A Quantitative Analysis of Schacht's Hadith Backgrowth Theory

Alex Ditzel
aditzel@ucdavis.edu

Donald Pinckney
djpinckney@ucdavis.edu

May 18, 2017

UC Davis Department of Religious Studies
Undergraduate Writing Prize Submission

Professor Mairaj Syed

A Quantitative Analysis of Schacht's Hadith Backgrowth Theory

ALEX DITZEL and DONALD PINCKNEY, University of California, Davis

1. INTRODUCTION

Hadith are reports of something the Islamic Prophet Muhammad said, did, or was seen saying or doing. They were passed down orally for centuries before being written down and are not a part of the Quran, but they do hold significant weight in Islamic society and even in legal rulings because of the way that they show what is just through the example of the Prophet. These reports consist of two parts: the matn, which is the story about the Prophet which is being passed on, and the isnad, which is a chain of narrators who communicated the story until it was written down. The matn is important for legal scholars and theologians, but the isnad can be of particular interest to specialists in the field of hadith, as it provides a mechanism to possibly ensure that hadith are legitimate. Falsified hadith are a well documented phenomenon and there are many possible reasons why falsification could occur including the desire for a certain law or moral value supported by the falsifier.

In the 20th century Joseph Schacht developed and demonstrated his theory of isnad backgrowth for some hadith. Backgrowth is when a narrator is added to the beginning of an already established isnad to connect it directly to the Prophet and legitimize it. Schacht postulated that originally many hadith had isnads which did not go all the way back to the Prophet, but were used as the basis of legal rulings in the ancient Islamic world. This was initially fine, however, significantly after the death of the Prophet a new legal movement was created which altered the standards of which hadith were considered valid for legal rulings. Only hadith with isnads that could be traced back to the Prophet directly were considered valid. This caused problems with laws which relied upon hadith that had incomplete isnads, so in order to prevent the upheaval of many laws there was a backgrowth of the isnads where an earlier narrator was added to lengthen

an isnad and allow it to reach back completely to the Prophet. Schacht speculated that this phenomenon was widespread, but he had no way of knowing how common it was beyond manually checking many hadith. However, using databases of hadith data and modern data analysis techniques we were able to estimate how many hadith may fit a suspicious pattern similar to what Schacht described.

Schacht demonstrated backgrowth in his book The Origins of Muhammadan Jurisprudence. In one example he determined that Muwatta i. 371: "Malik-Muhammad b. "Abdalrahman b. Sa'd b. Zurara - Hafsa killed a mudabbar slave of hers who had bewitched her" was backgrown from Muwatta Shaib. 359 and Tr. III, 93 which read "Malik - Abdul-Rijal Muhammad b. 'Abdalrahman [b. Jariya] - his mother 'Amra - 'A'isha sold a mudabbar slave of hers who had bewitched her." He says that one is clearly modeled on the other and that the isnads were backgrown and fabricated after the hadith was put into circulation in the generation before Malik.

2. METHODS

In order to perform our analysis there were two main tasks which needed to be completed. The hadith needed to be sorted into groups of hadith which have similar enough matn to each other to count as different versions of the same hadith and then the isnads within those groups needed to be examined for any patterns indicative of backgrowth. First, we utilized data from islamweb.net to find groups of hadith with similar matn, and then we used custom Mathematica code to search for isnad patterns of backgrowth.

2.1 Similar Hadith Grouping

The groups of similar hadith is based on data available on islamweb.net. We started with a SQLite database which contains both isnad data and islamweb.net links for a large dataset of hadith across multiple book collections. The isnad data is used for backgrowth detection (see below), while the islamweb.net links are used for hadith groups. Each hadith in the database contains a URL that links to a list of hadith with similar matn, for example: http://library.islamweb.net/hadith/hadithServices.php?type=1&cid=2963&sid=720.

This page lists the hadith which have similar matn to the hadith associated with that URL. So, to determine all of the groups, we needed to systematically open every URL, scrape data from each of the pages, and use this to form the groups. Since there are over 600,000 hadith in our database, this is clearly infeasible to do manually, so we created a custom web scraper to complete this task automatically.

We wrote the web scraper in Python, which is commonly regarded as one of the most convenient programming languages for doing miscellaneous tasks such as web scraping. Specifically, we used the lxml (http://lxml.de) Python toolkit to easily perform the parsing of the HTML of the islamweb.net web pages. Then, the Python code exported the computed groups to the JSON file format (http://www.json.org), which Mathematica can easily read.

Finally after importing the group data into Mathematica, we can view the results of the web scraping. For example, the very first group scraped and imported into Mathematica is shown in Figure 1. Note that the `book_number` and `hadith_number` fields are only for identifying information.

```
{
    <|"book_number" -> 19, "hadith_number" -> 472|>,
    <|"book_number" -> 121, "hadith_number" -> 24689|>,
    <|"book_number" -> 121, "hadith_number" -> 24737|>,
    <|"book_number" -> 121, "hadith_number" -> 25613|>,
    <|"book_number" -> 22, "hadith_number" -> 169|>,
    <|"book_number" -> 68, "hadith_number" -> 251|>
}
```

Fig. 1. Sample group of hadith by matn data from islamweb.net

However, just because two hadith are grouped together does not mean that there is backgrowth. For example, both the hadith (book ID 121, hadith ID 128) and (52, 28) are in the same group. The text for (121, 128) is:

> 'Affan narrated to me: Hammam narrated to me: Qatada narrated to me: Abu al- 'Aliya narrated to me: from ('an) Ibn 'Abbas, who said: "A pleasing people, among whom was 'Umar, and the most well-pleasing of whom in my eyes is 'Umar, witnessed that the God's Messenger said: "there is no salat after daybreak (subH), until the sun has [fully] appeared, and there is no salat after 'asr until the sun has [fully] set." (Syed)

Clearly the isnad for (121, 128) is:

> 'Umar → Ibn 'Abbas → Abu al- 'Aliya → Qatada → Hammam → 'Affan

And the text for (52, 28) is:

> Hammam narrated to us: from Qatada: from Abu al-'Aliya: from Ibn 'Abbas, who said: "A pleasing people, among whom was 'Umar, and the most well-pleasing of whom in my eyes is 'Umar, witnessed that the God's Messenger forbade salat after 'Asr until the sun [fully] sets, and salat after the morning prayer until the sun [fully] appears." (Syed)

Which has an isnad of:

> 'Umar → Ibn 'Abbas → Abu al- 'Aliya → Qatada → Hammam

Clearly there is no backgrowth before 'Umar between these two hadith. Therefore, similar groups do not guarantee backgrowth, and so we need a custom method for detecting backgrowth within these groups.

2.2 Backgrowth Detection

To determine if backgrowth was likely to have occurred we considered two main factors: an overlap in the isnads and that overlap being present only at the beginning of one of the isnads. As an example to illustrate the algorithm, consider the isnads in Figure 2. Since they overlap with $B \rightarrow C \rightarrow D$, which occurs at the beginning of one but not the second, the algorithm classify these isnads as suspicious.

In addition, even if the ends of the isnads differ, the algorithm can still detect suspicious isnads. It would also consider the isnads in Figure 3 suspicious.

$$B \to C \to D \text{ (Before backgrowth)}$$

$$A \to B \to C \to D \text{ (After backgrowth)}$$

Fig. 2. Sample isnads before and after backgrowth

$$B \to C \to D \to E \text{ (Before backgrowth)}$$

$$A \to B \to C \to D \to F \text{ (After backgrowth)}$$

Fig. 3. Sample isnads before and after backgrowth

Note that it is crucial that the overlap occurs at the beginning of one isnad and *not* at the beginning of the other. Thus, the following pair in Figure 4 is *not* classified as suspicious, due to the fact that the overlap is at the beginning of both isnads.

$$A \to B \to C \to D$$

$$A \to B \to E \to F$$

Fig. 4. Sample isnads with common prefix but not backgrowth

In order to search within the groups of hadith with similar matn and find suspicious isnad pairs as described above, we utilized a relatively simple algorithm shown in Figure 5 to sort and compare the isnads.

The algorithm iterates through each group of similar hadith found and creates a new set, `isnadPairs`, which contains the identity of any hadith that match, the length of the overlap, and the isnads of each of the two hadith. See the comments embedded in the code for details on the role of each part of the function. This algorithm was applied to each group of similar hadith and the collection of the `isnadPairs` for each group was collected.

```
DoBackgrowths[hadithAssoc_] := (isnadPairs = {};
    hadithKeys = Keys[hadithAssoc];
    (* Creates a set of keys for the groups of hadith examined so they can be identified
    later in the algorithm *)
    For[hadOne = 1, hadOne <= Length[hadithAssoc], hadOne++,
    (* Iterates through the isnad of the first hadith *)
        For[hadTwo = 1, hadTwo <= Length[hadithAssoc], hadTwo++,
        (* Iterates through the isnad of the second hadith *)
            If[hadithAssoc[[hadOne]] == {} || hadithAssoc[[hadTwo]] == {}
                || hadOne == hadTwo || hadithAssoc[[hadOne, 1]] == hadithAssoc[[hadTwo, 1]],
                Continue[],
            (* Skips comparison if one isnad is empty or they both have the same narrator *)

            For[hadOnePos = 1; hadTwoPos = 1; matchCounter = 0,
                hadTwoPos <= Length[hadithAssoc[[hadTwo]]], hadTwoPos++,
                If[hadithAssoc[[hadOne, hadOnePos]] == hadithAssoc[[hadTwo, hadTwoPos]]
                    && hadTwoPos == Length[hadithAssoc[[hadTwo]]],
                    AppendTo[isnadPairs,
                        {hadithKeys[[hadOne]], hadithKeys[[hadTwo]], ++matchCounter,
                        hadithAssoc[[hadOne]], hadithAssoc[[hadTwo]]}]; Break[], Nothing]
                    (* Appends the pair to isnadPairs if the loop ends without
                    any differences, but there was at least one match earlier *)
                    If[hadithAssoc[[hadOne, hadOnePos]] == hadithAssoc[[hadTwo, hadTwoPos]],
                        matchCounter++;
                        If[Length[hadithAssoc[[hadOne]]] > hadOnePos,
                            hadOnePos++,
                            Continue[]];
                        Continue[],
                    If[matchCounter >= 1,
                        AppendTo[isnadPairs,
                            {hadithKeys[[hadOne]], hadithKeys[[hadTwo]], matchCounter,
                            hadithAssoc[[hadOne]], hadithAssoc[[hadTwo]]}]; Break[]
         (* Appends the pair to isnadPairs if the currently examined narrators don't match,
         but there was a match earlier between them *)
         ]]]]]];
  isnadPairs)
```

Fig. 5. Suspicious isnad comparison algorithm.

3. RESULTS

First, we obtained some interesting results about the islamweb.net similar hadith groups. In particular, the sizes of the groups are interesting to analyze. Figure 6 displays a histogram of the sizes of the group. We can clearly see that the size of a group drops of fairly rapidly. In fact, statistical analysis in Mathematica shows that the group size most closely follows a negative binomial distribution with parameters NB(2, 0.405234).

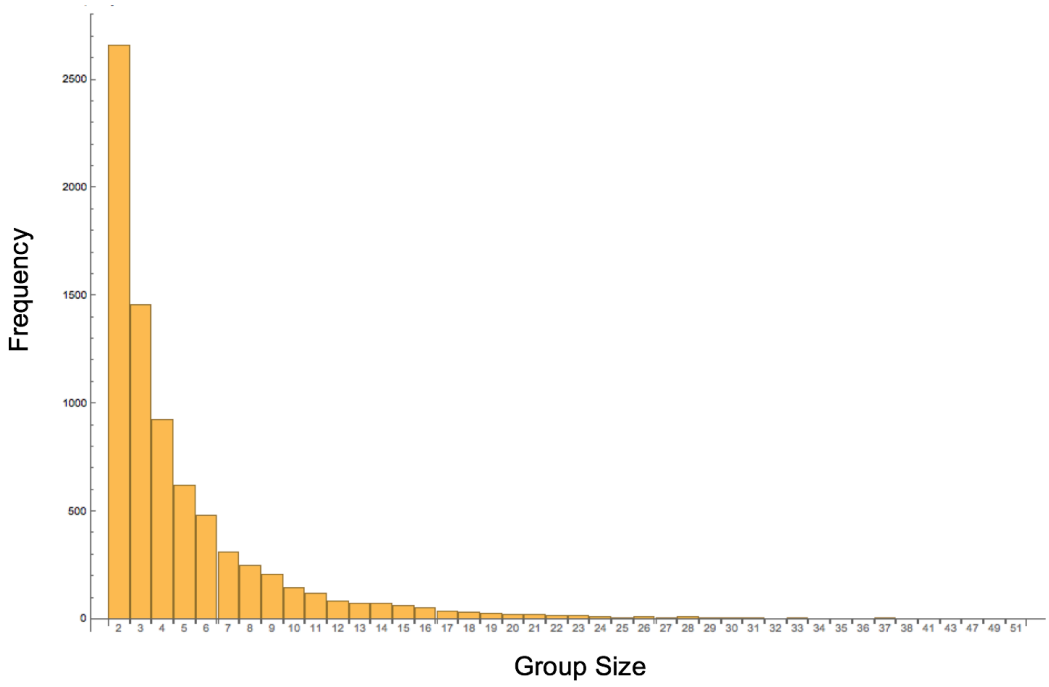

Fig. 6. Histogram of sizes of similar matn groups

The next significant result is the actual percentage of backgrown hadith. Out of a total of 14,519 total hadith analyzed, our method found 946 cases of possibly suspicious backgrowth, which gives an estimated 6.51% backgrowth rate, as shown in Figure 7. However, as will be discussed below a variety of issues with our method cause 6.51% to be an estimated upper bound on the actual backgrowth percentage. Some of the detected cases of backgrowth are not actually backgrowth, so the rate of actual backgrowth is likely less than 6.51%.

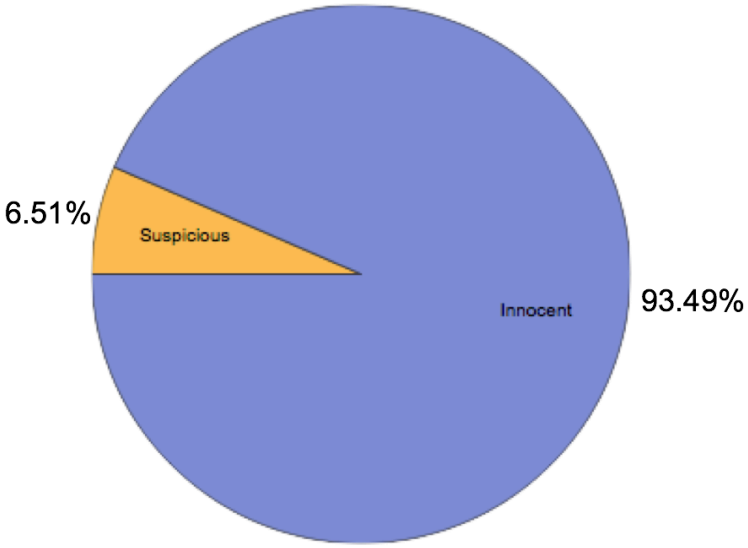

Fig. 7. Frequency of suspicious hadith

Certainly deciding backgrowth can be somewhat ambiguous. For example the hadith (60, 1057) has the matn:

> That a woman asked the Prophet a question: “O God’s messenger, does a woman see in dreams what a man sees?” Umm Salama interjected: “does a woman have wet dreams?”. He replied: “Yes, through it her son resembles her.” (Syed)

In addition, the hadith (19, 114) has the matn:

> That Umm Sulaym asked the Prophet: “Do women see in dreams things similar to what men see, such that they would need to do the ritual bath?” God’s Messenger replied: “Yes, she must do the ritual bath.” ᶜĀʾisha said to her: “You’re kidding me, do women really see that?” God’s Messenger replied: “May your oath turn to dust, how else is there resemblance?” (Syed)

Clearly, these two hadith are related by matn content. In the first hadith, an unnamed woman asks the Prophet a question, while the second hadith names this woman as Umm Sulaym. This retroactive naming might be classified as backgrowth in Schacht's theory.

There were a number of problems in our data and methods which could have led to inaccuracy in our results. The first and perhaps most serious concern is the quality of the hadith database that we used. There are examples of hadith with multiple isnad where they are considered to only have one isnad, but they are appended to each other in the database, for example the hadith (121, 267) is considered to have a single isnad, with the format of narrator id 1 $\rightarrow$ narrator id 2 $\rightarrow \cdots$:

$$4883 \rightarrow 2949 \rightarrow 6458 \rightarrow 18 \rightarrow 5653 \rightarrow 5913 \rightarrow 4883 \rightarrow 2949 \rightarrow 6458 \rightarrow 8097 \rightarrow 5653$$

This clearly does not make sense, narrators 4883, 2949, and 6458 are repeated and it seems to really be two isnads:

$$4883 \rightarrow 2949 \rightarrow 6458 \rightarrow 18 \rightarrow 5653$$

$$5913 \rightarrow 4883 \rightarrow 2949 \rightarrow 6458 \rightarrow 8097 \rightarrow 5653$$

The fact that they are not differentiated in the database may be indicative of other minor problems in the data. Additionally, in the same hadith there seems to be two different isnads that could demonstrate backgrowth due to the overlap and additional narrator at the beginning of one of them. However, examination of the actual text yields a different result: the first isnad is incorrectly documented and narrator 5913 ('Umar) was mistakenly dropped from the isnad. In that isnad narrator 4883 (Ibn 'Abbas) narrates the story in a way that narrator 5913 could be interpreted as either a participant or a narrator of the story. The text of the hadith as translated by Mairaj Syed is displayed below:

> 'Affan narrated to me: Hammam narrated to me: Qatada narrated to me: Abu al- 'Aliya narrated to me: from ('an) Ibn 'Abbas: "A pleasing people, among whom [*one time 'Affan said: a pleasing people, the most well-pleasing of whom in my eyes was 'Umar] was 'Umar, narrated to me that the God's Messenger said: "there is no salat after two salats: after day break until the sun has [fully] appeared, and after 'asr until the sun has [fully] set." (Syed)

This incorrect documentation of the isnad by islamweb.net makes it possible that there were other errors in the data used. This could result in either incorrect categorization of a hadith that does not show backgrowth as suspicious, as is shown above, or the categorization of a hadith that underwent backgrowth as being not suspicious, if the oldest narrator was not documented on the original version of a backgrown hadith. This leads to some skepticism of the quality of the source's documentation of the isnads.

Additionally, there is reason to believe that there could be issues with the categorization of the hadith into groupings of variants of the same hadith. The groups produced by the website contain no more than 100 hadith per group. And the distribution looking at all books (not only the ones examined in this paper) shows that there are significantly more groups with exactly 100 members than would be expected. This causes a concern of technical limitations or some other outside factor possibly limiting the size of the groups and making them smaller than they otherwise would be. Additionally after consultation with Professor Mairaj Syed he expressed concerns over the judgement used to determine similarity of hadith by the website when he examined individual hadith we suspected of backgrowth. Together, these things bring additional skepticism of the groups created by the website, which would influence our results.

Other limitations restrict us to categorizing suspicion rather than definitive detection of backgrowth here too. One is that we can not automatically date the hadith in the pairs, so it is possible that the longer chain came first, which would not be an example of backgrowth.

There are a couple future directions this research could be taken to expand it's scope and improve its accuracy. The books analyzed could be expanded from the selected texts that we looked at to include other books, potentially expanding the amount of backgrowth found. The algorithm used to detect suspicious isnads could be refined to filter out more unwanted inclusions and include other hadith that we may have missed with our current version. The grouping and isnad formation could be done again through the use of natural language processing (NLP) to detect the narrators in an isnad and to group the texts by similarity of matn. However, this is very difficult as the texts have not been translated into English and NLP in the Arabic language is not well developed, resulting in less accurate results than would be desired. Ideally this would allow us to create a new data set that does not rely on islamweb's possibly problematic methodology.

4. CONCLUSION

Overall, computational methods of using Mathematica to evaluate prevalence of Schacht's backgrowth theory in hadith literature were highly effective. By using similar hadith group data from islamweb.net and a custom isnad comparison algorithm in Mathematica, we were able to objectively give an estimated upper bound of a backgrowth rate of 6.51%. While this is not a negligible amount of backgrowth, it does question the supposed widespread phenomenon of backgrowth claimed by Schacht. However, in this work we can not strongly reject Schacht's claim: as discussed above a variety of issues caused us to obtain only an approximate upper bound of 6.51%.

Bibliography

Brown, Jonathan. Hadith: Muhammad's Legacy in the Medieval and Modern World. Oxford: Oneworld, 2011. Print.

Islamweb.net. Government of Qatar, February 2017. www.islamweb.net Lutz, Mark. Programming Python. 3rd ed.: O'Reilly Media, 2006. Print.

McMahon, David, and Daniel M. Topa. A Beginner's Guide to Mathematica.: Chapman & Hall/CRC, 2006. Print.

Mitchell, Ryan. Web Scraping with Python : Collecting Data from the Modern Web.: O'Reilly Media, 2015. UC Davis University Library. Web.

Schacht, Joseph. "A Revaluation of Islamic Traditions." Studies on Islam. Web.

Schacht, Joseph. The Origins of Muhammadan Jurisprudence. Oxford: Clarendon, 1967. Print.

Smith, Ben. Beginning JSON:. Berkeley, CA: Apress, 2015. UC Davis University Library. Web.

Syed, Mairaj. Personal Interview. March 2017

Weinberg, Paul N., James R. Groff, and Andrew J. Oppel. SQL, The Complete Reference. New York: McGraw-Hill, 2010. Print.

Wolfram, Stephen. An Elementary Introduction to the Wolfram Language.: Wolfram Media, 2016. Print.