This paper features my first application of KITAB’s text reuse detection methods to a manuscript transcription. It was delivered Nov. 22, 2015 at the Middle East Studies Annual Meeting in Denver for a panel entitled “Imagining the Ummah.” The panel papers focused on the composition, meaning, and reception of sources pertaining to the first centuries of the Muslim community (600-1000). The paper’s contents (which reflect oral delivery, the audience, and the 20 minute time limit) will be revised, expanded, and incorporated into my 2017 University of Leiden LUCIS lectures. I have cut and pasted the slides into this document.

Not for citation.

**CULTURAL MEMORY IN THE PRE-MODERN MIDDLE EAST:**

**KÖPRÜLÜ 01589, A LIBRARY IN MINIATURE**

Sarah Bowen Savant
INTRODUCTION

Consider books. Book titles. Book in catalogues. Books as physical, bound realities on bookshelves. How we, as historians, work with books. Before print – the manuscripts, bearing this or that title, that we as a field curate, edit, annotate, publish in critical editions. A book and its manuscript tradition, the iterations copied, and the modern scholarly tradition that identifies originals and derivatives. “Great” books, canonical ones, ones often cited, ones rarely so.

Now consider an assumption which underlies virtually all of our work on books – an assumption about their unity. The very word “fragment” points to a now-lost whole. A title is a pointer, to a text, with a beginning and an end, and its contents—narrative, expository, explanatory, encyclopaedic, or otherwise. I would like to propose, instead, three points.

1. **Breakdown.** Books were frequently chopped up and received only in fragmentary forms. These fragments, for audiences, often were the book. This means that, from the viewpoint of readers and reception, many modern editions are anachronisms.
2. **Selection.** Parts of some books were salient and copied, others were not. Readers found different parts of books to be meaningful at different points in time.
3. **Reformulation.** Pieces of older books were incorporated into new ones, which themselves were chopped up and reconstituted again, in a remarkable churning process that partly explains the sheer size of the Arabic tradition.

In what follows, I would like to insist that these creative processes should inform our own reading of the texts that we use when we write about history and topics such as “Imagining the Ummah.”

I have chosen a perhaps unusual case to explore how memory of the early Muslim community was processed through texts: a 14th-century Arabic and Persian manuscript containing more than 100 textual fragments that is held in the Fazil Ahmed Pasha collection of the Köprülü library in Istanbul – because I believe that it illustrates particularly well how our texts were frequently repurposed. It stands on a spectrum that includes also very literal and complete copying of texts such as the Qur’an (although I would note that Qur’an manuscripts were transmitted in parts). I have just started to work on this multi-text compilation (majmūʿa) with a former student of mine – Majid Montazermahdi – and hope that you will accept what follows as more of a project proposal and critical intervention than a finished paper. My ambition is to tell the story of this manuscript, including its makeup, script, textual apparatus and decoration, these shedding light on scribal intent and also on reader reception. To these aspects of *l’archéologie du livre* (Delaissé),¹ I will be adding digital text reuse detection and graphic visualisation to widen our view of the manuscript as a “cultural portal,” or “a prime access point into the dynamics of human communication, in a matrix that extends vertically across time and horizontally across space” (Michelle Brown, 2013). This, I believe, is a community’s memory created in culturally meaningful books.

**THE MANUSCRIPT**

First, let’s consider the manuscript itself. The 40 main texts that Köprülü 01589 contains were assembled in Shiraz in the middle of the fourteenth century and run over 436 folios (though from folio 384 onwards is reproduced with gaps in the pdf produced by the library—but accounted for in the catalogue). I say “main” to distinguish them from those in the margins. Based on the colophons, these 40 texts were written onto what appears to have been one main batch of paper (and possibly a few others) over the course of 56 years (753/1353 to 811/1409).

The state of the Köprülü manuscript

The state of the Köprülü manuscript
Nearly all of the pages—which bear signs of water damage, scuffing, and deterioration at the edges—have consistently rendered boxes and triangular excisions; several seem to have been remounted onto newer paper stock. The first of the 40 texts is dated with a colophon at 754/1354, in Shiraz. The seventeenth is from a year earlier, in 753; the fortieth and last text in the manuscript is dated at 811/1409—the latest date for any of the texts. Given the relatively close dates, it seems most likely that the manuscript was assembled in the fourteenth or early fifteenth century. There is a lot more work to do on the manuscript as a physical object, including, I hope, on its paper and binding.

There are 67 further, texts copied with somewhat more ornamentation into the manuscript’s margins, suggesting later additions, possibly when the manuscript had gained in value. The largest such text covers the margin of 36 folios. ²

In terms of handwriting, the manuscript is the work of several copyists—at this point, we have not subjected it to serious palaeographic analysis, but one can say that main body of the manuscript and the margins are written in naskh and riq’a scripts, in multiple, relatively clear hands. The manuscript also contains numerous seals, including from the Köprülü, but also much earlier ones Majid and I are seeking to identify.

The most important point to make now is that we do not seem to be dealing with an originally high value object here. My working hypothesis is that one main reason that such majmūʿât survived and indeed became incorporated into not just any collection, but a royal library, is because of the comparative rarity of the fragments they contained and their antique aura. This rarity and evocation of antiquity should bear on how we understand and situate such manuscripts and their contents as cultural products. There are several, similar such unattractive manuscripts containing rare and interesting finds in the Köprülü—they are distinct from what might have been more “usable,” pragmatic majmūʿât—examples of these being the many miscellanies of hadith now being studied by Konrad Hirschler from Damascus’s Qubbat al-khazna.

THE MANUSCRIPT’S CONTENTS

In terms of contents, as with all such collections, we are dealing with a unique miscellany of Arabic and Persian fragments. The library staff have impressively catalogued the manuscript—providing descriptive titles for all but three of the fragments—and identified authors for well over half of the texts’ excerpts. Judging from the titles, the contents pertain to the Qur’an, mysticism, philosophy, medicine, astrology, poetry, and other topics.

² Meşariku’l-envari’n-nebeviyye min sihahi’l-ahbari’l-mustafevi/es-Sagani-Radiyyun Hasan b. Muhammed. All transcription from the catalogue follows the Köprülü’s transliteration.
A treatise on astrology in Köprülü 01589

In a handful of cases, the Köprülü catalogue lists an author’s name attached to multiple fragments, for example, 9 fragments in Arabic and Persian bunched together towards the end of the manuscript and each bearing the title “risāla,” all seeming to treat Sufi topics and attributed to Majd al-Dīn al-Baghdādī, a Sufi master of Rumi (d. 1273).

There are also very rare texts pertaining to Iranian history and memory, such as a Persian translation of the Middle Persian Jāmāsp-nāmag, a Persian treatise on Pahlavi, and a rather long section of the popular animal fable Kalīla wa-Dīmna.

Now let me make 3 points based on a very preliminary investigation of these contents.

a. By looking at a manuscript such as ours, we can learn a lot about transmission and the cultural profile of a reading community. Shiraz, situated in Fārs, is sometimes credited a leading role in the revival of a specifically Iranian consciousness; according to Dorothea Krawulsky, only from the thirteenth or fourteenth centuries onward. The fourteenth century, when our main texts came together, is indeed a fascinating chapter in the history of Shiraz – which found itself enmeshed in the internal conflicts that followed the
The main body of the manuscript (the only part we can date with confidence) was initiated in the very same year, 1353, that Shiraz was re-captured by a former Mongol governor, Sharaf al-Mużaффar (from İnjü rivals—who had taken it from this Muzaffarid earlier); the manuscript was continued during the reign of his son, Shāh-i Shujā’ (who participated in the blinding and deposition of his father in 760/1359); it was continued still further during the reign of Shāh-i Shujā’’s own son, Zayn al-‘Ābidīn, and completed after Timur extinguished the independent dynasties of western Iran in 1393. Shiraz was the central Muzaffarid city, producing Hafiz (d. 792/1390, or a year later, and who was patronized by Shah-i Shujā’). So what we are dealing with is a reading community that produced Hafiz and that participated in cultural production just prior to the Timurids (themselves famous for gorgeous manuscripts).

b. For the study of cultural meaning and memory, the marginal texts should be of interest – for example, here, we see my working layout of the manuscript, and the series of text excerpts that ran alongside the Persian translation of Kalīla wa-Dimna by Naṣr Allāh b. Muḥammad, who was a Persian author and statesman born at Ghazna in a family which came from Shiraz. His fame rests on his version of Kalīla wa-dimna, which he completed between 1143 and 1146, and dedicated to the Ghaznavid sultan Yamīn al-Dawla Bahramshāh (r. 1117-1152).

Kalīla wa-dimna (main body) running alongside other texts (margins) in the Köprülü manuscript

We can explore the associations that readers made with texts in the main body and ask why they assembled and annotated a manuscript in the ways that they did.
c. **In terms of language**, I think it is also noteworthy that we see here real bilingualism in Arabic and Persian – both are used for the main body and margins — so, we should be able to explore diglossia. Such bilingualism was not unusual, if we think of many authors or famed courtiers. But some titles seem to be translations from Arabic into Persian – you do not seem to have the reverse (Persian being translated into Arabic).

**TRACING MEMORY: THE ABŪ MUSIM FRAGMENT**

Thirdly, I would like to propose – just briefly – that we can explore the fragments individually from different angles.

In terms of fragments, recent scholarship on the medieval West has taken an interest in the importance of anthologizing for literary culture, and considered, also, the material importance of miscellanies and florilegia as testimonies to cultural identity (e.g. the 2010-2013 European project, “The Dynamics of the Medieval Manuscript: Text Collections from a European Perspective”). For the medieval Islamic world, complex works such as Köprülü 01589 have frequently been mined but until fairly recently rarely studied in their own right, with new work now on “Muslim encyclopaedism” (e.g., Endress, 2006; Muhanna, 2013); anthologies (Orfali, 2015); and multi-text compilations (Hirschler, 2016; Savant, 2014).

I am just initiating a project called KITAB – Knowledge, Information Technology, and the Arabic Book – which uses algorithms that operate much like anti-plagiarism software to detect and align common passages between books and, indeed, within the entirety of the Arabic textual tradition.³ In the first 10 months of this year, we ran a pilot during which we created a digital corpus of about 4,500 unique texts in TXT format from the period 750–1500; collected basic metadata relating to these texts; experimented with, adapted, and ran one algorithm to detect text reuse on the entirety of the corpus (comparing every book to every other); created a database of indexed results; and designed a prototype interactive website. One of our overall purposes is to consider how new books were assembled from fragments – books such as Köprülü 01589, but also less dramatically fragmentary works (e.g. any history from the period), and also the processes of breakdown, selection, and reformulation outlined above. The premise, to be explored, is that new meanings were generated within our texts, and that in each unique work we can understand the cultural identity of the compiler or commissioner and probe how cultural and social heritage were passed on to new generations.

Part of this work involves following the trail of fragments, such as, in the case of Köprülü 01589, an example here of about 1100 words, copied onto the 101 and 102 folios of the

³ KITAB’s team includes Maxim Romanov, of the Perseus Digital Library; Ahmad Sakhi and Malik Merchant, two professional programmers who work as volunteers for the project as part of their service to the Aga Khan Development Network; and Sohail Merchant, of the AKU-ISMTC, who has a programming and project management background. Our partners include David Smith, of Northeastern University, and Marco Büchler, of the Göttingen Center for Digital Humanities, as well as the Perseus Digital Library at Tufts and Leipzig Universities, with Gregory Crane (Editor-in-Chief). Crane and Perseus have more than 25 years of experience in Digital Humanities and expertise in many areas of research and development. So far our data has been run off of the supercomputing cluster at Perseus and is hosted at the AKU and Perseus.
manuscript. I chose it because of today’s panel’s thematic focus on the creation of community – as it tells the lifestory of Abū Muslim – the major missionary recruiter of the ʿAbbasid movement in Khurasan –from his obscure origins as the son of a slave belonging to a leader from the Arab tribe al-ʿĪjlī (ʿUmayr b. Buṭayr al-ʿĪjlī) through to his successes for the Abbasids, but final death at the instigation of the second Abbasid caliph, al-Manṣūr. With Abū Muslim, I would argue, we have an example of what Pierre Nora has termed a “site of memory” – in the textual tradition, the people, places, and events that have a capacity for metamorphosis, “an endless recycling of their meaning and an unpredictable proliferation of their meaning” (Nora, 1989). Such sites of memory historically supported Muslims’ sense of a shared past and identity.

So, a few brief points about the fragment:

a. First, my earlier mentioned student, Majid Montazermahdi, transcribed it and then my team, including Maxim Romanov of the Perseus Digital Library, ran our pilot text reuse algorithm, comparing the fragment to the entire corpus, through which we discovered that the fragment came from a heavily excerpted section of Abū Ḥanīfa al-Dīnawarī’s al-Akhbār al-ṭiwāl.
The cataloguers in Istanbul had not worked this out – understandably since the 1100 words come from a more than 50-page section of the *Akhbār*, in which al-Dīnawari is more interested in narrating conflicts among Arab tribes, apparently as a way to explain the fall of the Umayyad dynasty (r. 661-750). In fact, I myself had entirely by coincidence very closely read this part of al-Dīnawarī’s text a year earlier as part of one of my Arabic reading groups and had no idea that the fragment came from it. It really is remarkable that our algorithms could identify the origins.

What you have here is the entirety of the 1100 words running along the x-axis in 11 one-hundred word chunks and where within that the *Akhbār al-tiwal* overlaps. The arrow in the graph points to the alignment of texts displayed below. We set the threshold to detect reuse rather high – so chunks of less than 50 words did not register (otherwise, when looking over 4500 works, you would have had a lot of noise). With something like a 10 word threshold, more of this graph would be red. In fact, just looking at the texts, there is more reuse than shown here. Nonetheless, as a discovery tool, the algorithm was powerful and correctly led me to al-Dīnawarī’s text. It could have come to us through any meandering path of transmission (copy of incomplete copy, etc.).
Another part of the graph and its texts, displayed below

We also determined – fascinatingly – that no other work in our entire 4,500 work corpus seems to have copied from this portion of al-Dīnawarī’s book. What we seem to have then is a rather singular and unique use of Abū Ḥanīfa’s book. This is the case even though al-Dīnawarī’s Akhbār was used by subsequent authors, as our data shows (not displayed here).

b. In terms of how we understand the cultural meaning of the fragment – we have a complex hermeneutics, where we have an unknown scribe selectively editing a previous work into a marginal notation that lies within a manuscript containing pieces of multiple books copied out by multiple copyists; the assembly of the work is still unclear (by one or more agents), and it was also held by multiple owners. These are the agents of a textual community.

One clear point for consideration will be the relationship of the fragment to its neighbors. It frames the start of another, only slightly longer Persian text, the ‘Iqd al-jawāhir fi ansāb al-akābir by a certain Aḥmad b. Abī As’ad al-Dīwwānī al-Mīṣkātī.

Conclusion

Ideally, now I would say a lot more about the contents of the manuscript, but I hope this introduction to my way of approaching the archaeology of the book supports an argument relating to this panel, but also the study of Islamic history generally: which is that if we wish to understand how the Ummah was imagined historically, or the creation of community in
Islam, we need to include the textual communities that sustained it, including how they read and used texts and generated through them cultural meaning and memory.

**Works cited**

