, he gives Muḥammad b. Isḥāq's name and confirms that he is also one of Shuʿba's teachers. See the *Tahdhīb al-Kamāl* tab in biographical information for each narrator. See footnotes 42 and 43 above for the url.

meet this criterion. The rest may require some other criteria or manual investigation of the relevant *isnād*s.

Another option is to reverse those infrequent relationships that occur in hadiths that have otherwise come down to us through *isnād*s that confirm the frequent counterpart relationship. This refers to cases where the same narrative has multiple *isnād*s, one of them attesting the infrequent direction while one or more of the others attest the frequent one. There are 366 cases where the infrequent relationship is thus weakened by the corroboration of the frequent direction. 84 of them have frequency ratios less than 0.8 and greater than 0.5. These represent an extra 84 cases to be reversed even though their frequency ratios are less than 0.8—84, that is, on top of the 812 eliminated thanks to frequency ratios of 0.8 or above.

Ultimately, how the researcher chooses such thresholds or decides how to treat cases that fall in gray zones may depend on the purpose at hand. For example, for some applications in social network analysis, we have found that failing to remove false bidirectional relationships is much more damaging than mistakenly removing true ones, so we adjust our procedures to err on the side of eliminating bidirectionality. ▪