

Maxim Romanov, U Leipzig

Brill, December 3-4, 2015

# Of graphs, maps, and 30,000 Muslims

Premodern Arabic texts &  
the Digital humanities



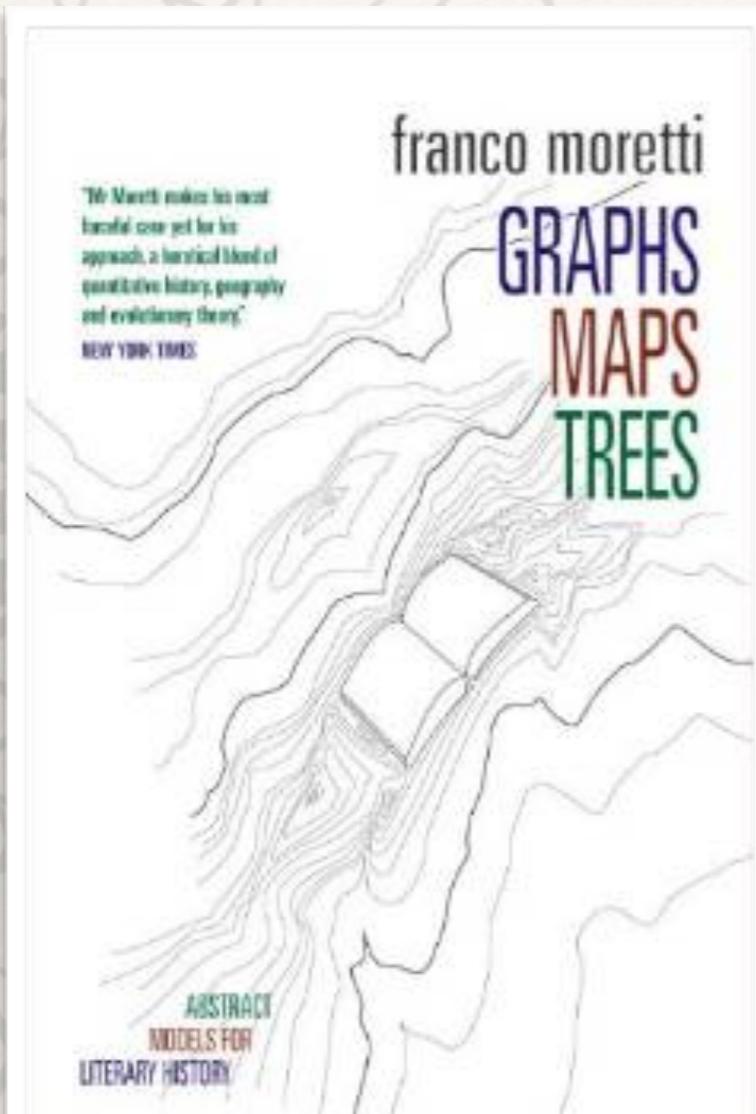
Digital Humanities  
UNIVERSITÄT LEIPZIG

# Digital Humanities

Engaging in a kind of humanistic research that *cannot be done without* digital methods, where complexity and novelty of each research task would vary from impossible ‘without 500 monks at hand’\* to otherwise unthinkable. Such research would rely on large volumes of data (e.g., corpora of full-text primary sources) and a variety of digital approaches that allow converting raw data into meaningful information and then exploring it with different visualization techniques in order to trace long-term and large-scale developments.

\* *The phrase is from:* Mathisen, Ralph W. “Where Are All the PDBs?: The Creation of Prosopographical Databases for the Ancient and Medieval Worlds.” In *Prosopography Approaches and Applications: A Handbook*, 95–126. University of Oxford, Linacre College Unit for Prosopographical Research, 2007.

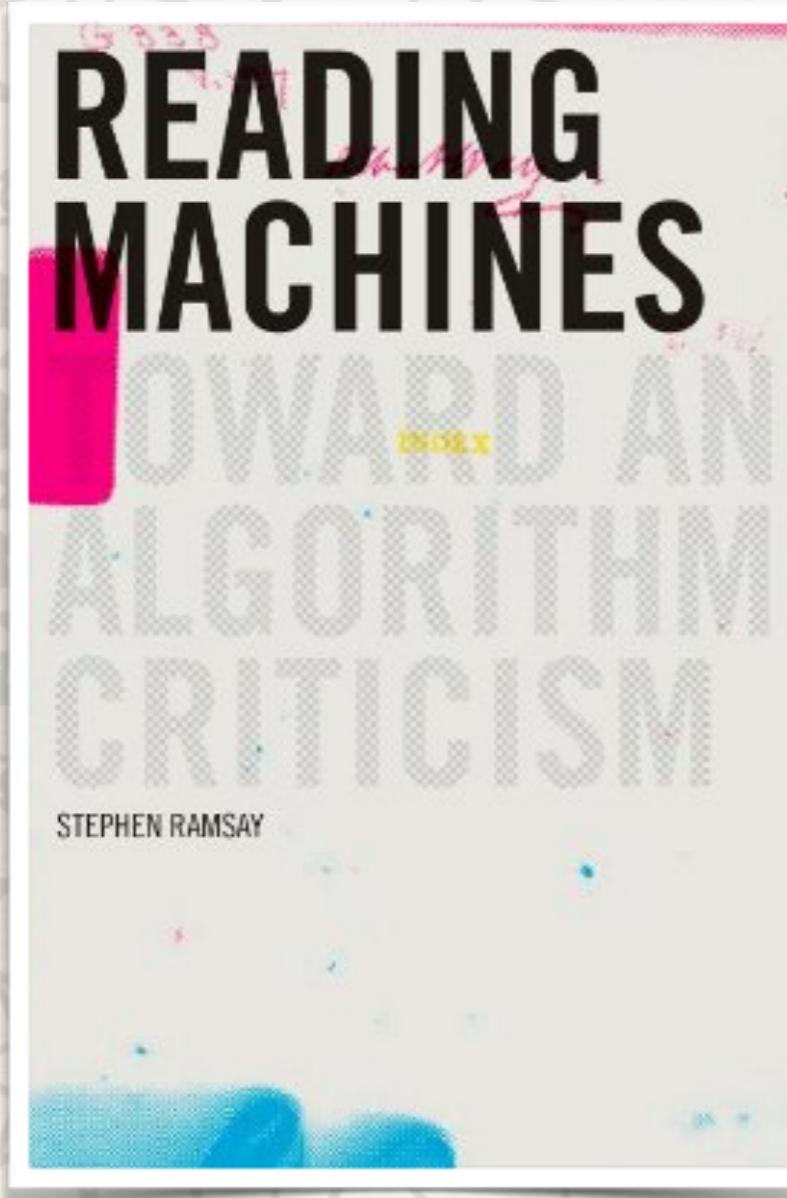
# Method: *Distant Reading*



'Distant reading', I have once called this type of approach, where distance is however not an obstacle, but a specific form of knowledge: *fewer elements, hence a sharper sense of their overall interconnection*. Shapes, relations, structures. Forms. Models.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for Literary History*. Verso, 2007.

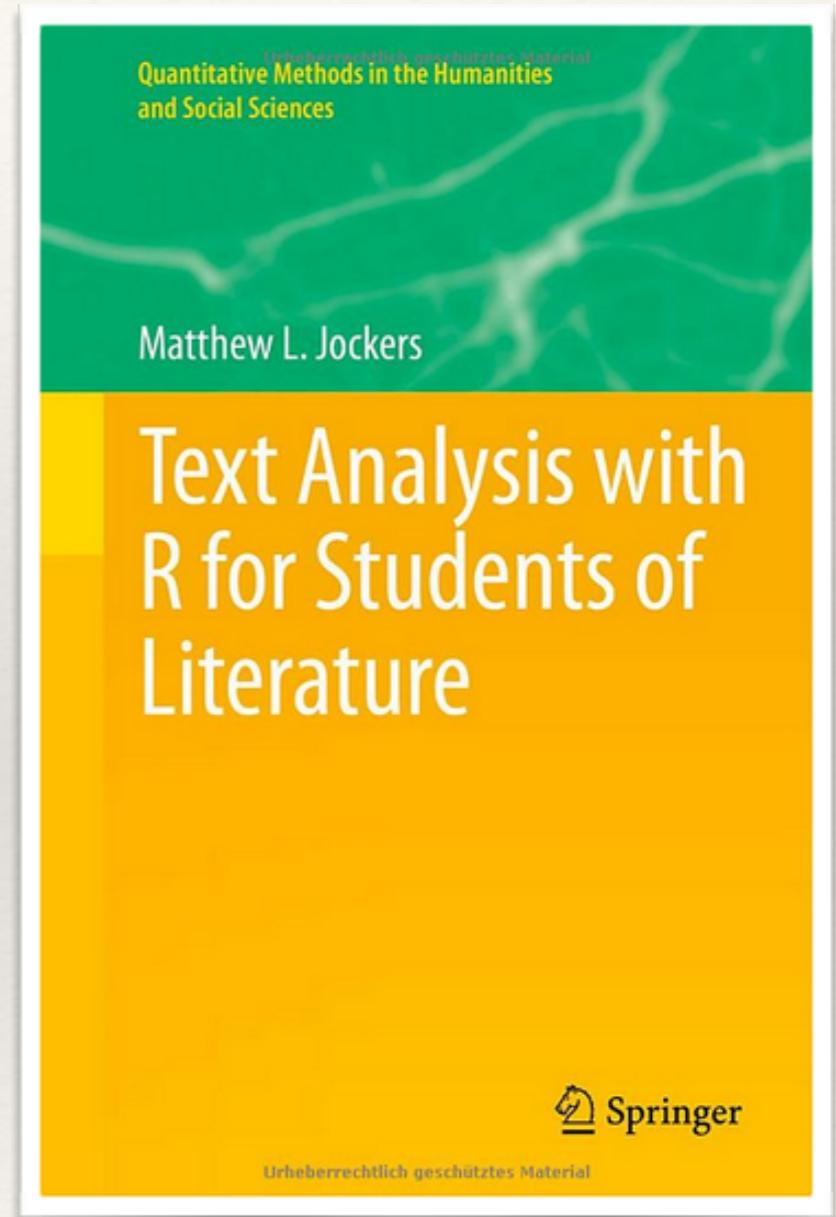
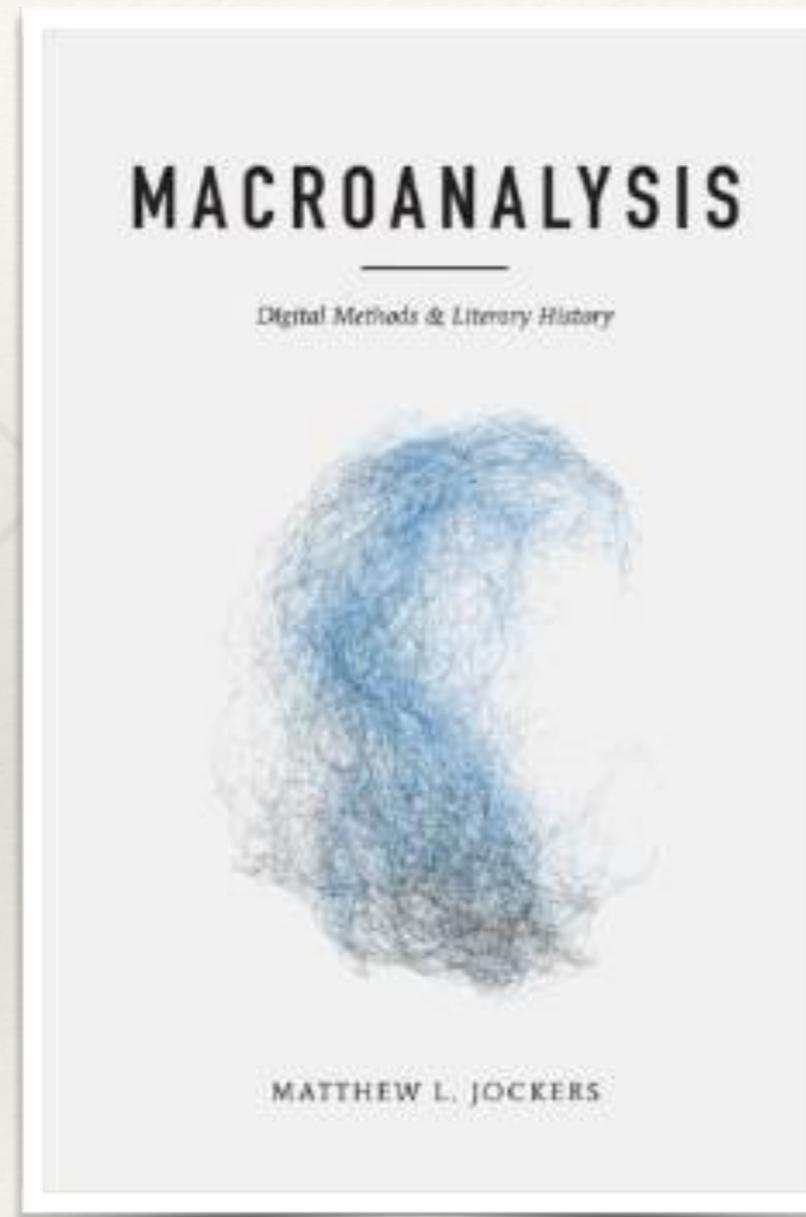
# Method: *Algorithmic Deformation*



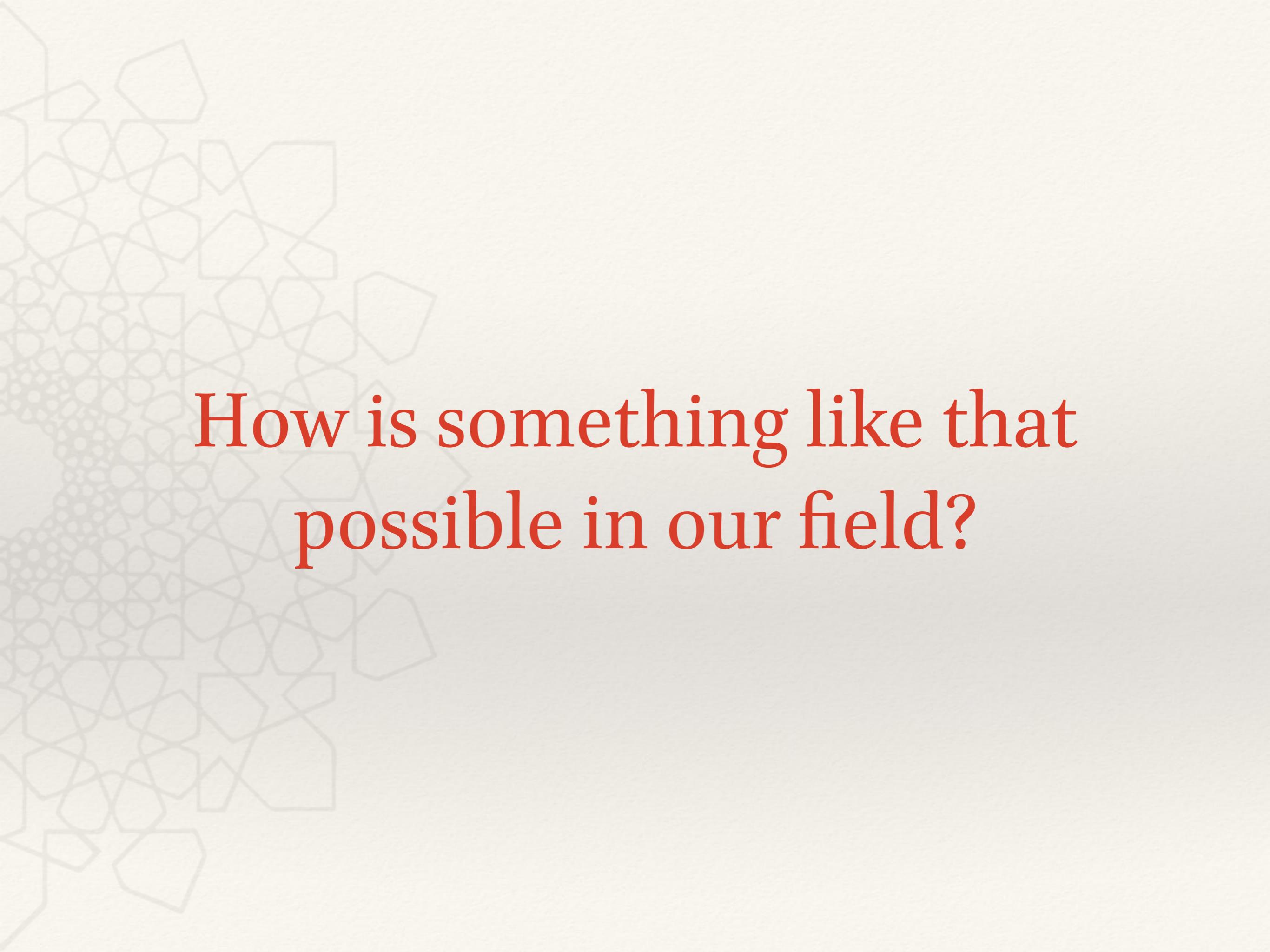
“Algorithmic criticism is easily conceived as the form of engagement that results when imperative routines are inserted into the wider constellation of texts stipulated by critical reading. But it is also to be understood as the creation of interactive programs in which readers are forced to contend not only with deformed texts, *but with the ‘how’ of those deformations.*”

Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. 1st Edition. University of Illinois Press, 2011.

# Method: *Micro-, Meso-, and Macroanalysis*



<http://www.matthewjockers.net/>



How is something like that  
possible in our field?

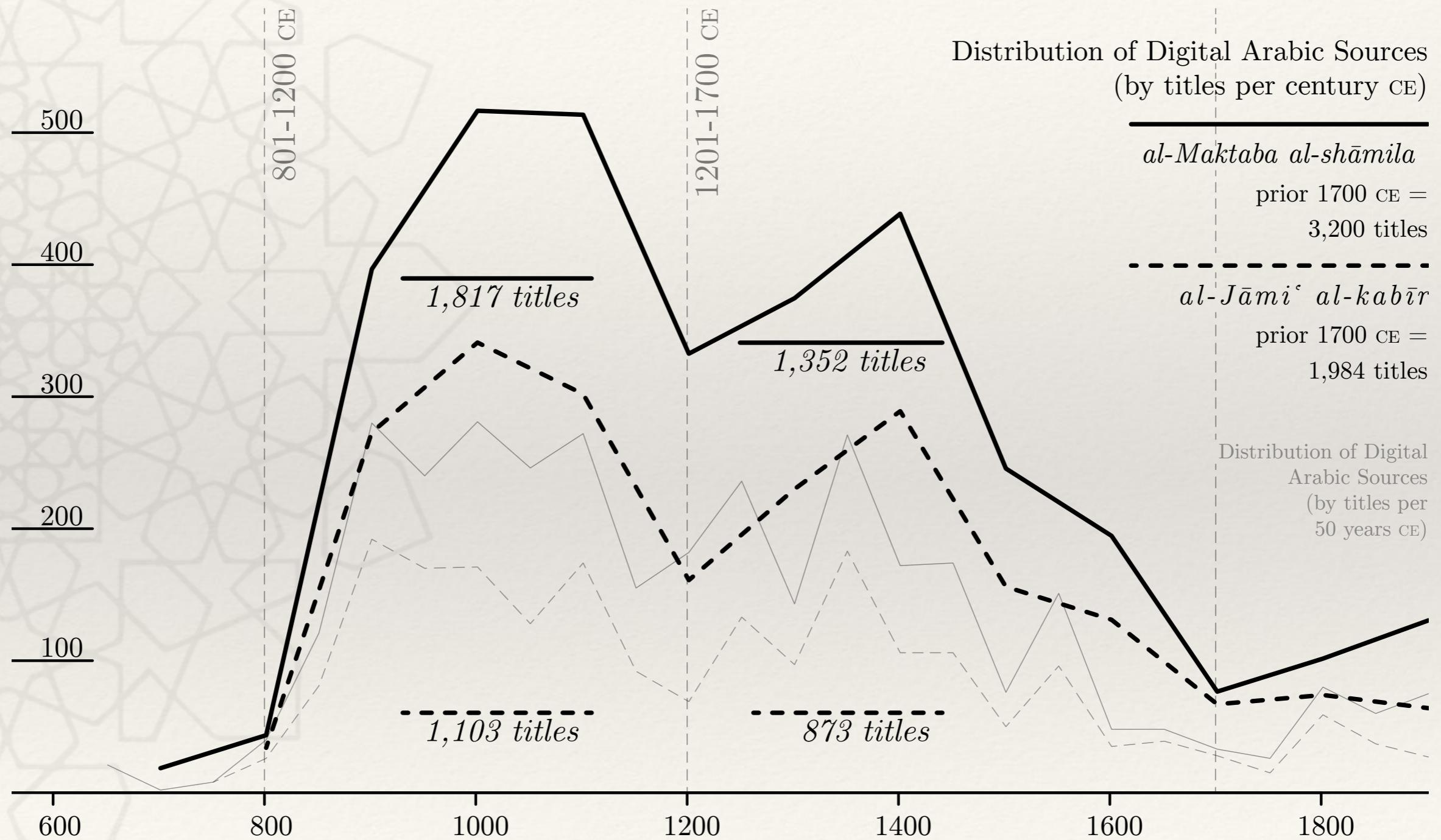
# Digital Collections of Arabic Texts

Library	Media	Titles	Vols	Words
<i>al-Jāmi‘ al-kabīr</i>	HDD, Windows 95	2,400	5,550	~400 mln
	<a href="http://www.shamela.ws">www.shamela.ws</a>	6,300	—	~820 mln
	<a href="http://www.almeshkat.net">www.almeshkat.net</a>	7,300	—	—
	<a href="http://www.said.net">www.said.net</a>	10,000	—	—
	<a href="http://www.alwaraq.com">www.alwaraq.com</a>	860	—	—
<i>al-Mujam al-fiqhī</i>	DVD	1,130	3,000	—
	<a href="http://www.shiaonlinelibrary.com">www.shiaonlinelibrary.com</a>	1,970	4,175	~280 mln
<i>al-Maktaba al-shāmila</i>				
<i>al-Mishkāt</i>				
<i>Sayd al-fawā’id</i>				
<i>al-Warrāq</i>				

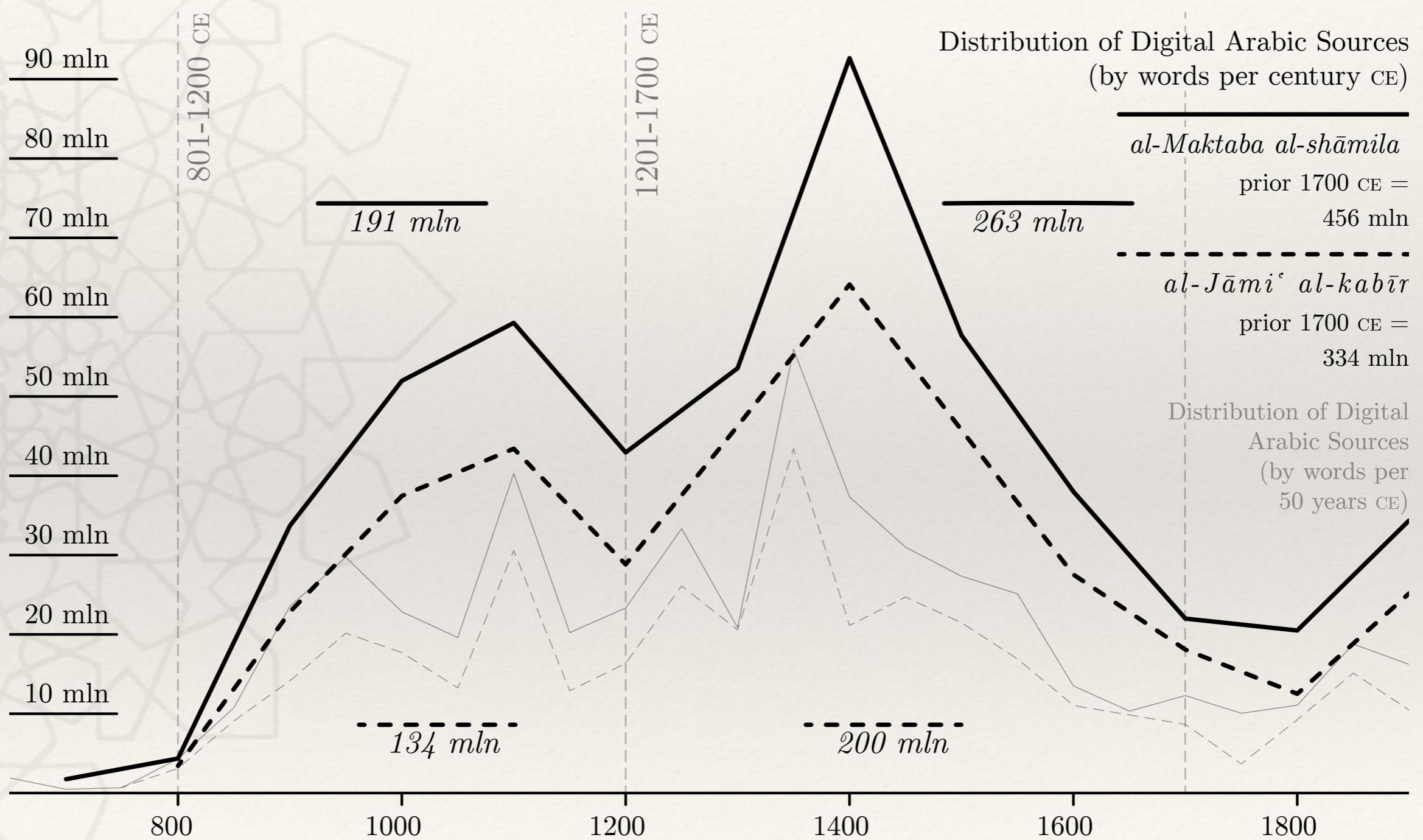
# Digital Collections: Problems

- ❖ Thematic/subject biases (“salafi” collections): almost no scientific, philosophical, medical and other texts
- ❖ Spotty chronological and geographical coverage
- ❖ Difficult to assess their overall coverage

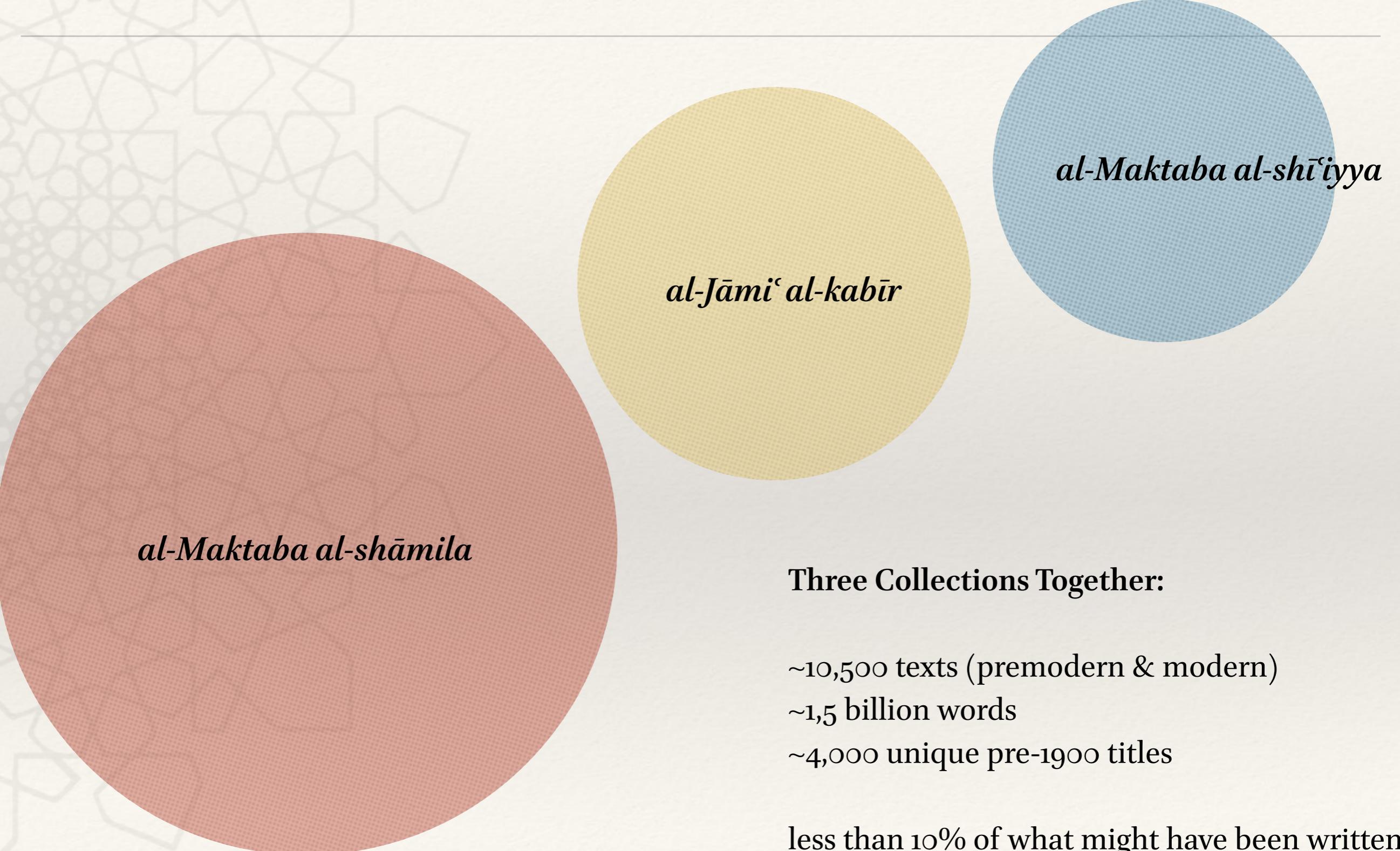
# Digital Collections: *Coverage Issues*



# Digital Collections: *Coverage Issues*



# Digital Collections of Arabic Texts



*al-Maktaba al-shāmila*

*al-Jāmi‘ al-kabīr*

*al-Maktaba al-shī‘iyya*

## Three Collections Together:

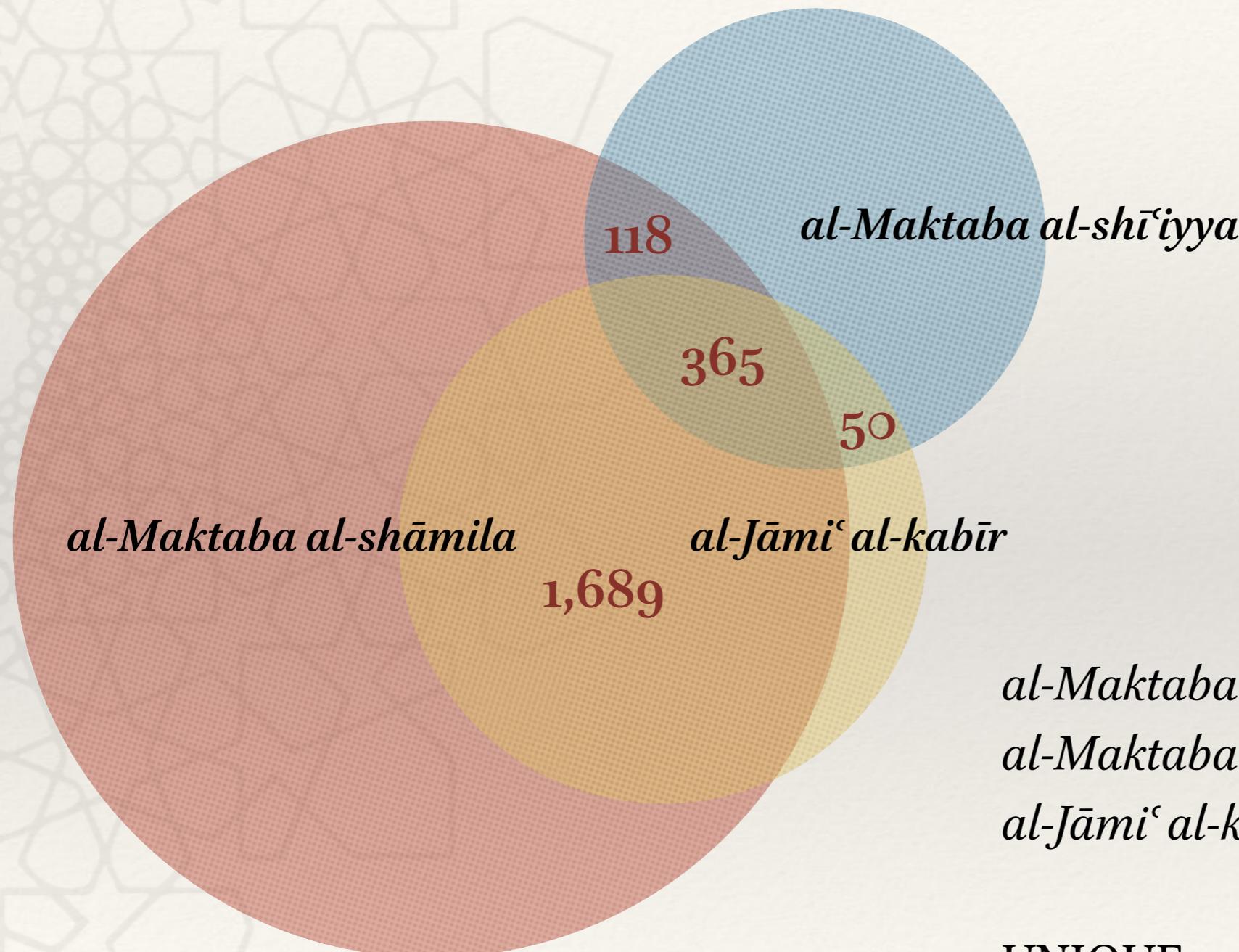
~10,500 texts (premodern & modern)

~1.5 billion words

~4,000 unique pre-1900 titles

less than 10% of what might have been written

# Digital Collections of Arabic Texts



*al-Maktaba al-shī‘iyya*: 1,810 titles  
*al-Maktaba al-shāmila*: 5,999 titles  
*al-Jāmi‘ al-kabīr*: 2,364 titles

UNIQUE: 7,895 titles (~1.1 billion words)

# Digital Collections: Major Genres & Forms

التفسير

Interpretation of the Qur'ān

الحديث

"Prophetic sayings"

أصول الفقه ومسائله

Legal writings

التاريخ

Chronicles

الترجم والطبقات

Biographical collections

النحو والصرف

Arabic language, grammar & morphology

الشعر والأدب

Poetry & fine literature

المعاجم

Various dictionaries & references

# Digital Collections: Major Genres & Forms

التفسير

Interpretation of the Qur'ān

الحديث

"Prophetic sayings"

أصول الفقه ومسائله

Legal writings

التاريخ

Chronicles

الترجم والطبقات

Biographical collections

النحو والصرف

Arabic language, grammar & morphology

الشعر والأدب

Poetry & fine literature

المعاجم

Various dictionaries & references

# Digital Collections: *Biographical Collections*

~300-400  
~250  
structure  
coverage  
biographies

> 400,000

have been written  
available in digital format  
chronological, generational, alphabetical  
social and religious, geographical  
well-structured (linguistic formulae):  
onomastic section, birth [date &] place,  
teachers, students, contribution,  
miscellanea, dear: date & place, reaction of  
the community

biographies?

# Algorithmic Analysis

# Digital Collections: *Structures*

L C F R	\$ # الهروي	237
L C F R	# أبو سعيد إبراهيم بن طهمان بن شعيب من قرية باشان نزيل نيسابور	238
L C F R	~~ سافر إلى مكة ومات بها كان فقيها محدثاً توفى سنة 163 ثلاث وستين	239
L C F R	~~ ومائة صنف تفسير القرآن سنن الفقه كتاب العيددين	240
L C F R	~~ المناقب	241

# Digital Collections: *Structures*

L C F R	\$ # الهروي	237
L C F R	# أبو سعيد إبراهيم بن طهمان بن شعيب من قرية باشان نزيل نيسابور	238
L C F R	~~ سافر إلى مكة ومات بها كان فقيها محدثاً توفى سنة 163 ثلاث وستين	239
L C F R	~~ ومائة . صنف تفسير القرآن . سنن الفقه . كتاب العيددين . كتاب المناقب	240
L C F R	~~ المناقب .	241

## # \$ al-Harawī

# Abū Sa‘īd Ibrāhīm b. Ṭahmān b. Šu‘ayb, from the village of Bashan, a resident of Nishapur.

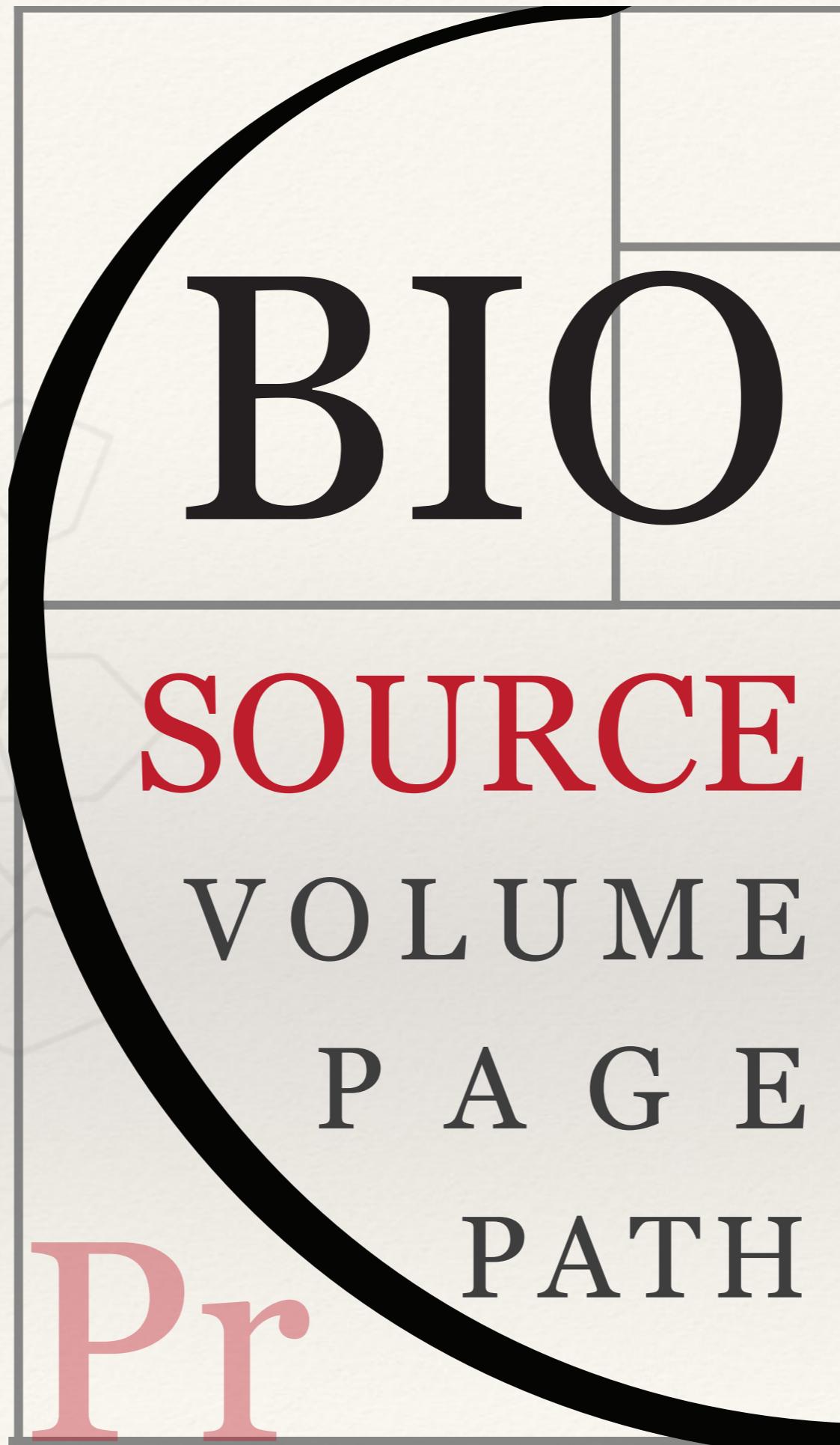
~~ He traveled to Mecca and died there. He was a jurist, transmitter of Hadith. He died in 163.

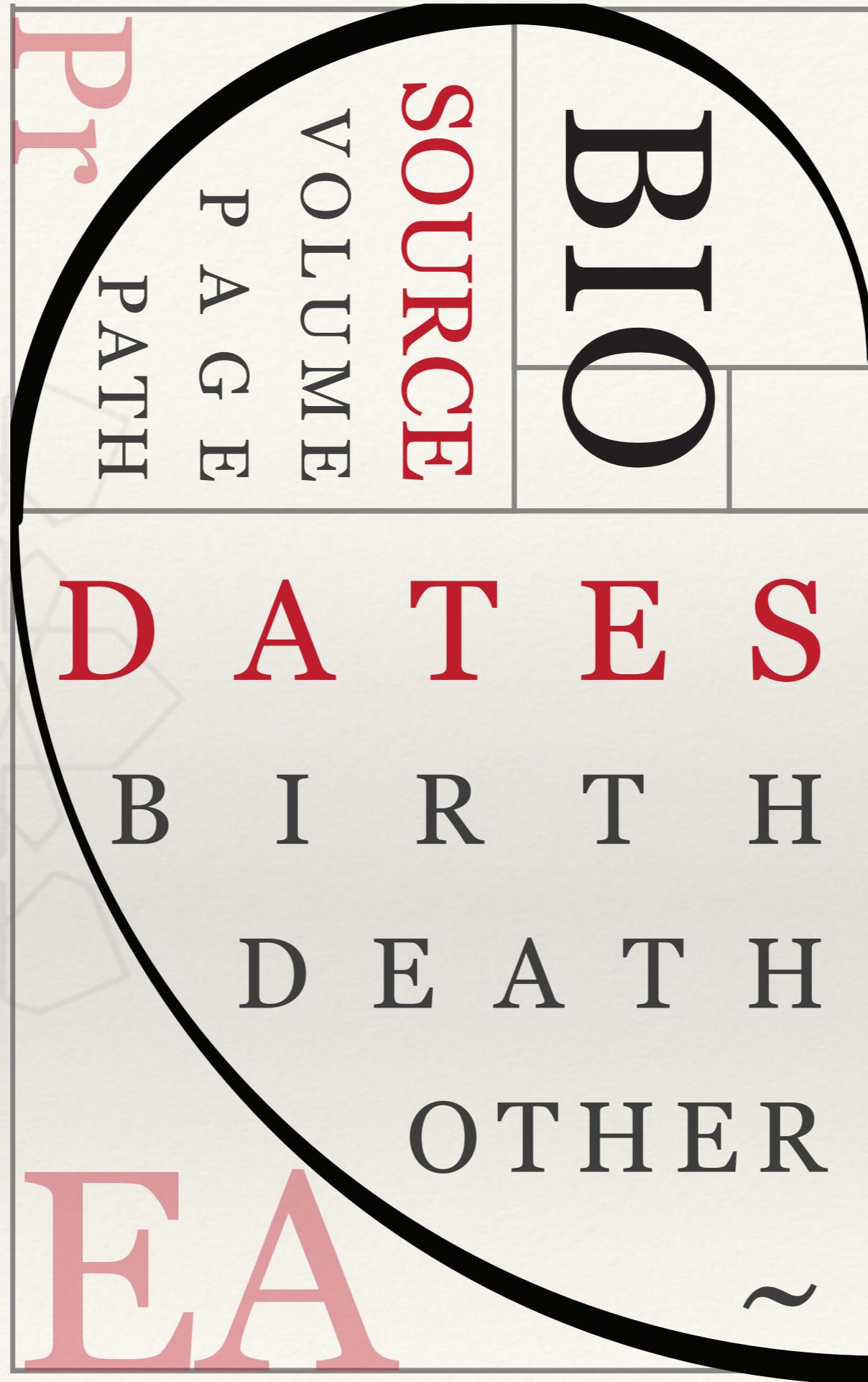
~~ ... **He composed** The Exegesis of the Qur'an, Legal hadith, The Book of Two Celebrations, The Book of Virtues.

# Digital Collections: *Structures*

L C F R	\$ . الهروي #	237
L C F R	أبو سعيد إبراهيم بن طهمان بن شعيب من قرية باشان نزيل نيسابور #	238
L C F R	سافر إلى مكة ومات بها كان فقيها محدثاً توفى سنة 163 ثلاثة وستين #	239
L C F R	ومائة . صنف تفسير القرآن . سنن الفقه . كتاب العيددين . كتاب المناقب . #	240
L C F R	المناقب . #	241
		242
L C F R	-----#META#	243
L C F R	163 «META#YEAR##	244
L C F R	تفسير القرآن «0163 «HA0006 «META##BOOKTITLE##	245
L C F R	سنن الفقه «0163 «HA0006 «META##BOOKTITLE##	246
L C F R	كتاب العيددين «0163 «HA0006 «META##BOOKTITLE##	247
L C F R	كتاب المناقب «0163 «HA0006 «META##BOOKTITLE##	248
L C F R	META#BOOKSTOTAL# 4#	249

*Algorithmic deformation (with Python)*

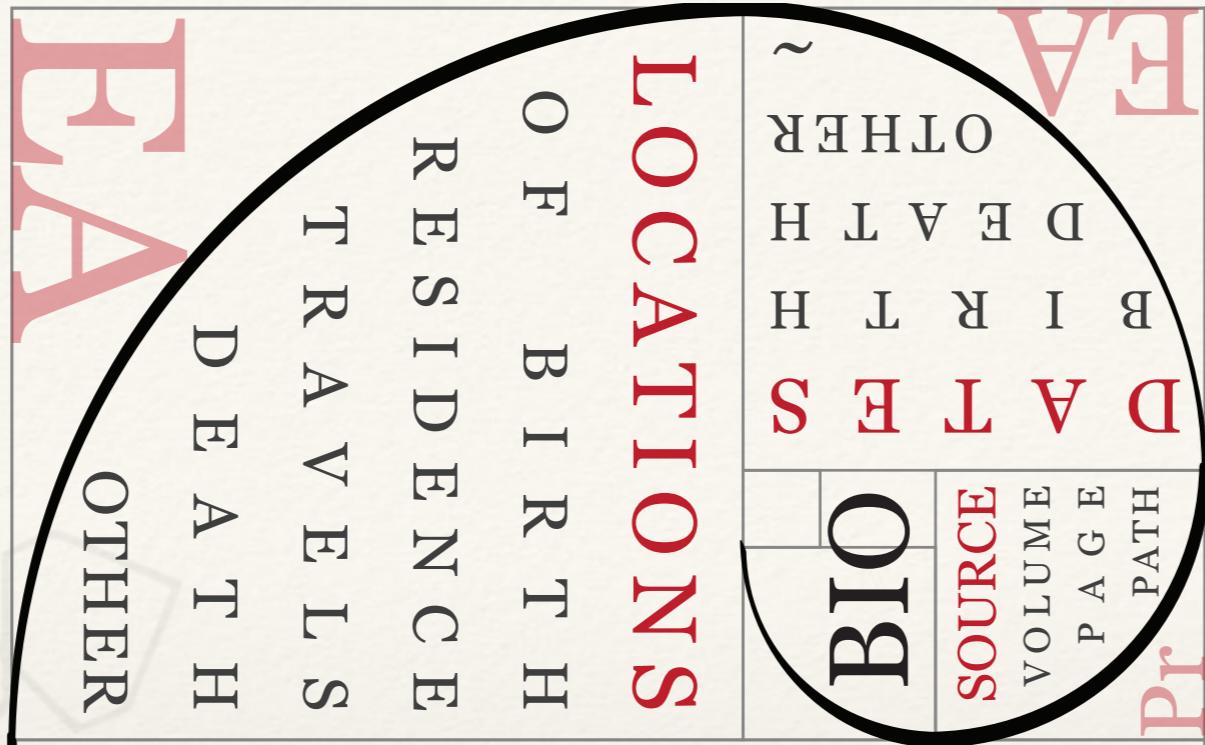






LOCATIONS  
OF BIRTH  
RESIDENCE  
TRAVELS  
DEATH

EA OTHER



## DESCRIPTIVE NAMES

RELIGIOUS SPECIALIZATION

RELIGIOUS AFFILIATION

PHYSICAL DESCRIPTION

GEOGRAPHICAL

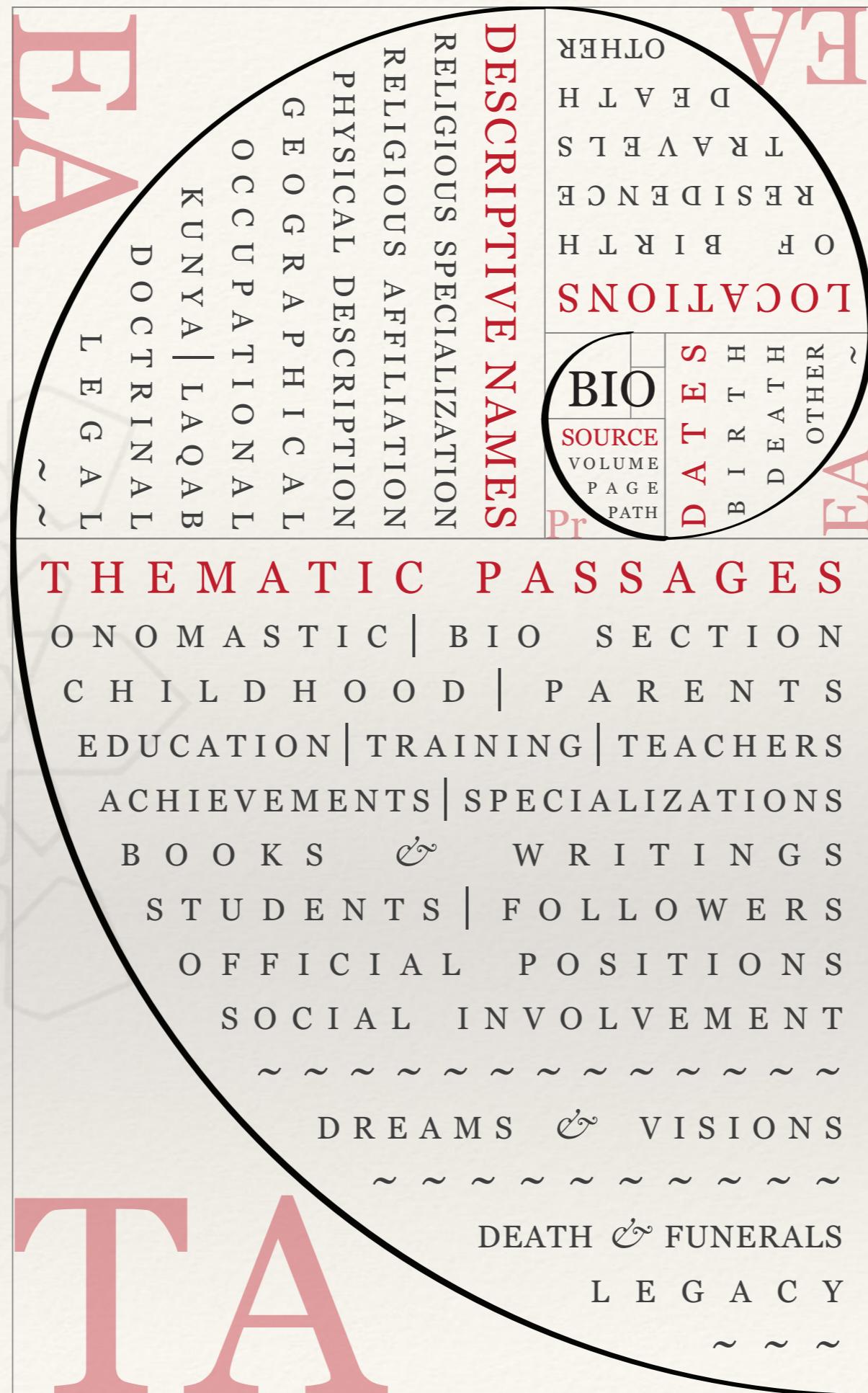
OCCUPATIONAL

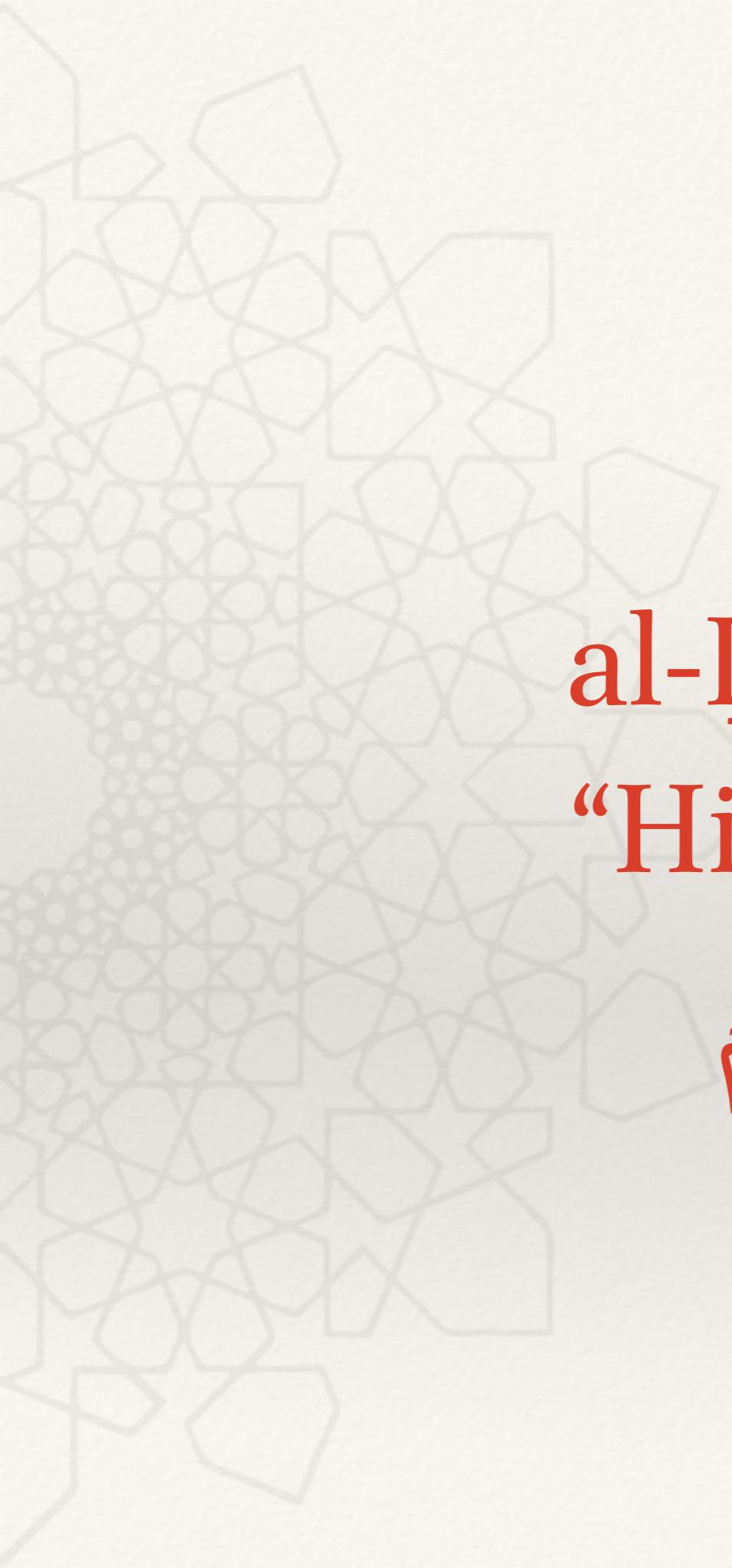
KUNYA | LAQAB

DOCTRINAL

LEGAL

EA



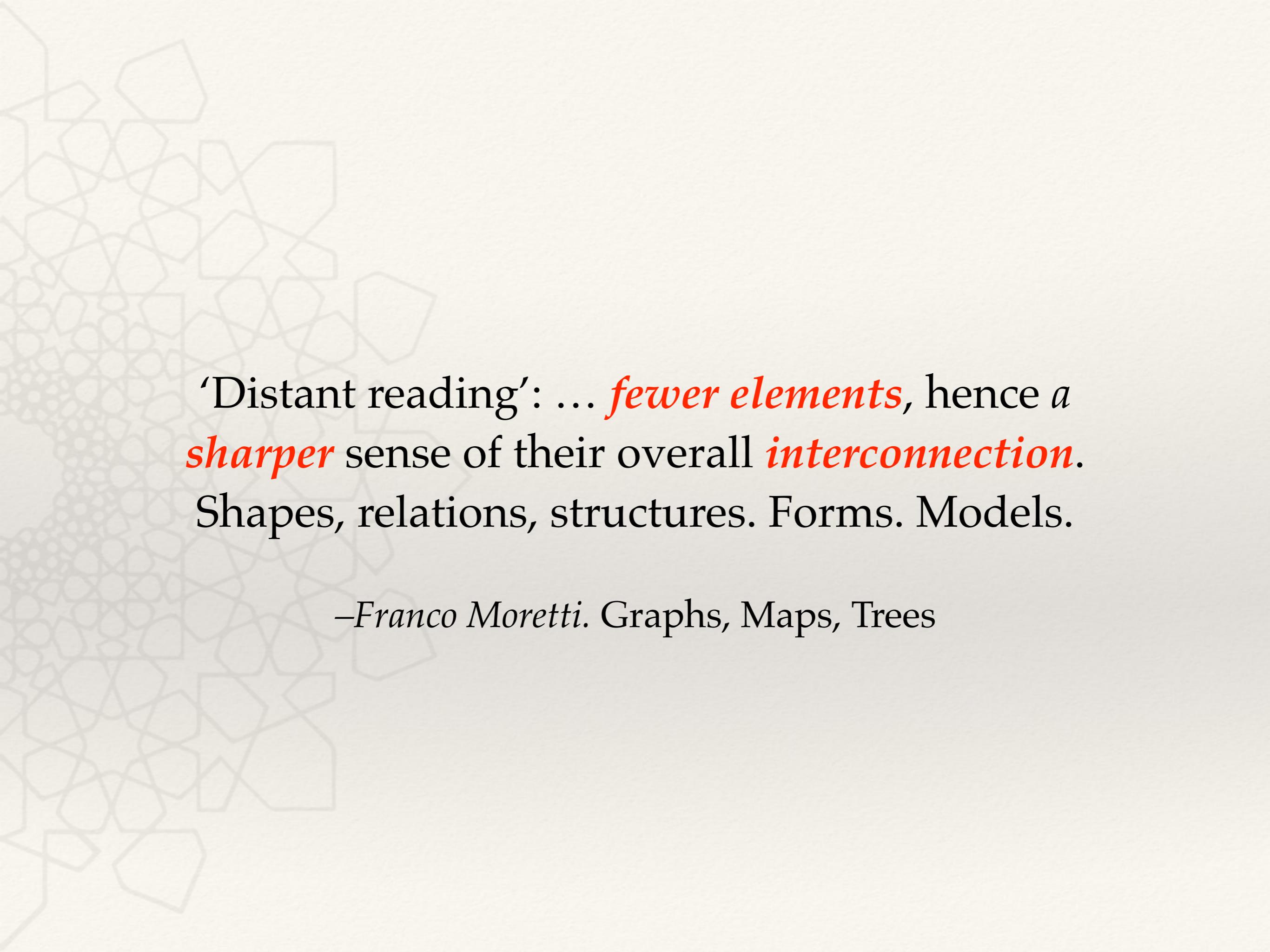


# al-Dahabī and his “History of Islam”

تاریخ اسلام

# al-Dahabī (d. 748/1347 CE)

- ❖ a Damascene religious scholar, Šāfi‘ī jurist and historian
  - ❖ “The History of Islam” (*Ta’rīh al-islām*)
    - ❖ “first draft” in 714/1314 CE
    - ❖ 50 volumes (here: 4-50)
    - ❖ 7 centuries (here: 41-700/661-1300 CE)
    - ❖ over 30,000 biographies (here: ~29,100)
    - ❖ ~ 3,2 mln. words



‘Distant reading’: ... *fewer elements*, hence *a sharper* sense of their overall *interconnection*.  
Shapes, relations, structures. Forms. Models.

—Franco Moretti. Graphs, Maps, Trees

# “The History of Islam”: Distant Reading

- ❖ Death dates
- ❖ Place names / Toponyms
- ❖ “Descriptive names” (*nisbats*)
  - ❖ 700 unique (frequency 10 and higher)
  - ❖ 70,000 total
  - ❖ ~ social profile (issue with their meaning\*)

# “The History of Islam”: Distant Reading

Traditional Arab Name: Example

‘Abd al-Rahmān ibn ‘Alī ibn Muḥammad ibn ‘Ubayd  
Allāh ibn ‘Abd Allāh ibn Ḥamadá ibn Muḥammad ibn  
Ğa‘far ibn ‘Abd Allāh ibn al-Ķāsim ibn al-Nađar ibn al-  
Ķāsim ibn Muḥammad ibn ‘Abd Allāh ibn ‘Abd al-  
Rahmān ibn al-Ķāsim ibn Muḥammad ibn Abī Bakr  
al-Şiddīk (*may God be pleased with him!*) al-Ķurašī al-  
Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-  
Hanbalī al-Wā‘iz al-Şaffār, Ğamāl al-Dīn, Abū-l-Farağ,  
known as Ibn al-Ğawzī (d. 597/1201 ce)

# “The History of Islam”: Distant Reading

Traditional Arab Name: Example

‘Abd al-Rahmān ibn ‘Alī ibn Muḥammad ibn ‘Ubayd  
Allāh ibn ‘Abd Allāh ibn Ḥamādā ibn Muḥammad ibn  
Ğa‘far ibn ‘Abd Allāh ibn al-Ḳāsim ibn al-Naḍar ibn al-  
Ḳāsim ibn Muḥammad ibn ‘Abd Allāh ibn ‘Abd al-  
Rahmān ibn al-Ḳāsim ibn Muḥammad ibn Abī Bakr  
al-Şiddīk (*may God be pleased with him!*) al-Ḳurašī al-  
Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-  
Hanbalī al-Wā‘iz al-Şaffār, Ğamāl al-Dīn, Abū-l-Farağ,  
known as Ibn al-Ğawzī (d. 597/1201 ce)

# “The History of Islam”: Distant Reading

Traditional Arab Name: Example

‘Abd al-Rahmān ibn ‘Alī ibn Muḥammad ibn ‘Ubayd  
Allāh ibn ‘Abd Allāh ibn Ḥamādā ibn Muḥammad ibn  
Ğa‘far ibn ‘Abd Allāh ibn al-Ķāsim ibn al-Naḍar ibn al-  
Ķāsim ibn Muḥammad ibn ‘Abd Allāh ibn ‘Abd al-  
Rahmān ibn al-Ķāsim ibn Muḥammad ibn Abī Bakr  
al-Şiddīk (*may God be pleased with him!*) al-Ķurašī al-  
Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-  
Hanbalī al-Wā‘iz al-Şaffār, Ğamāl al-Dīn, Abū-l-Farağ,  
known as Ibn al-Ğawzī (d. 597/1201 ce)

# “The History of Islam”: Distant Reading

Traditional Arab Name: Example

‘Abd al-Rahmān ibn ‘Alī ibn Muḥammad ibn ‘Ubayd  
Allāh ibn ‘Abd Allāh ibn Ḥamadá ibn Muḥammad ibn  
Ğa‘far ibn ‘Abd Allāh ibn al-Ķāsim ibn al-Nađar ibn al-  
Ķāsim ibn Muḥammad ibn ‘Abd Allāh ibn ‘Abd al-  
Rahmān ibn al-Ķāsim ibn Muḥammad ibn Abī Bakr  
al-Şiddīk (*may God be pleased with him!*) al-Ķurašī al-  
Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-  
Hanbalī al-Wā‘iz al-Şaffār, Ğamāl al-Dīn, Abū-l-Farağ,  
known as Ibn al-Ğawzī (d. 597/1201 ce)

# “The History of Islam”: Distant Reading

Traditional Arab Name: Example

‘Abd al-Rahmān ibn ‘Alī ibn Muḥammad ibn ‘Ubayd  
Allāh ibn ‘Abd Allāh ibn Ḥamadá ibn Muḥammad ibn  
Ğa‘far ibn ‘Abd Allāh ibn al-Ķāsim ibn al-Nađar ibn al-  
Ķāsim ibn Muḥammad ibn ‘Abd Allāh ibn ‘Abd al-  
Rahmān ibn al-Ķāsim ibn Muḥammad ibn Abī Bakr  
al-Şiddīk (*may God be pleased with him!*) al-Ķurašī al-  
Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-  
Hanbalī al-Wā‘iz al-Şaffār, Ğamāl al-Dīn, Abū-l-Farağ,  
known as Ibn al-Ğawzī (d. 597/1201 ce)

# “The History of Islam”: Distant Reading

Traditional Arab Name: Example

... *al-Kurašī al-Taymī al-Bakrī al-Bağdādī al-Ḥāfiẓ al-Mufassir al-Ḥanbalī al-Wā’iz al-Şaffār* ...

*al-Kurašī*

member of the tribe of Quraysh (tribal)

*al-Taymī*

member of the clan of Taym (tribal)

*al-Bakrī*

descendant of Abū Bakr al-Şiddīq (ancestral)

*al-Bağdādī*

resident/native of Baghdad (toponymic)

*al-Ḥāfiẓ*

“Preserver” of the Tradition (religious specialization)

*al-Mufassir*

exegete of the Qur’ān (religious specialization)

*al-Ḥanbalī*

jurist of the Ḥanbalī legal school (religious affiliation)

*al-Wā’iz*

public preacher (religious specialization)

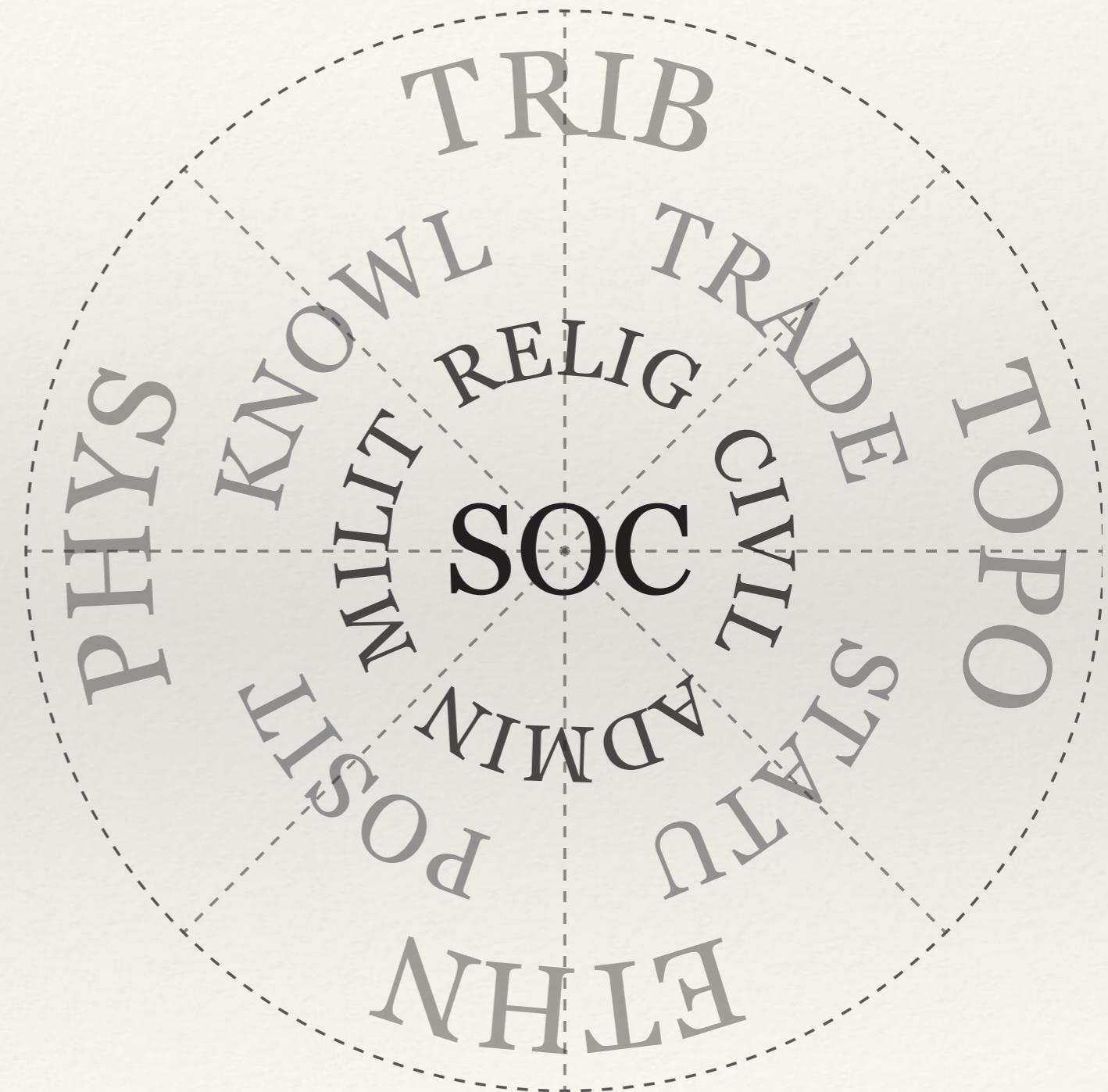
*al-Şaffār*

seller of copper/brass utensils (occupational)

# Classification of *Nisbats* from the Social Perspective

Previous work on *nisbats*  
(mostly occupational):

- Hayyim Cohen
- Carl Petry
- Maya Shatzmiller



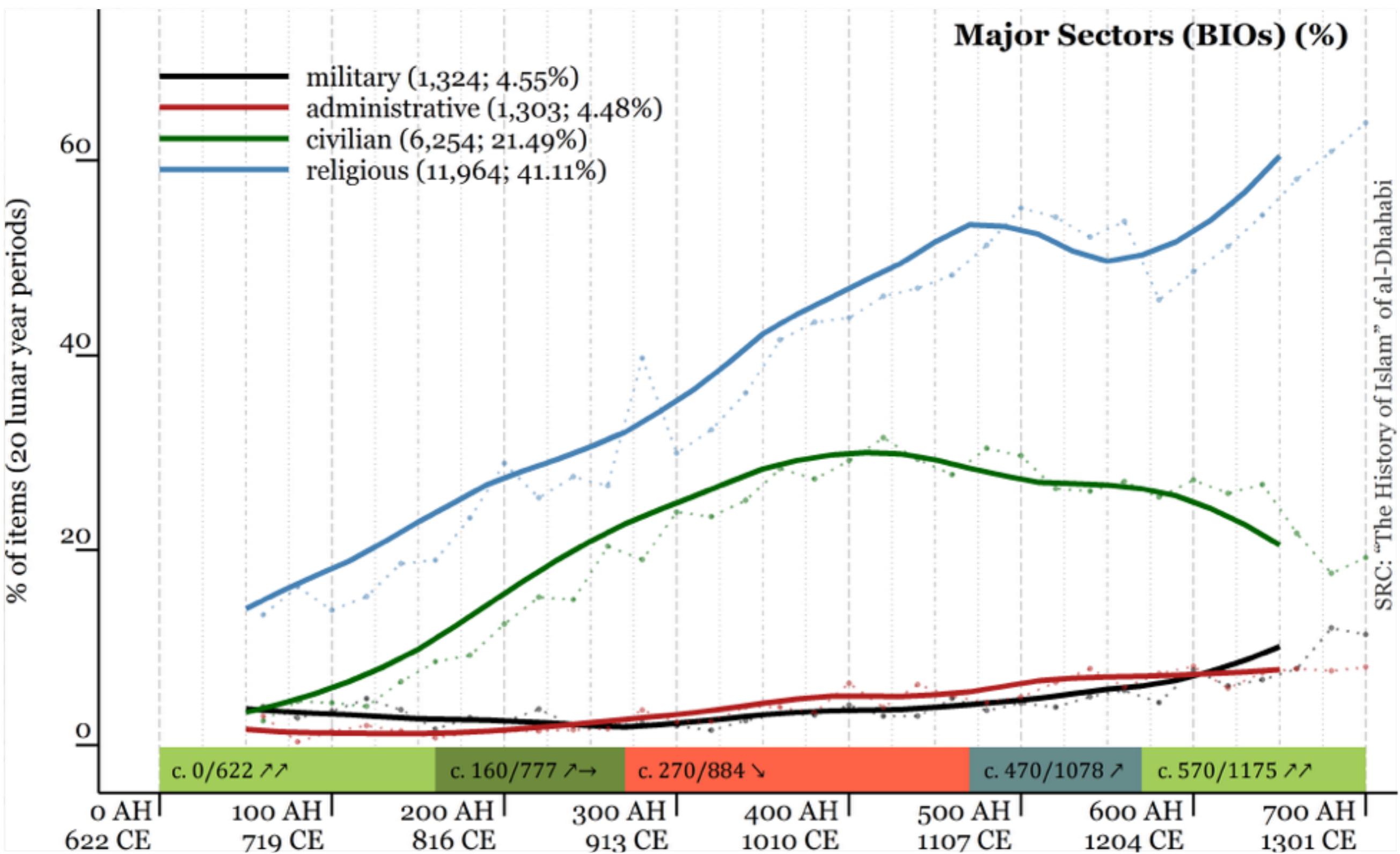
# Social transformations

De-tribalization

Militarization

Professionalization / Institutionalization

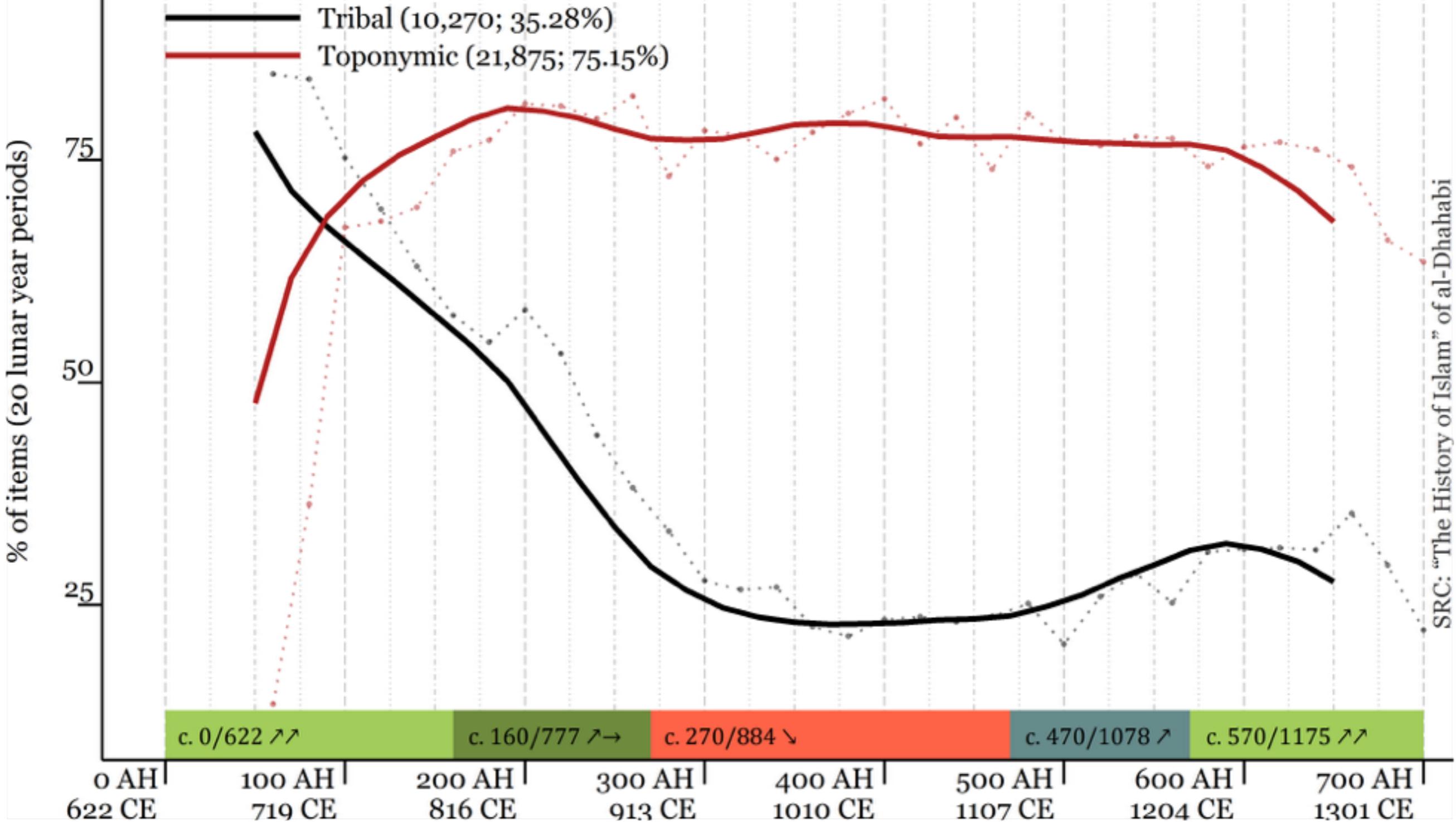
## Major Sectors (BIOS) (%)



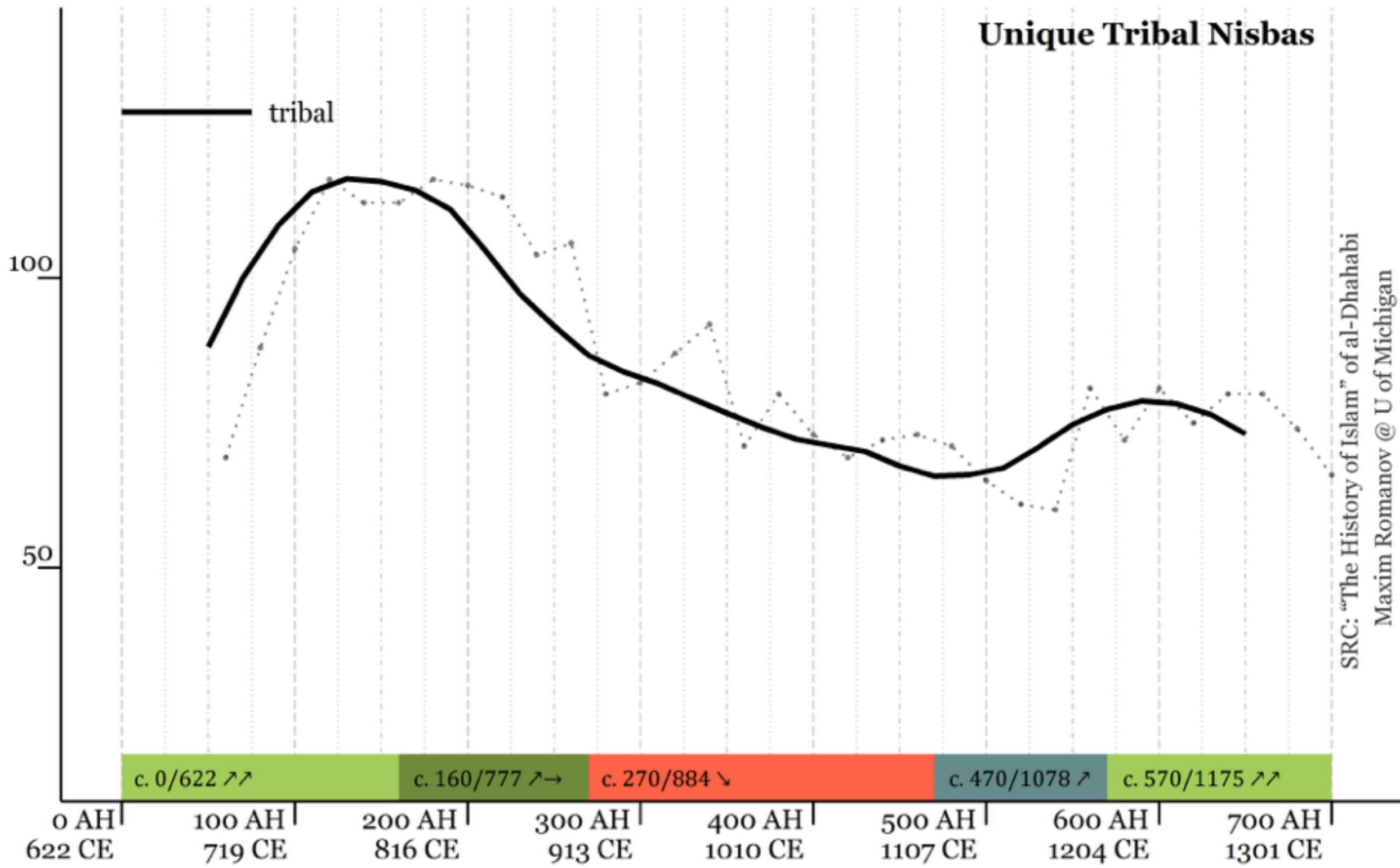


# De-tribalization

## Tribal and Toponymic Identities : BIO (%)



# of items (20 lunar year periods)



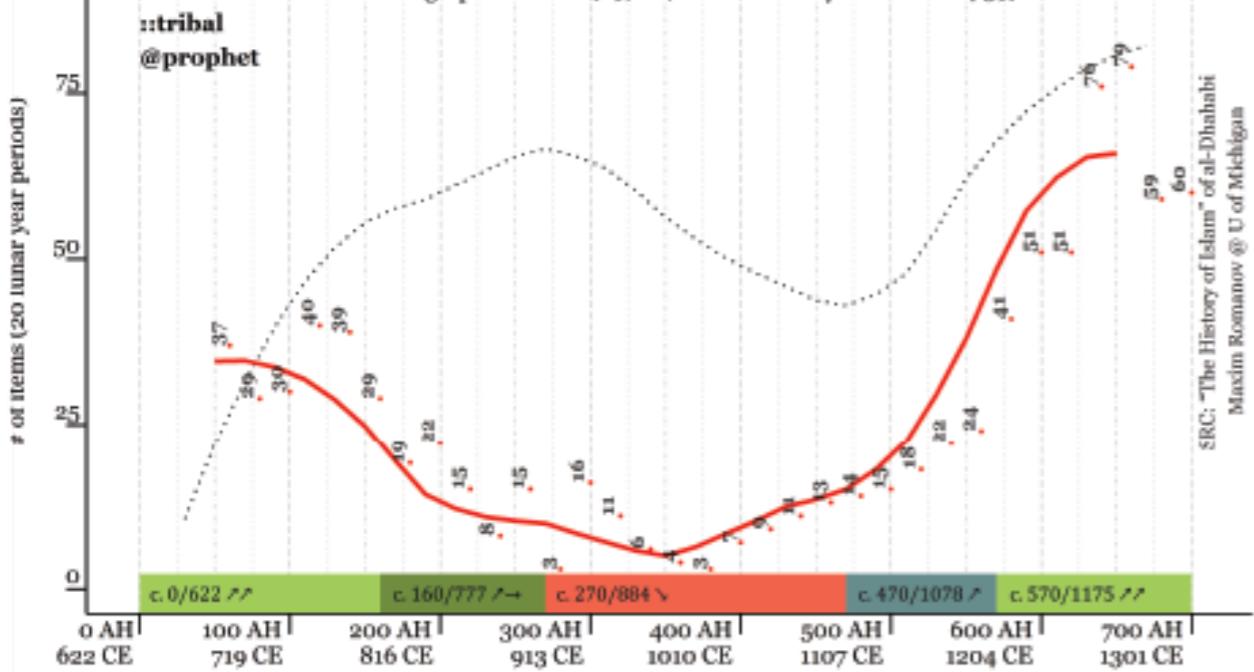
## Unique Tribal Nisbas

SRC: "The History of Islam" of al-Dhahabi

Maxim Romanov @ U of Michigan

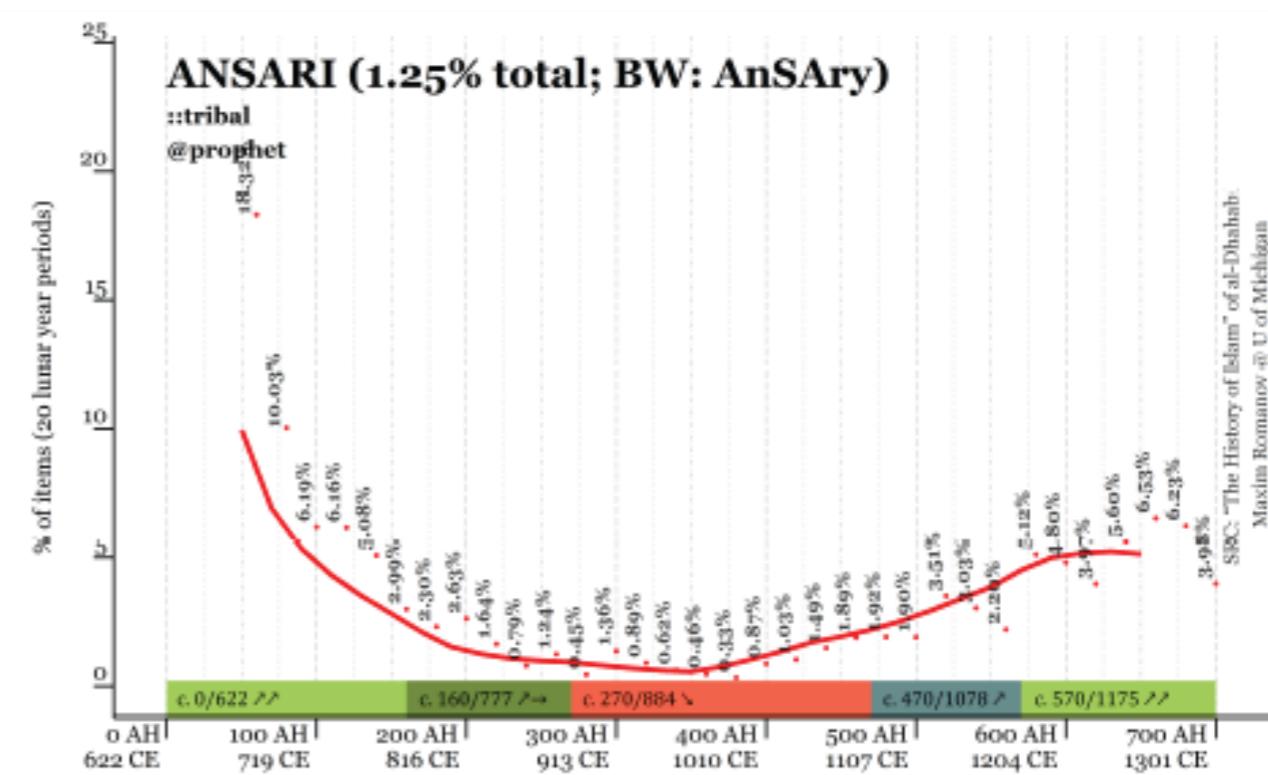
## ANSARI (876 total; BW: AnSAry)

Cumulative biographical curve (29,110; scaled down by a factor of 14.59)



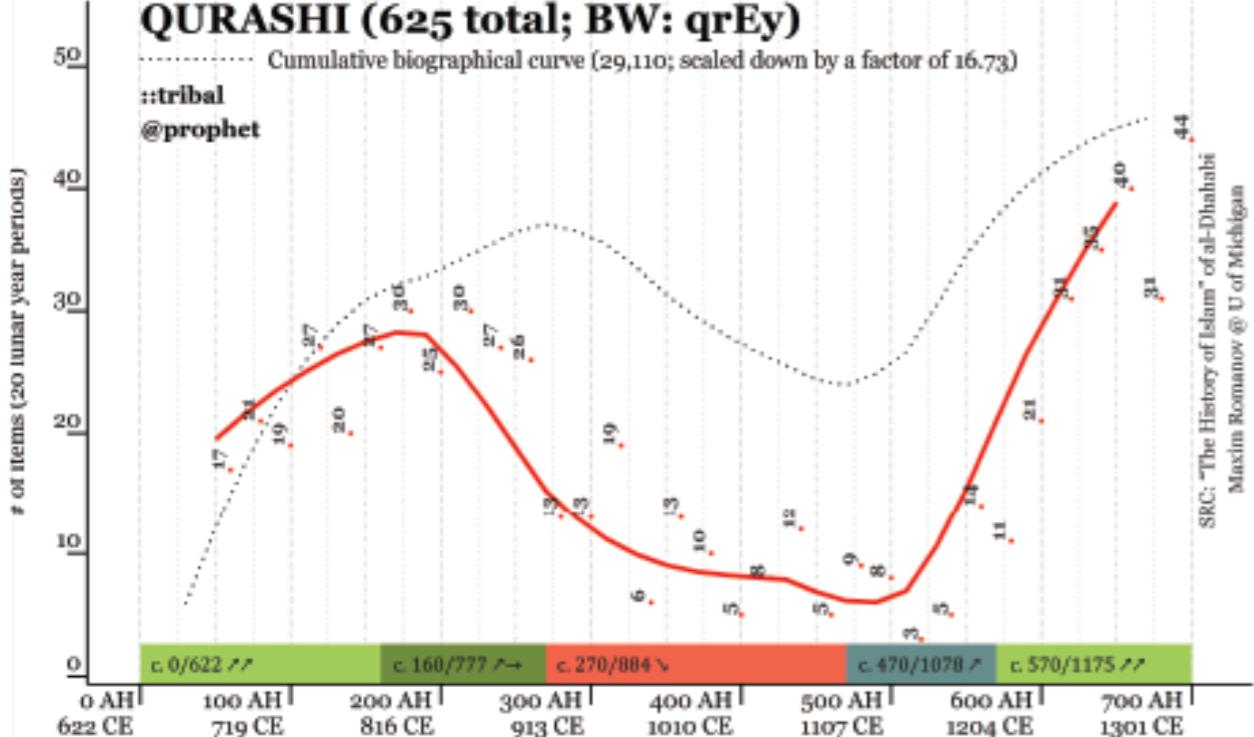
## ANSARI (1.25% total; BW: AnSAry)

::tribal  
@prophet



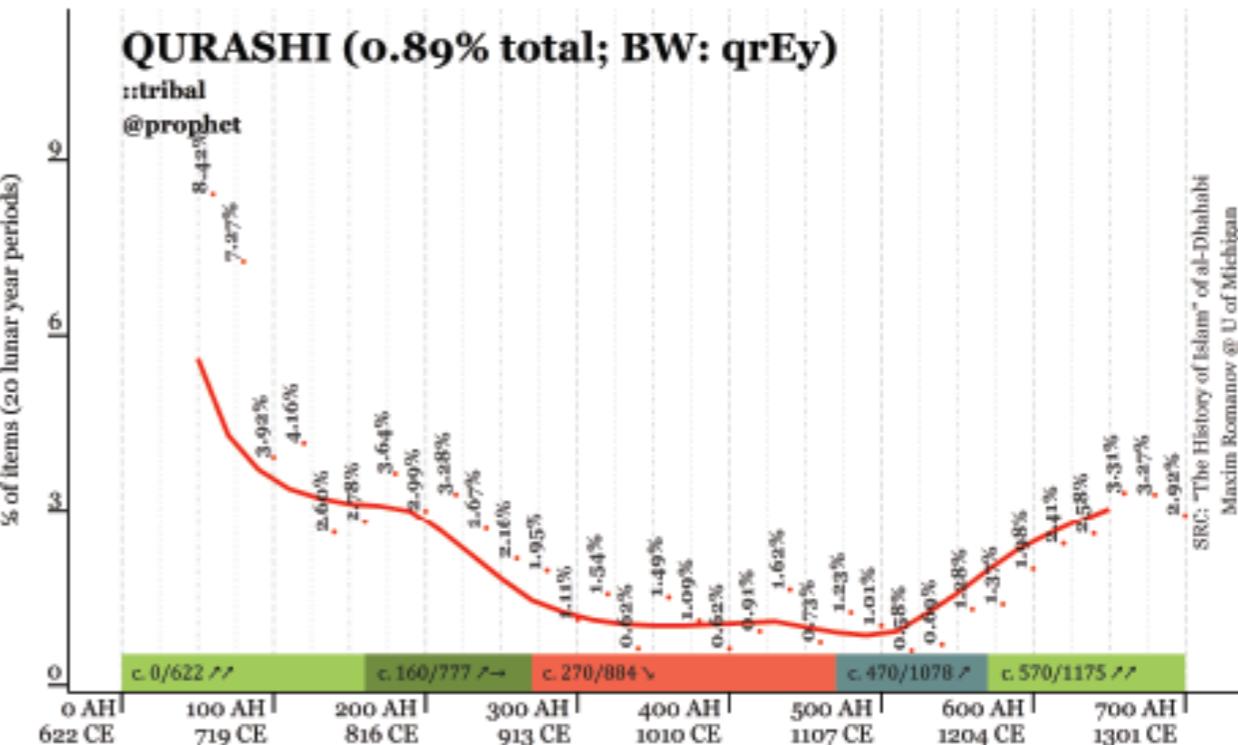
## QURASHI (625 total; BW: qrEy)

Cumulative biographical curve (29,110; scaled down by a factor of 16.73)



## QURASHI (0.89% total; BW: qrEy)

::tribal  
@prophet

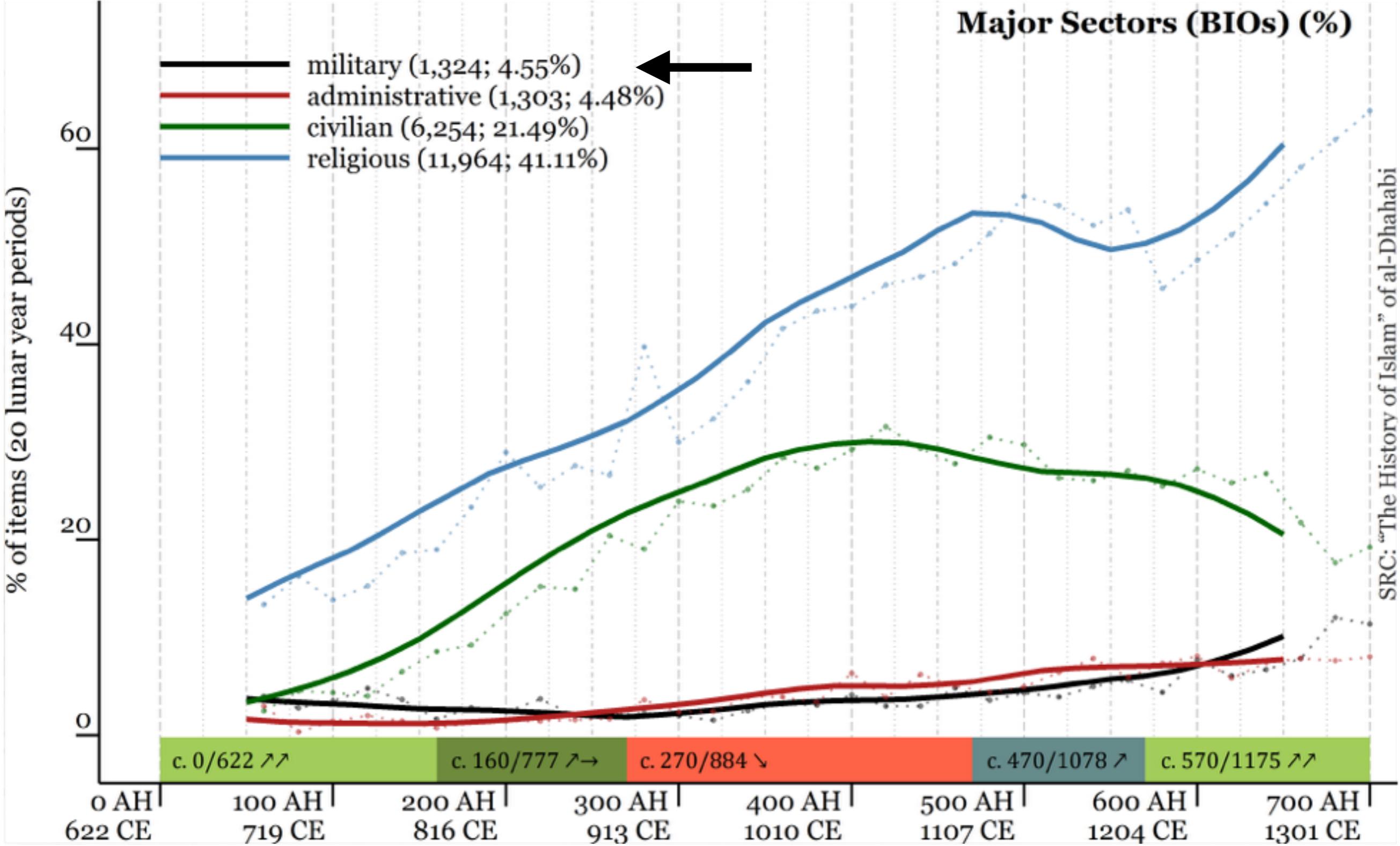


SRC: "The History of Islam" of al-Dhahabi  
Maxim Romanov © U of Michigan

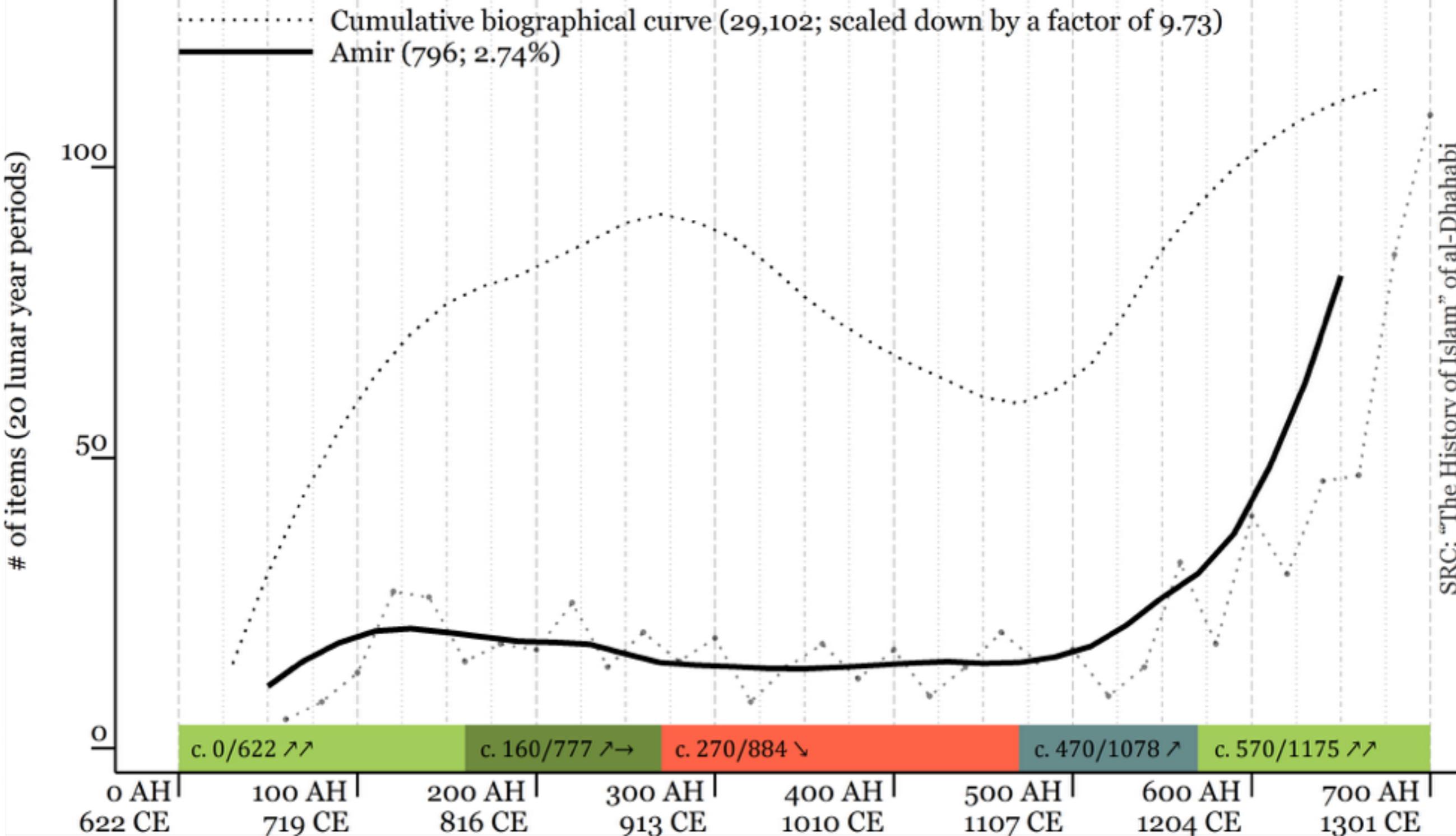
SRC: "The History of Islam" of al-Dhahabi  
Maxim Romanov © U of Michigan

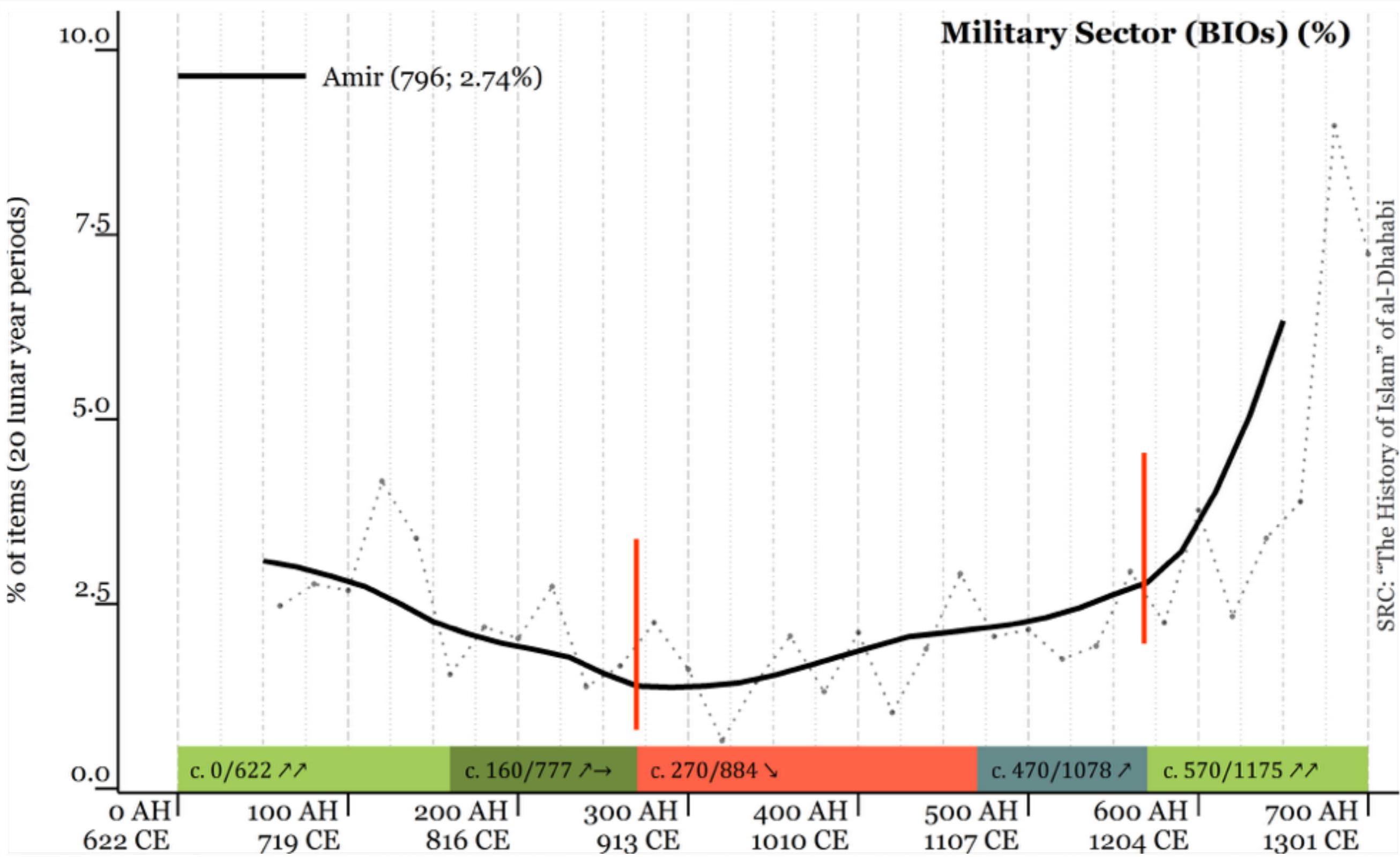
# Militarization

## Major Sectors (BIOS) (%)



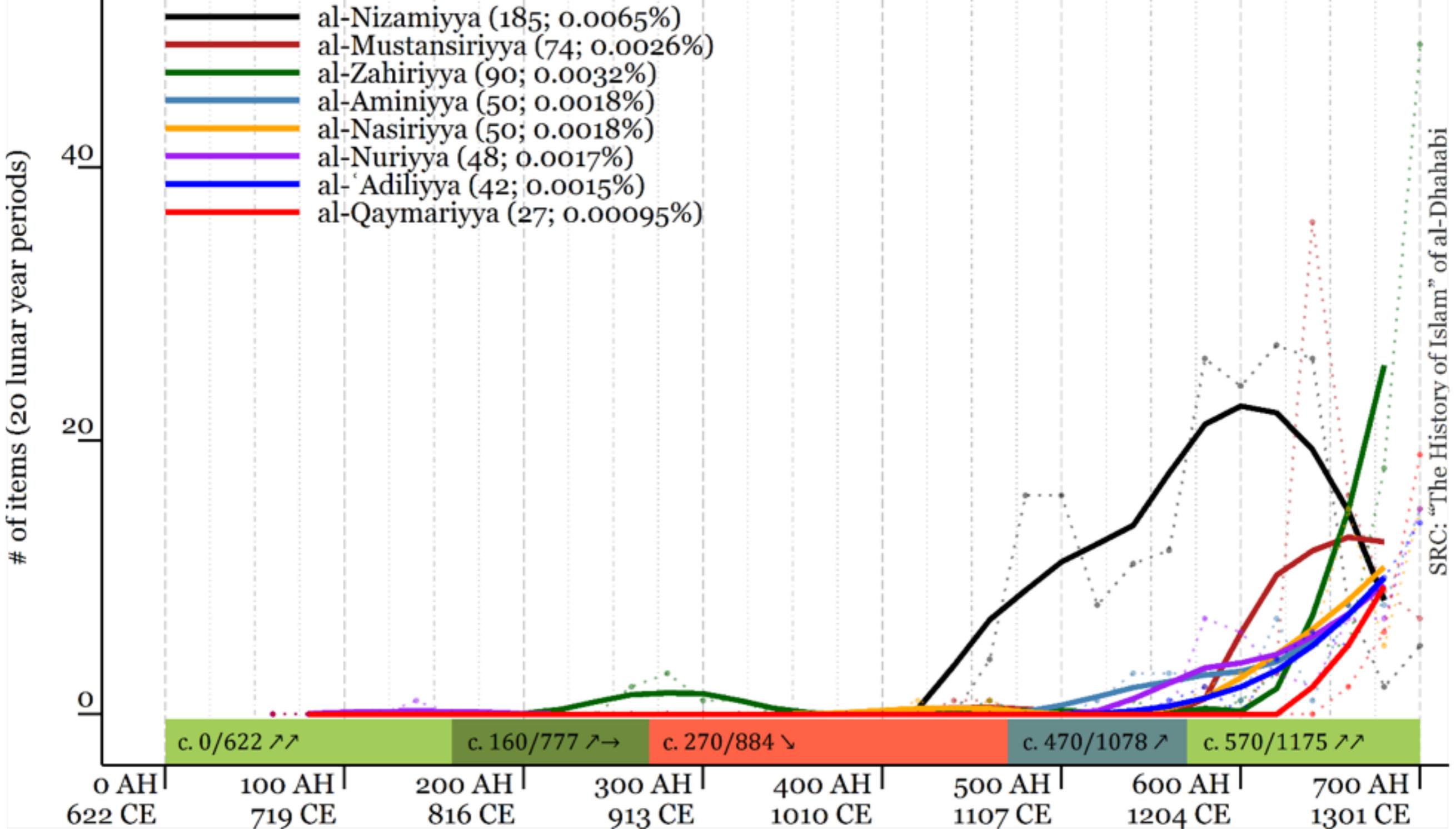
## Military Sector (BIOs)

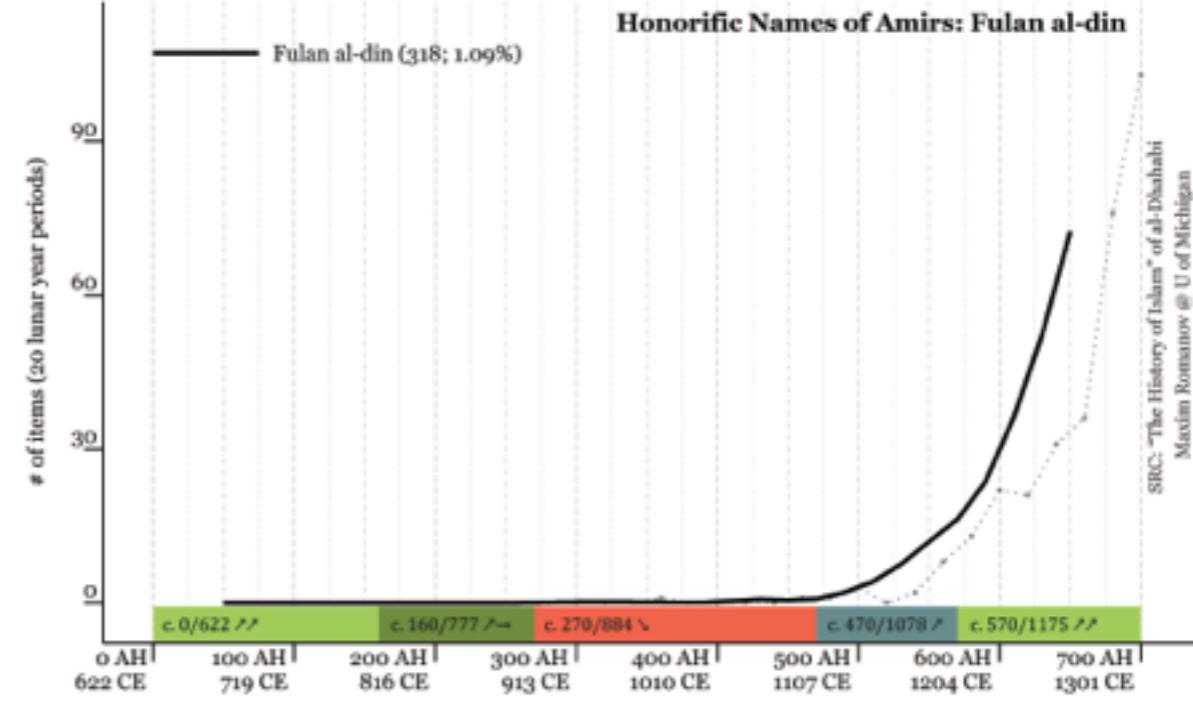
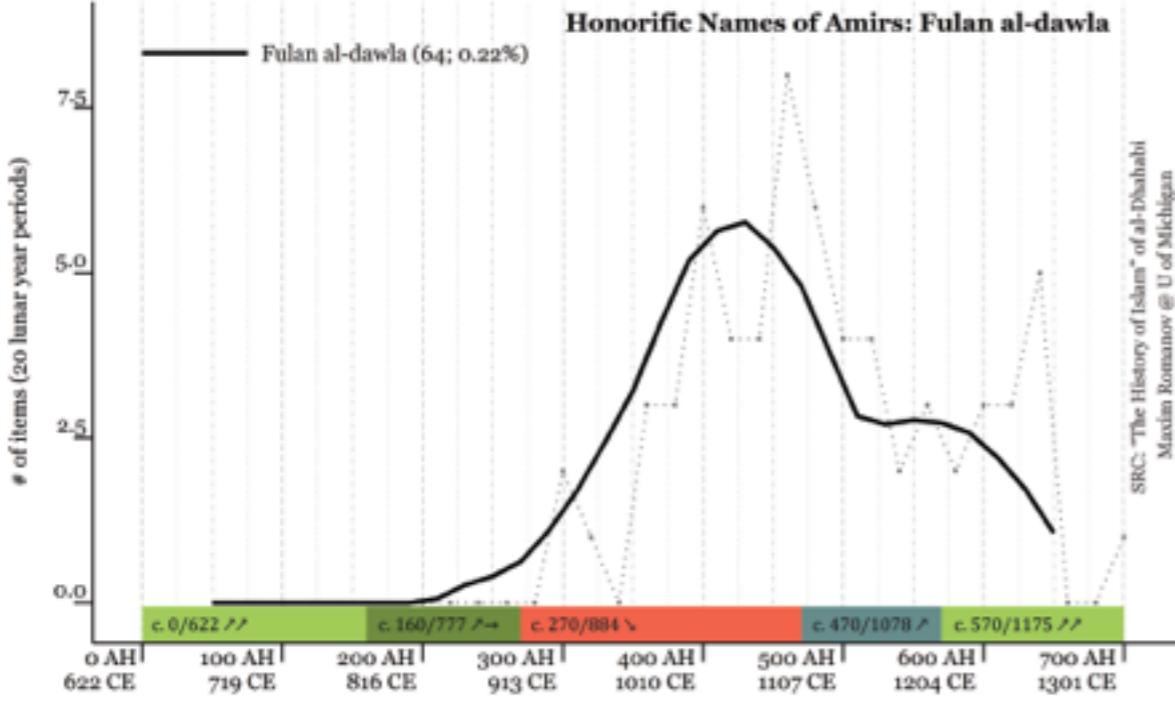




SRC: "The History of Islam" of al-Dhahabi

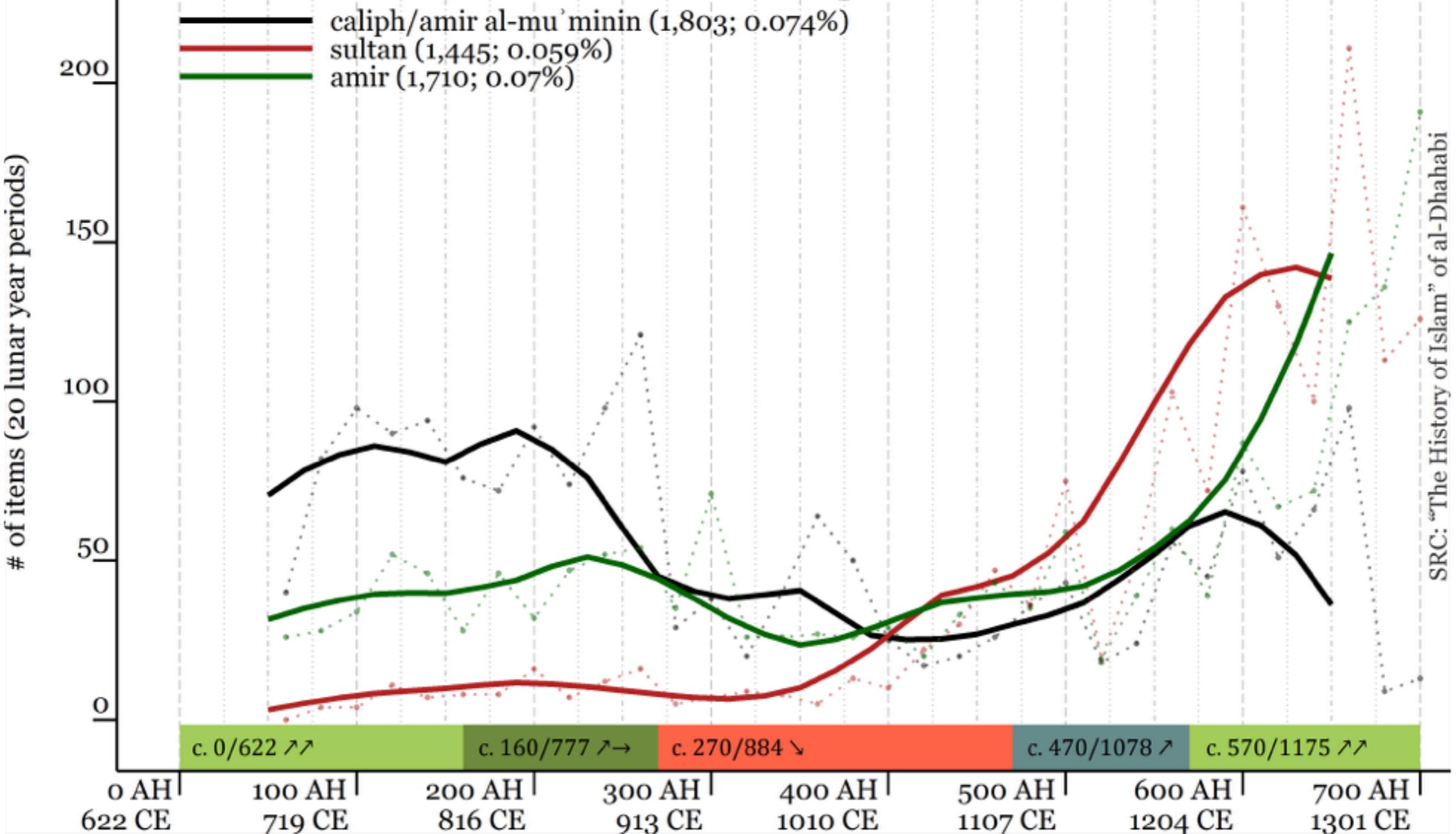
## Some of the Most Prominent Madrasas





Patterns of Military Honorific Names: Fulān al-dawla, the most common pattern in the middle period, gets replaced by Fulān al-dīn pattern in the later period

## Frequencies: CALIPH, SULTAN, AMIR

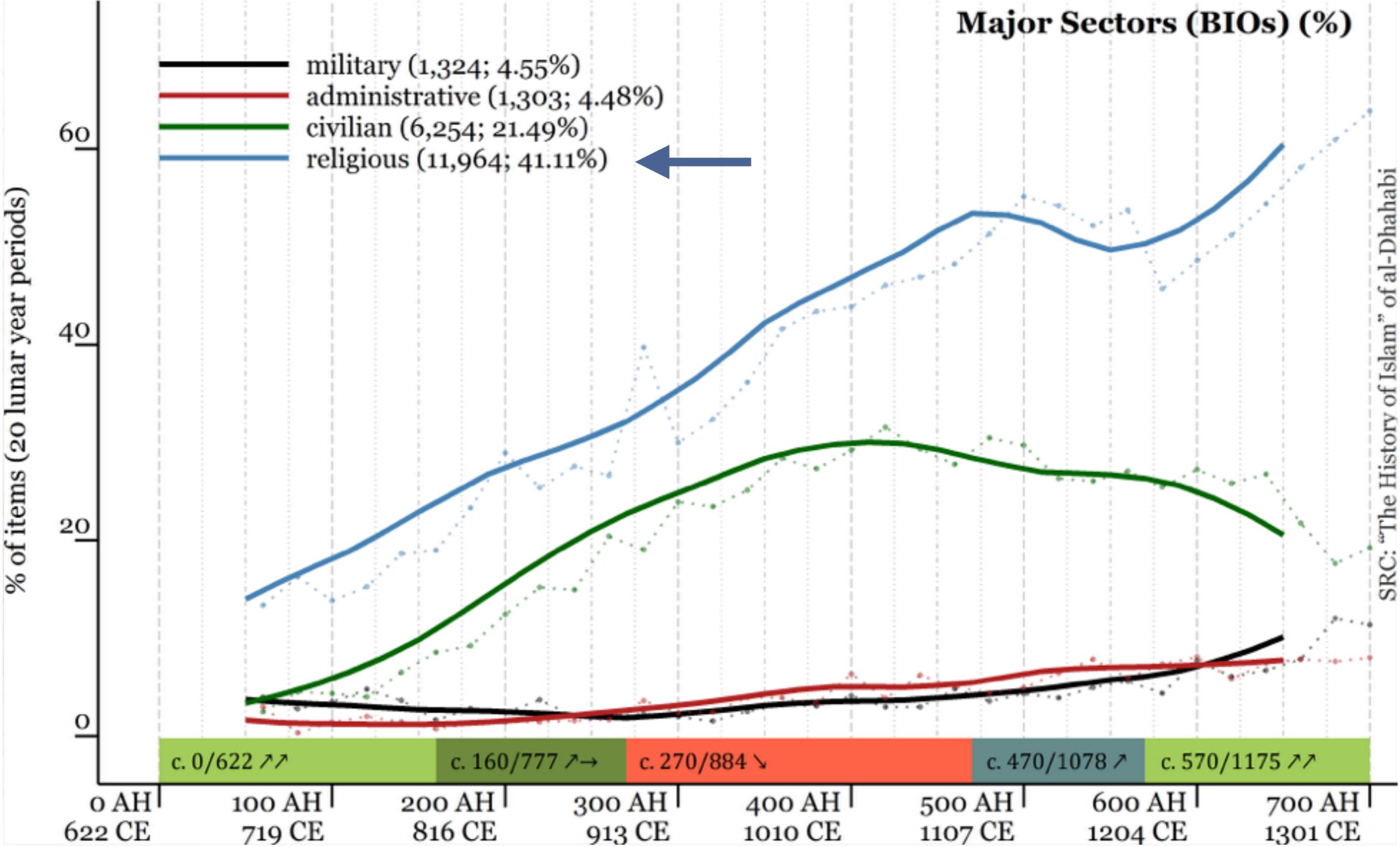




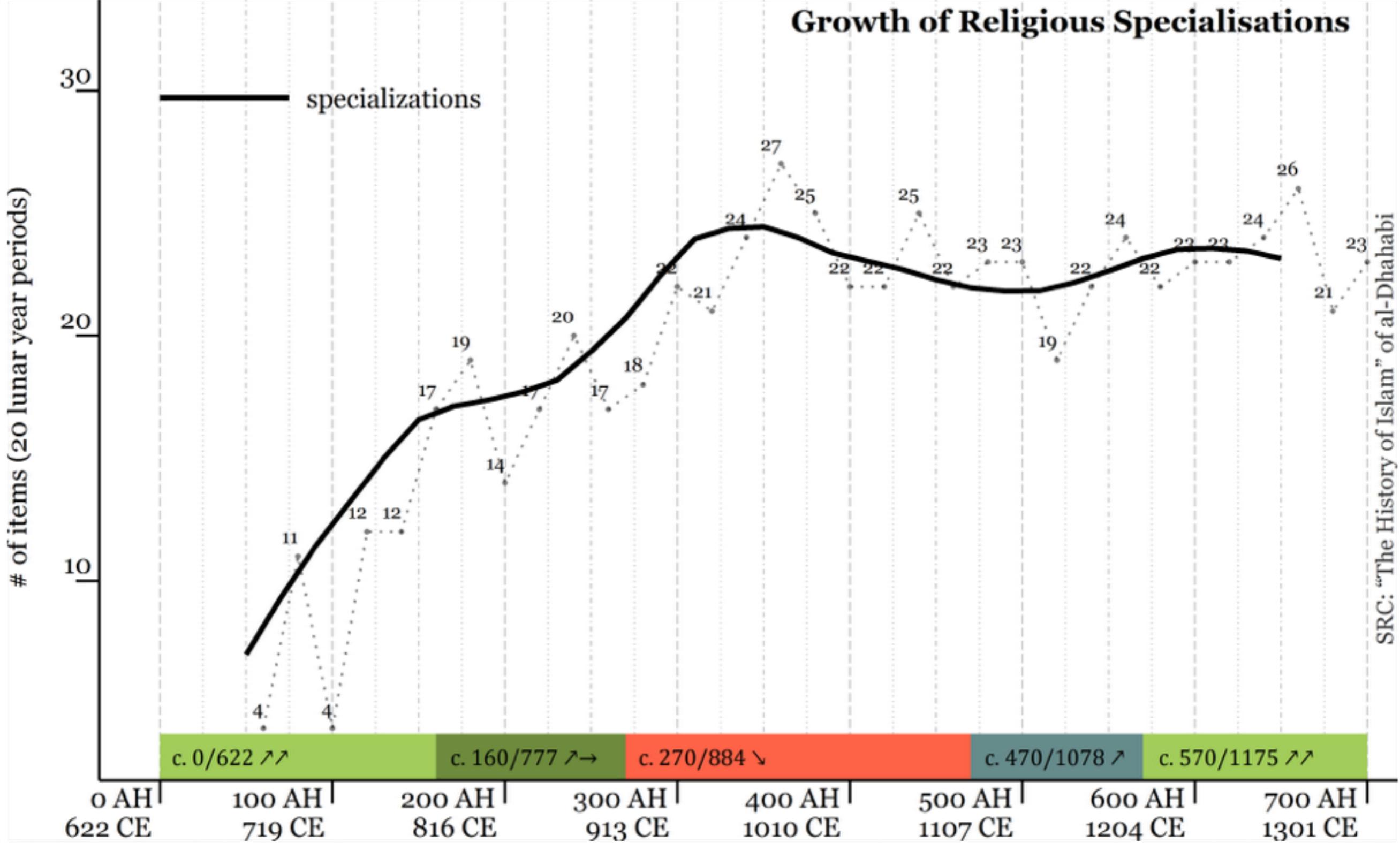
# Professionalization & Institutionalization of the Learned

العلماء

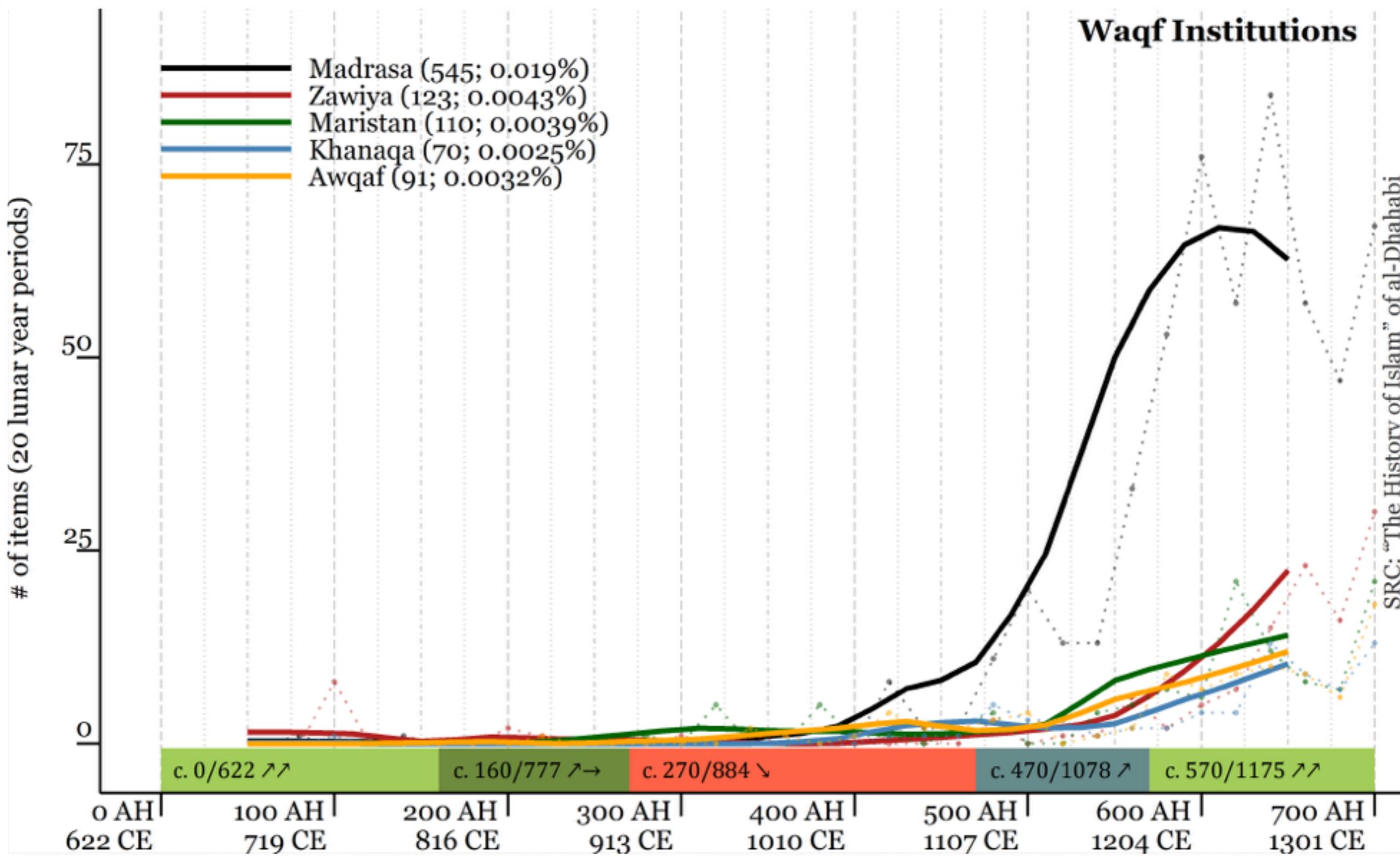
## Major Sectors (BIOS) (%)



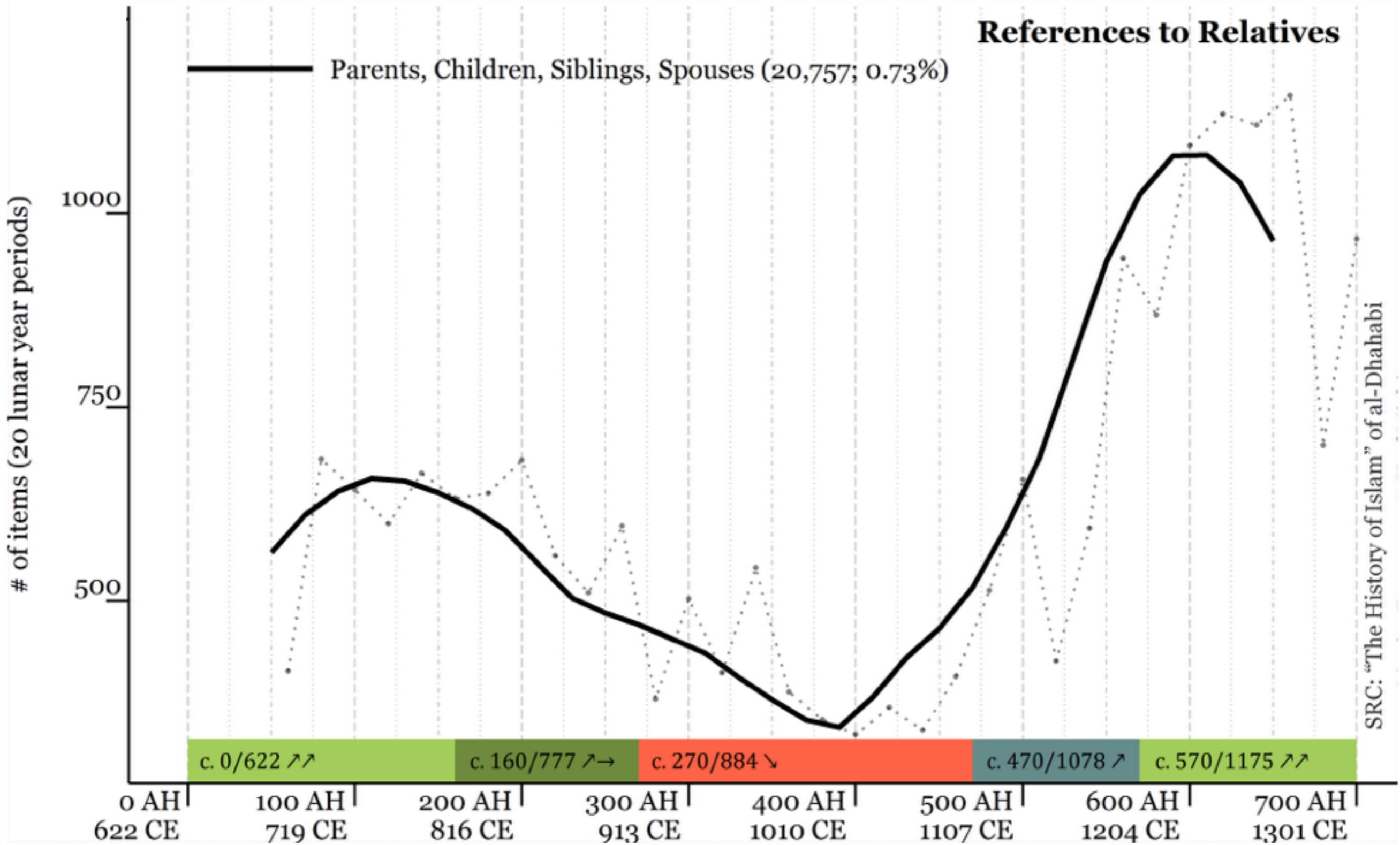
## Growth of Religious Specialisations



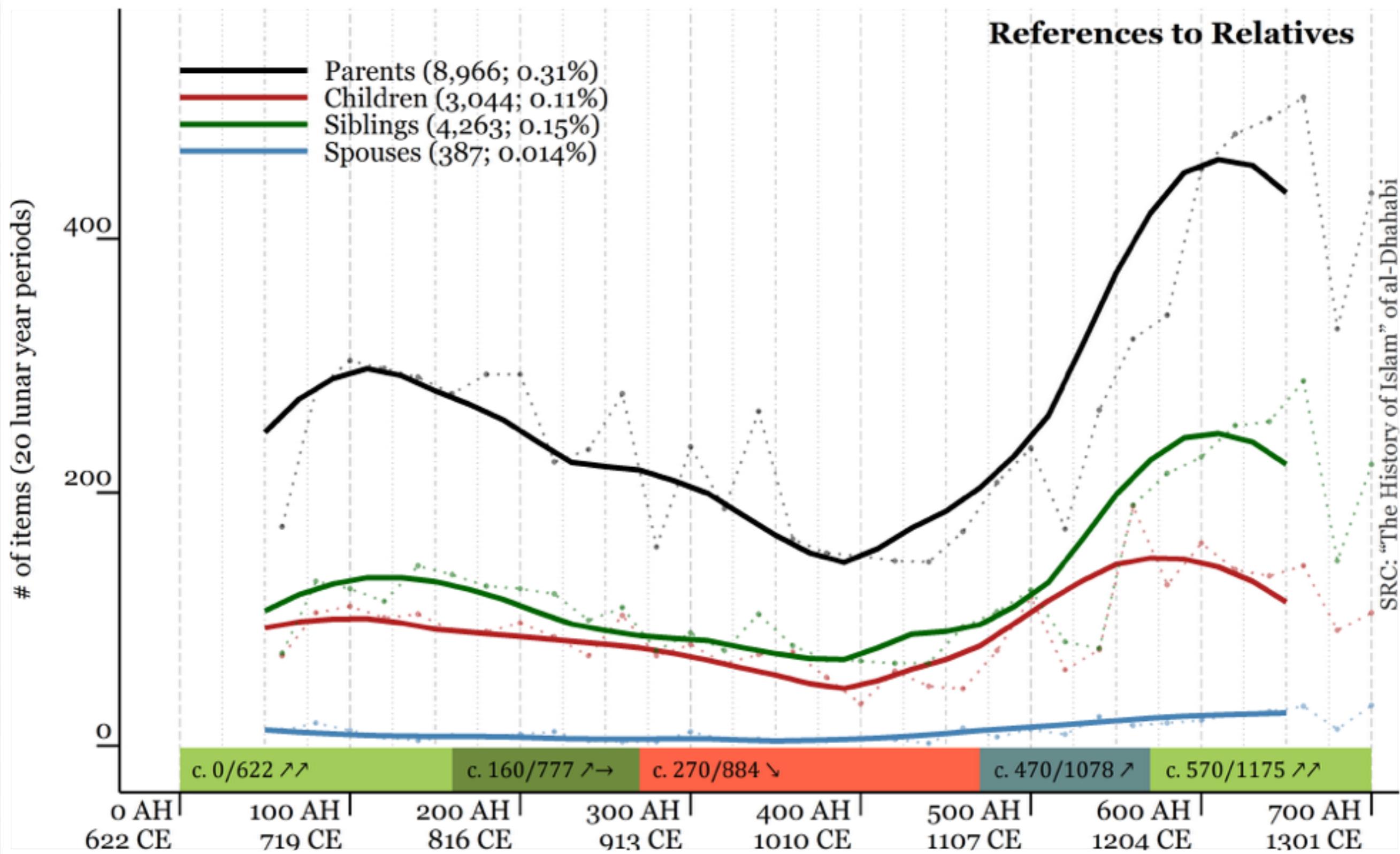
## Waqf Institutions



# Lineages & Households



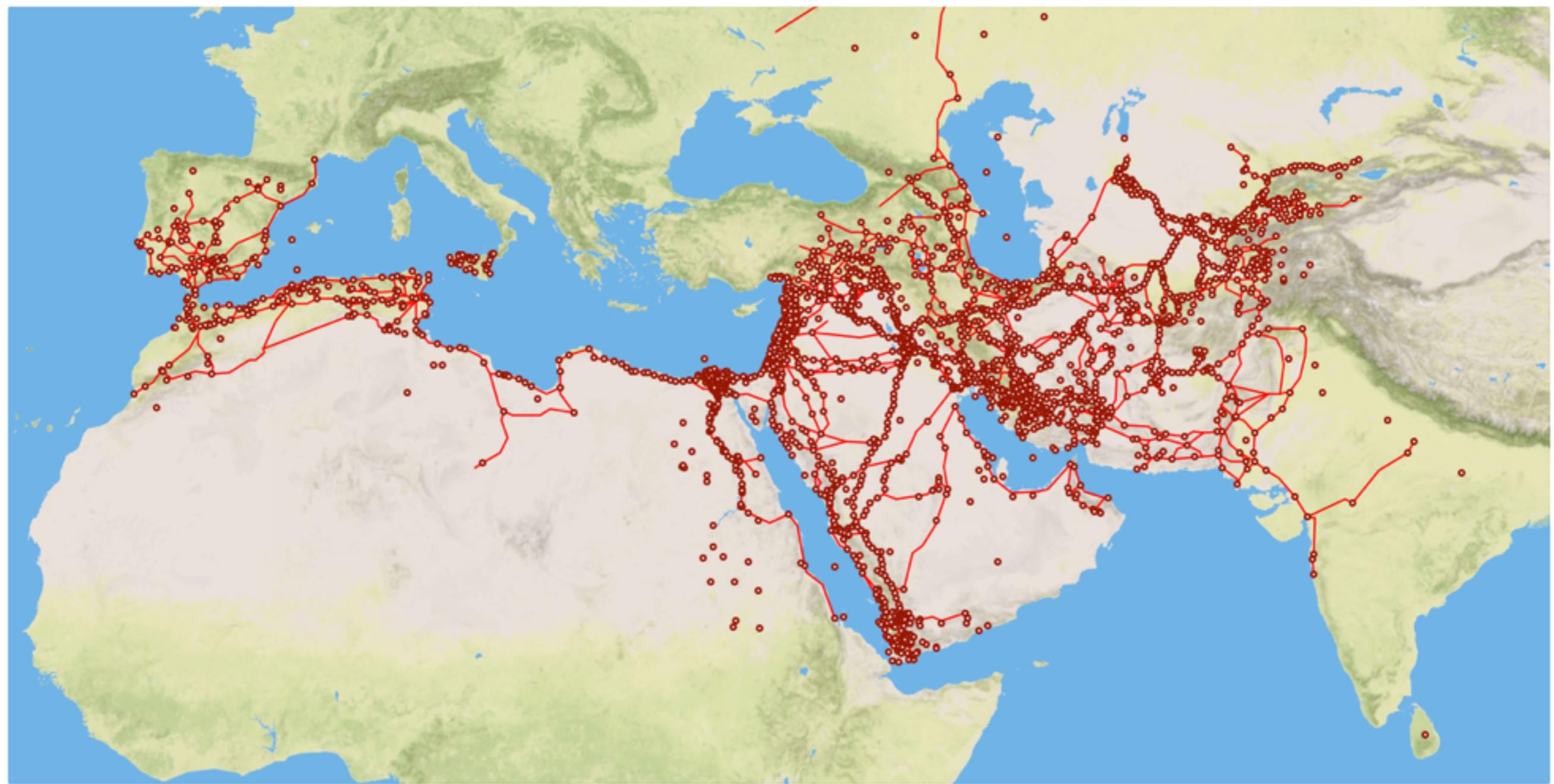
# Lineages & Households





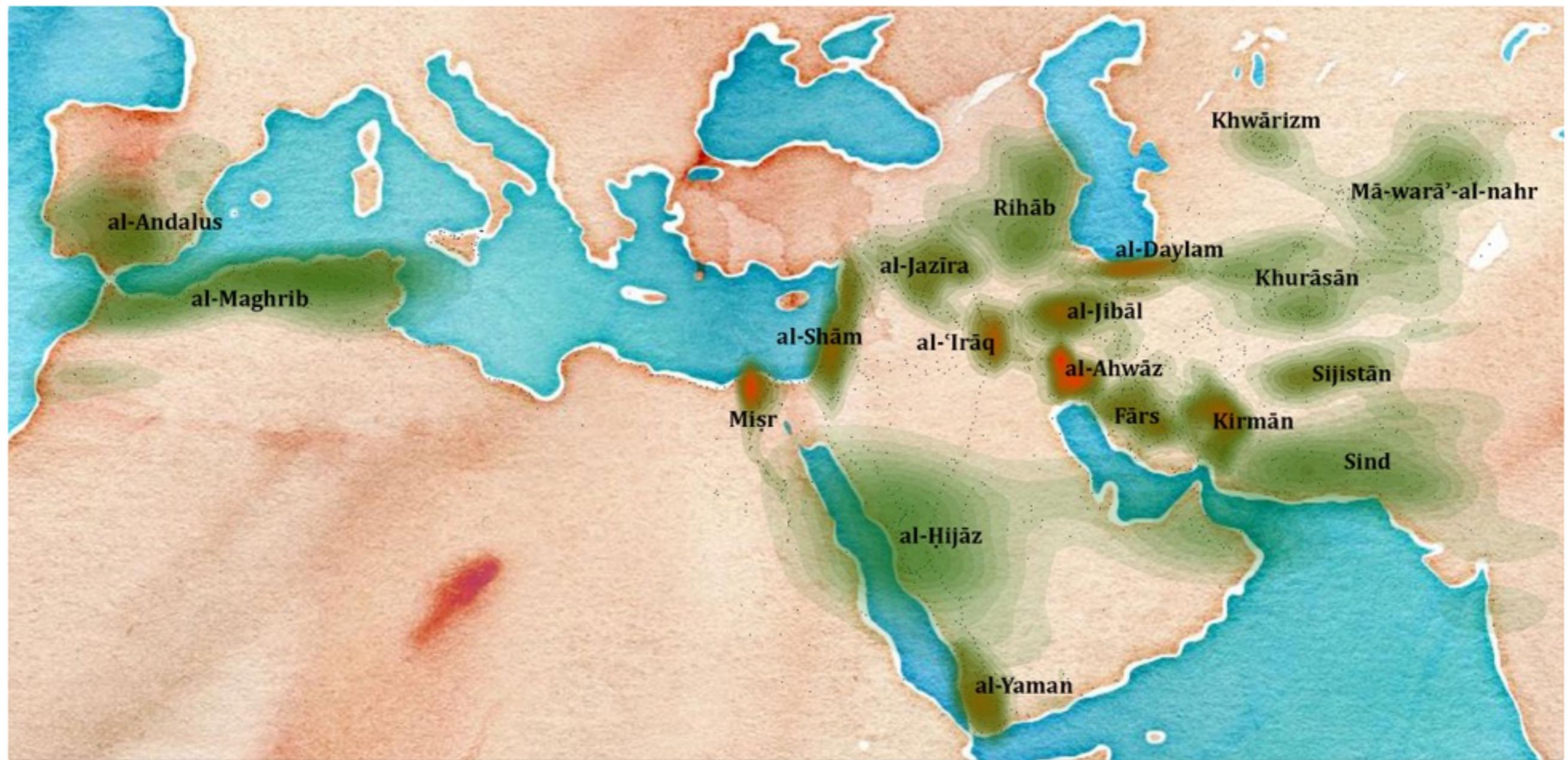
# Social Geography

# Charting Islamic World



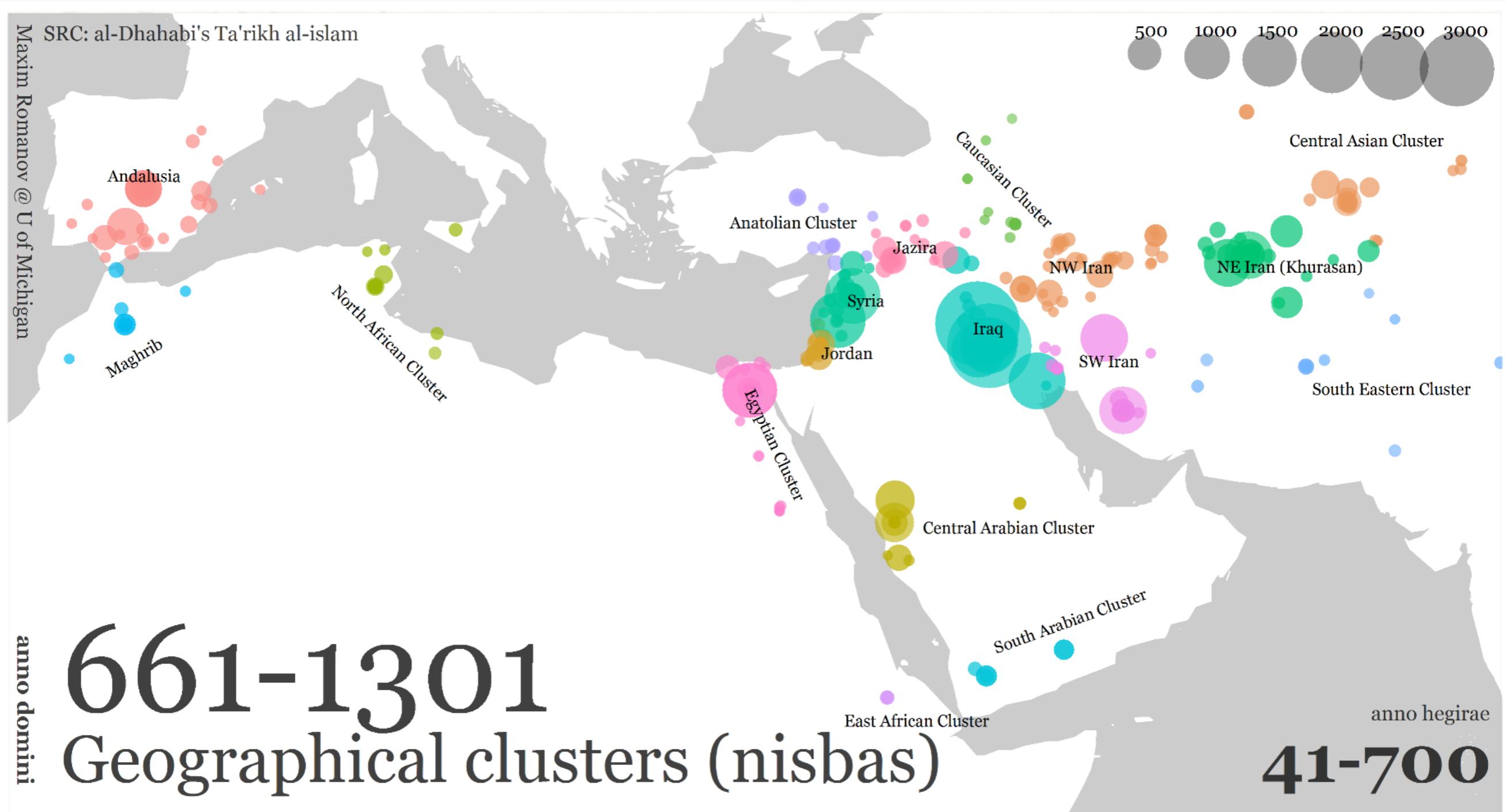
Source: Cornu, Georgette. *Atlas du monde arabo-islamique à l'époque classique: IXe-Xe siècles*. Leiden: E.J. Brill, 1983.

# Charting Islamic World

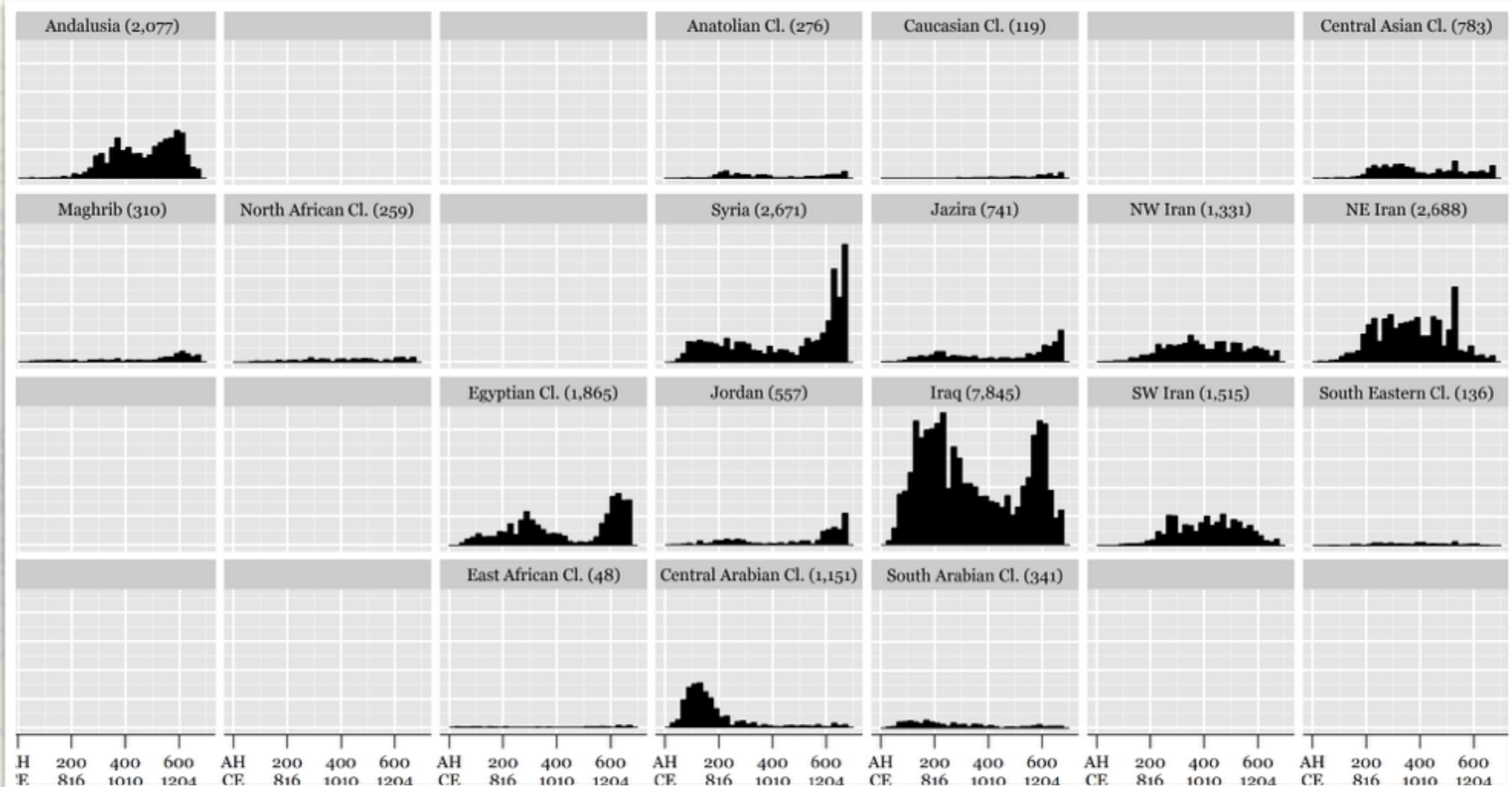


Source: Cornu, Georgette. *Atlas du monde arabo-islamique à l'époque classique: IXe-Xe siècles*. Leiden: E.J. Brill, 1983.

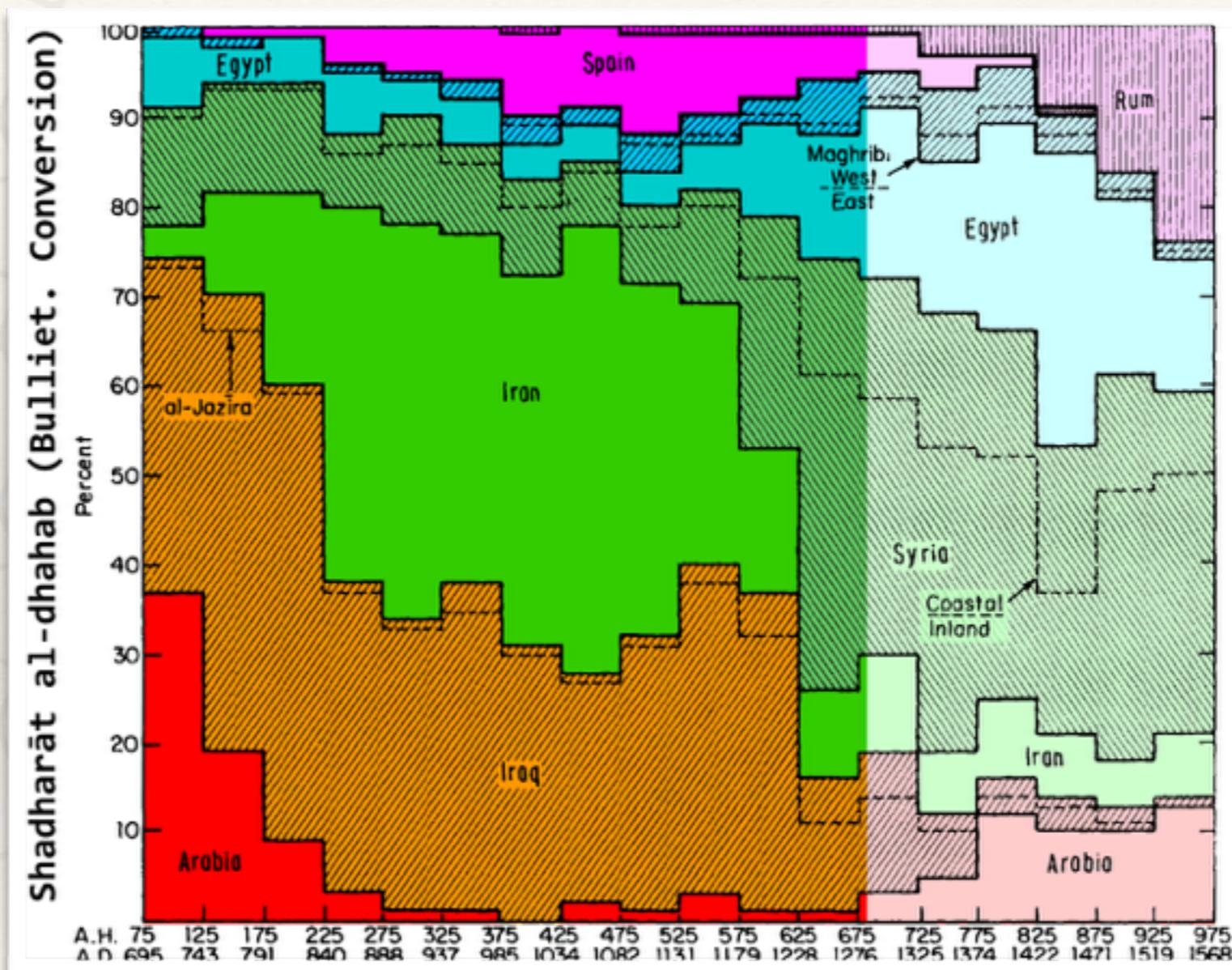
# Geographical Coverage



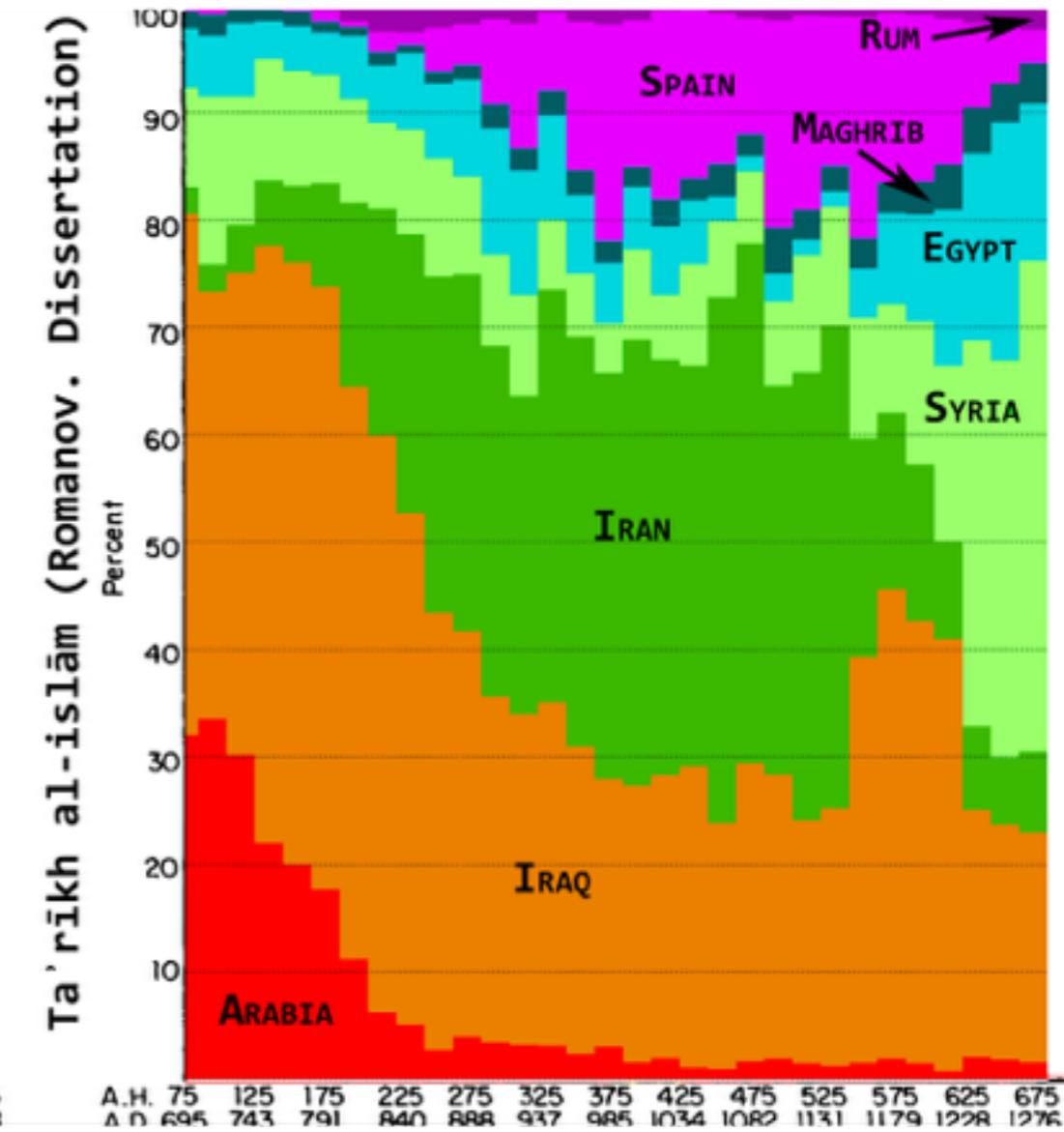
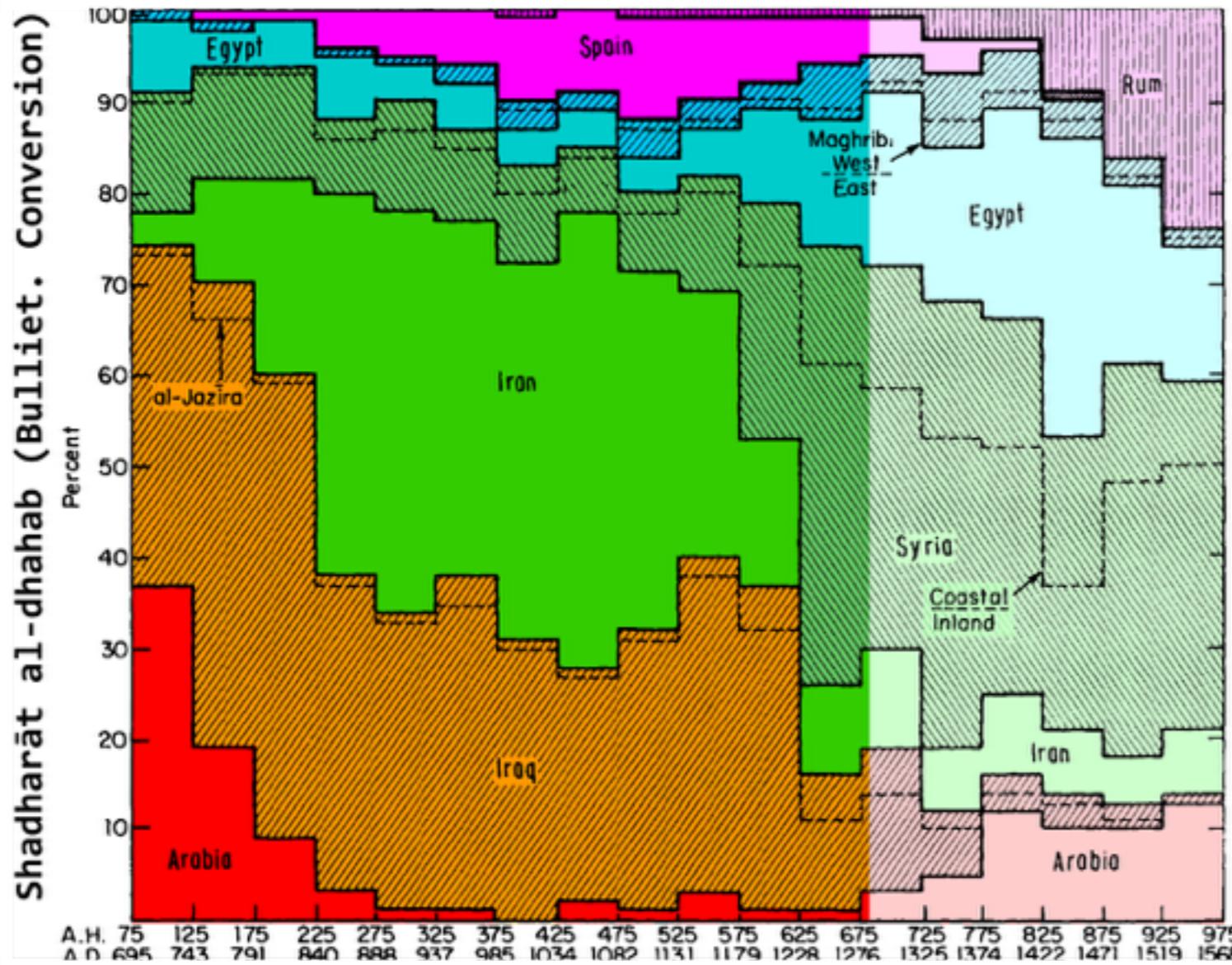
# Geo-Chronological Coverage



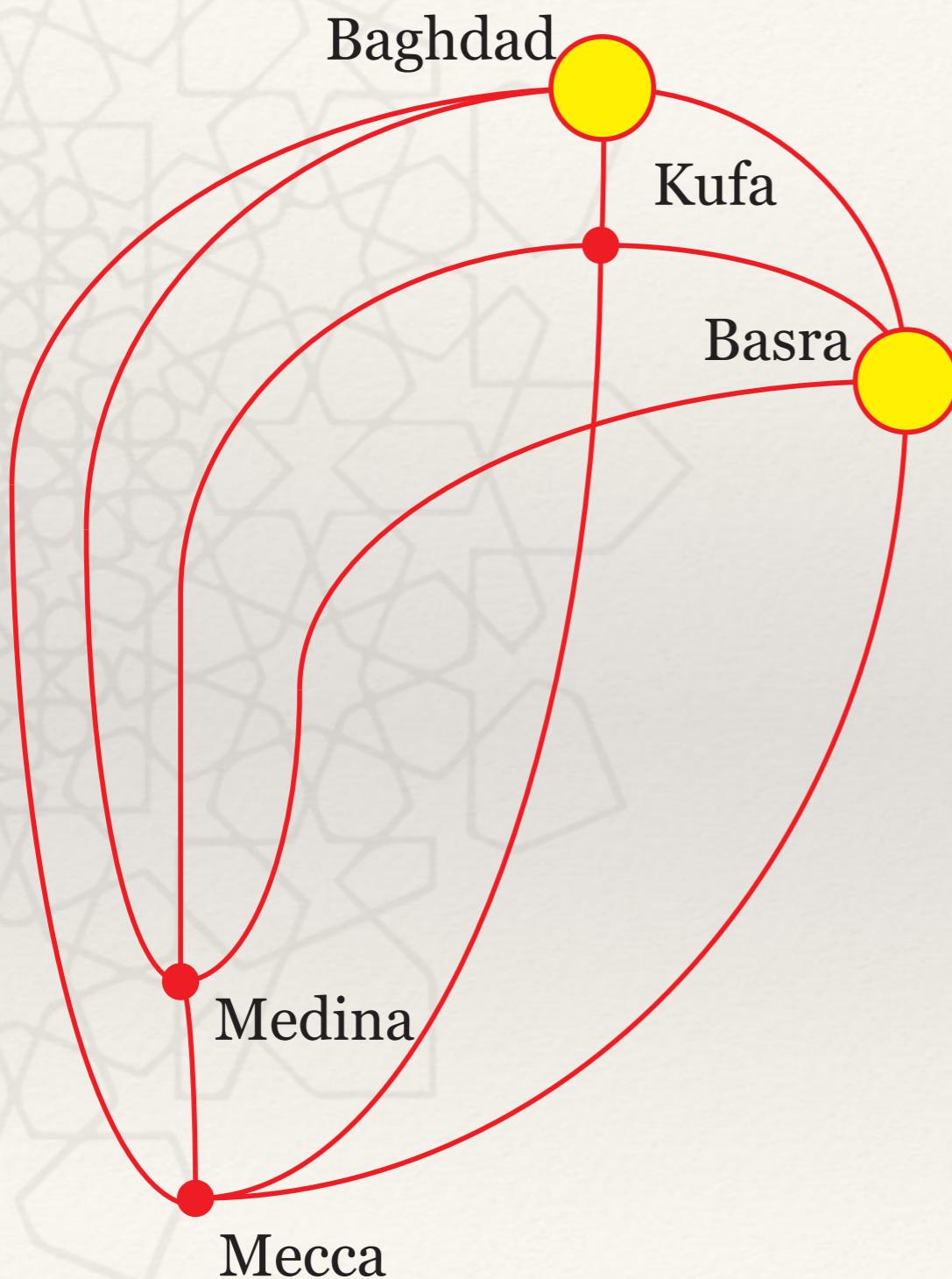
# Geo-Chronological Coverage: *Comparative Perspective*



# Geo-Chronological Coverage: Comparative Perspective



# Modeling Geographical Connections



- ❖ A cartogram of a person—*al-Bağdādī* and *al-Basrī*—whose biography mentions Baghdad, Kufa, Basra, Medina and Mecca
- ❖ Such data can be grouped to show particular groups and/or periods



# The Core

# 0632-0680 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 40-90AH/661-709CE

380 760 1140 1530 1910 2290



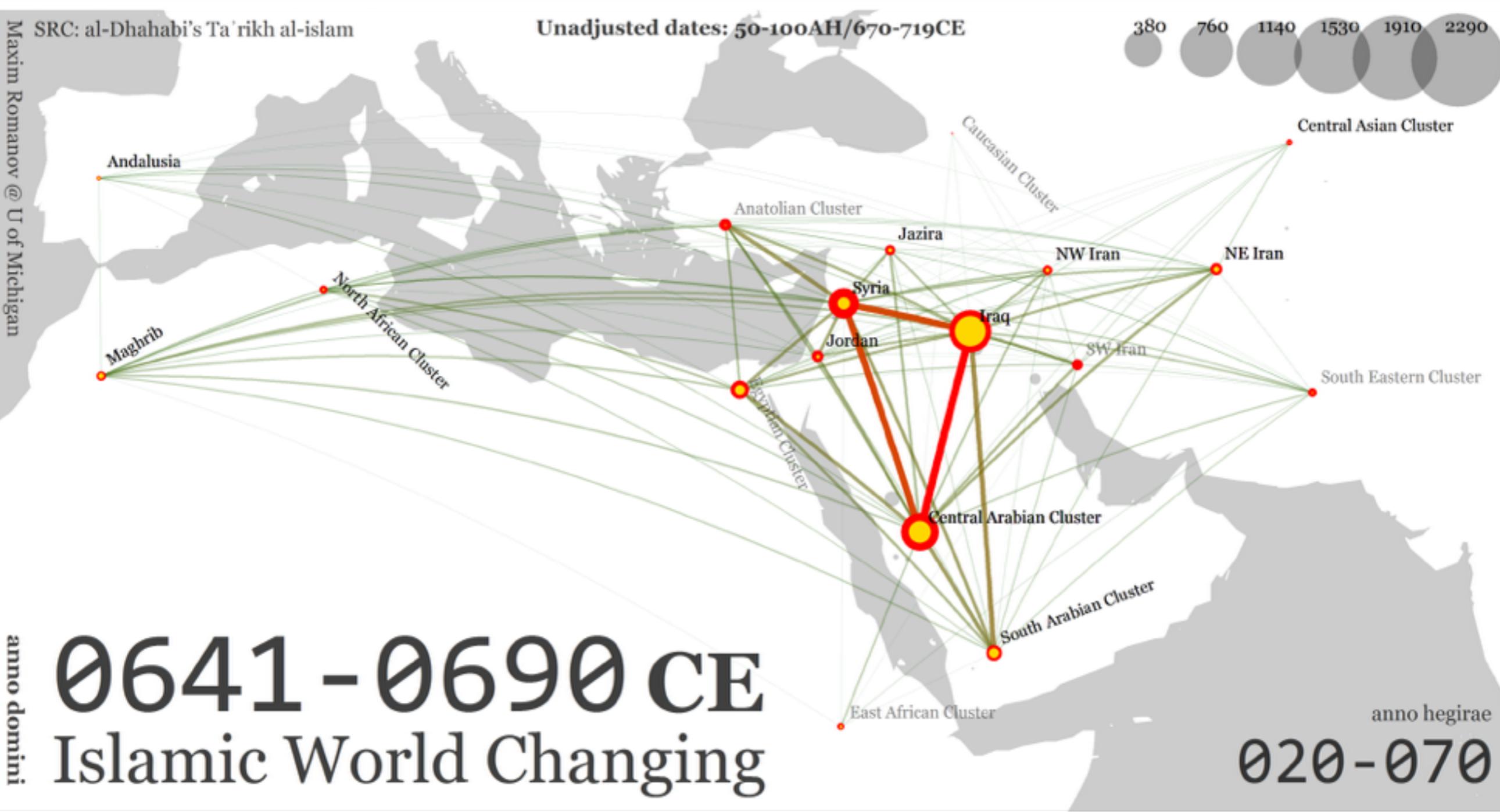
anno hegirae

010-060

# 0641 - 0690 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 50-100AH/670-719CE



anno hegirae

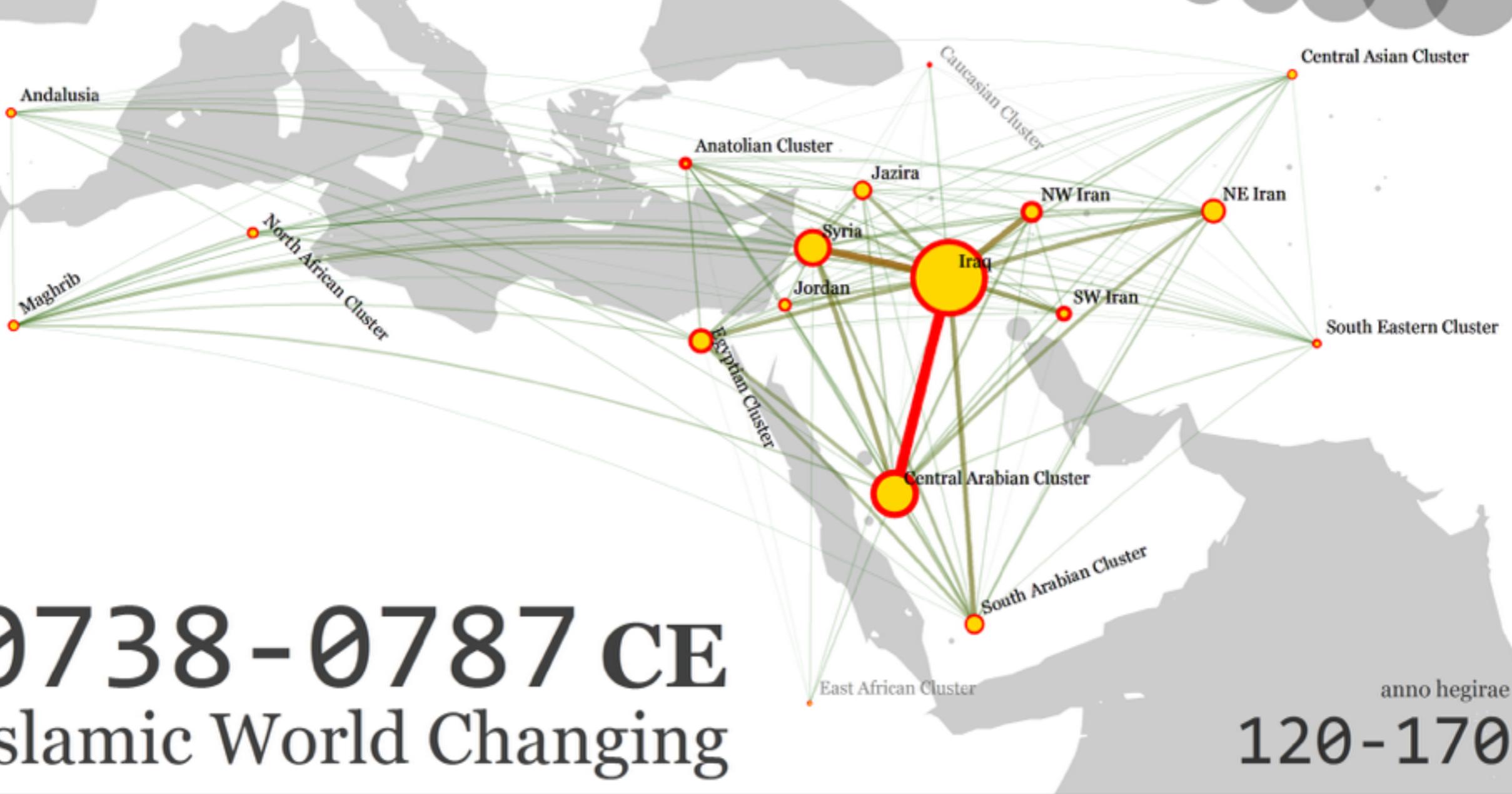
020-070

0738-0787 CE  
Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 150-200AH/767-816CE

380 760 1140 1530 1910 2290



anno hegirae

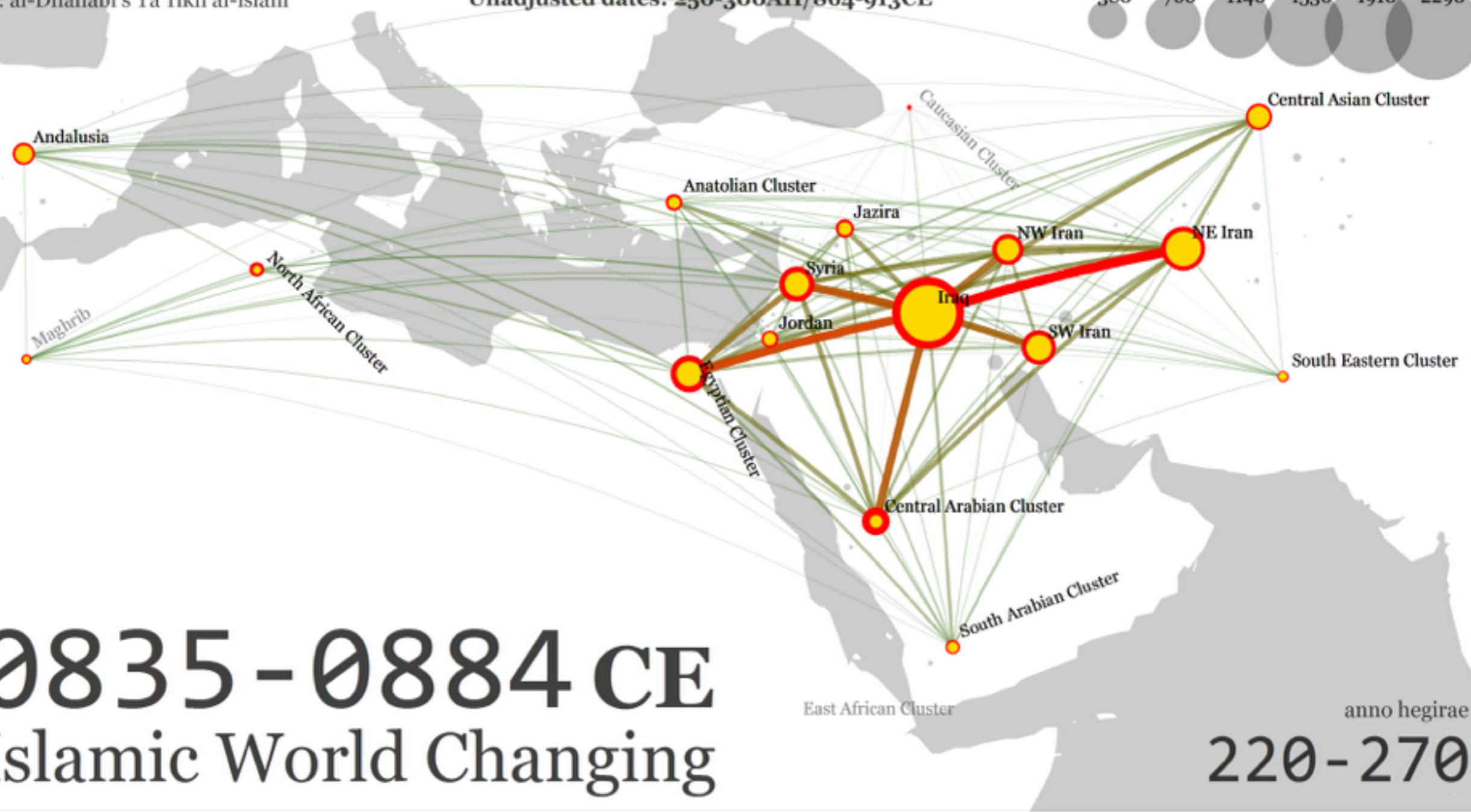
120-170

# 0835 - 0884 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 250-300AH/864-913CE

380 760 1140 1530 1910 2290



anno hegirae

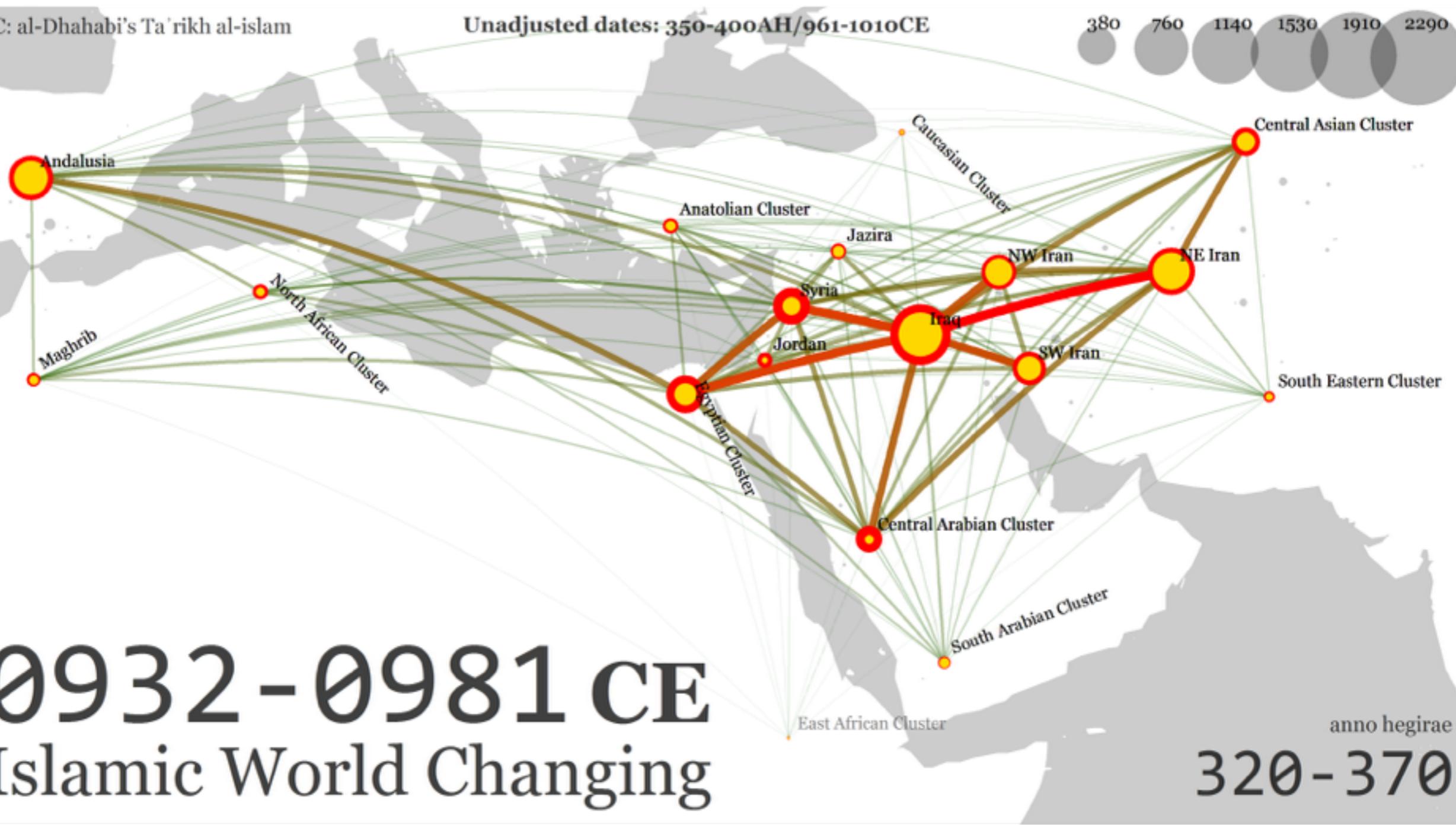
220-270

# 0932-0981 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 350-400AH/961-1010CE

380 760 1140 1530 1910 2290



anno hegirae

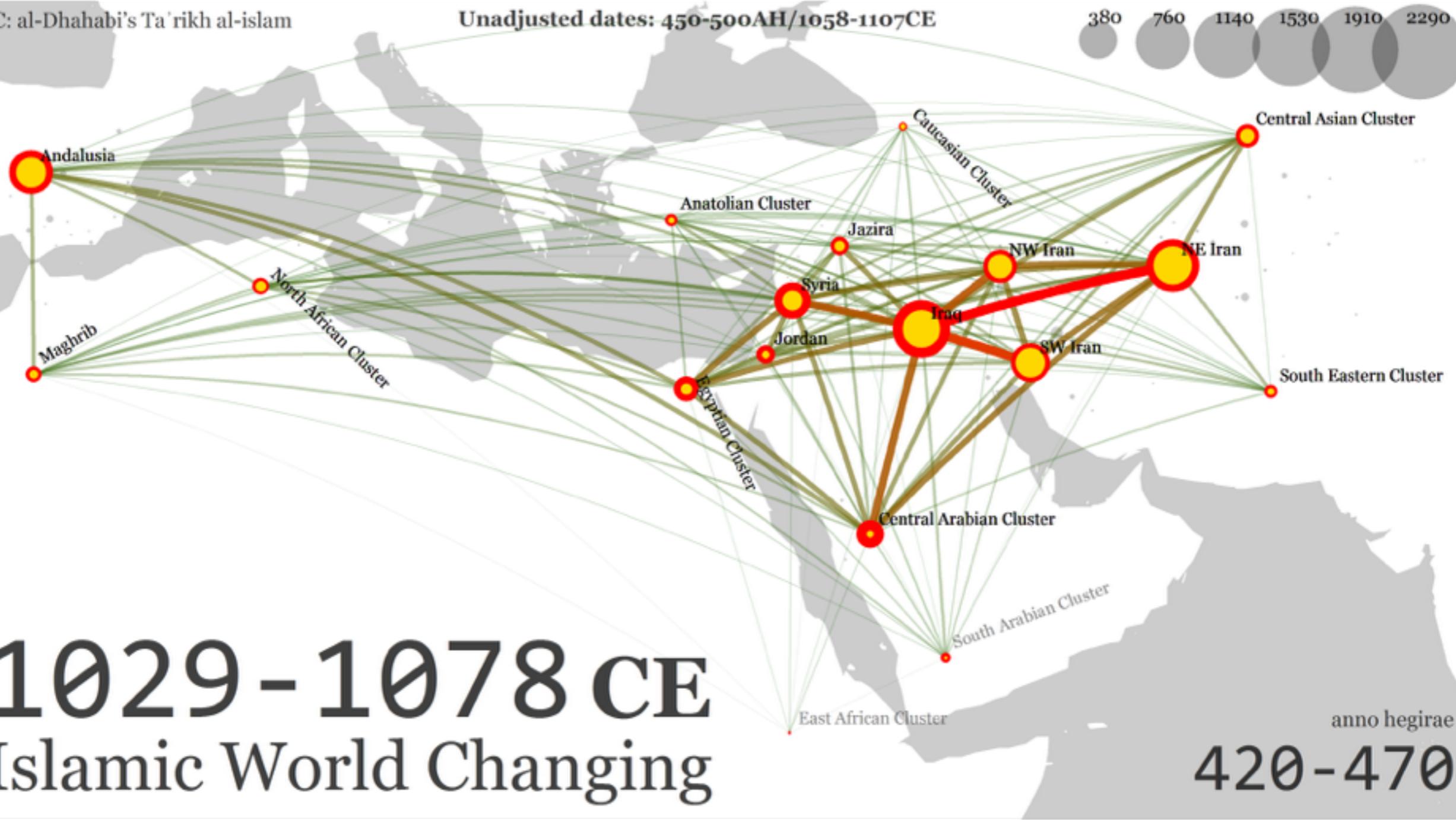
320-370

# 1029-1078 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 450-500AH/1058-1107CE

380 760 1140 1530 1910 2290



anno hegirae

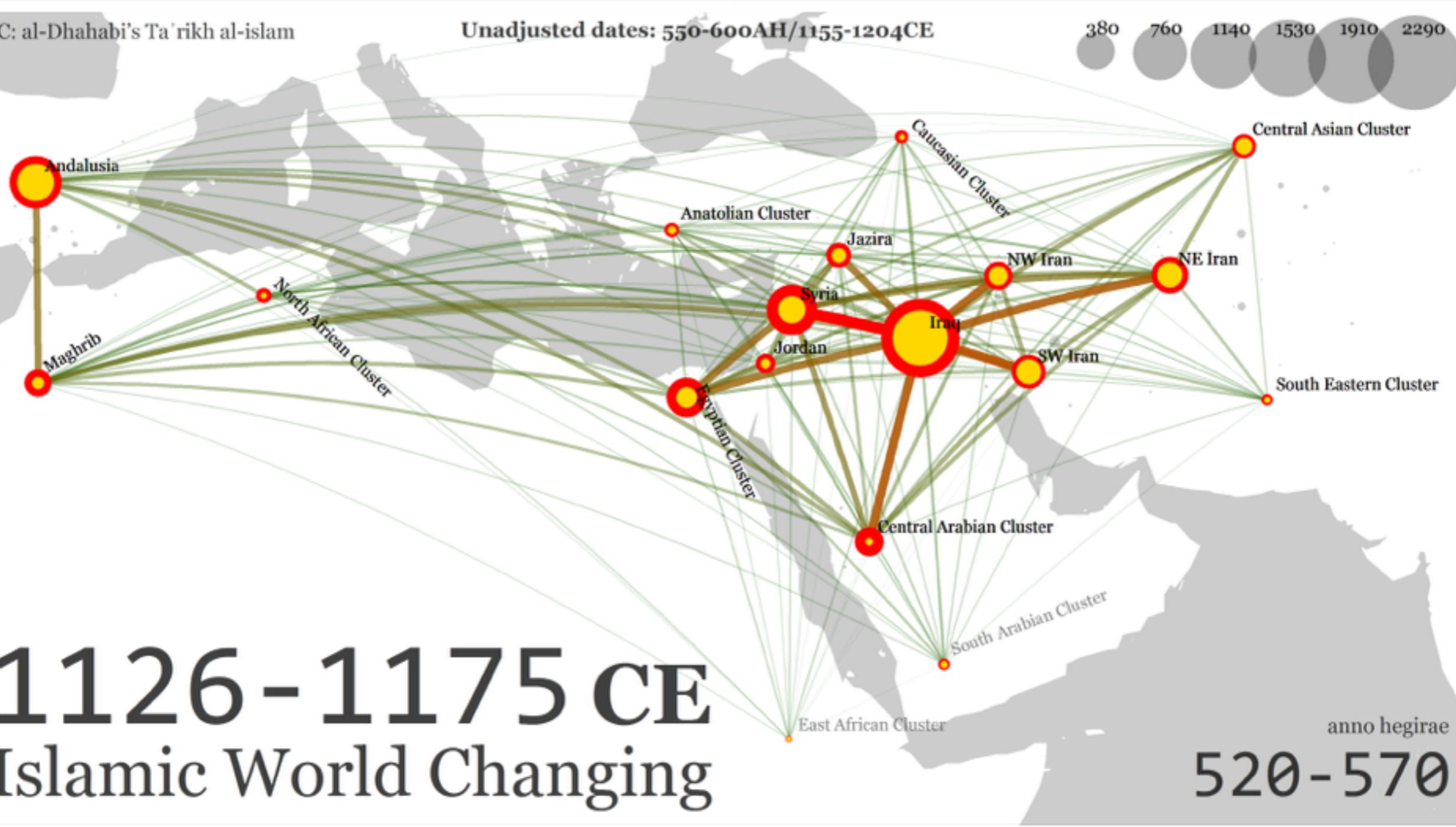
420-470

# 1126-1175 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 550-600AH/1155-1204CE

380 760 1140 1530 1910 2290



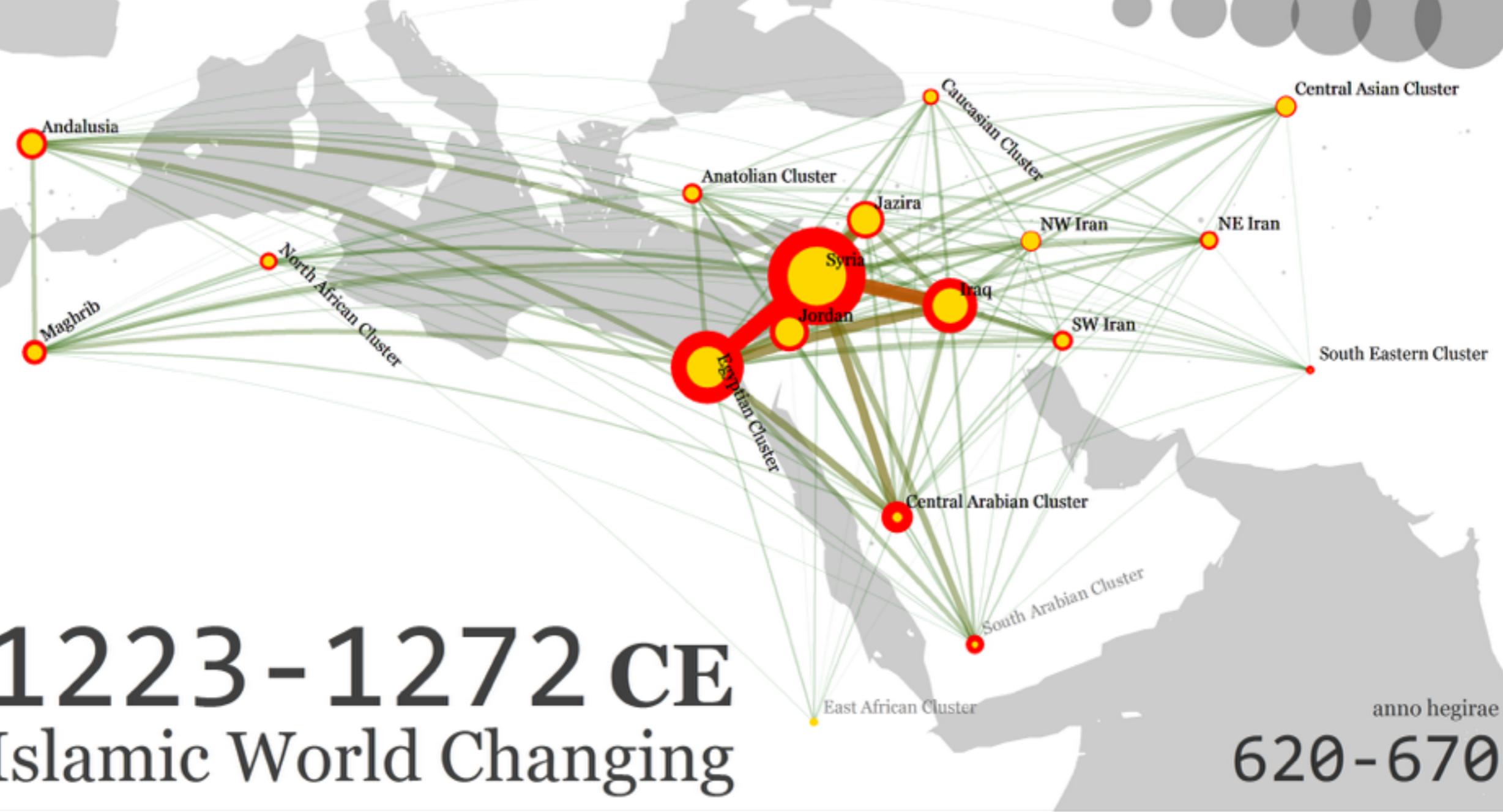
anno hegirae

520-570

# 1223-1272 CE Islamic World Changing

SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 650-700AH/1252-1301CE



anno hegirae

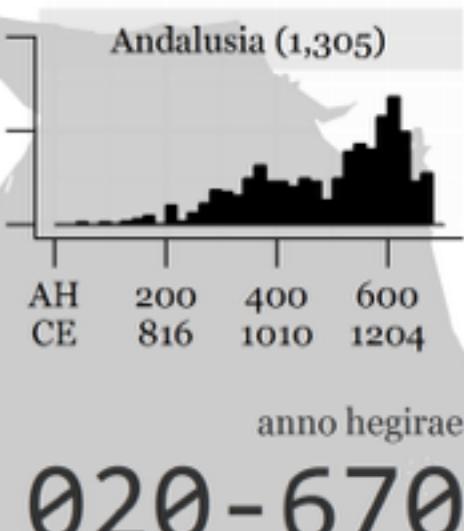
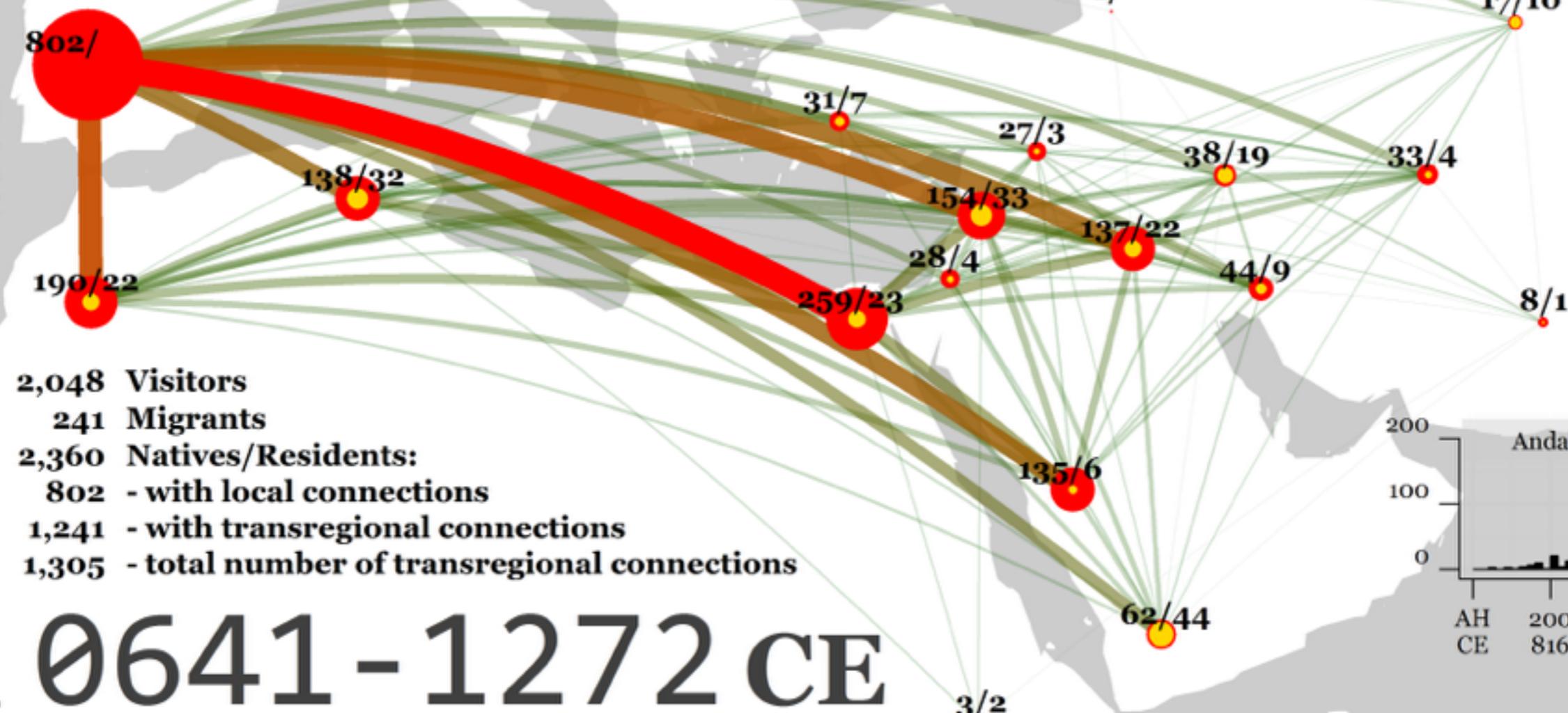
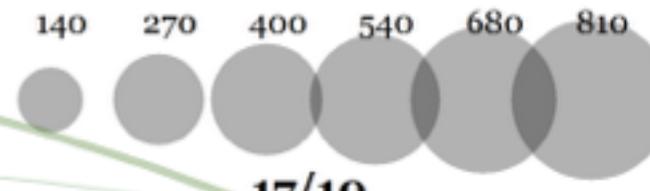
620-670



Regional Integration  
Social and religious groups  
Urban centers

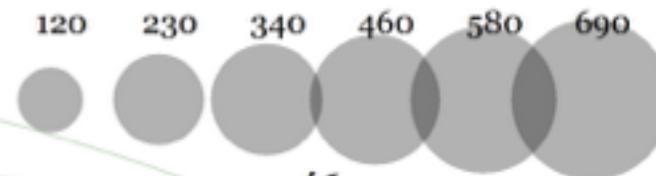
SRC: al-Dhahabi's Ta'rikh al-islam

Unadjusted dates: 50-700AH/670-1301CE



SRC: al-Dhahabi's Ta'rikh al-islam

Unadjusted dates: 50-700AH/670-1301CE



14/4

8/3

47/5

**2,312** Visitors**526** Migrants**2,736** Natives/Residents:  **452** - with local connections  **1,398** - with transregional connections  **2,077** - total number of transregional connections

38/7

143/15

249/81

54/10

39/10

684/192

214/30

20/2

181/42

138/42

452/

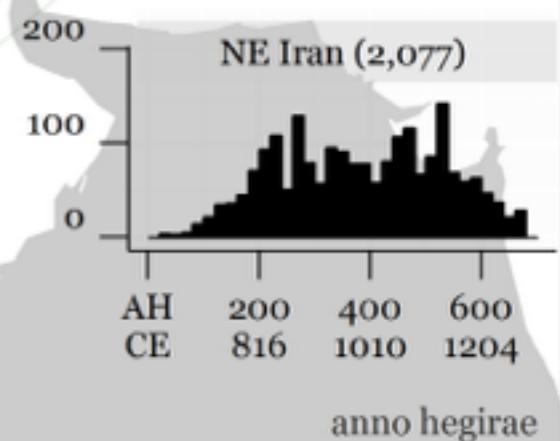
150/67

49/13

47/2

**0641-1272 CE**

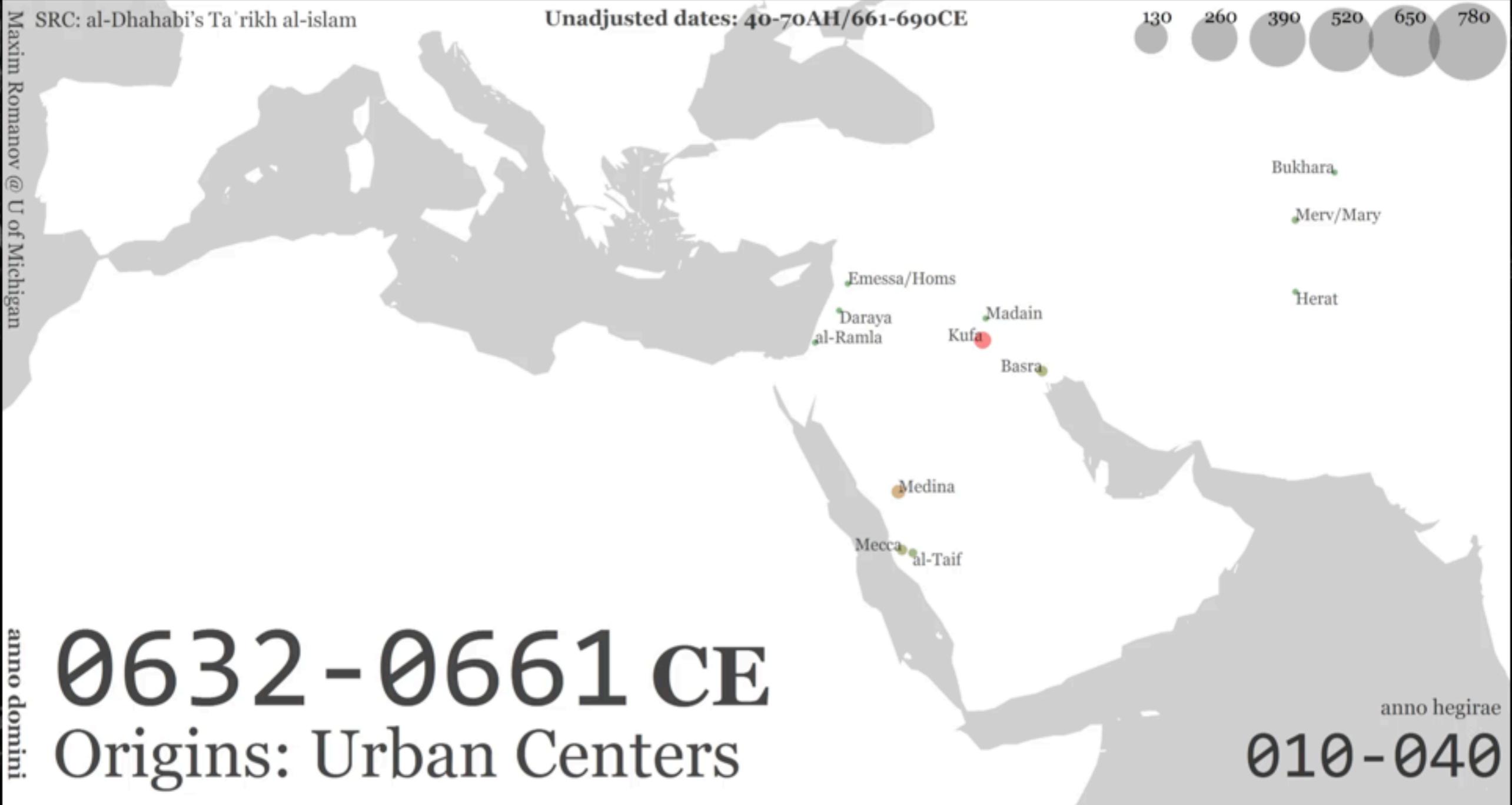
Connections: NE Iran

**020-670**



SRC: al-Dhahabi's *Ta'rikh al-islam*

Unadjusted dates: 40-70AH/661-690CE





Can we actually do  
this kind of  
thing?

# Reactions...

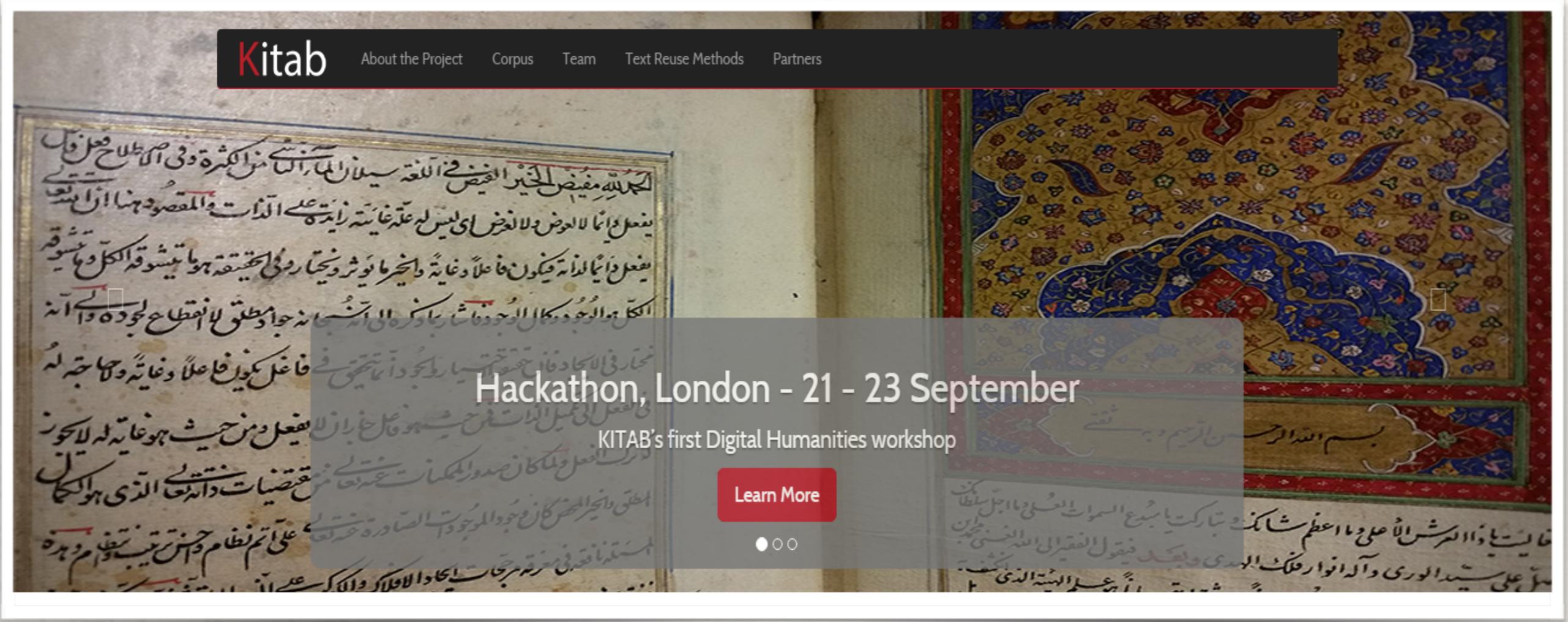
- ❖ Quite a few outraged “NO!”s
  - ❖ The *Ta’rīh al-islām* is al-Dahabī’s vision of Islamic history from the 13th century
- ❖ Or is it?!
  - ❖ 50 volumes
  - 700 year of Islamic history
  - 30,000 biographies
  - ~ 3,2 mln. words
  - “compilation of earlier sources”



# “Compilation” ?

- ❖ How does one write a book like that w/o copy-paste?!
  - ❖ Did he paraphrase his sources?
    - ❖ changing their language, and thus bringing 13th century language into the descriptions of earlier periods
  - ❖ Did he “stitch” the TI from his sources?
    - ❖ thus preserving the language of earlier sources

# KITAB Project: [kitab-project.org](http://kitab-project.org)



Agha Khan University, U Leipzig/Tufts U,  
Northeastern U, Göttingen Centre for DH

# KITAB Project: [kitab-project.org](http://kitab-project.org)



**Sarah Savant**, Associate Professor, AKU-ISMC, and Principle Investigator of KITAB. Dr. Savant is a cultural historian specialising in the Middle East and Iran ca. 600-1100. Her work on KITAB is motivated by a desire to write a cultural history of the Arabic book using digital tools. Her publications include *The New Muslims of Post-Conquest Iran: Tradition, Memory, and Conversion* (Cambridge University Press, 2013; winner of the Said-Sirjani book award given by the International Society for Iranian Studies on behalf of the Persian Heritage Foundation).



**Ahmad Sakhi** is an IT professional, specializing in Enterprise Information and Content Management space. He is currently consulting for Capgemini, a global consulting firm. Mr. Sakhi has a degree in Computer Information Systems and has completed various projects in the Finance, Telecom, Banking and Energy sectors in North America, including designing, building architecture and implementing end-to-end solutions. Mr. Sakhi is from Toronto, Canada and is involved in the KITAB project as a TKN volunteer.



**Malik Merchant** is a Software Consultant, specializing in Microsoft Technologies. He currently is working for Avanade, a joint venture between global consulting firm, Accenture and Microsoft. Mr. Merchant has a Bachelor's degree in Computer Science from Mumbai, India and has been involved in several projects across several industries with focus in Oil & Gas and Insurance domain. He now lives in Calgary, Canada and is involved in the KITAB project as a TKN volunteer.



**Sohail Merchant** is currently working as the Assistant Registrar for the Aga Khan University. Prior to working at AKU-ISMC, he has worked with various organisations like City Hampers, Child-to-Child – University of London and Blue Systems Inc. where he has designed, managed and implemented various solutions and web applications. Sohail is a Microsoft Certified Professional and holds a Master of Science degree in Computing from London Metropolitan University.

# KITAB Project: [kitab-project.org](http://kitab-project.org)



Gregory Ralph Crane

Alexander von Humboldt Professor of Digital Humanities



Digital Humanities  
UNIVERSITÄT LEIPZIG

Greg completed his doctorate in classical philology at Harvard University and subsequently worked there as an assistant professor. From 1985, he was involved in planning the Perseus Project as a co-director and is now its Editor-in-Chief. He was associate professor at TUFTS University. He has received, among others, the Google Digital Humanities Award 2010 for his work in the field.



**David Smith** is an Assistant Professor in the College of Computer and Information Science at Northeastern University and a founding member of the NULab for Texts, Maps, and Networks, Northeastern's center for the digital humanities and computational social sciences. Previously, he was a research faculty member at the University of Massachusetts' Center for Intelligent Information Retrieval, a Ph.D. student in computer science at Johns Hopkins University, and the head programmer at the Perseus Digital Library Project. His research focuses on building statistical models of human language, with applications to information retrieval, machine translation, the humanities, and social sciences. Most recently, he has been working on inference for social networks from textual evidence, in collaboration with colleagues in English, history, and political science, under the aegis of the Proteus and Viral Texts projects.

<http://viraltexts.org/> Ryan Cordell, David Smith (Code), and others



# How to address this issue?

Passage1: JK000982\_000292  
Passage2: Shamelah0023775\_003346  
Begin Position: 1  
End Position: 58

Text reuse: computational methods of tracing long quotations, paraphrases, allusions, etc.

al-Ta‘alibī’s *Timār al-qulūb fī-l-mudāf wa-l-mansūb* and one of its sources

Text1:

عند عبيد الله بن زياد اذ ادخل علي----- جرذا ابيض فتعجب منه ف-قال --  
-----لعبد الله----- يا ابا صالح هل رأيت اعجب من هذا-----  
--- واذا عبد الله قد تفاءل----- كانه فrex واصف-----ر كانه جرادة---  
- فقال عبيد الله ابو صالح يعصي الرحمن ويتهاون **بالسلطان** ويقبض على  
الثعبان ويمشي الي الاسد الورد ويلقي الرماح بوجهه والسيوف بيده وقد  
اعتراه من

Text2:

عند عبيد الله بن زياد اذ ادخل علي عبد الله جرذا - ابيض ليعجب منه فا قبل عبيد  
الله علي عبد الله فقال هل رأيت يا ابا صالح----- اعجب من هذا الجرذا فقط  
و اذا عبد الله قد تفاءل حتى صار كانه فrex واصفر حتى صار كانه جرادة ذكر  
فقال عبيد الله ابو صالح يعصي الرحمن ويتهاون **باليشيطان** ويقبض على الثعبان  
ويمشي الي الاسد ويلقي الرماح بوجهه ----- وقد اعتراه من

# Graphing Method



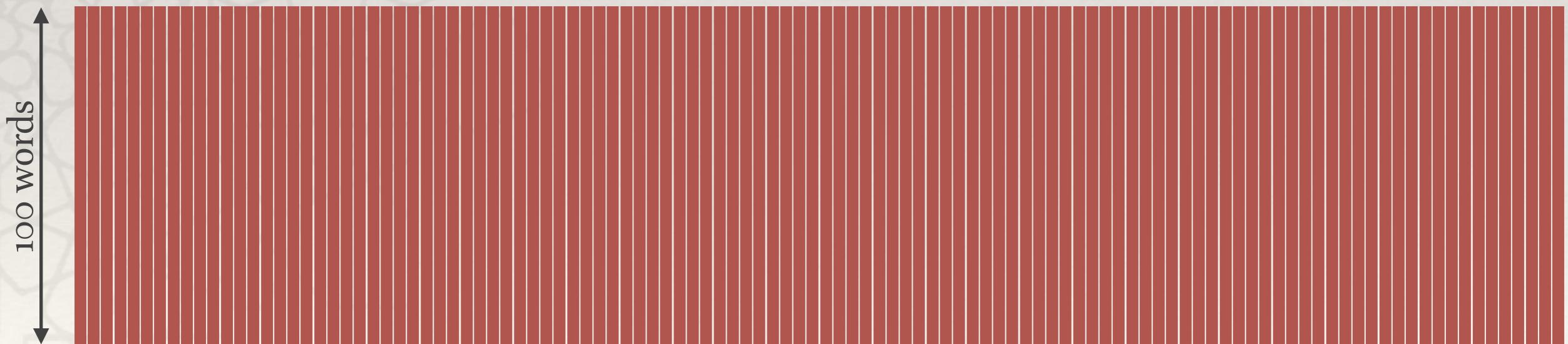
Book A

# Graphing Method



Beginning of the book

End of the book



Book A

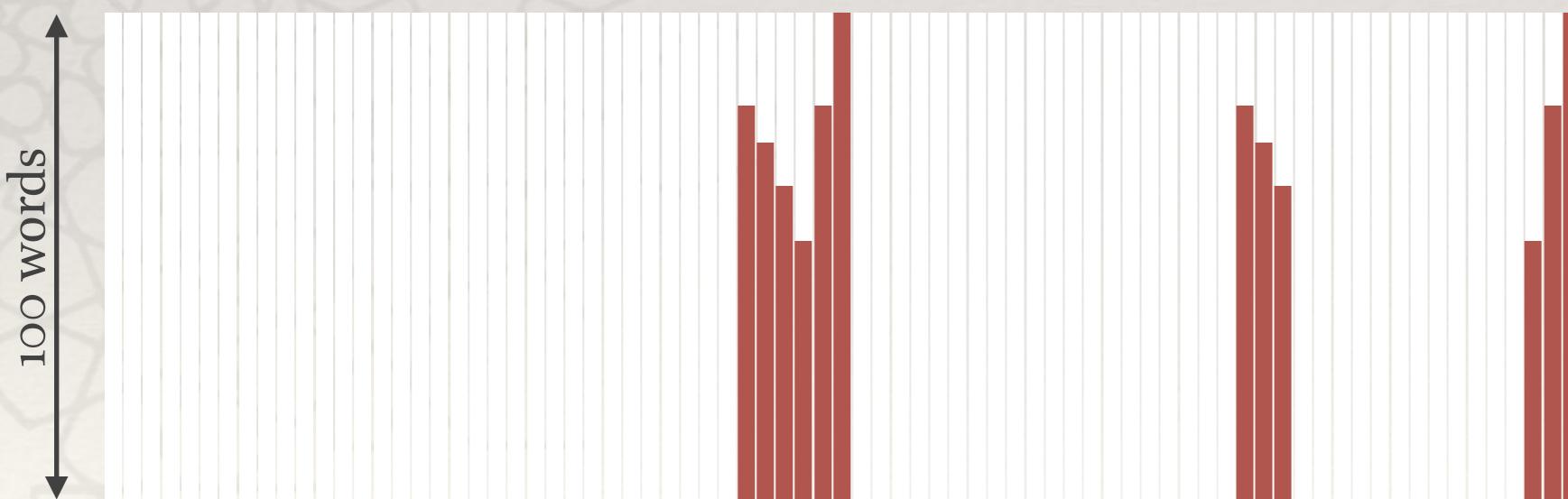
Book A Sliced

# Graphing Method



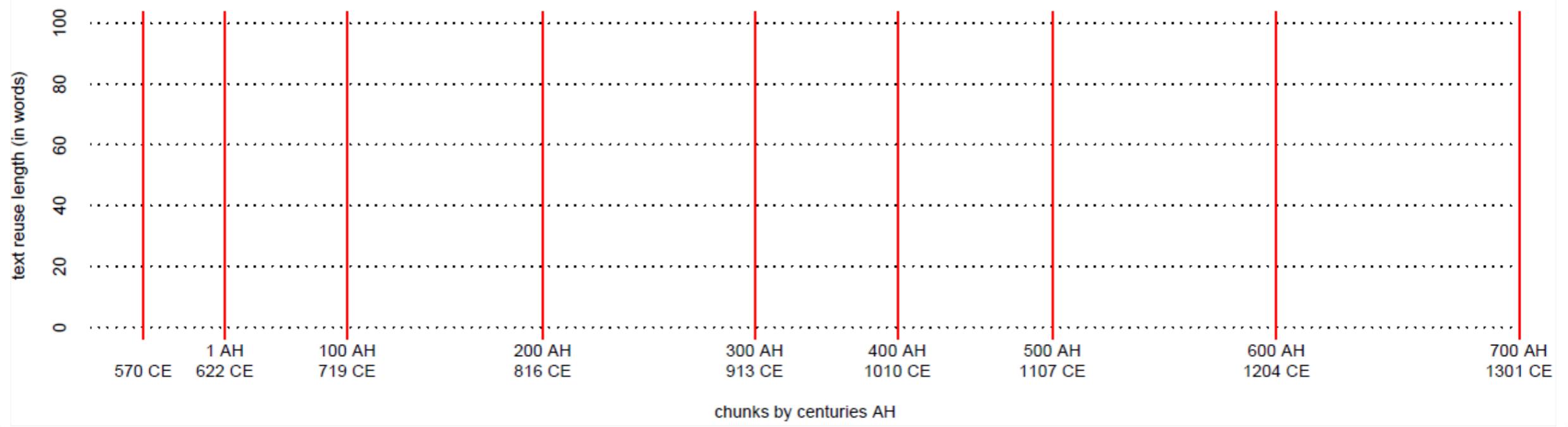
Beginning of the book

End of the book

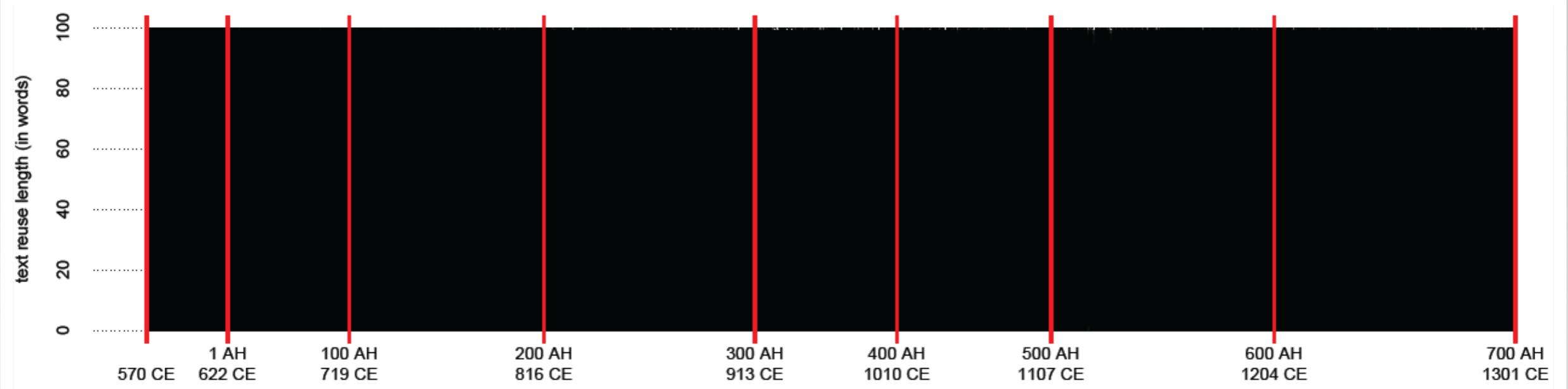


# Text Reuse in the *Ta'rīh al-islām*

*Ta'rīh al-islām* chunked and divided into periods

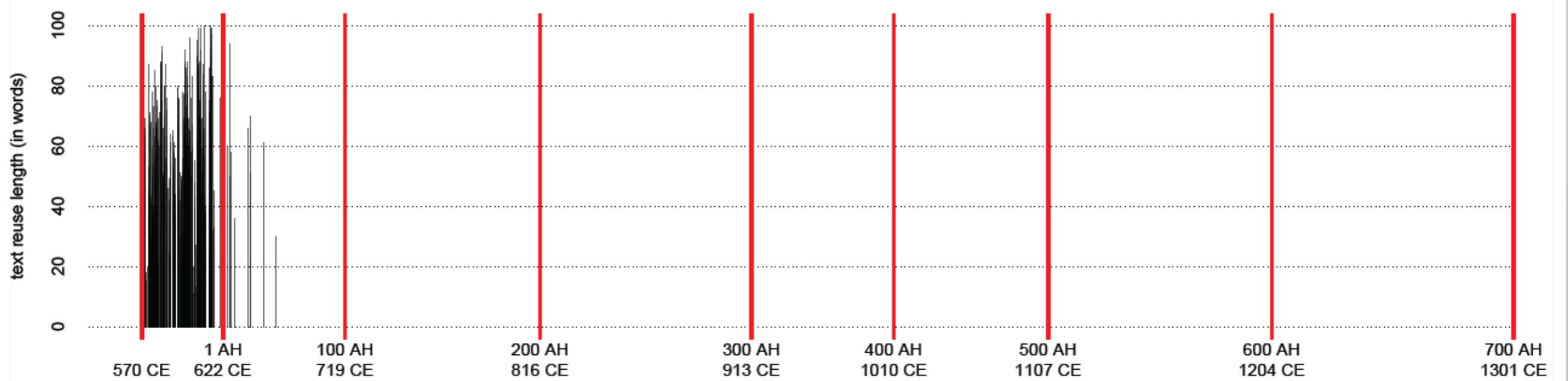


*Ta'rīh al-islām* compared with itself

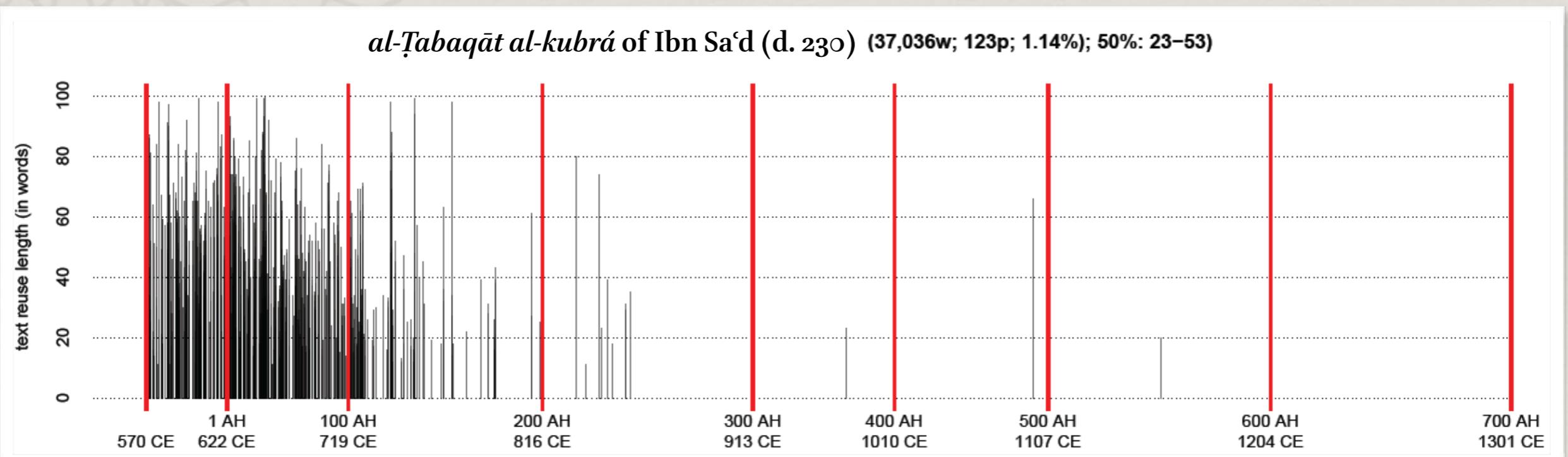


# Text Reuse in the *Ta'rīh al-islām*

*al-Sīrat al-Nabawīyat* of Ibn Hišām (d. 213) (28,339w; 94p; 0.872%); 50%: 30–63)

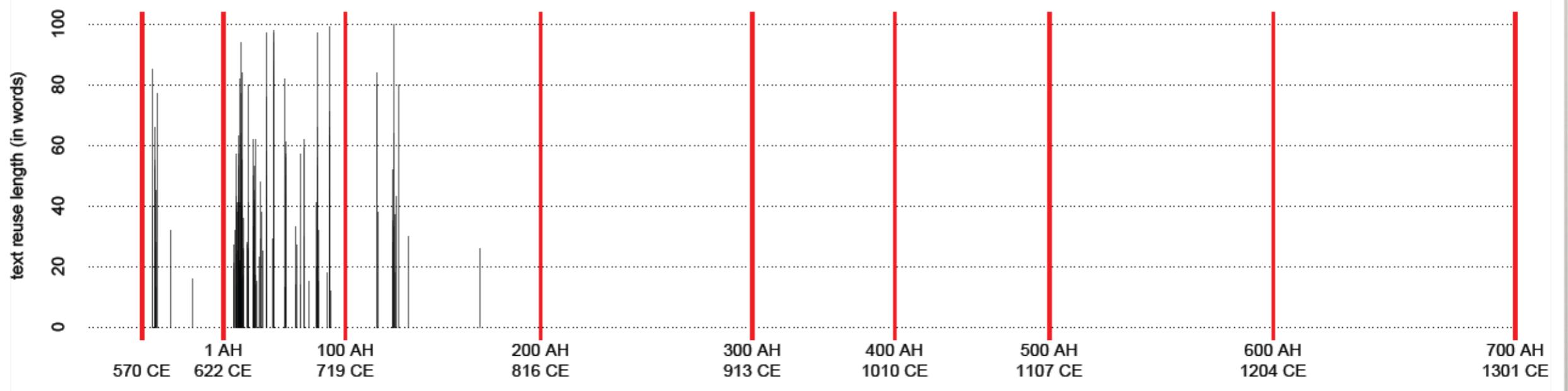


*al-Tabaqāt al-kubrā* of Ibn Sa'd (d. 230) (37,036w; 123p; 1.14%); 50%: 23–53)

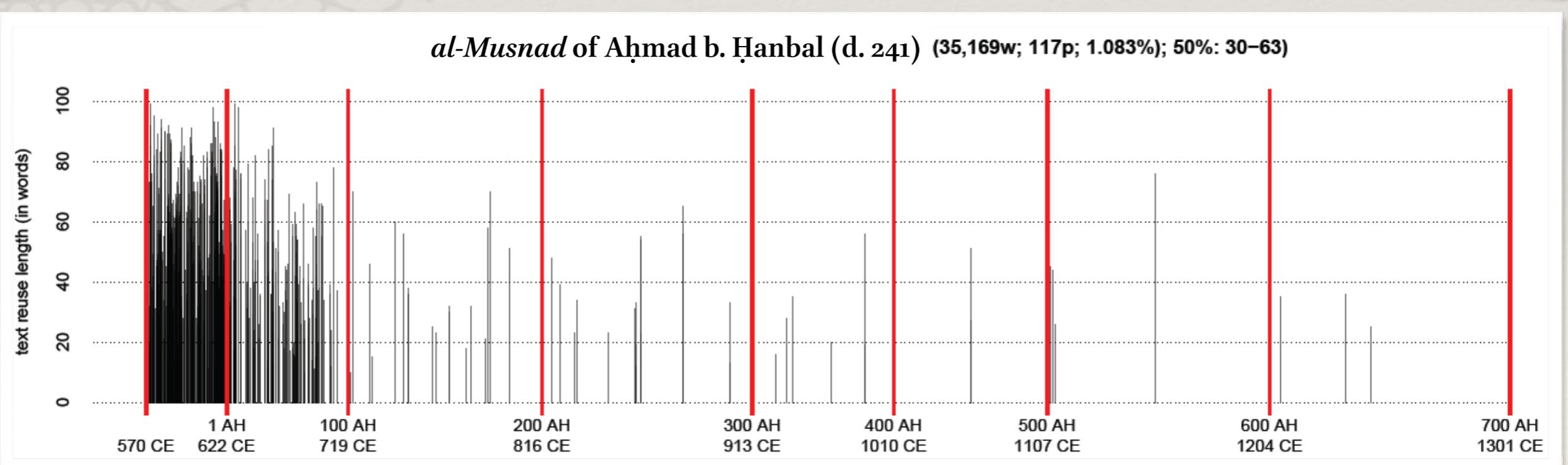


# Text Reuse in the *Ta'rīh al-islām*

*al-Ta'rīh* of Ḥalīfat b. Ḥayyāt (d. 240) (7,696w; 25p; 0.237%); 50%: 25–57)

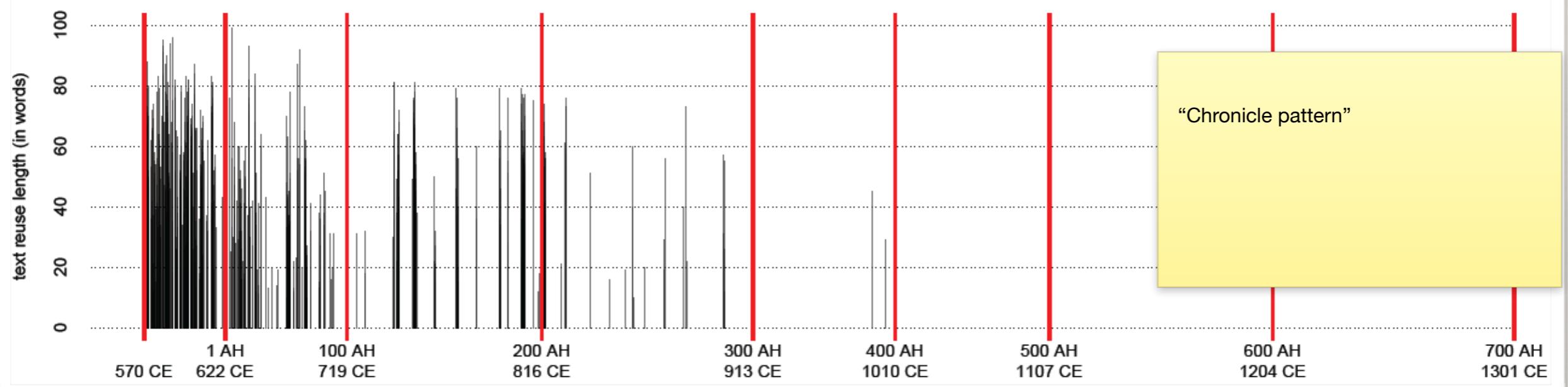


*al-Musnad* of Ahmad b. Ḥanbal (d. 241) (35,169w; 117p; 1.083%); 50%: 30–63)

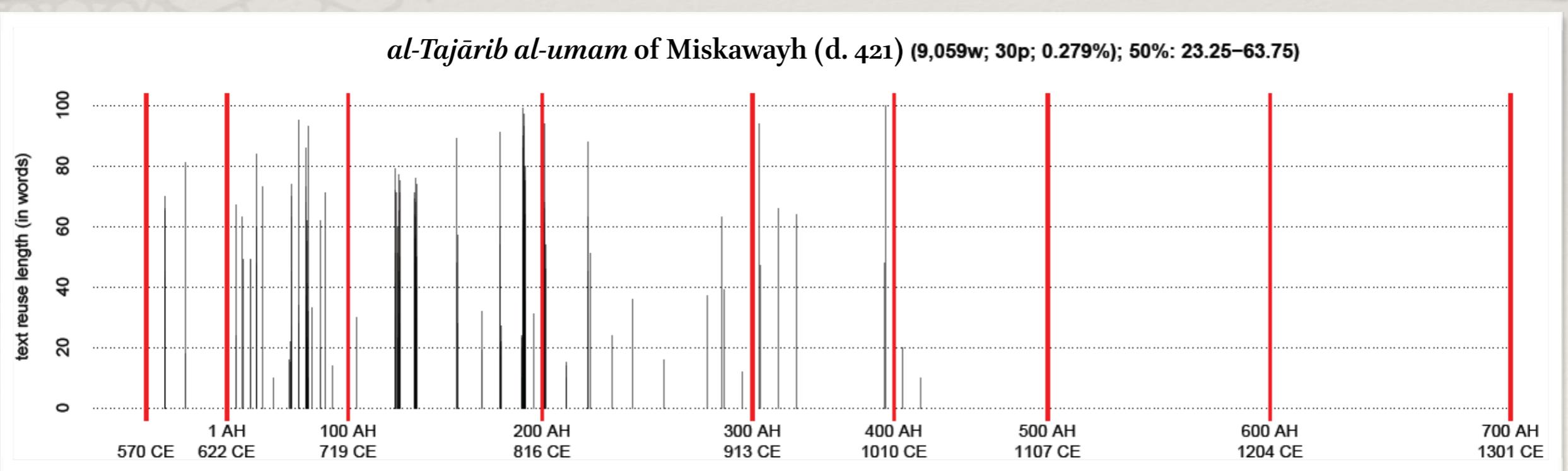


# Text Reuse in the *Ta'rīh al-islām*

*al-Ta'rīh* of al-Ṭabarī (d. 310) (37,390w; 124p; 1.151%); 50%: 26–57)

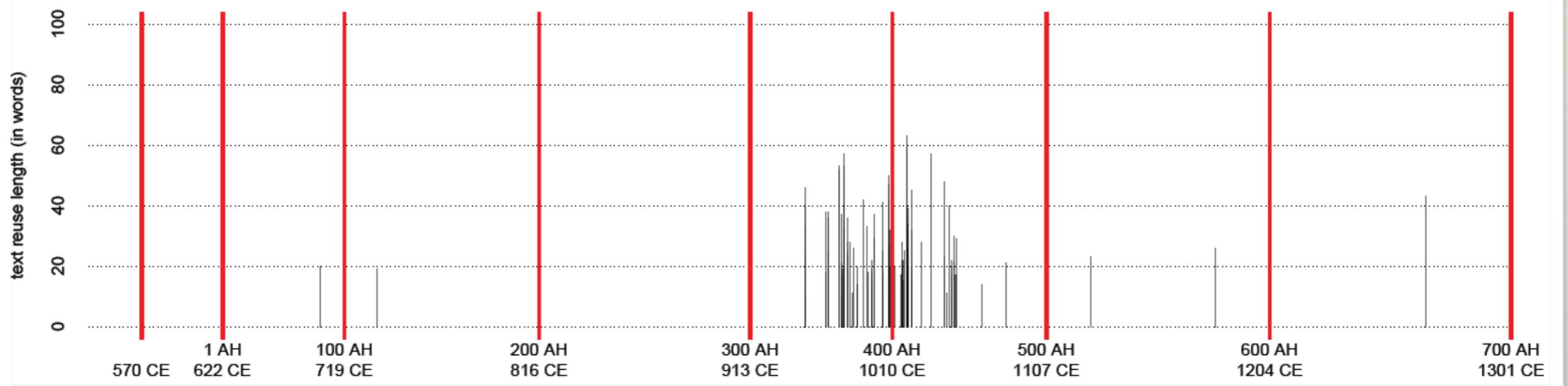


*al-Tajārib al-umam* of Miskawayh (d. 421) (9,059w; 30p; 0.279%); 50%: 23.25–63.75)

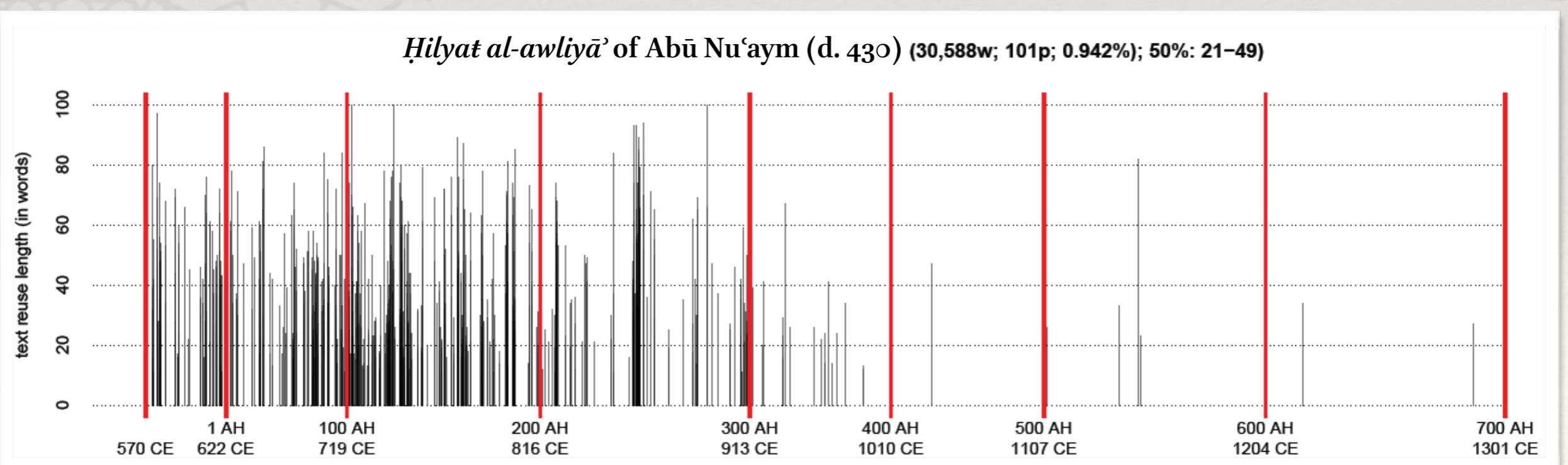


# Text Reuse in the *Ta'rīh al-islām*

*Yatīmat al-dahr* of al-Ta'ālibī (d. 429) (3,192w; 10p; 0.098%); 50%: 17–33

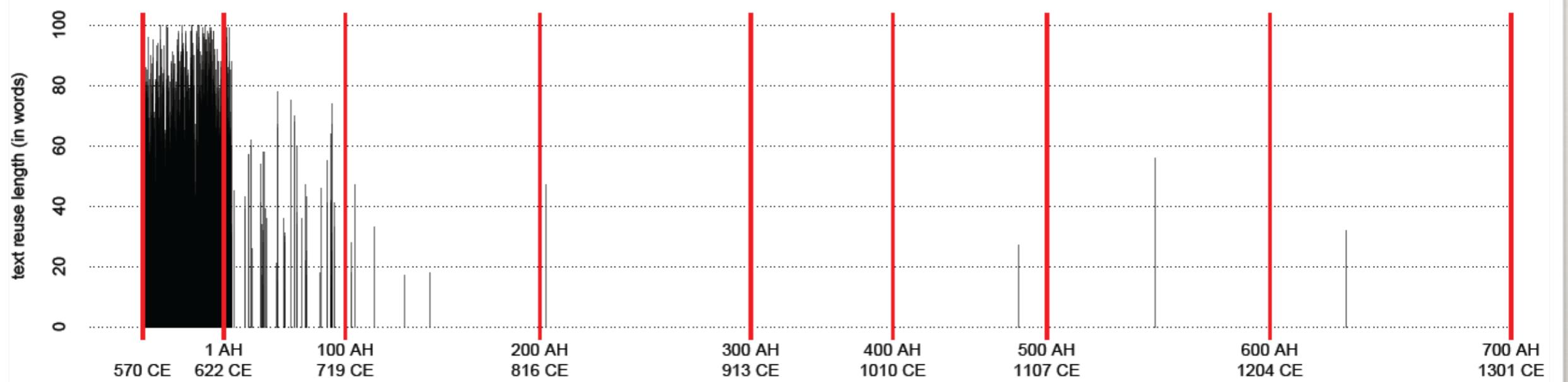


*Hilyat al-awliyā'* of Abū Nu'aym (d. 430) (30,588w; 101p; 0.942%); 50%: 21–49

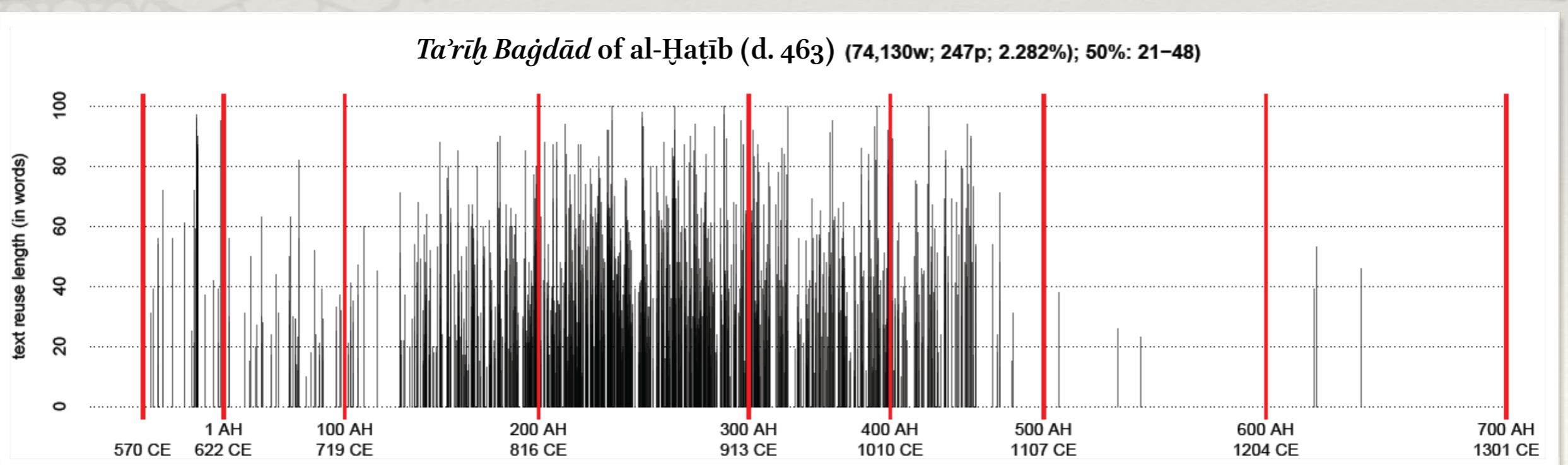


# Text Reuse in the *Ta'rīh al-islām*

*Dalā'il al-nubuwat* of al-Bayhaqī (d. 458) (111,436w; 371p; 3.431%); 50%: 28–61)

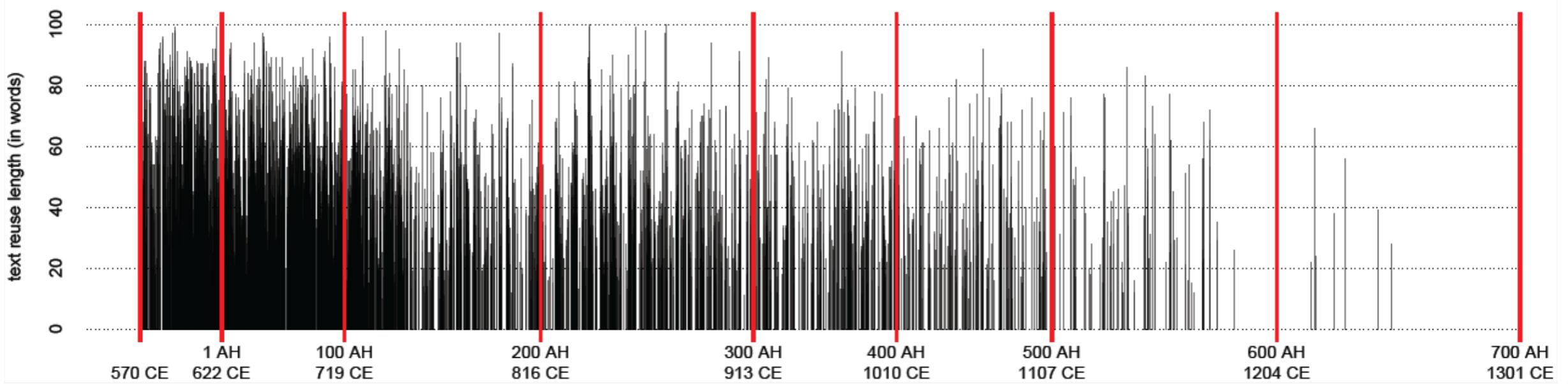


*Ta'rīh Bağdād* of al-Ḥaṭīb (d. 463) (74,130w; 247p; 2.282%); 50%: 21–48)

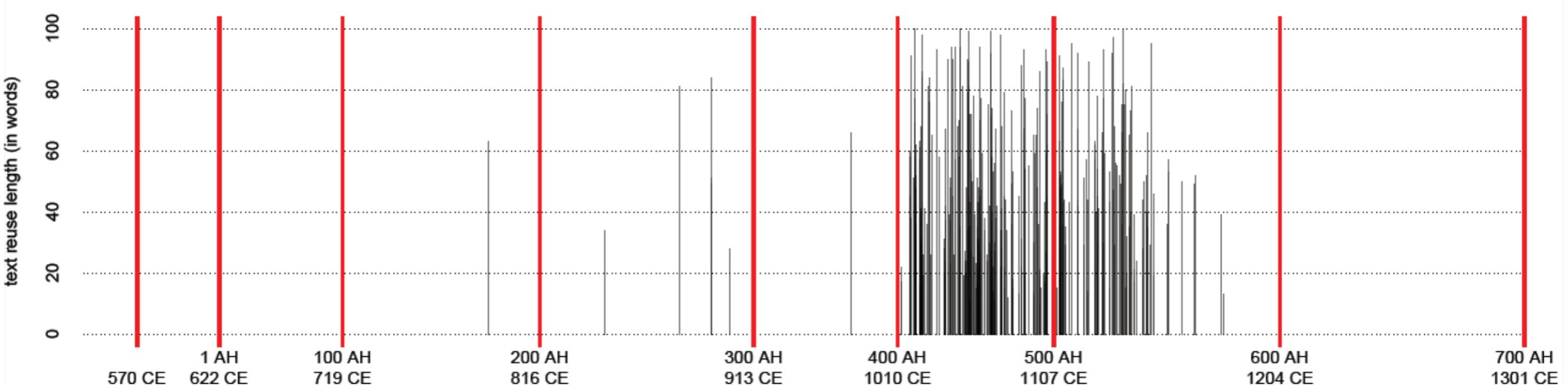


# Text Reuse in the *Ta'rīh al-islām*

*Ta'rīh Dimašq* of Ibn 'Asākir (d. 571) [245,161w; 817p; 7.547%]; 50%: 22–48)

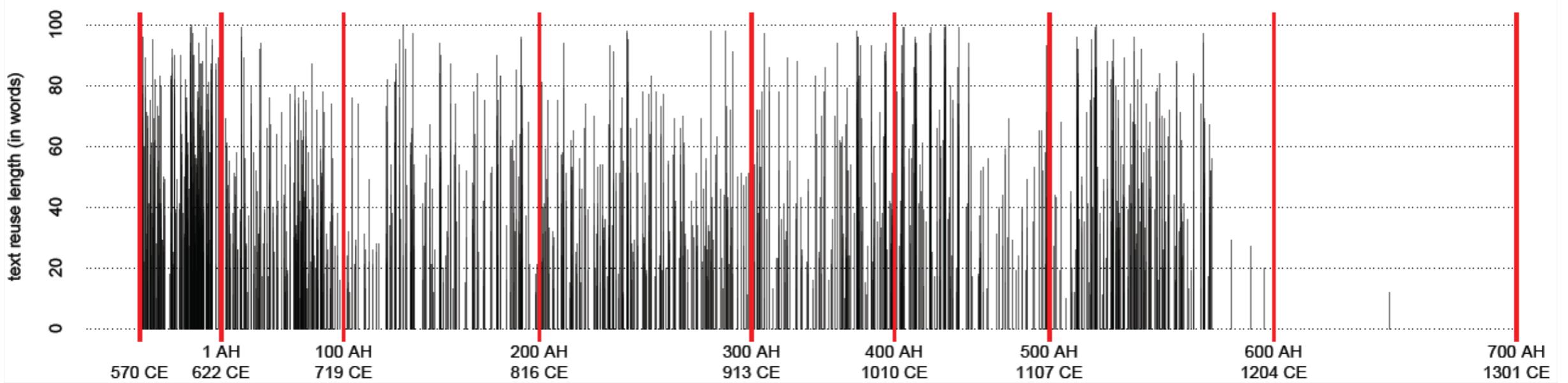


*al-Silat* of Ibn Baškuwāl (d. 578) (15,648w; 52p; 0.482%); 50%: 27–65)

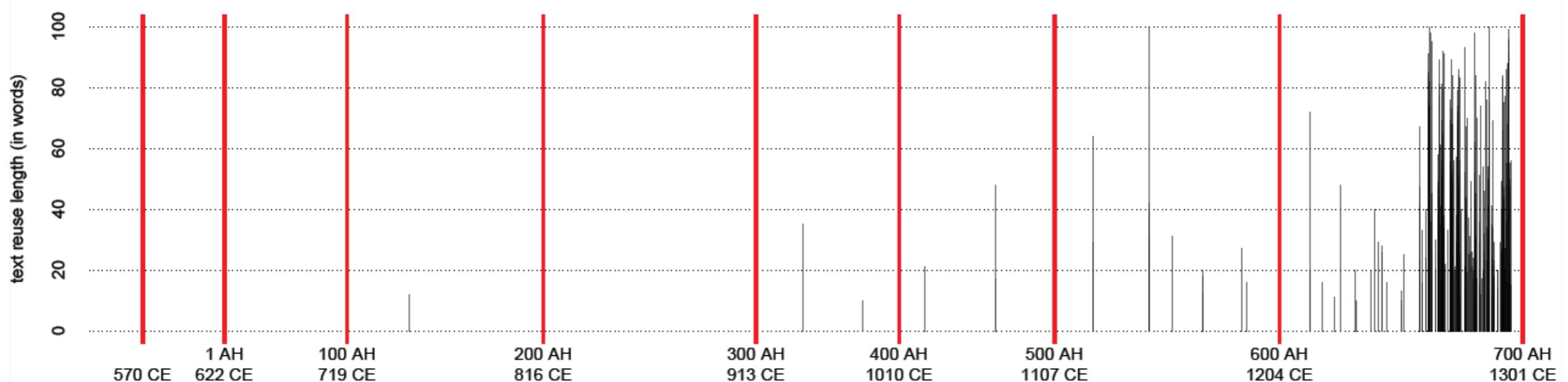


# Text Reuse in the *Ta'rīh al-islām*

*al-Muntażam* of Ibn al-Ǧawzī (d. 597) (83,828w; 279p; 2.581%); 50%: 25–60



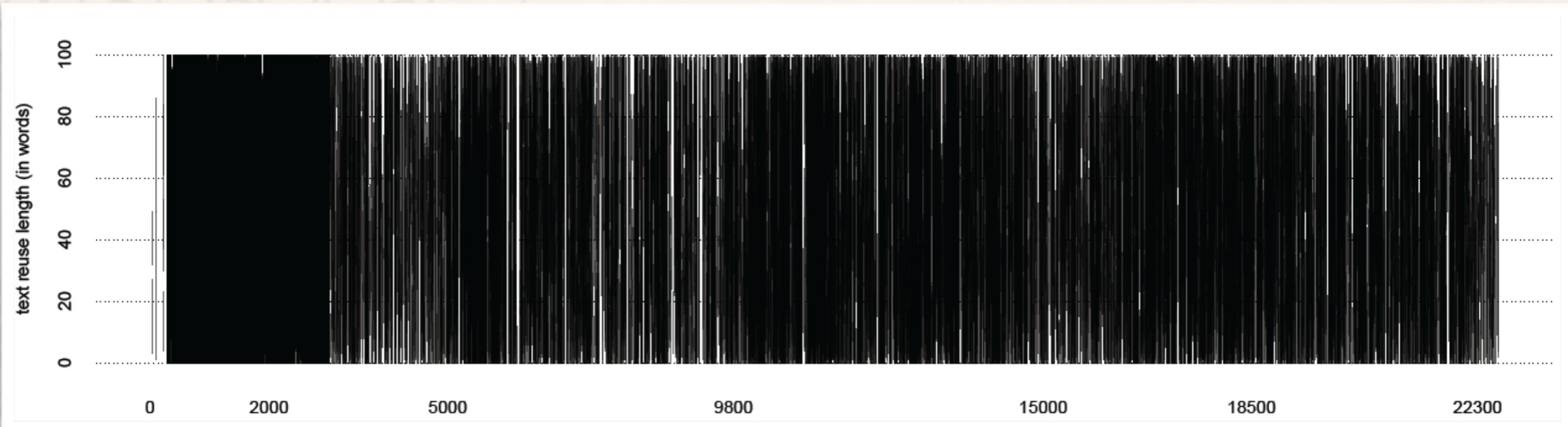
*Dayl mir'at al-zamān* of al-Yūnīnī (d. 726) (14,738w; 49p; 0.454%); 50%: 21–56



# The Scale of Text Reuse

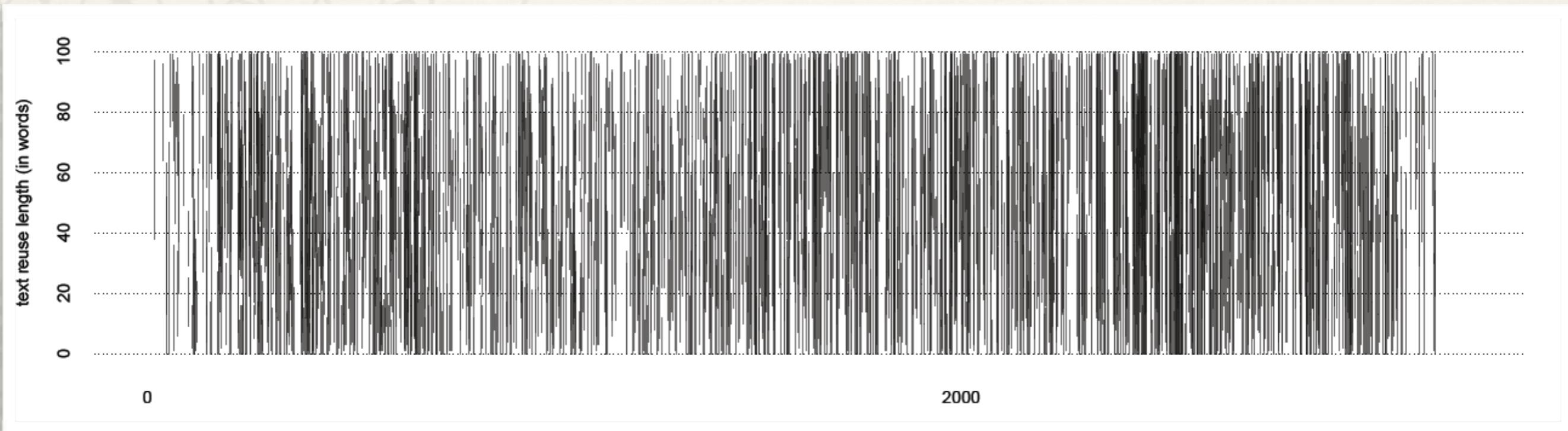
- ❖ ~11,000 pp. in the *Ta'rīh al-islām* (300 w/p)
- ❖ ~ 3,200 pp. from 15 major sources (100-800 pp.)
  - ❖ 817 pp. > *Ta'rīh Dimašq* of Ibn 'Asākir (d. 571)  
[passages 22-48 words]
  - ❖ 413 pp > *Dalā'il al-nubuwat* of al-Bayhaqī's (d. 458)  
[passages 28-62 words]
  - ❖ 279 pp > *al-Muntazam* of Ibn al-Ǧawzī (d. 597)  
[passages 25-60 words]
  - ❖ 270 pp. > *Tahdīb al-kamāl* of al-Mizzī (d. 742)  
[passages 21-46 words]
  - ❖ 247 pp > *Ta'rīh Bağdād* of al-Ḥaṭīb (d. 463)  
[passages 21-48 words]

# *Siyar a'lām al-nubalā'*



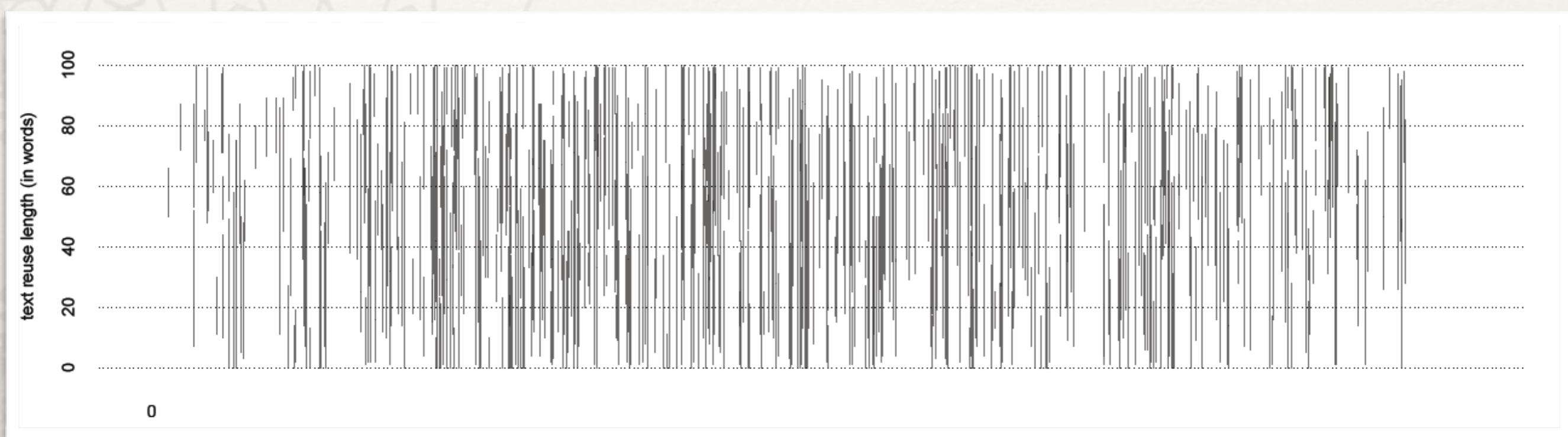
30-35% from the *Ta'rīh al-islām*  
50% of instances are 26-66 words

# *Tadkīrat al-ḥuffāz*



25-26% from the *Ta'rīh al-islām*  
50% of instances are 26-66 words

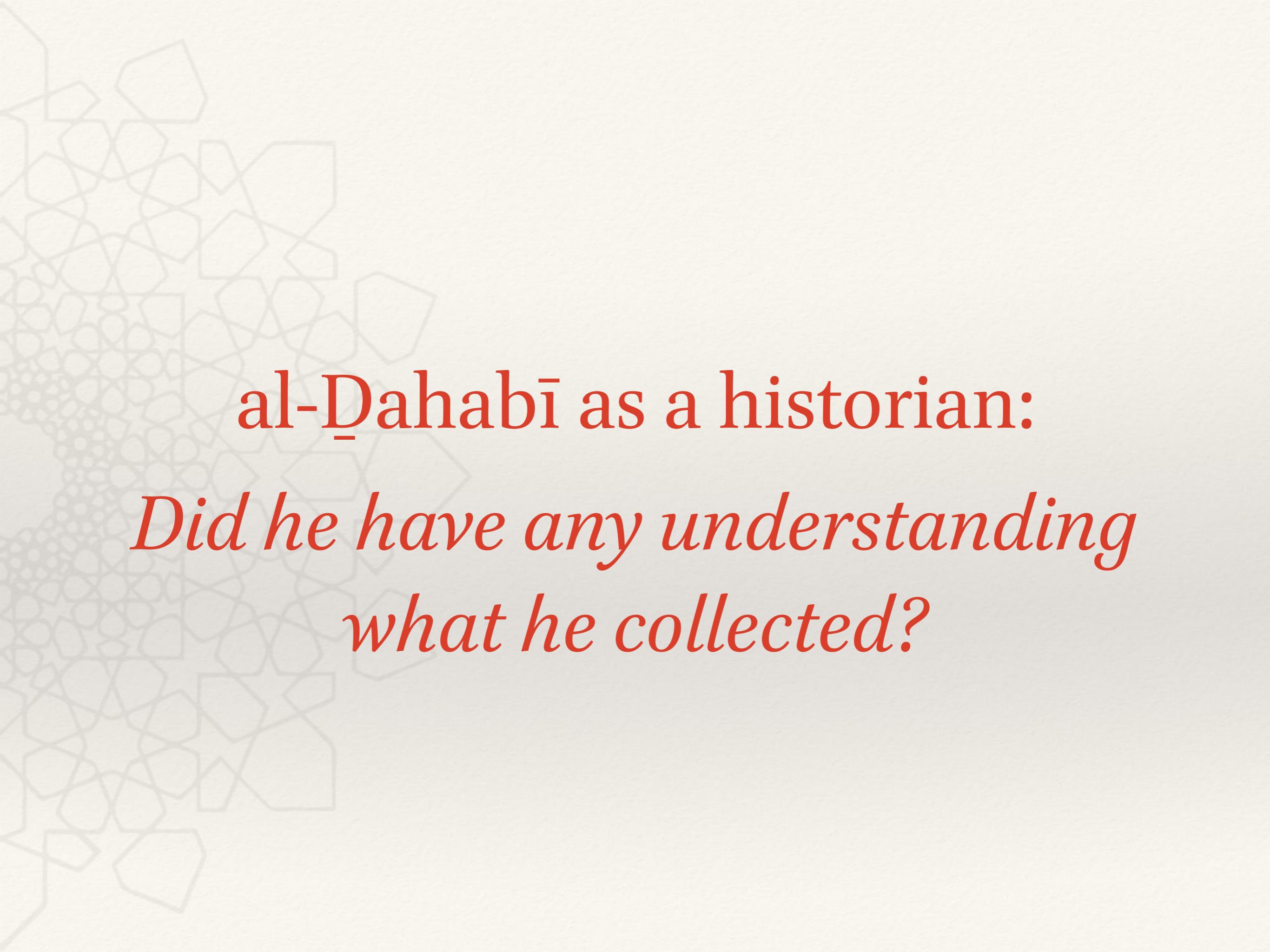
# *Ma‘rifat al-qurrā’ al-kibār ‘alá al-tabaqāt wa-l-aṣār*



22-24% from the *Ta’rīh al-islām*  
50% of instances are 23-55 words

# Working theory:

- ❖ The *Ta'rih al-islām* is a databank for his other books
- ❖ Way to prove: experimentally
- ❖ Generate his “smaller” collections computationally
  - >> generate and compare visualizations and various statistics (*to be continued*)



al-Ḏahabī as a historian:

*Did he have any understanding  
what he collected?*

# al-Dahabī as a historian

- ❖ “Cities and Ports for Hearing the Reports”\*  
*(al-Amṣār ḍawāt al-āṭār)*

\* Translation by Michael Cooperson

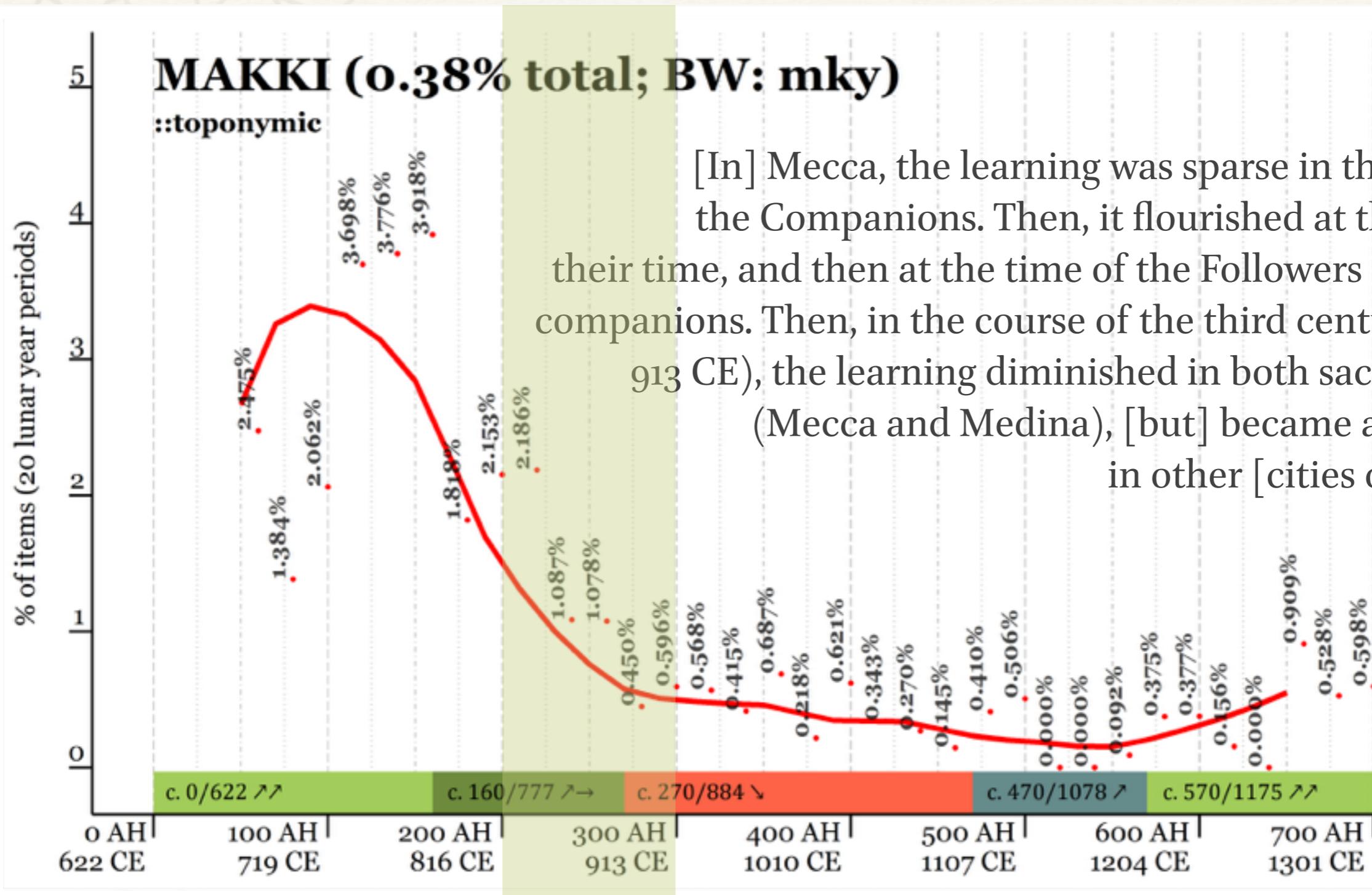
- ❖ 4-folio epistle
- ❖ focus on cities and provinces (~80)
- ❖ *statements about the role of the regions  
of the Islamic world in Ḥadīt learning*

# Mecca / Makkat

مكة ... كان العلم بها يسيرا في زمن الصحابة ثم  
كثر في أواخر عصر الصحابة وكذلك في أيام  
التابعين وزمن أصحابهم ... ثم في أثناء المائة  
الثالثة تناقص علم الحرمين وكثير بغيرهما

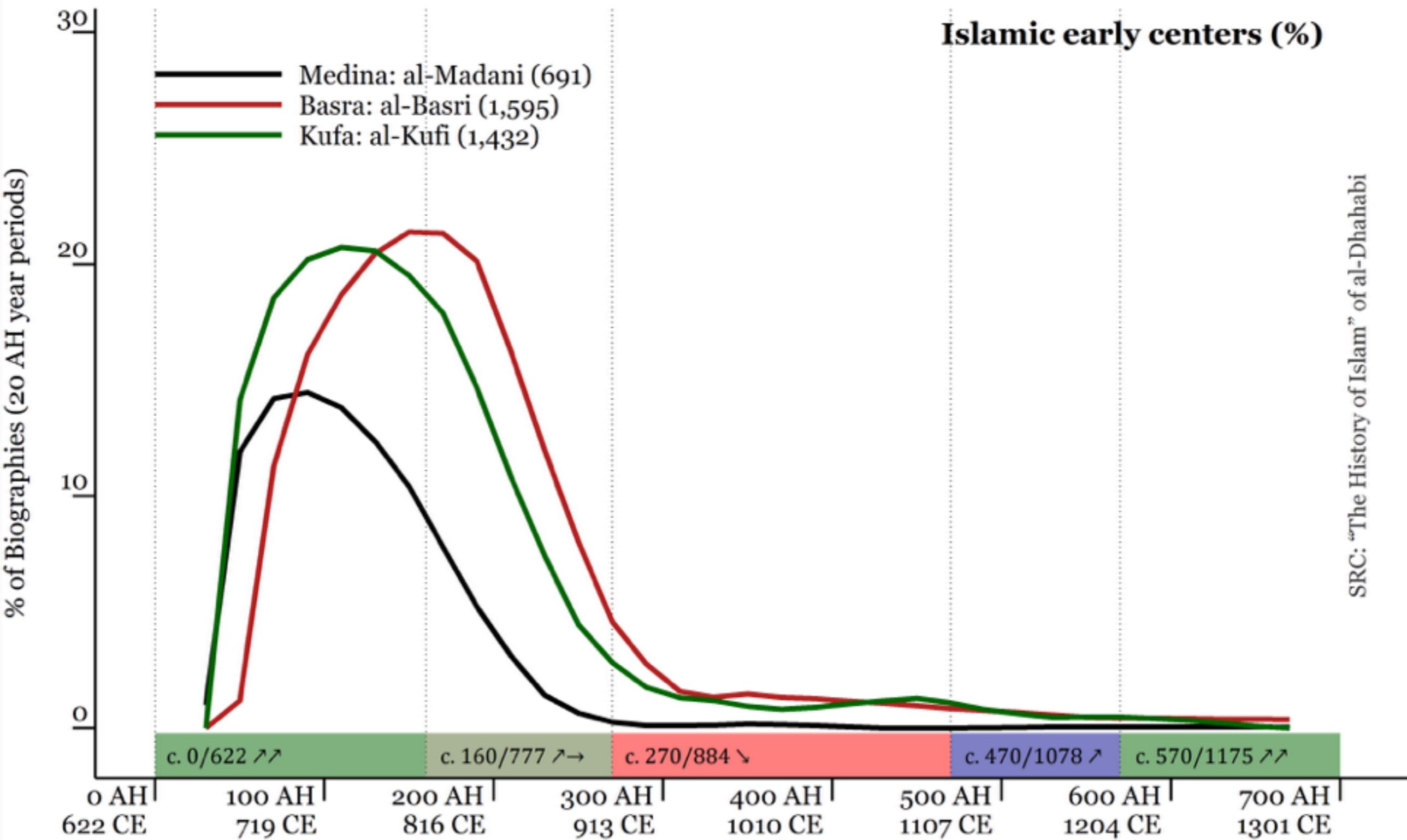
[In] Mecca, the learning was sparse in the time of the Companions. Then, it became abundant at the end of their time, and then in the time of the Followers and their companions. Then, in the course of the third century (816–913 CE), the learning diminished in the two sacred cities (Mecca and Medina), [but] became abundant in other [cities of Islam].

# Mecca / Makkat

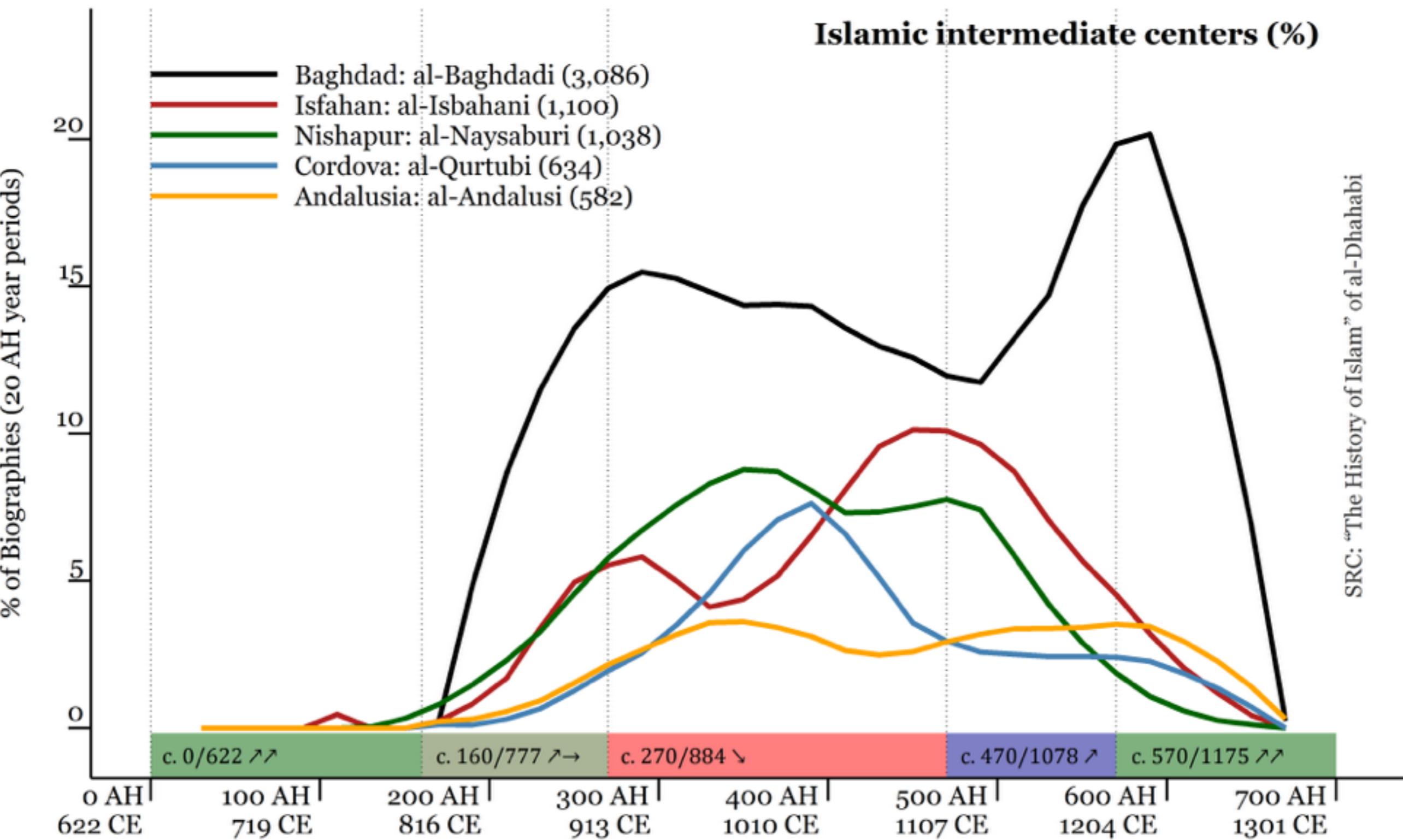


[In] Mecca, the learning was sparse in the time of the Companions. Then, it flourished at the end of their time, and then at the time of the Followers and their companions. Then, in the course of the third century (816–913 CE), the learning diminished in both sacred cities (Mecca and Medina), [but] became abundant in other [cities of Islam].

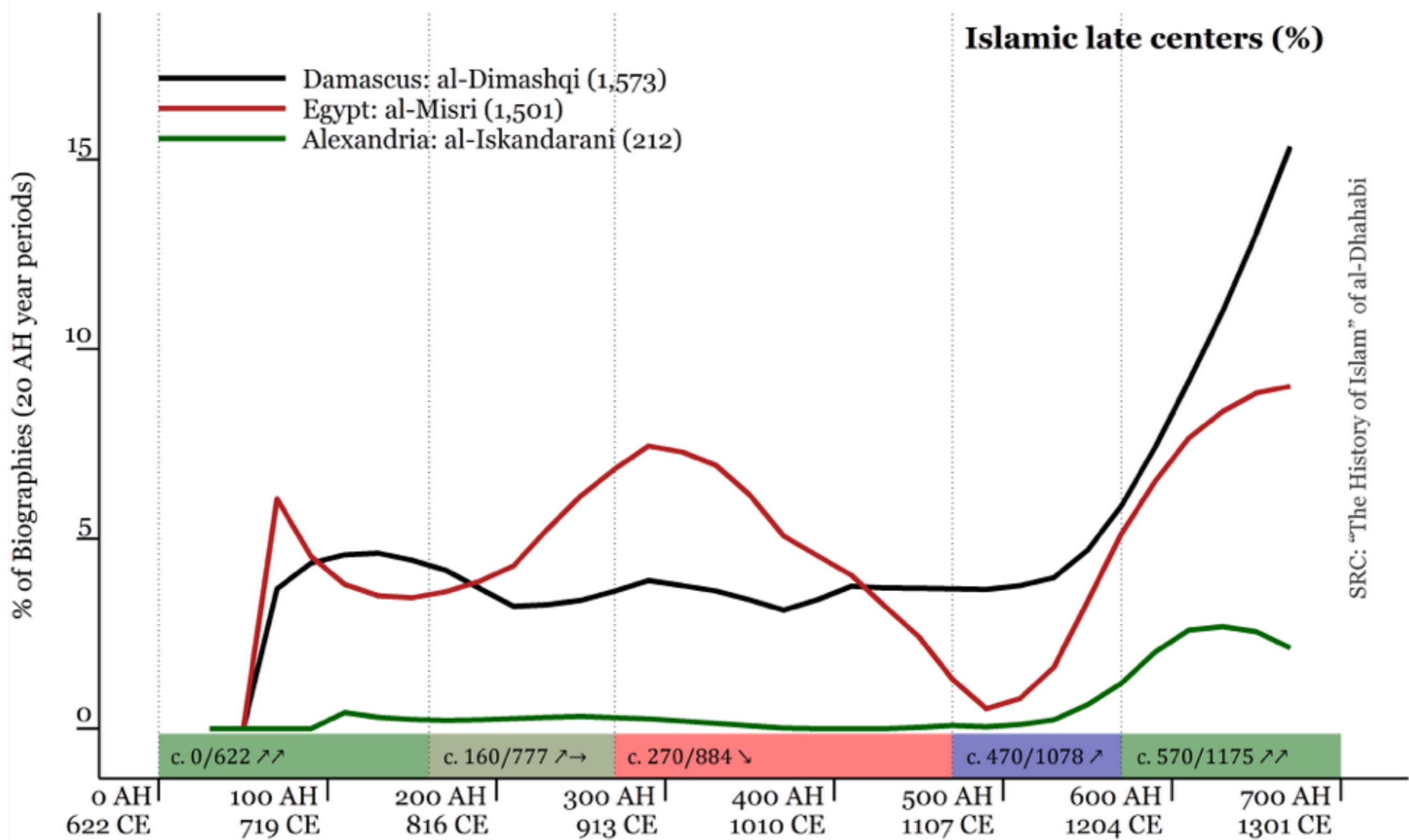
## Islamic early centers (%)



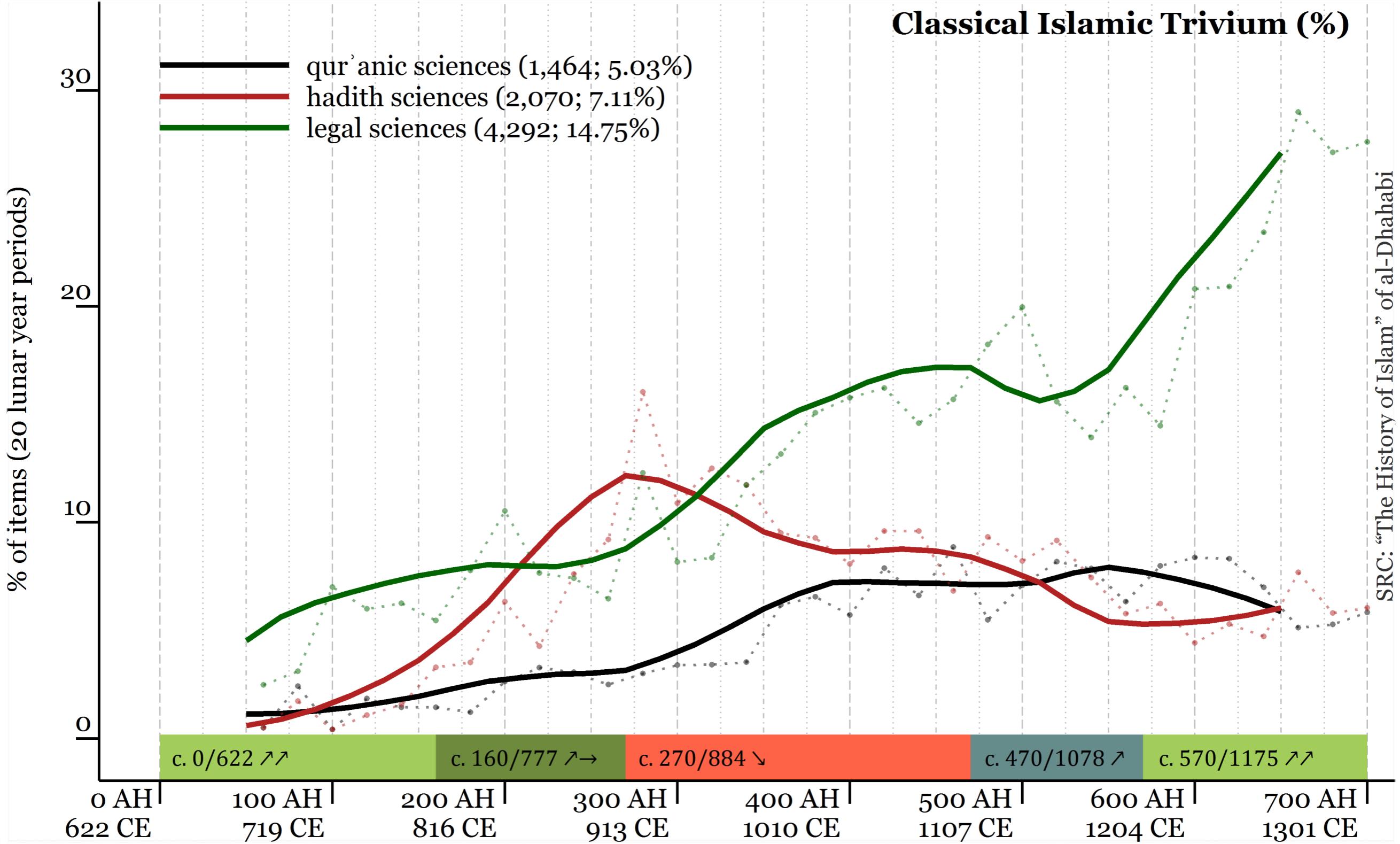
## Islamic intermediate centers (%)



## Islamic late centers (%)



## Classical Islamic Trivium (%)



# *al-Amṣār ḫawāt al-āṭār*

- ❖ Certain statements > about large groups in the *Ta'rīh al-islām*  
(at least 100 people, but usually much more)
- ❖ Uncertain statements > about small groups in the *Ta'rīh al-islām*
- ❖ Did he count?!

# Did al-Dahabī count?



## Reference:

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.

# Did al-Dahabī count?



# Did al-Dahabī count?

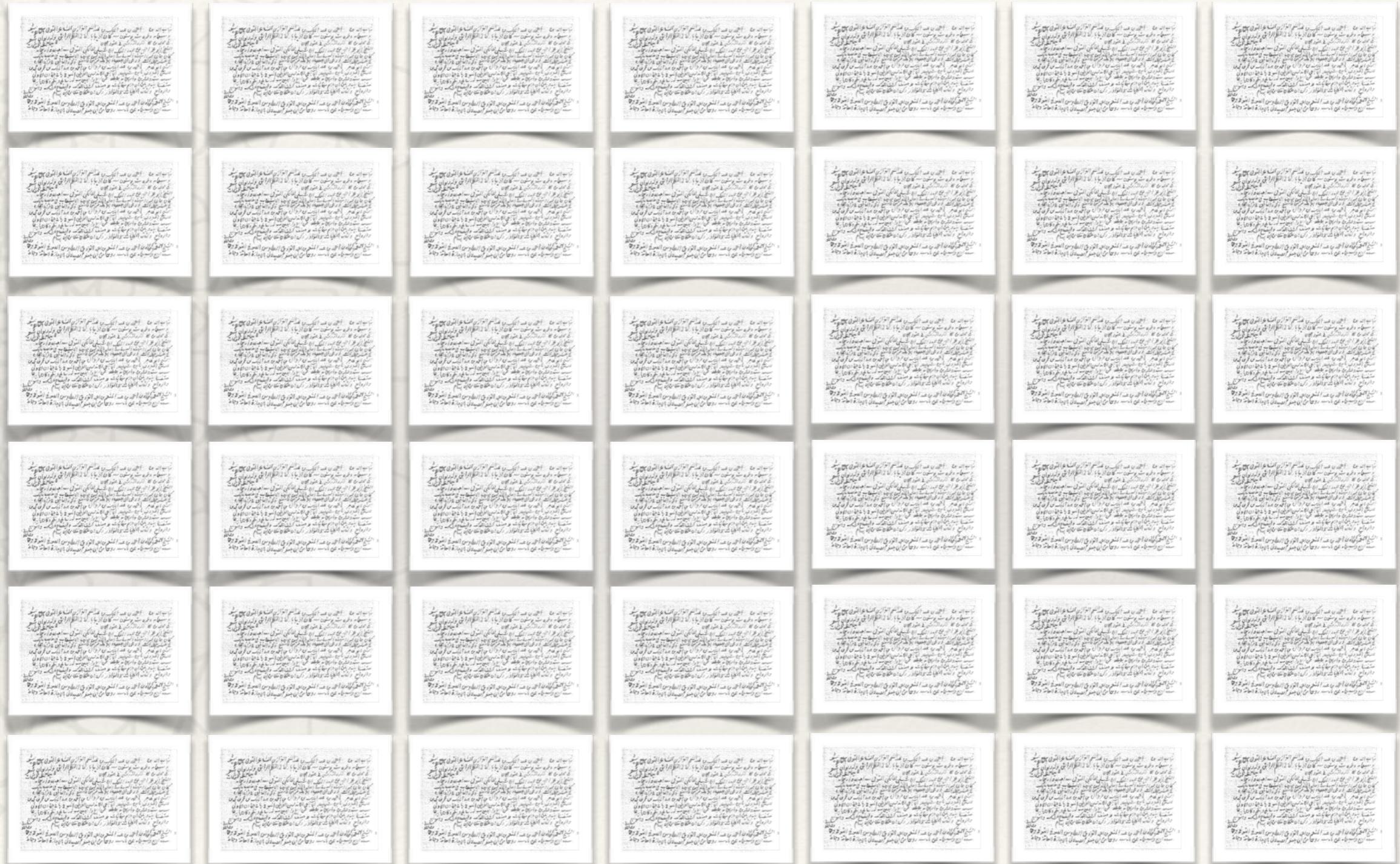
میتوانند بخواهند که اینکس میتواند از این امور ایجاد شود  
و همان‌طور که در متن مشاهده شد، میتواند با این اتفاق این امر را در میان افراد  
که میتوانند این اتفاق را در میان افراد داشته باشند، ایجاد کنند و این اتفاق را در میان افراد  
که میتوانند این اتفاق را در میان افراد داشته باشند، ایجاد کنند و این اتفاق را در میان افراد  
که میتوانند این اتفاق را در میان افراد داشته باشند، ایجاد کنند و این اتفاق را در میان افراد  
که میتوانند این اتفاق را در میان افراد داشته باشند، ایجاد کنند و این اتفاق را در میان افراد

لما سباهه الله تعالى أطلقوا على ذلك المذهب عبارة المذهب العثماني أو المذهب العثماني العثماني،  
وهو مذهب عثماني شرقي ينبع من العقيدة العثمانية التي اتبعتها الدولة العثمانية في إسلامها ودولتها ودينها.  
وهو مذهب عثماني شرقي ينبع من العقيدة العثمانية التي اتبعتها الدولة العثمانية في إسلامها ودولتها ودينها.

بـِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ  
أَلْهَدَنِي اللَّهُ أَكْبَرُ وَعَلَى الْوَالِدَيْنِ بِكَانَتْ  
بِسْمَهَا وَلَمْ يَكُنْ لَهُ شَرُوطٌ مَنْ كَانَ لِلْأَبِي بِالْأَنْوَارِ  
أَنْ يَكُونَ لَهُ شَرُوطٌ مَنْ كَانَ لِلْأَمْمَةِ  
أَنْ يَكُونَ لَهُ شَرُوطٌ مَنْ كَانَ لِلْأَمْمَةِ

لابد أن يكون أثليتين بعد المكبوت من هذة المعاشرات كالتالي: سبعة وعشرين سنتاً على سطح سفن — كافى لإنزال المركب إلى الماء في المحيط الأطلسي — ولذلك فإن سعر المركب يزيد على السعر المذكور على عقد المركب.

# Did al-Dahabī count?

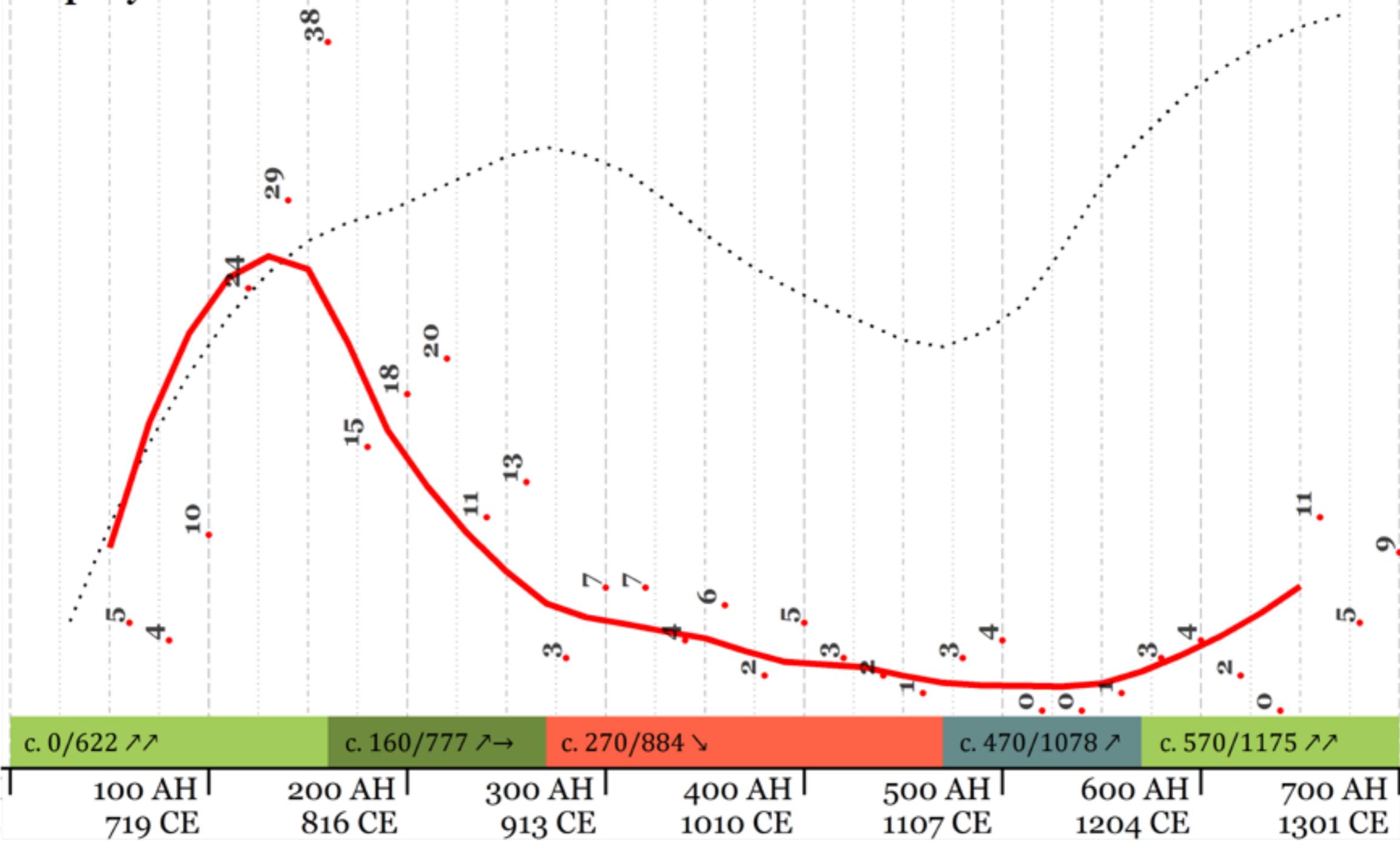


# Did al-Dahabī count?

**MAKKI (269 total; BW: mky)**

..... Cumulative biographical curve (29,110; scaled down by a factor of 38.03)

::toponymic



There's a graph  
for that

