

Maxim Romanov, U Leipzig

Harvard University, Nov 17, 2016

# From Text To Map

Premodern Arabic Texts  
& Digital Humanities



Digital Humanities  
UNIVERSITÄT LEIPZIG

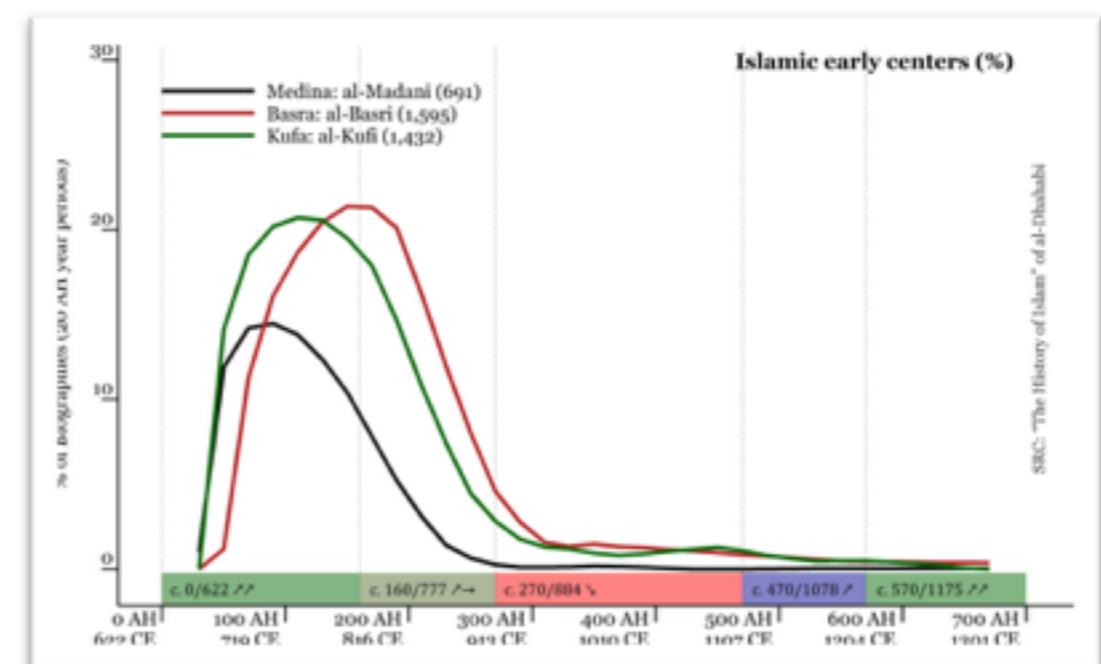
# Overview

- ❖ Collections of Arabic texts
- ❖ Analysis of a biographical collection (focusing on cultural geography)
- ❖ Recent work on Islamic geography:
  - ❖ Developing a Gazetteer
  - ❖ Modeling geography from primary texts

# In Practical Terms...



?

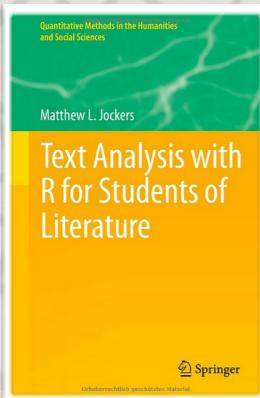
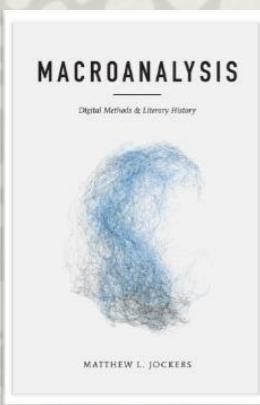
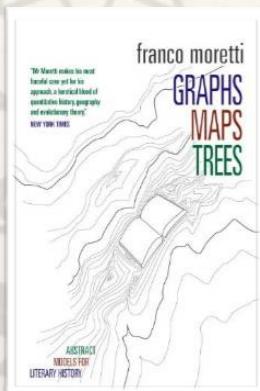


# Digital Humanities

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.

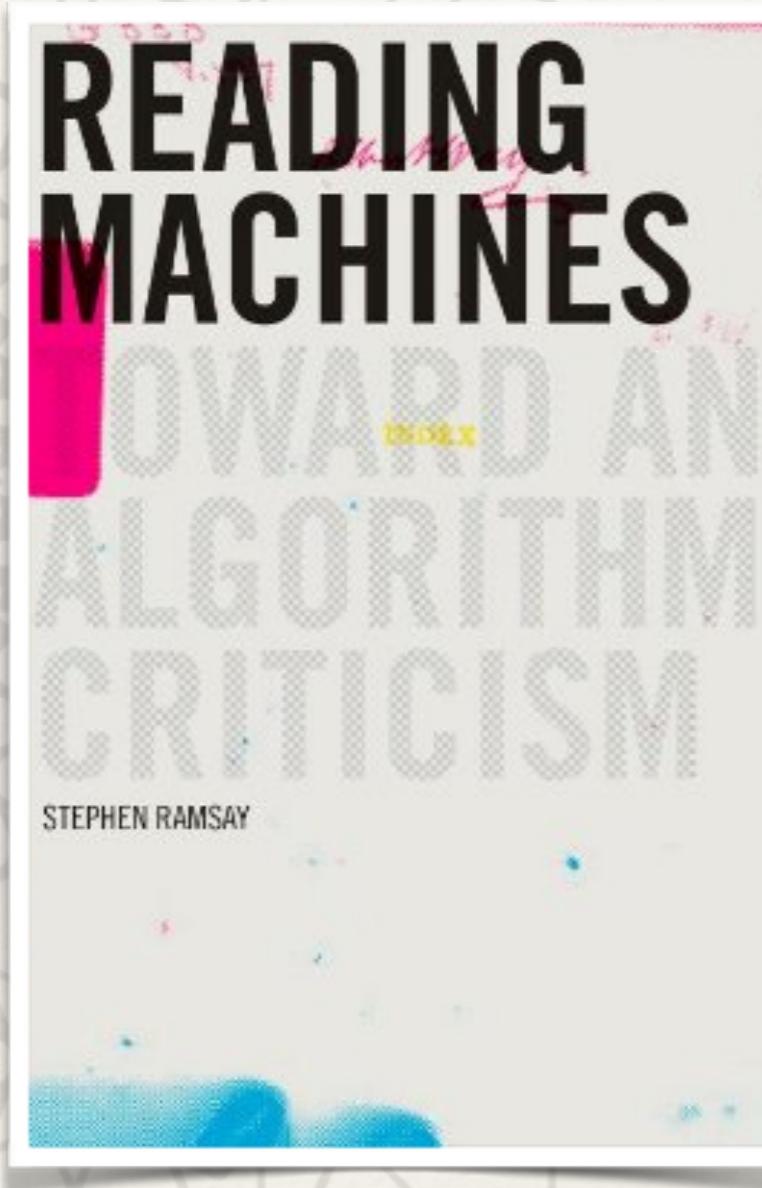
# Digital Humanities



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: *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.

# Overview

- ❖ **Collections of Arabic texts**
- ❖ Analysis of a biographical collection (focusing on cultural geography)
- ❖ Recent work on Islamic geography:
  - ❖ Developing a Gazetteer
  - ❖ Modeling geography from primary texts

# Collections of Arabic Texts

# Digital Collections of Arabic Texts

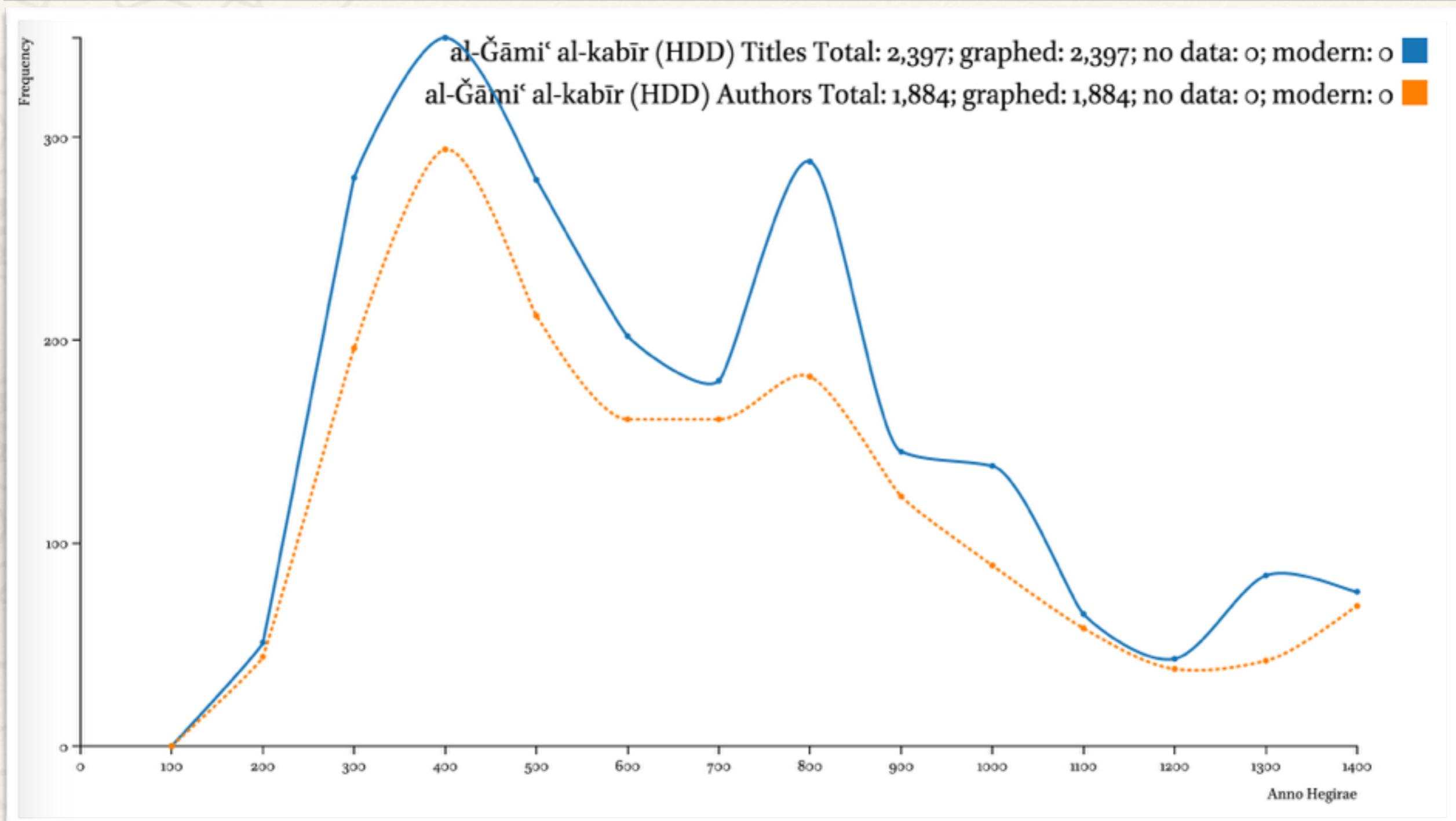
Library	Media	Titles	Vols	Words
<i>al-Jāmi‘ al-kabīr</i>	HDD, Windows 95	2,400	5,550	~400 mln
	<a href="http://www.shamela.ws">www.shamela.ws</a>	6,300	—	~820 mln
	<a href="http://www.almeshkat.net">www.almeshkat.net</a>	7,300	—	—
	<a href="http://www.said.net">www.said.net</a>	10,000	—	—
	<a href="http://www.alwaraq.com">www.alwaraq.com</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

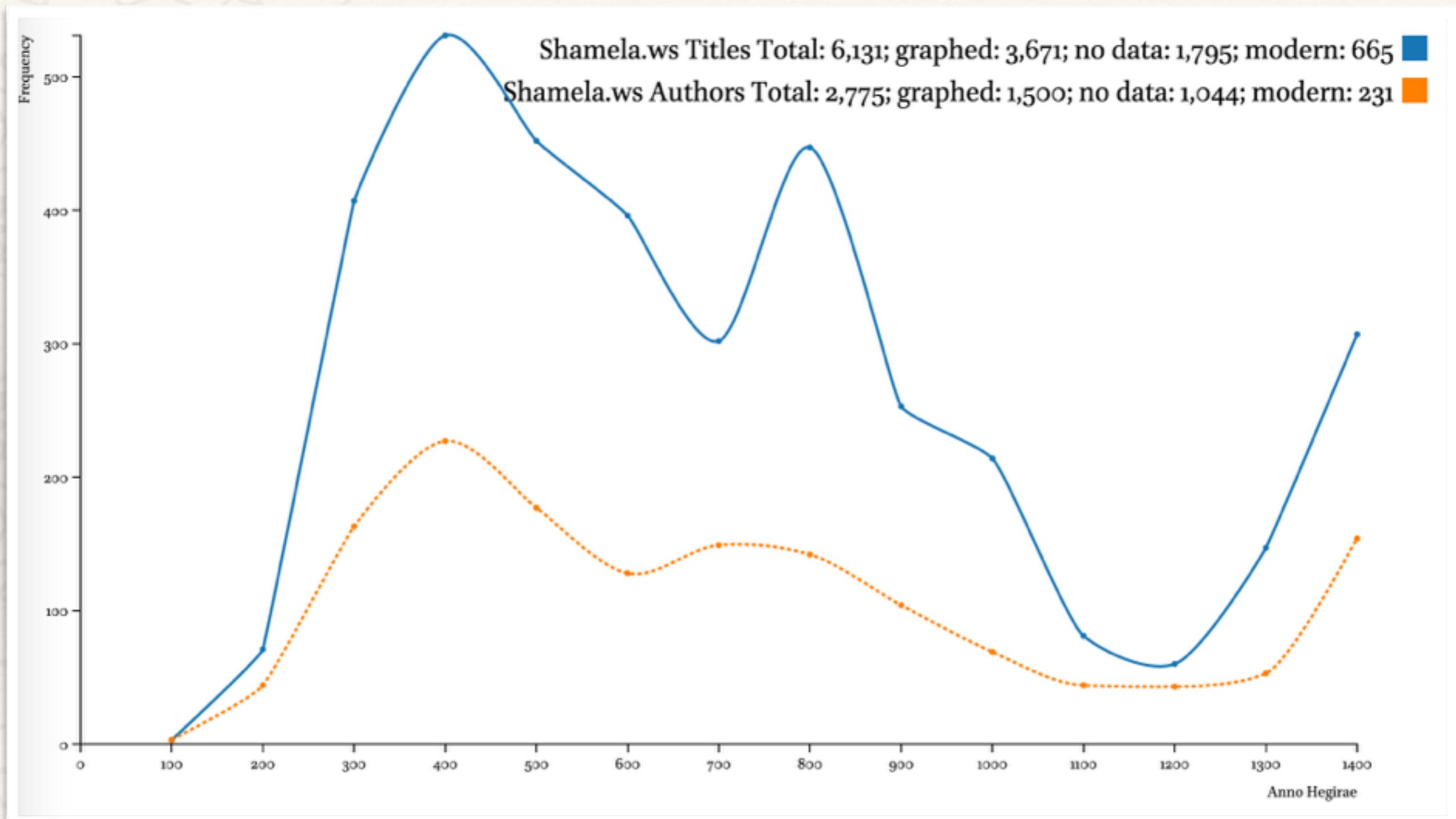
- ❖ Thematic/subject biases (“salafī” collections): almost no scientific, philosophical, medical and other texts
- ❖ Spotty chronological and geographical coverage
- ❖ Difficult to assess their overall coverage

NB: This is not a critique of these collections!  
We, academics, need to step in and fill in the blanks

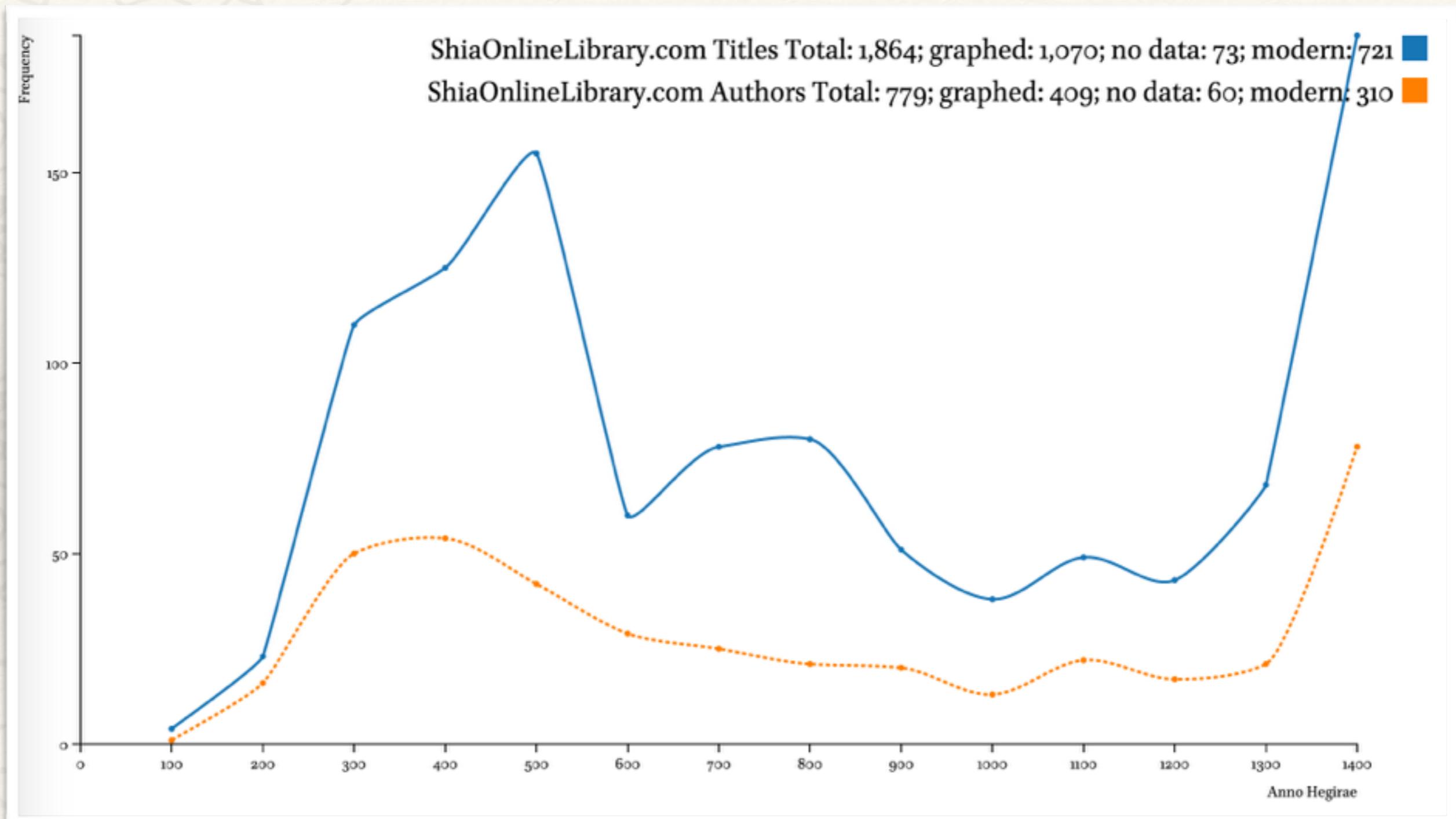
# Digital Collections: *Coverage Issues*



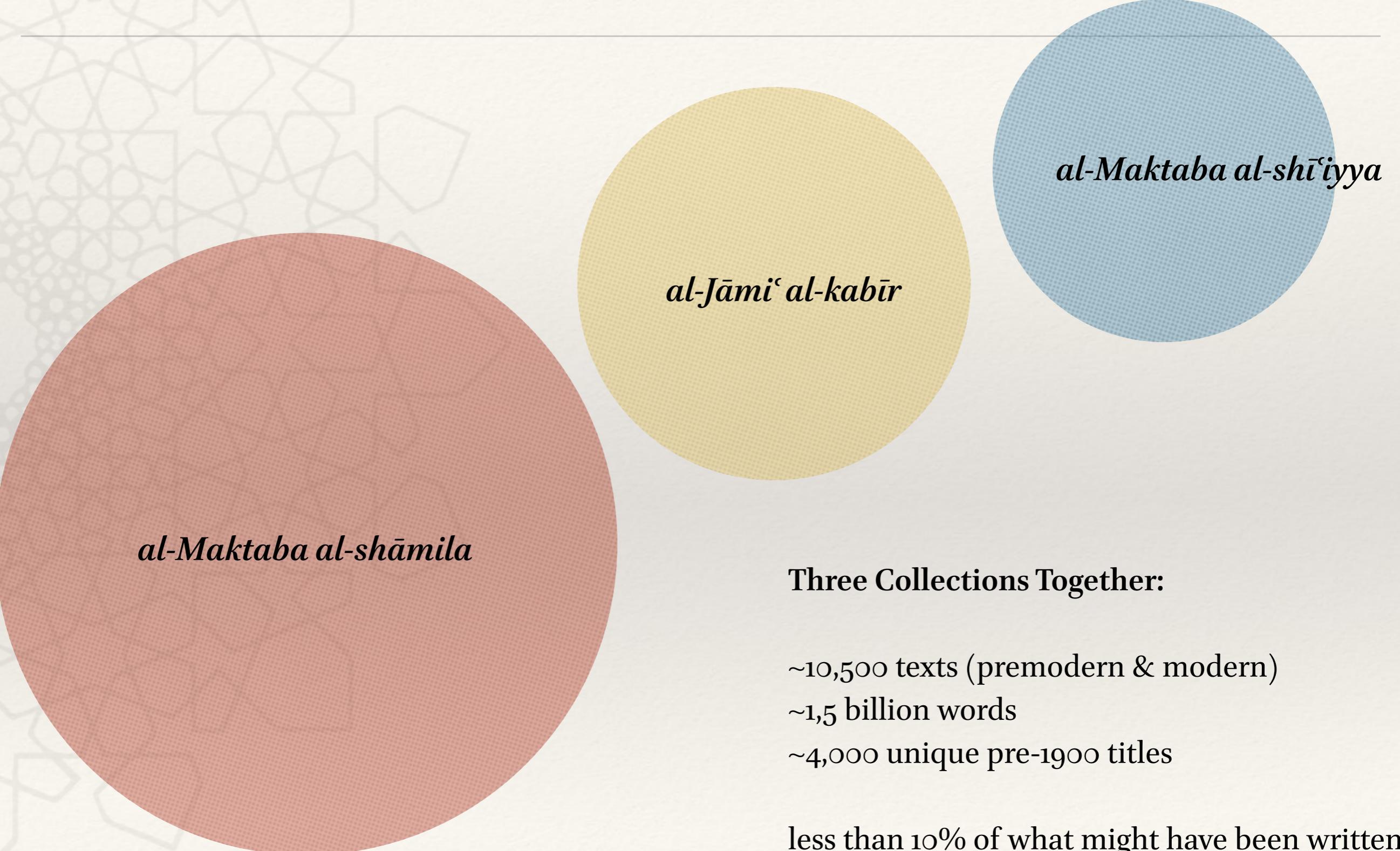
# Digital Collections: *Coverage Issues*



# 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:

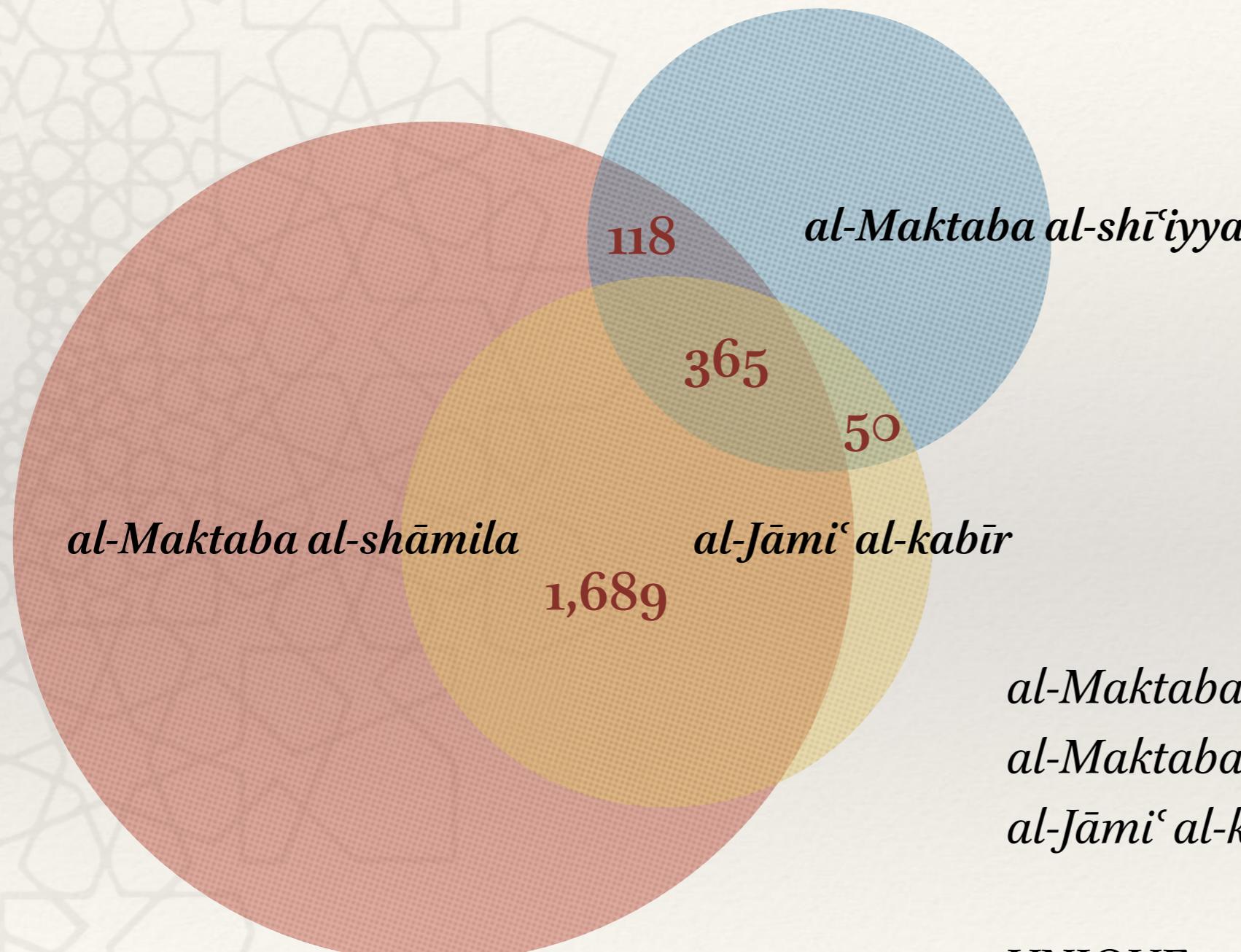
~10,500 texts (premodern & modern)

~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  
الحديث  
أصول الفقه ومسائله  
التاريخ  
الترجم والطبقات  
النحو والصرف  
الشعر والأدب  
المعاجم

Interpretation of the Qur'ān

"Prophetic sayings"

Legal writings

Chronicles

Biographical collections

Arabic language, grammar & morphology

Poetry & fine literature

Various dictionaries & references



How does this fit into  
the Arabic written  
tradition?

# “The Gift to the Knowledgeable”

Ismā‘īl Bāshā al-Bābānī al-Baghdādī (d. d. 1339/1920)

إِسْمَاعِيلْ باشا الْبَابَانِي

(١٣٣٩ - ١٩٢٠ هـ = ٠٠٠ - ٠٠٠ م )

إسماعيل بن محمد أمين بن مير سليم  
الباباني البغدادي : عالم بالكتب ومؤلفيها .  
باباني الأصل ، بغدادي المولد والمسكن .  
أقام زمناً في « مقرى كوي » بقرب  
الآستانة ، مشتغلًا باكمال كتابه « إيضاح  
المكتون في الذيل على كشف الظنون - ط »  
مجلدان . وله « هدية العارفين ، أسماء  
المؤلفين وآثار المصنفين - ط » في مجلدين<sup>(١)</sup> .



إسماعيل « باشا » بن محمد أمين الباباني

وفيما يلي خطه :

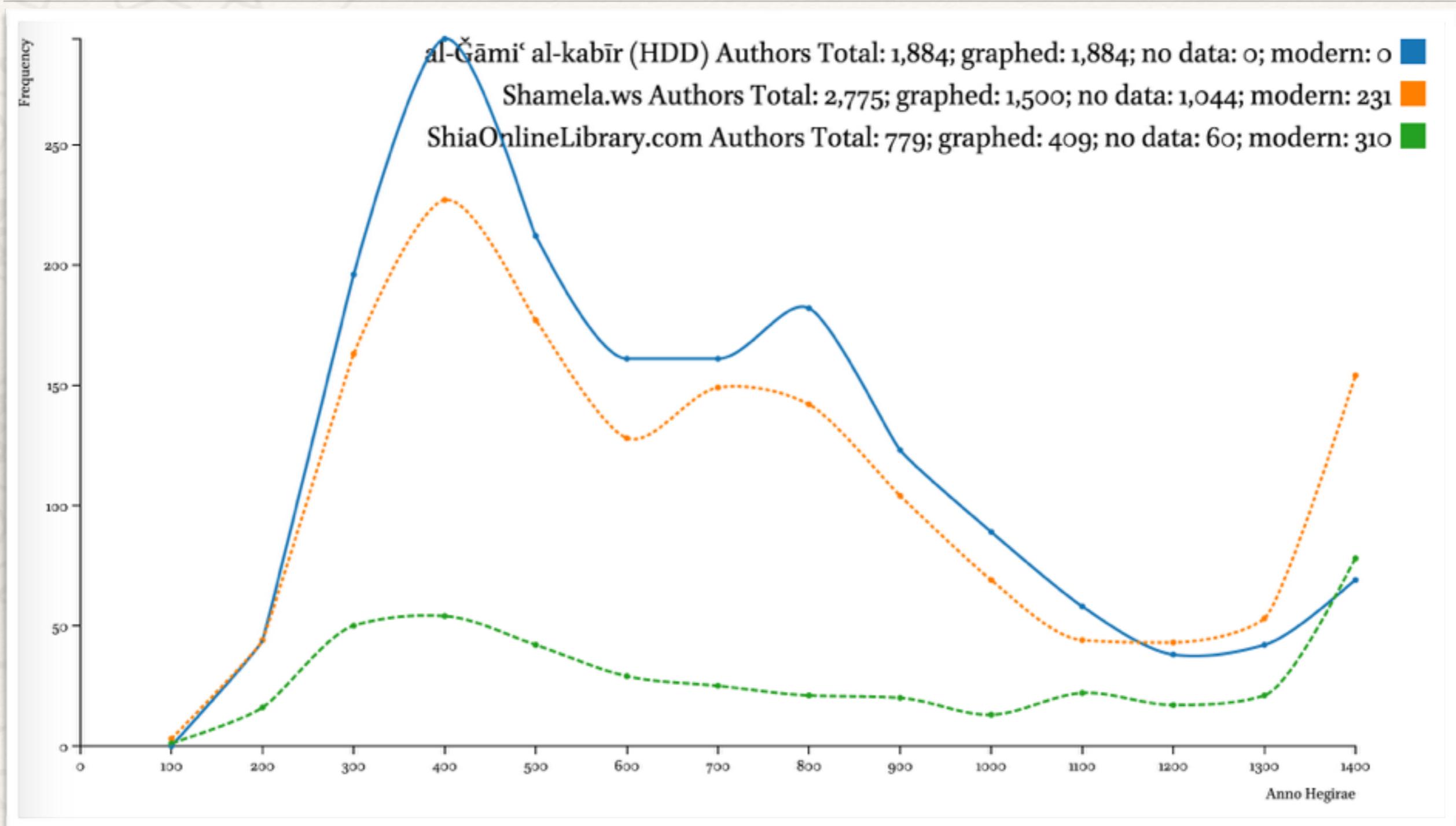
عن نهاية المجلد الأول من كتاب « كشف الظنون »، طبعة  
استانبول سنة ١٩٤١

فِرْدَوْسُ الْقُرْآنِ مِنْ جِبْرِيلٍ فِي ، تَسْلِيمُ بِحْرَفِهِ  
سَهْلَكْتُهُ عَلَيْهِ مَا سَعَى لِغَيْرِهِ  
اسْرَارُهُ مُحَمَّدٌ بِهِ بَشِّيرٌ

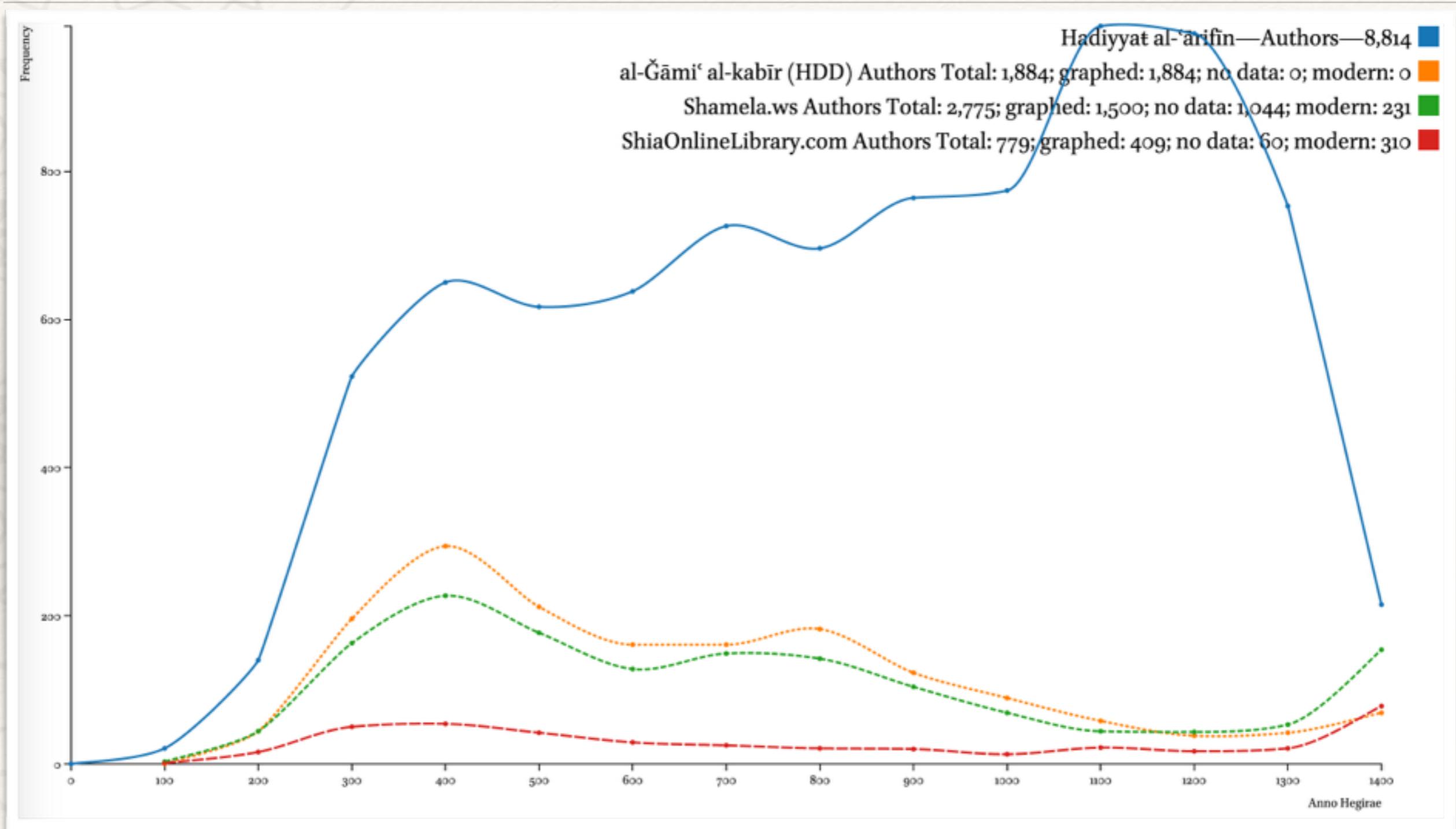
■ Source:

Khayr al-Dīn al-Ziriklī, *al-A'lām*, Beirut 1979, 326.

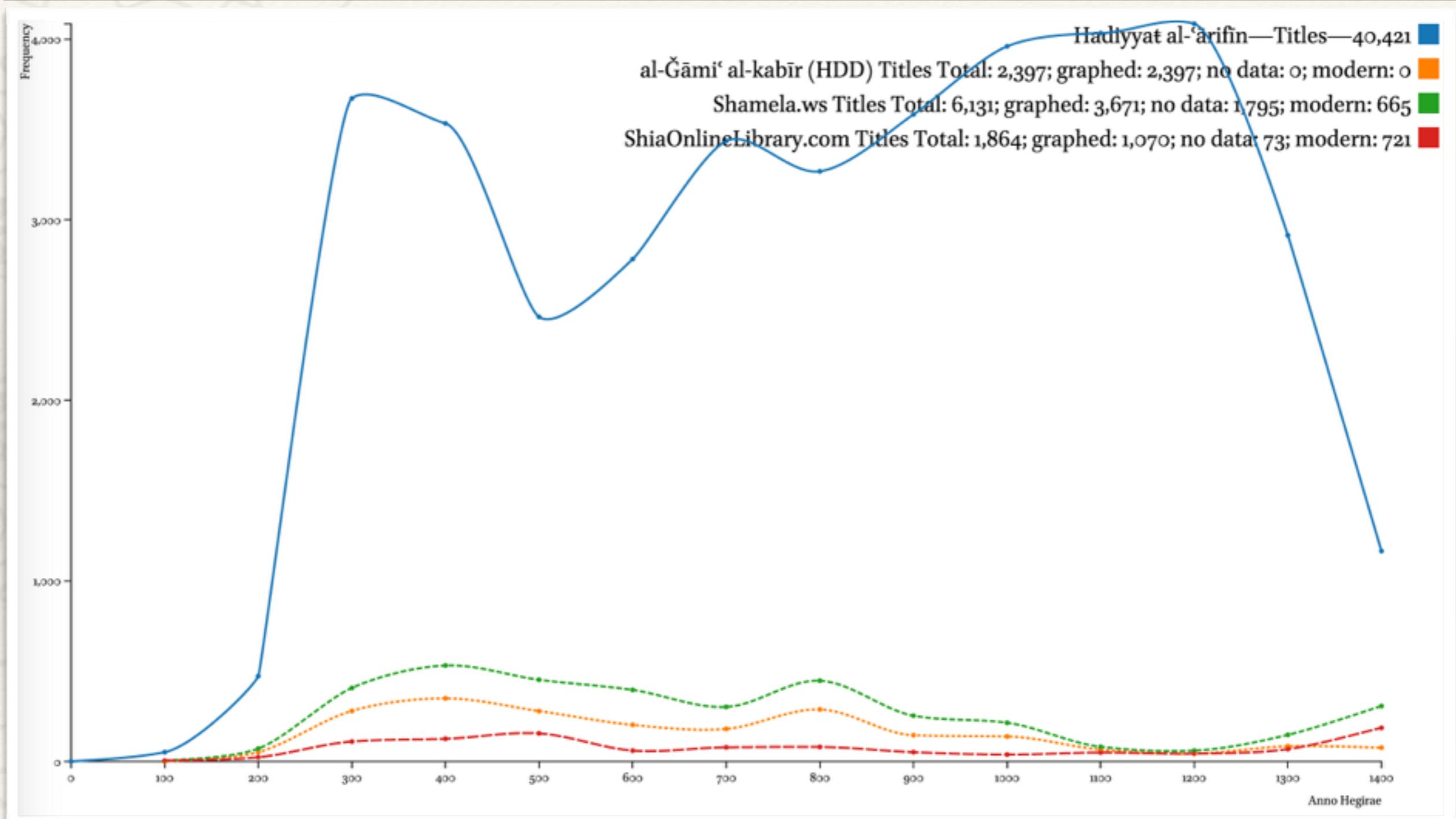
# Digital Collections and Tradition



# DC and Tradition: Authors



# DC and Tradition: Titles



# [Open]ITI

• ITI ISLAMICATE TEXTS INITIATIVE

[ABOUT](#)   [IMPLEMENTATION](#)   [TEAM](#)   [CONTACT](#)   [OCR](#)   [BLOG](#)

ایشان اماده حالت شدند از همه احتیاج علی‌اسا و موافقت ایشان و متعینت احتیاج باشد ایشان  
متعینت ایشان ترتیب و اصل نفع مانند احتیاج در ویشان بود با این احتمان و معروف و طلب فضیلت صداقت کار  
نفع منظوم است مردمان ایشان باعث می‌کردند بر مشارکت در معاملات و معاشرت داشتند اما جمله و متعینت با  
کلیک و لجیت در ریاضات و صید و دعوات تابعی سخن‌گویی است و انسقلپس کردن منعجه ایشان از کسانی  
که از هدایت خود را بجا بگذول و وقایع ایشان و ذکر عورت و مغایر و انتقامات خلوان نیکی کی ای مومن تدو در  
خطاب ایشان آن که الحاده والغیر اینها اکتاب موقت و ایغیر لازم آن فضیلت از دانشمندان قابل وحیت  
**ISLAM**  
**IN**

# SLAMICATE TEXTS INITIATIVE (ITI)

## Creating the digital infrastructure for study of the premodern Islamicate world

مطابعه انتها گشته و مصعبی و هوای خوش سخن کو افست تماجی سخن درین اب **فصل ششم**  
در فضیلت صداقت و کیفیت معاشرت با اصدقاء جون و دم همی با القلم است و تماجی مساعدة از بزرگ اصدقاء



# [Open]ITI

● ITI ISLAMICATE TEXTS INITIATIVE

ABOUT

IMPLEMENTATION

TEAM

CONTACT

OCR

BLOG

FAQS

العربية

## ITI TEAM

### CO-PRINCIPAL INVESTIGATORS



**Sarah Bowen Savant**

Associate Professor @ Institute for the Study of Muslim Civilisations, Aga Khan University, London ([kitab-project.org](http://kitab-project.org))



**Maxim Romanov**

Research Fellow @ Alexander von Humboldt-Lehrstuhl für Digital Humanities, Institut für Informatik, Universität Leipzig



**Matthew Thomas Miller**

Roshan Institute Research Fellow @ Roshan Institute for Persian Studies  
Associate Director, Roshan Initiative in Persian Digital Humanities ([PersDig@UMD](mailto:PersDig@UMD)), University of Maryland, College Park

### SENIOR ADVISORS

<http://iti-corpus.github.io/>

# [Open]ITI: Building a Corpus

- ❖ The Islamicate Texts Initiative (ITI) is a multi-institutional effort to construct the first machine-actionable scholarly corpus of premodern Islamicate texts. Led by researchers at the Aga Khan University (AKU), Universität Leipzig (UL), and the Roshan Institute for Persian Studies at the University of Maryland (College Park) and an interdisciplinary advisory board of leading digital humanists and Islamic, Persian, and Arabic studies scholars, ITI aims to provide the essential textual infrastructure in Persian and Arabic for new forms of macro textual analysis and digital scholarship. In the process, ITI will enable new synergies between Digital Humanities and the inter-related Islamicate fields of Islamic, Persian, and Arabic Studies.
  - ❖ CURATED METADATA (dates, geographies, classifications, etc.)
  - ❖ HIGH-QUALITY TEXTS with logical structure
  - ❖ UNIFIED LOCATIONS for ease of reuse

# OpenArabic

<https://github.com/OpenArabic/Annotation>



This repository

Search

Pull requests Issues Gist



## OpenArabic / Annotation

Unwatch ▾

3

Star ▾

6

Fork ▾

2

Code

Issues 0

Pull requests 0

Projects 0

Wiki

Pulse

Graphs

Settings

Description of the project and the status of development — Edit

325 commits

2 branches

0 releases

5 contributors

Branch: master ▾

New pull request

Create new file

Upload files

Find file

Clone or download ▾

maximromanov more books added

Latest commit 4452ae3 8 days ago

meta

Update priority\_list.md

21 days ago

readme

more books added

8 days ago

searches

search results for biographical collections

3 months ago

.gitignore

automatic update

3 months ago

README.md

more books added

8 days ago

folders.txt

more books added

8 days ago

metadata\_no\_dates\_records

dates added

2 months ago

README.md

**OpenArabic Project (@ AvH Lehrstuhl für Digital Humanities,  
U Leipzig, led and curated by Maxim Romanov)**

**Contents**

<https://github.com/OpenArabic/Annotation>

# OpenArabic mARkdown

*Tagging Arabic Historical Texts*



The main goal of **OpenArabic mARkdown** is to provide a simple system for tagging structural elements in premodern and early modern Arabic texts that are being prepared within the framework of the [OpenArabic Project](#).

The use of **OpenArabic mARkdown** will allow one to engage in the computational analysis of classical Arabic texts in the same way as more complex and time-consuming tagging schemes (like TEI XML);

**OpenArabic mARkdown** will also facilitate the conversion of the large volume of Arabic texts into TEI XML, which is now the standard format for digital editions. In principle, **OpenArabic mARkdown** does not require any special editor, but the current implementation relies on EditPad Pro, which supports right-to-left languages, Unicode, and large files. However, it is the support of custom highlighting and navigation schemes that makes this text editor particularly convenient for **OpenArabic mARkdown**.

<https://github.com/OpenArabic/Annotation>

# DC: *Fixing Coverage Issues — OCR*

- ❖ Kraken ibn Ocropus (*a fork of Ocropus*)

- Benjamin Kiessling, U Leipzig
- Matthew Miller, U of Maryland
- Sarah Savant, Aga Khan U—London
- Maxim Romanov, U Leipzig

Tested on Arabic, and most recently, on Syriac\*:  
*both in the high nineties*

with George Kiraz





# DC: Fixing Coverage Issues

<b>Book*</b>	<b>Quality</b>	<b>Type</b>	<b>Model accuracy level</b>			
			<b>Size 100</b>	<b>Ar**</b>	<b>Size 200</b>	<b>Ar**</b>
<b>1</b> Ibn al-Athīr. <i>al-Kāmil</i>	<i>high***</i>	<b>training</b>	93.79	97.71	93.58	97.59
<b>2</b> Ibn Qutayba. <i>Adab al-kātib</i>	<i>high***</i>	<i>testing</i>	82.68	95.72	80.92	94.88
<b>3</b> al-Jāhīz. <i>al-Hayawān</i>	<i>high***</i>	<i>testing</i>	71.78	75.16	70.85	74.27
<b>4</b> al-Ya‘qūbī. <i>al-Ta‘rīkh</i>	<i>high***</i>	<i>testing</i>	79.67	84.40	78.12	82.21
<b>5</b> al-Dhahabī. <i>Ta‘rīkh al-islām</i>	<i>low****</i>	<i>testing</i>	90.68	95.95	90.37	95.78
<b>6</b> Ibn al-Jawzī. <i>al-Muntazam</i>	<i>low****</i>	<i>testing</i>	93.33	98.51	92.96	98.22

\* Information on the editions of these texts is supplied at the end of the report

\*\* Performance on Arabic only (excluding punctuation and spaces)

\*\*\* 300 dpi, grayscale; scanned specifically for the purpose of testing, with ideal parameters

\*\*\*\* 200 dpi, black-and-white, pre-binarized; both downloaded from [www.archive.org](http://www.archive.org) (via <http://waqfeya.org>)



# DC: Fixing Coverage Issues

<b>Book*</b>	<b>Quality</b>	<b>Type</b>	<b>Model accuracy level</b>			
			<b>Size 100</b>	<b>Ar**</b>	<b>Size 200</b>	<b>Ar**</b>
<b>1</b> Ibn al-Athīr. <i>al-Kāmil</i>	<i>high***</i>	<i>testing</i>	80.23	86.27	82.46	87.48
<b>2</b> Ibn Qutayba. <i>Adab al-kātib</i>	<i>high***</i>	<i>testing</i>	80.90	91.54	82.61	93.24
<b>3</b> al-Jāhīz. <i>al-Hayawān</i>	<i>high***</i>	<b>training</b>	94.86	97.59	94.82	97.41
<b>4</b> al-Ya‘qūbī. <i>al-Ta‘rīkh</i>	<i>high***</i>	<i>testing</i>	90.91	95.13	91.28	94.71
<b>5</b> al-Dhahabī. <i>Ta‘rīkh al-islām</i>	<i>low****</i>	<i>testing</i>	81.93	91.23	83.03	92.22
<b>6</b> Ibn al-Jawzī. <i>al-Muntazam</i>	<i>low****</i>	<i>testing</i>	84.07	93.58	86.26	94.20

\* Information on the editions of these texts is supplied at the end of the report

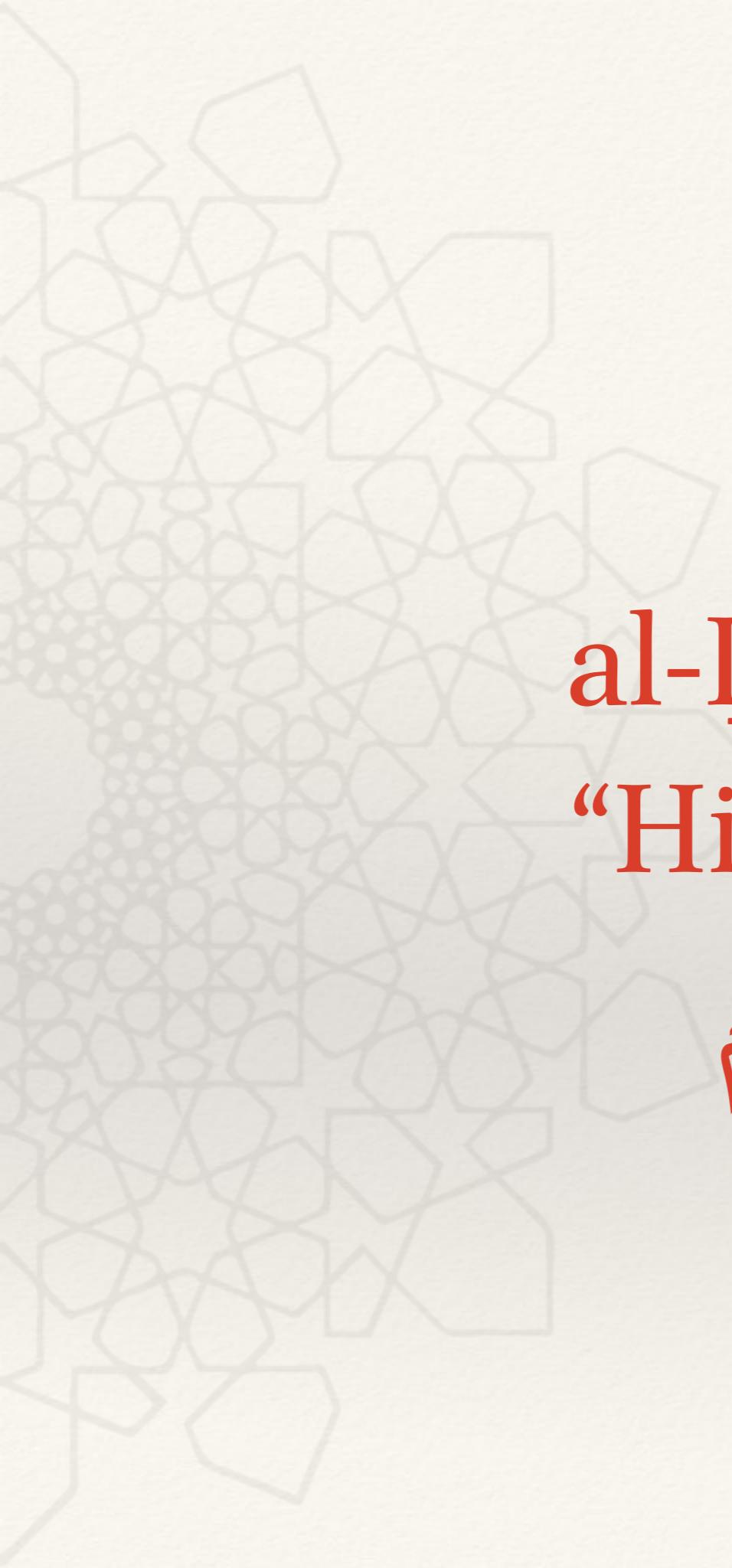
\*\* Performance on Arabic only (excluding punctuation and spaces)

\*\*\* 300 dpi, grayscale; scanned specifically for the purpose of testing, with ideal parameters

\*\*\*\* 200 dpi, black-and-white, pre-binarized; both downloaded from [www.archive.org](http://www.archive.org) (via <http://waqfeya.org>)

# Overview

- ❖ Collections of Arabic texts
- ❖ **Analysis of a biographical collection  
(focusing on cultural geography)**
- ❖ Recent work on Islamic geography:
  - ❖ Developing a Gazetteer
  - ❖ Modeling geography from primary texts

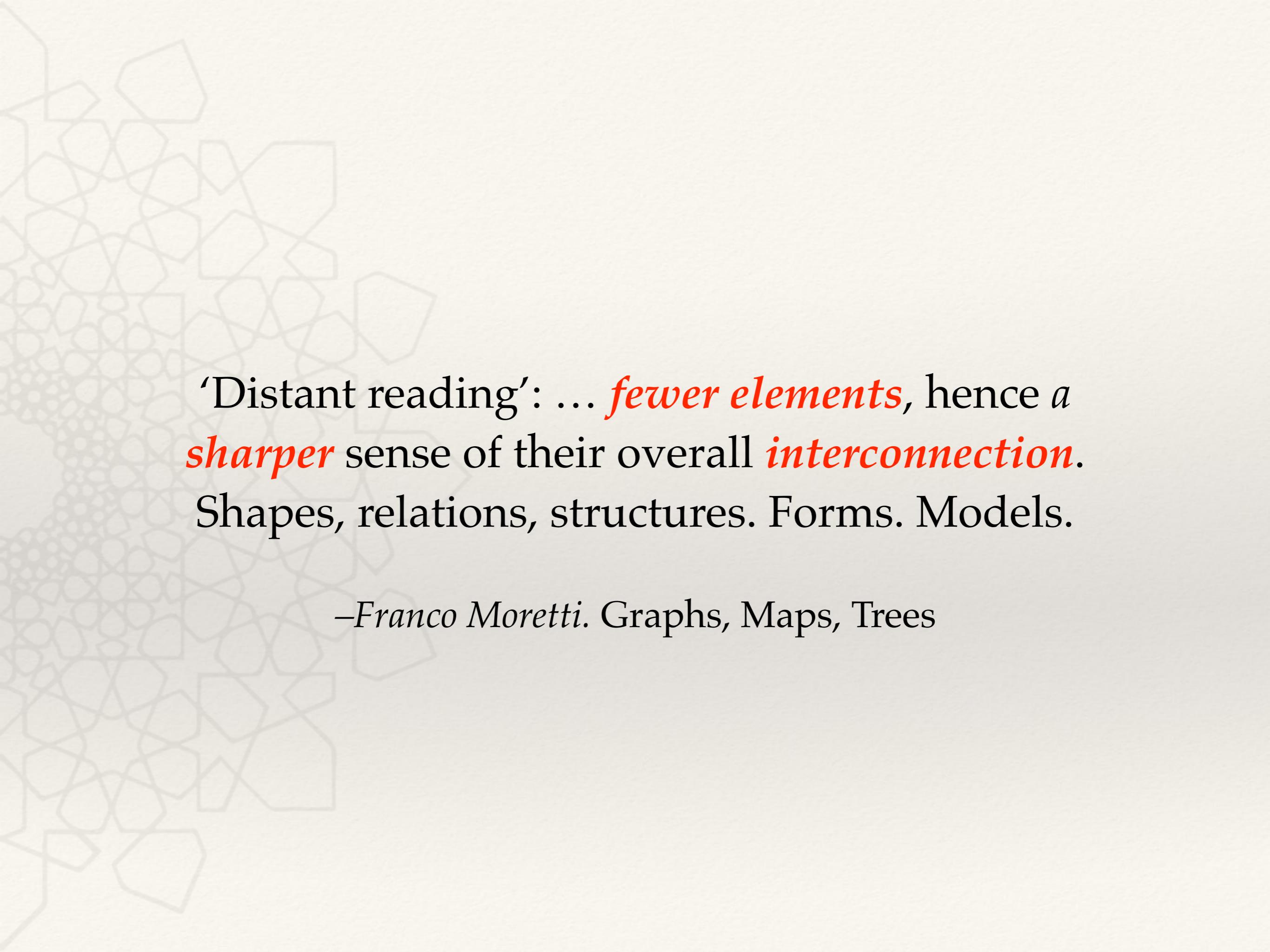


# al-Dahabī and his “History of Islam”

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)

# Algorithmic analysis: Tech Part

- ❖ Step-by-step algorithmic deformation of the original text
  - ❖ Tagging structure of initial texts
  - ❖ Processing biographies
  - ❖ Onomastic section > descriptive names
  - ❖ Dates > from structure
  - ❖ Toponyms > ngram search
- ❖ Mapping and graphing data

# Algorithmic analysis: Tech Part

- ❖ Frequency lists of suspects:
    - ❖ Descriptive names: morphological patterns and *auxiliary lists*\*
    - ❖ Toponyms: ngram search and *auxiliary lists*\*
    - ❖ Dates: from structural data; or with RE
- \* Auxiliary lists provide grouping into large categories (synsets: settlements > regions > provinces; toponyms > coordinates, etc.)

# Algorithmic analysis: Tech Part

1	00003086»	· \$baghdadi · #toponymic» «بغداد» MOD <sup>C L</sup> <sub>R F</sub>
2	00001935»	· \$shaykh · #status · #religious · @general» NA» general <sup>C L</sup> <sub>R F</sub>
3	00001930»	· \$faqih · #knowledge · #religious · @law» NA» law <sup>C L</sup> <sub>R F</sub>
4	00001595»	· \$basri · #toponymic» «البصرة» MOD <sup>C L</sup> <sub>R F</sub>
5	00001573»	· \$dimashqi · #toponymic» «دمشق» MOD <sup>C L</sup> <sub>R F</sub>
6	00001501»	· \$misri · #toponymic» «مصر» MOD <sup>C L</sup> <sub>R F</sub>
7	00001435»	· \$qadi · #position · #religious · @law» NA» law <sup>C L</sup> <sub>R F</sub>
8	00001432»	· \$kufi · #toponymic» «الكوفة» MOD <sup>C L</sup> <sub>R F</sub>
9	00001207»	· \$mawla · #tribal · @client · @nonarab · @islamic» NA» client <sup>C L</sup> <sub>R F</sub>
10	00001194»	· \$muqri' · #religious · #knowledge · @quran» NA» quran <sup>C L</sup> <sub>R F</sub>
11	00001161»	· \$imam · #status · #religious · #position · @general» NA» general <sup>C L</sup> <sub>R F</sub>
12	00001100»	· \$isbahani · #toponymic» «اصبهان» MOD <sup>C L</sup> <sub>R F</sub>
13	00001095»	· \$sahib · #noID» NA» MOD <sup>C L</sup> <sub>R F</sub>
14	00001081»	· \$hafiz · #religious · #knowledge · @hadith» NA» hadith <sup>C L</sup> <sub>R F</sub>
15	00001038»	· \$naysaburi · #toponymic» «نيسابور» MOD <sup>C L</sup> <sub>R F</sub>
16	00000990»	· \$shafi'i · #religious · #knowledge · #faction · @law» NA» law <sup>C L</sup> <sub>R F</sub>
17	00000983»	· \$salih · #status · #religious · @piety» NA» piety <sup>C L</sup> <sub>R F</sub>
18	00000885»	· \$zahid · #status · #religious · @piety» NA» piety <sup>C L</sup> <sub>R F</sub>
19	00000876»	· \$sansari · #tribal · @prophet» NA» prophet <sup>C L</sup> <sub>R F</sub>
20	00000815»	· \$amir · #status · #military · #position» NA» MOD <sup>C L</sup> <sub>R F</sub>
21	00000691»	· \$madani · #toponymic» «المدينة» MOD <sup>C L</sup> <sub>R F</sub>
22	00000651»	· \$sufi · #religious · #knowledge · #status · @piety» NA» piety <sup>C L</sup> <sub>R F</sub>
23	00000634»	· \$qurtubi · #toponymic» «قرطبي» MOD <sup>C L</sup> <sub>R F</sub>
24	00000625»	· \$qurashi · #tribal · @arab · @prophet» NA» prophet <sup>C L</sup> <sub>R F</sub>

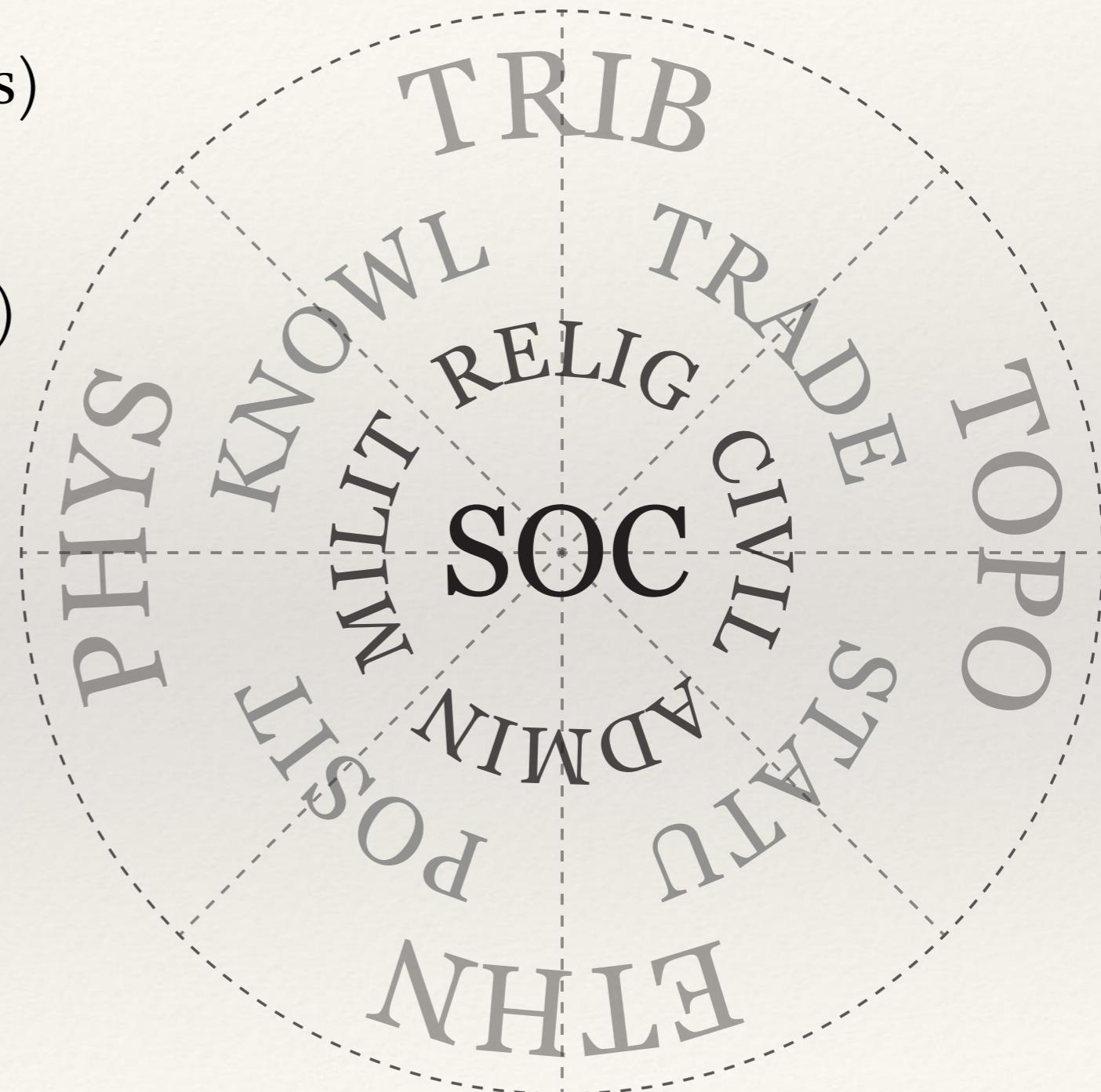
# Social transformations

# Classification of *Nisbats* from the Social Perspective

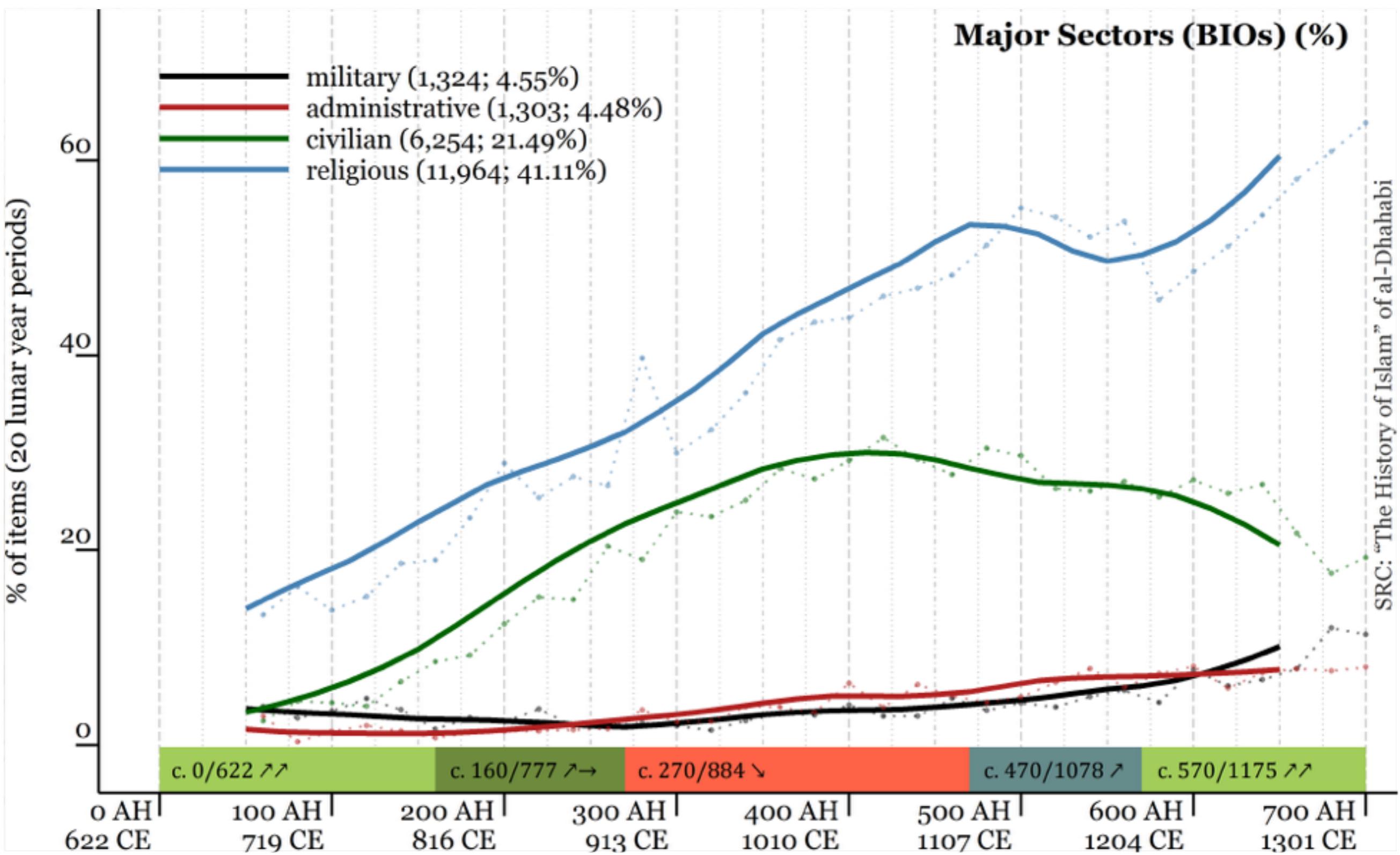
- ❖ “Descriptive names” (*nisbats*)
  - ❖ 700 unique  
(frequency 10 and higher)
  - ❖ 70,000 total

Previous work on *nisbats*  
(mostly occupational):

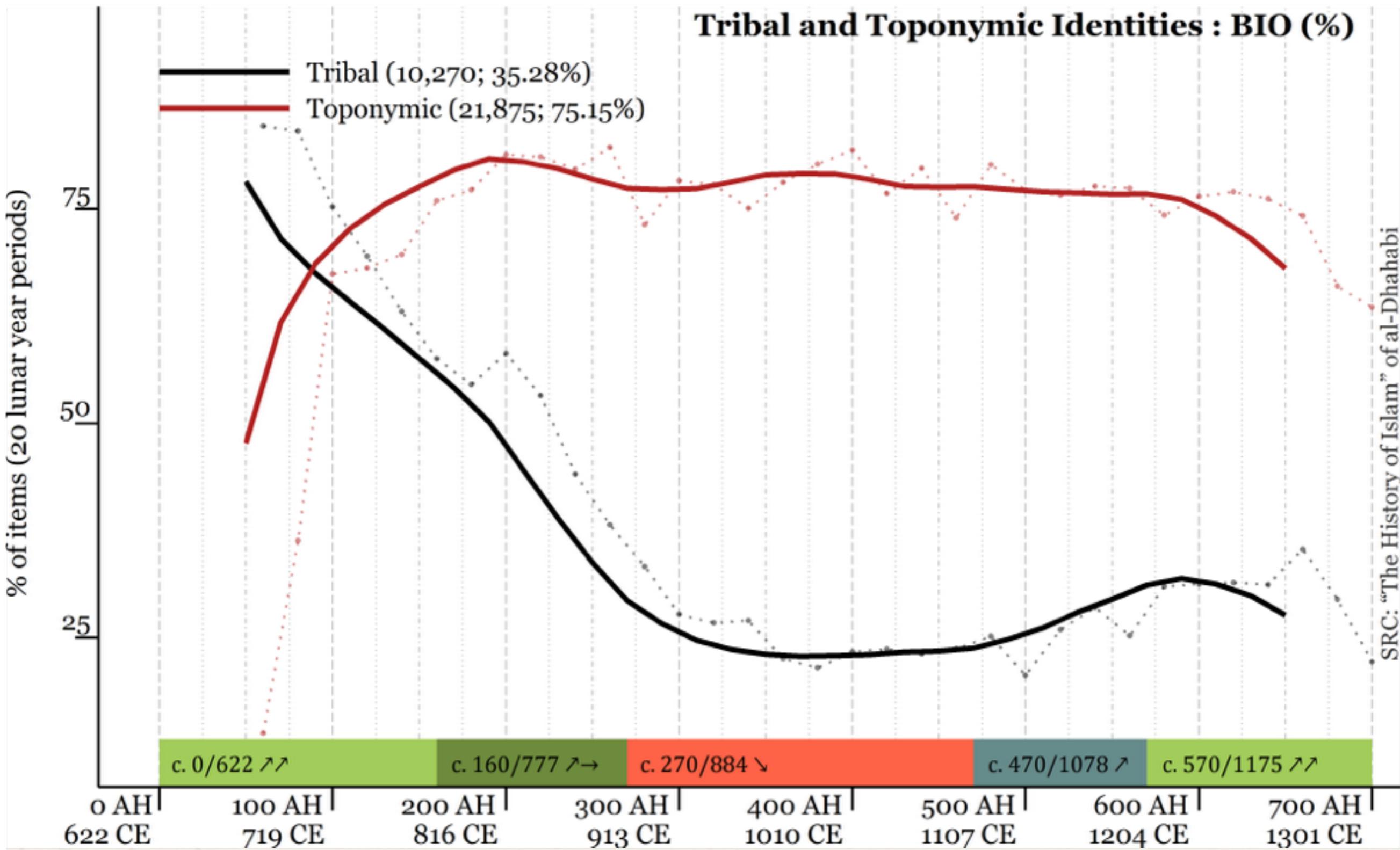
- Hayyim Cohen
- Carl Petry
- Maya Shatzmiller



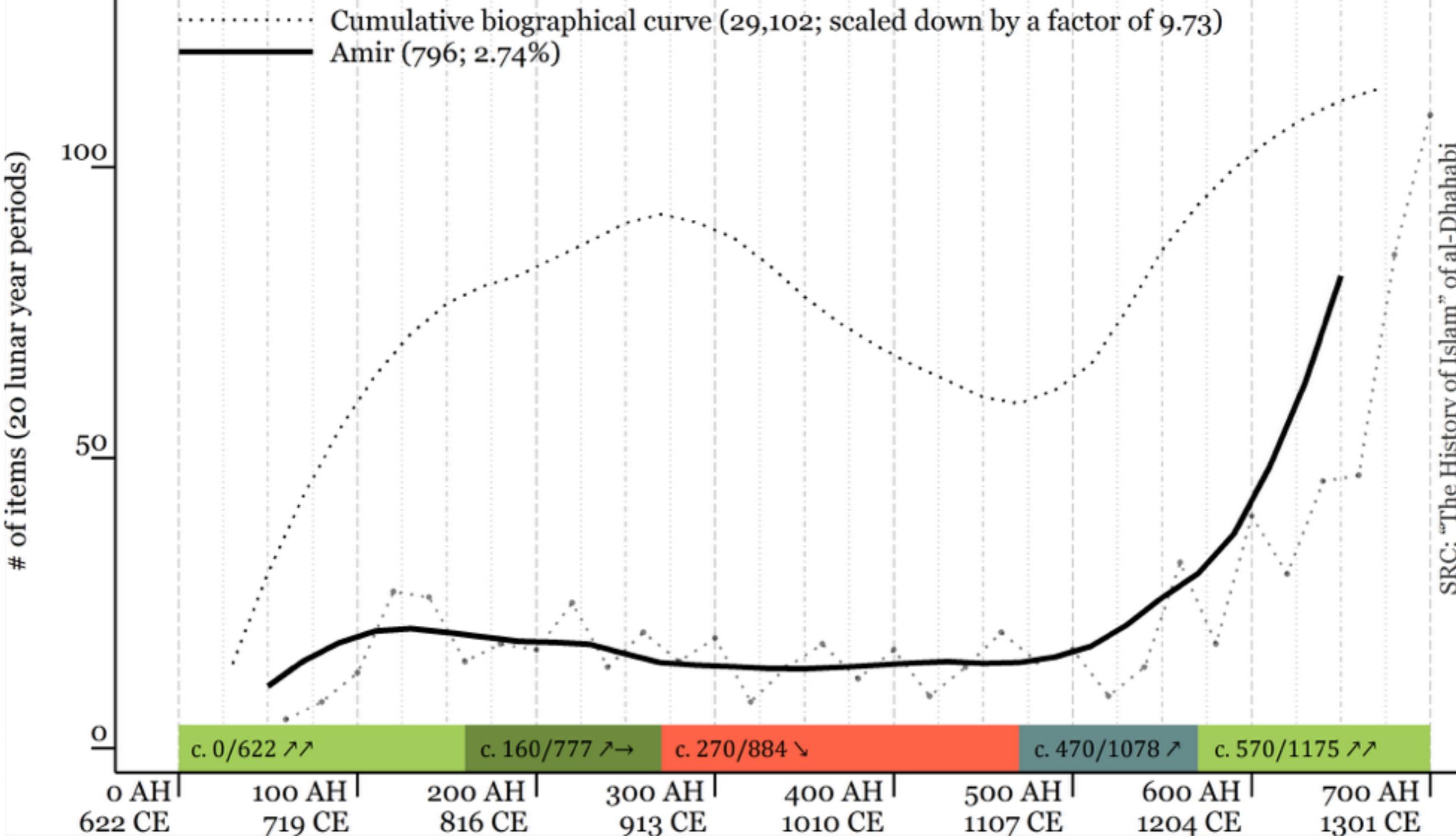
## Major Sectors (BIOS) (%)



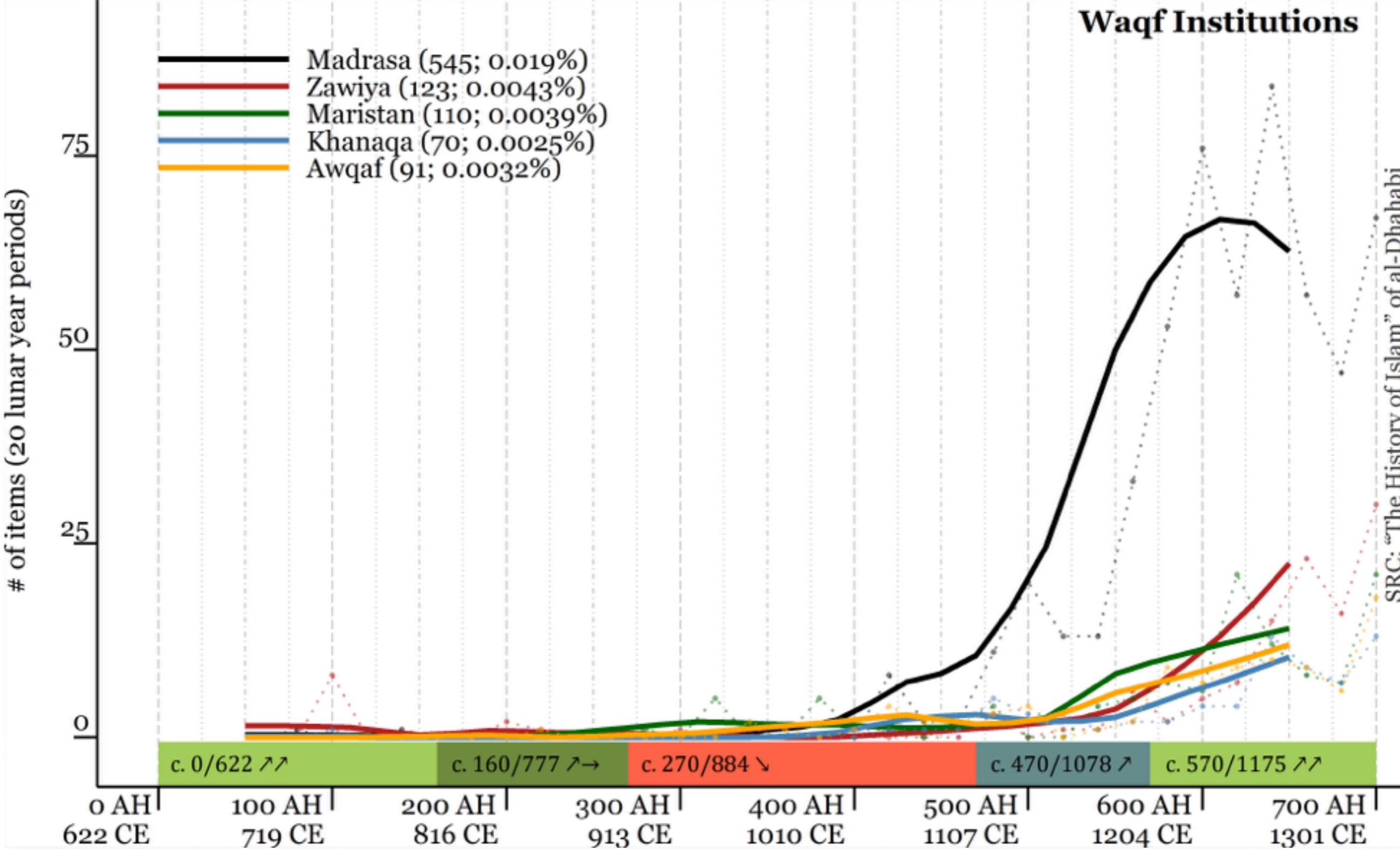
## Tribal and Toponymic Identities : BIO (%)



## Military Sector (BIOs)



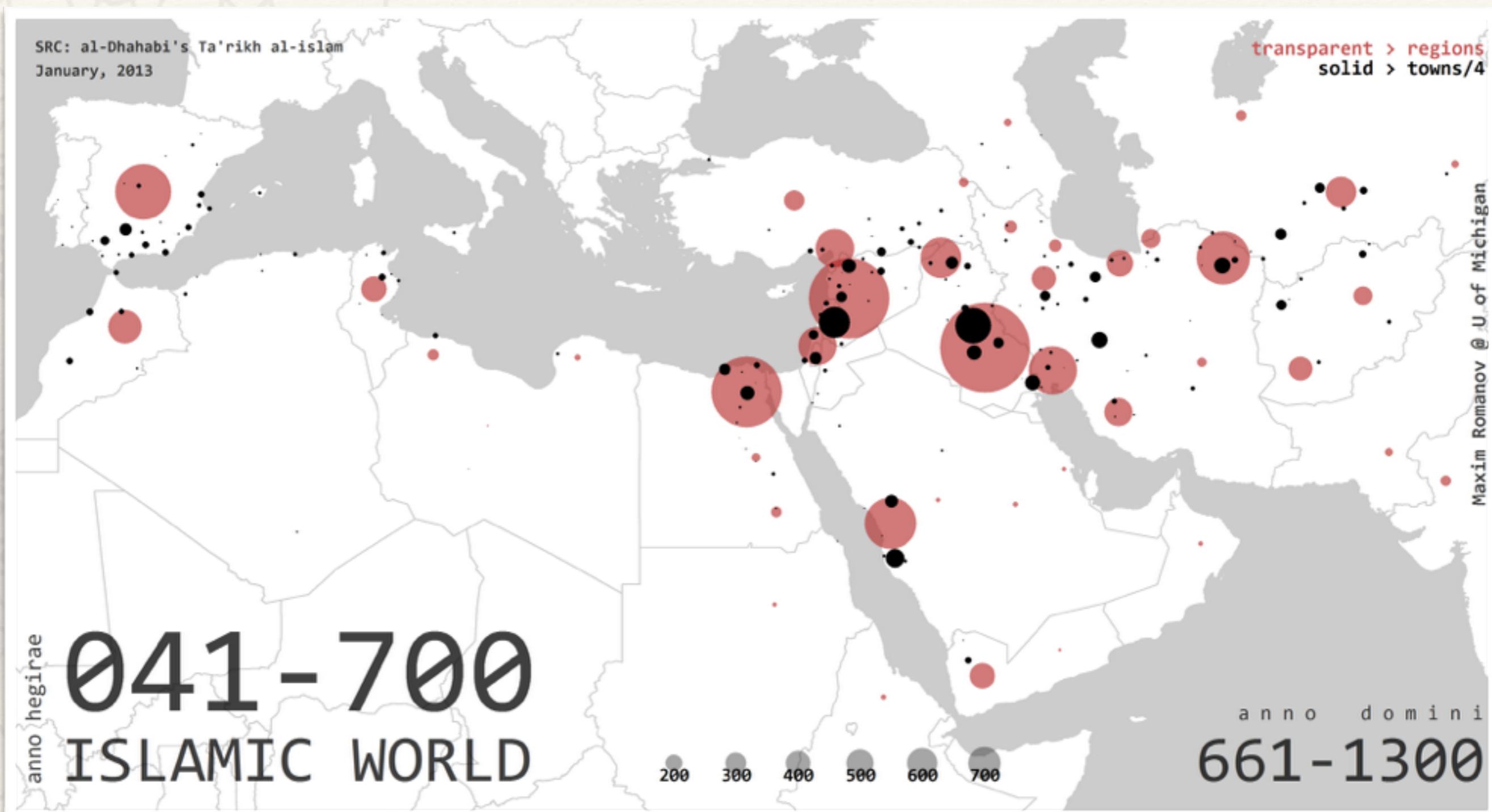
## Waqf Institutions



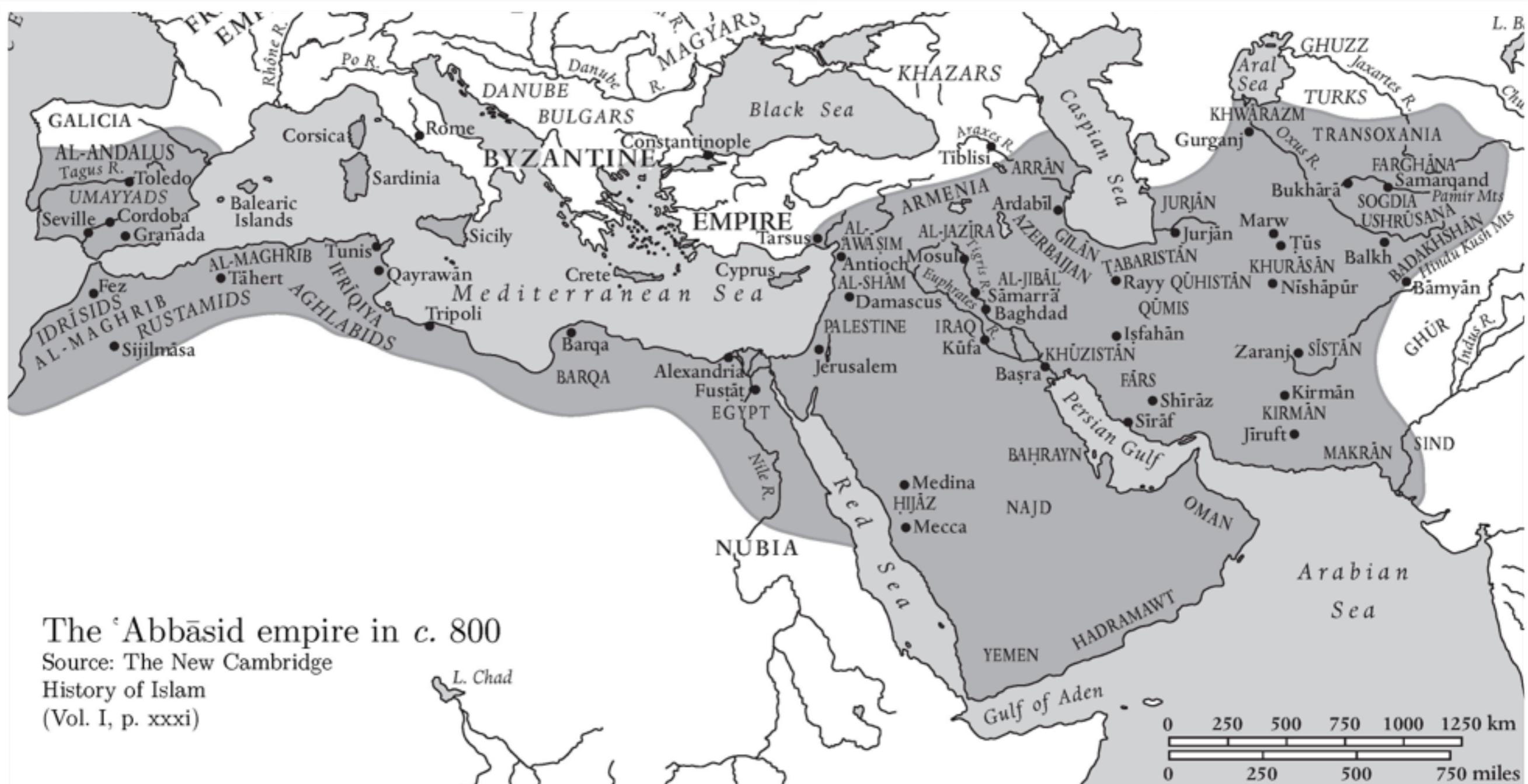


# Social Geography: Coverage

# Geographical Coverage

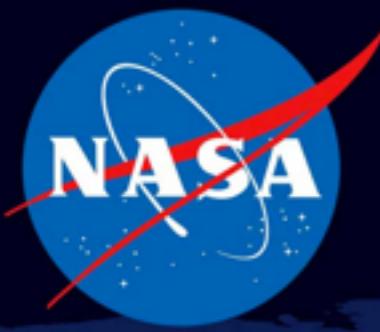
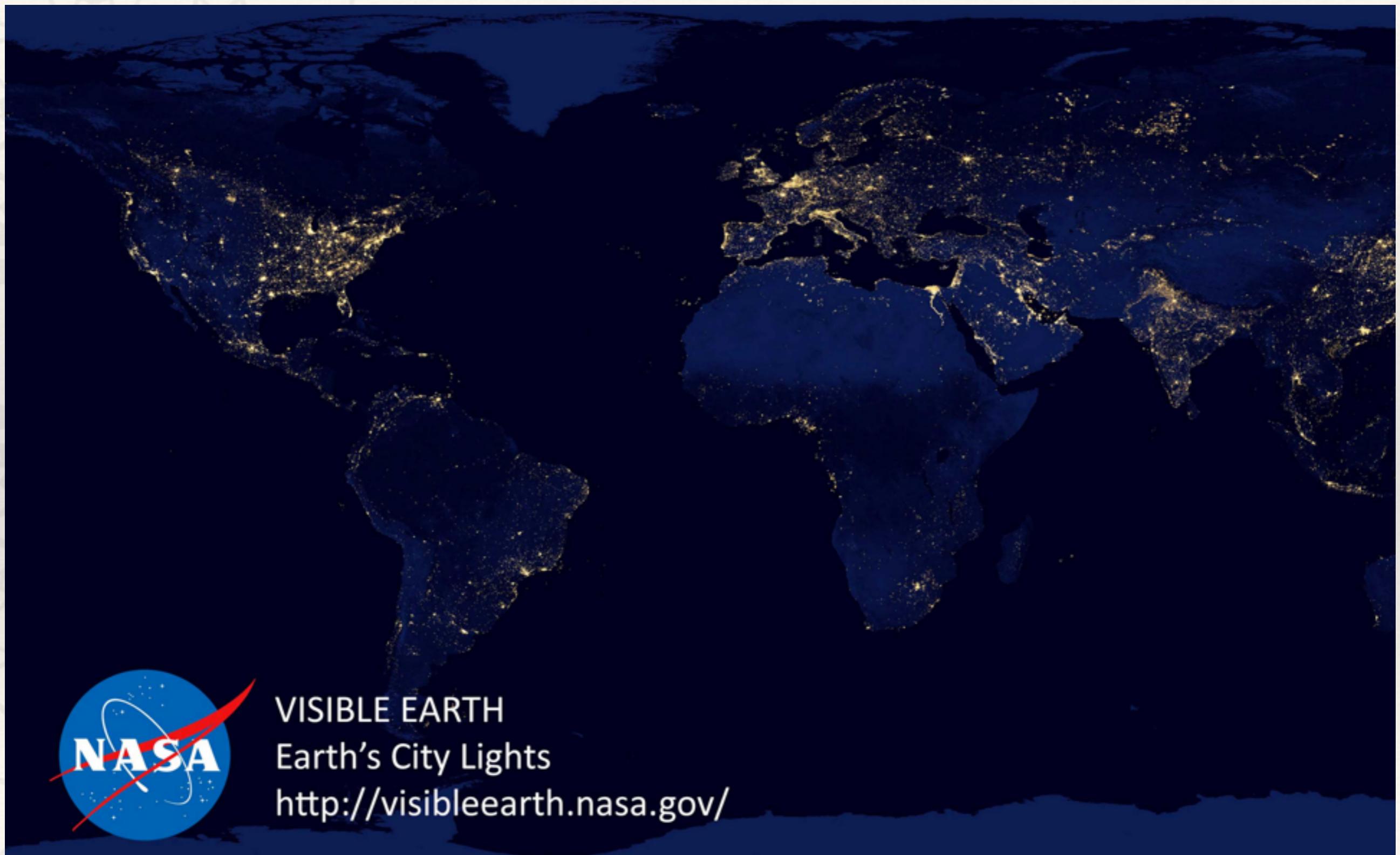


# Geographical Coverage



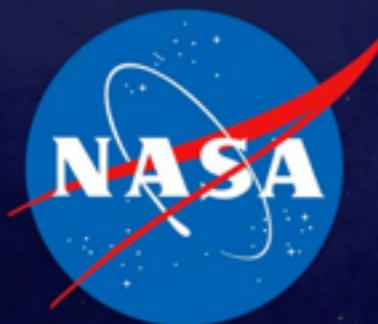
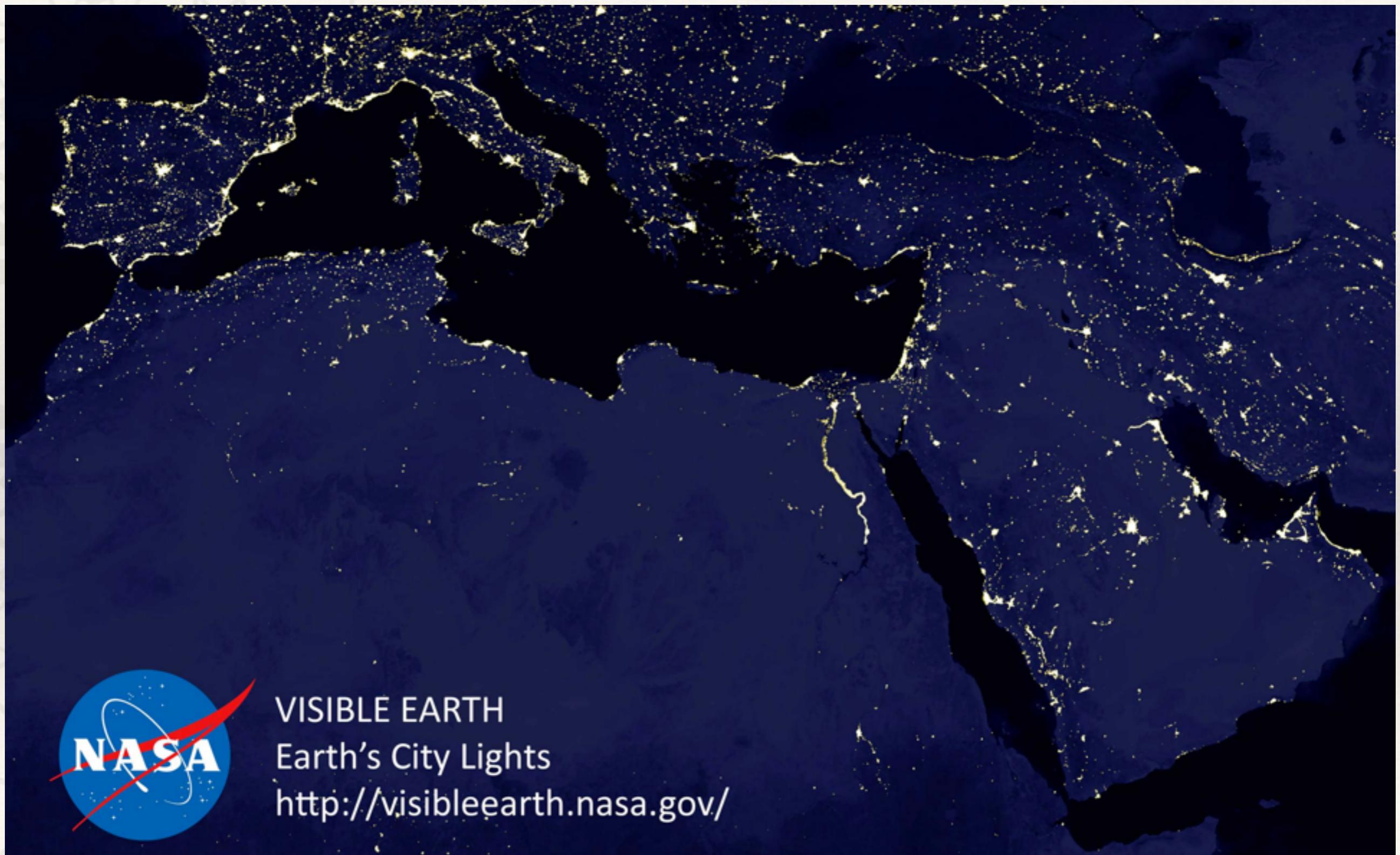
Source: Robinson, Chase F., ed. *The New Cambridge History of Islam*. 1st ed. Cambridge University Press, 2010.

# Geographical Coverage



VISIBLE EARTH  
Earth's City Lights  
<http://visibleearth.nasa.gov/>

# Geographical Coverage

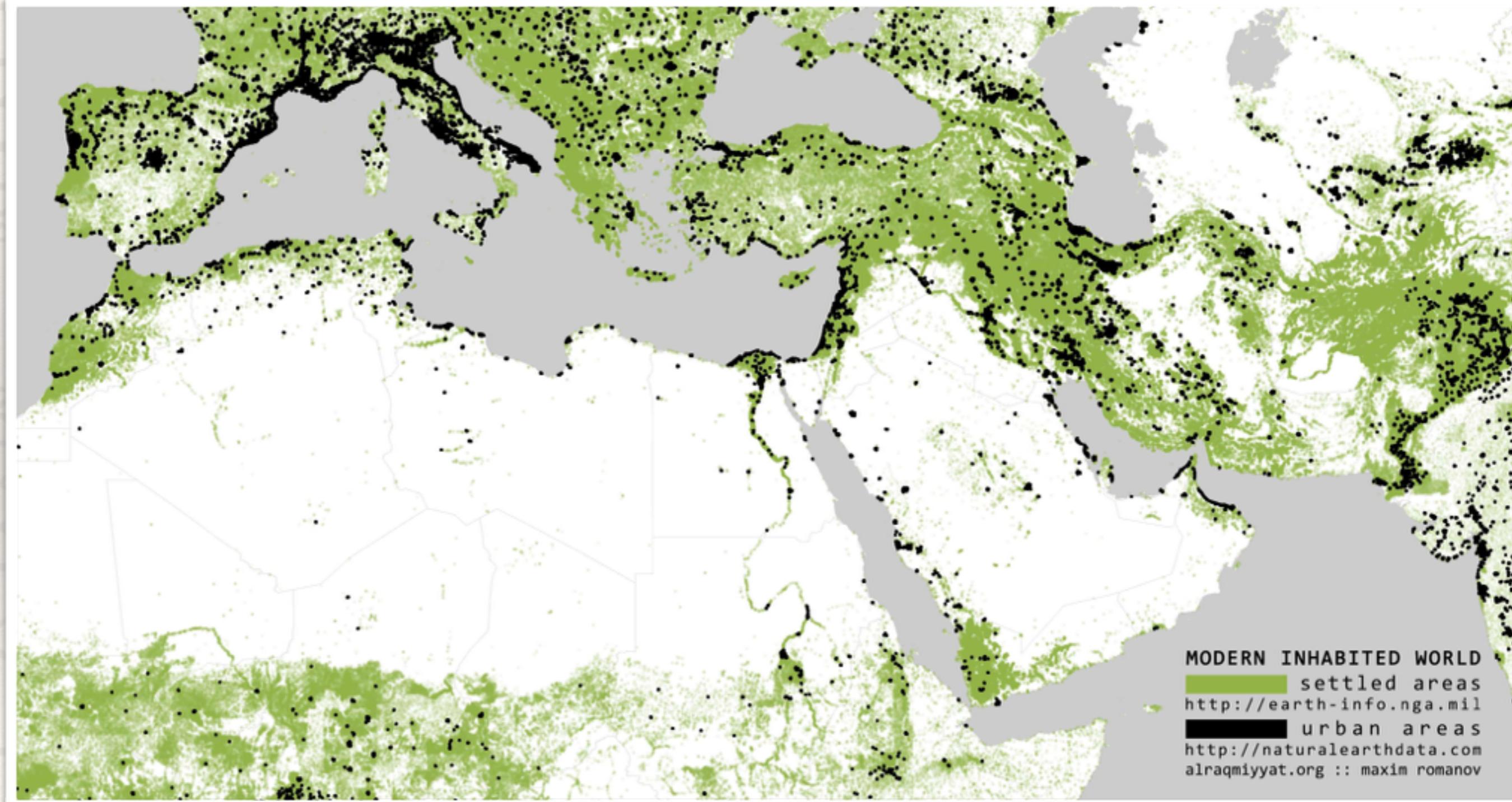


VISIBLE EARTH  
Earth's City Lights  
<http://visibleearth.nasa.gov/>

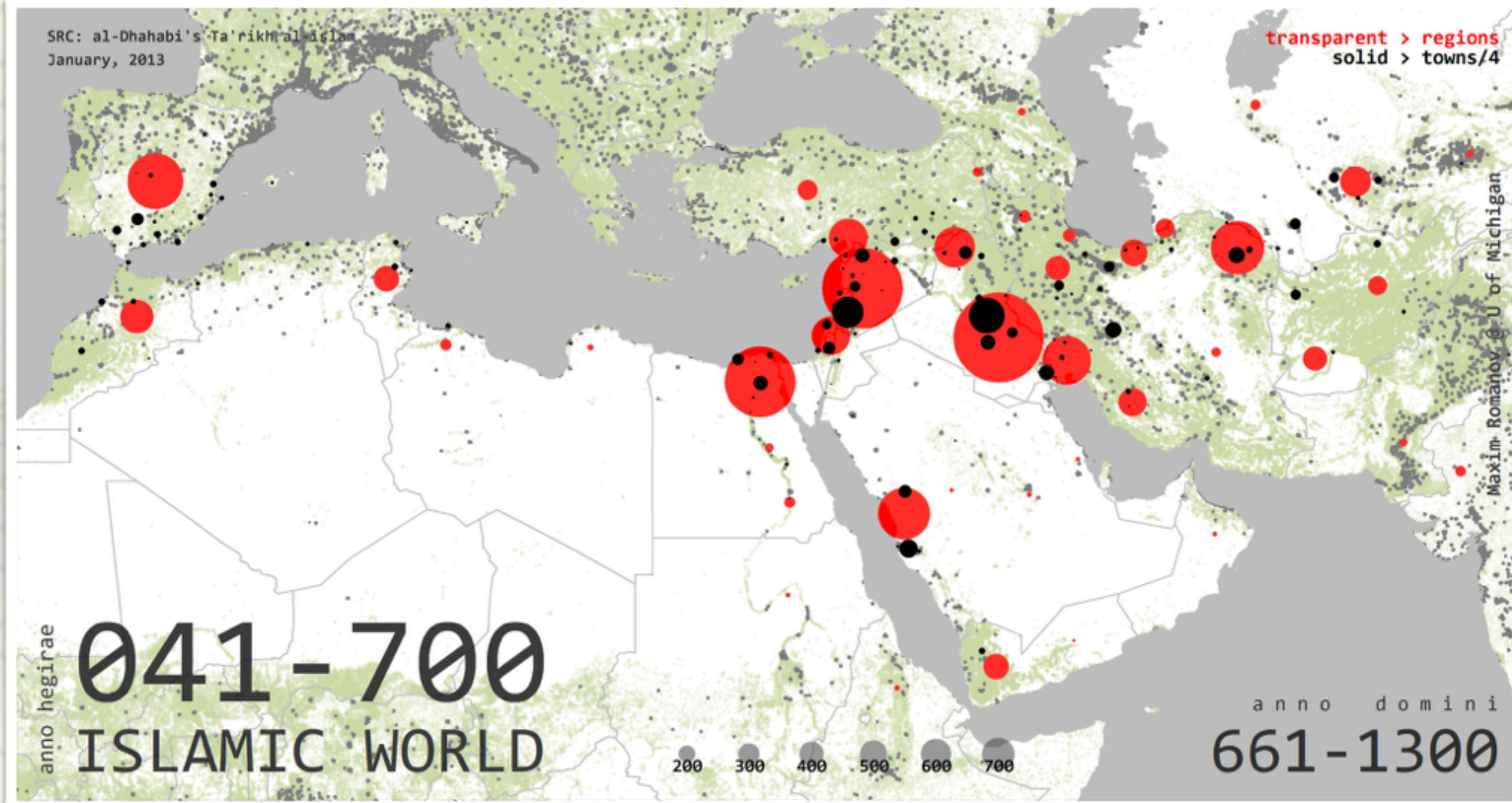
# Geographical Coverage



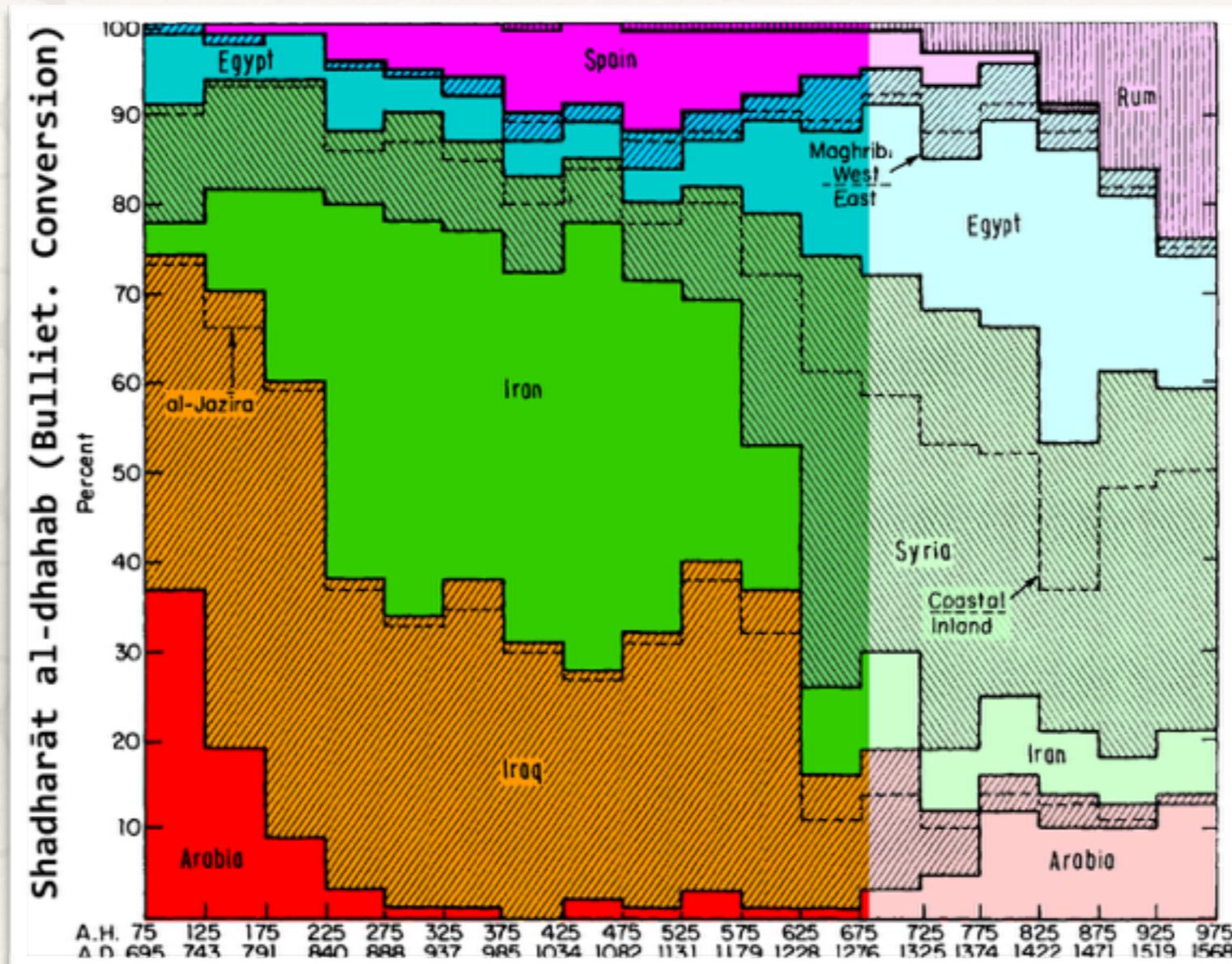
# Geographical Coverage



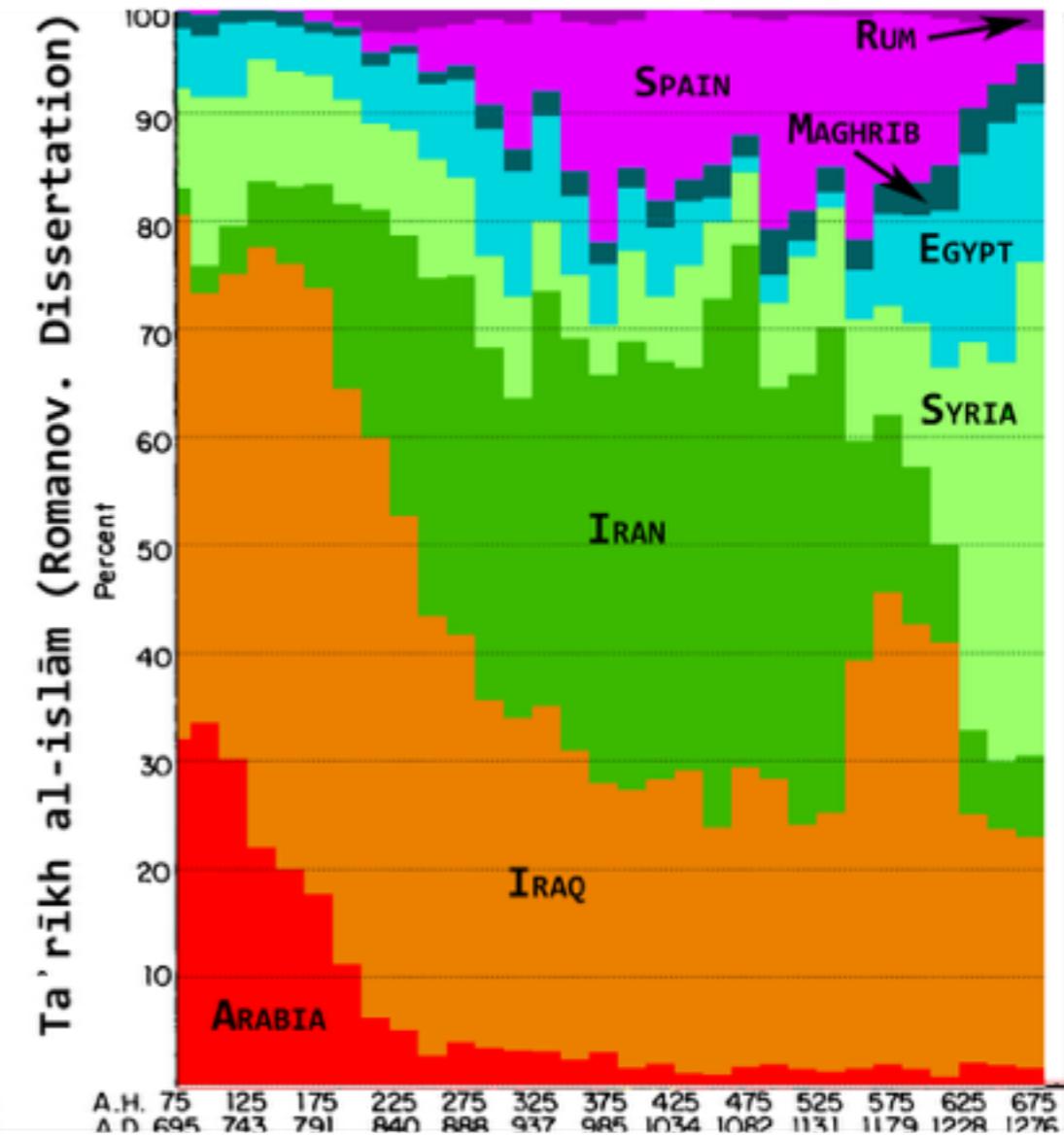
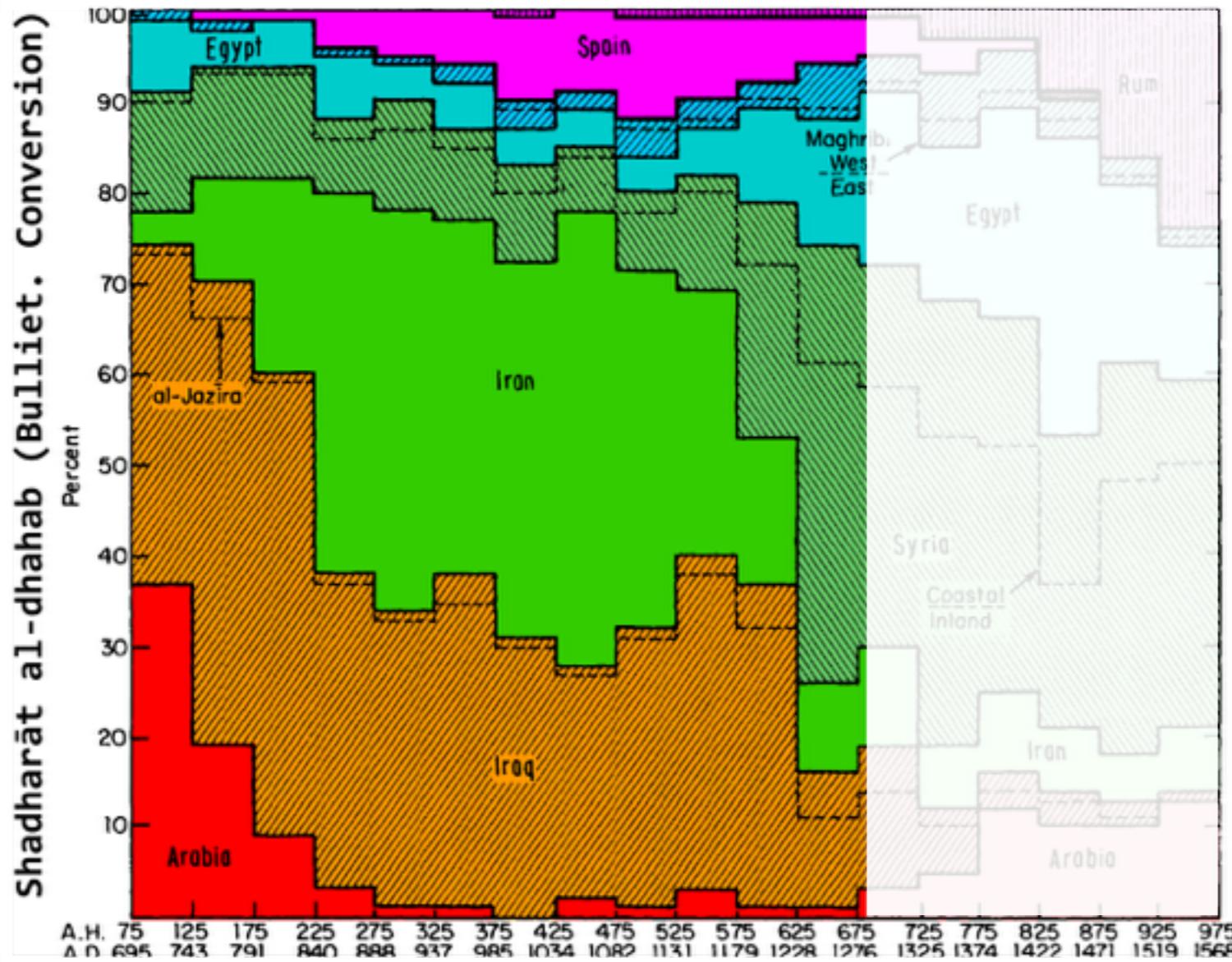
# Geographical Coverage



# Geo-Chronological Coverage: Comparative Perspective

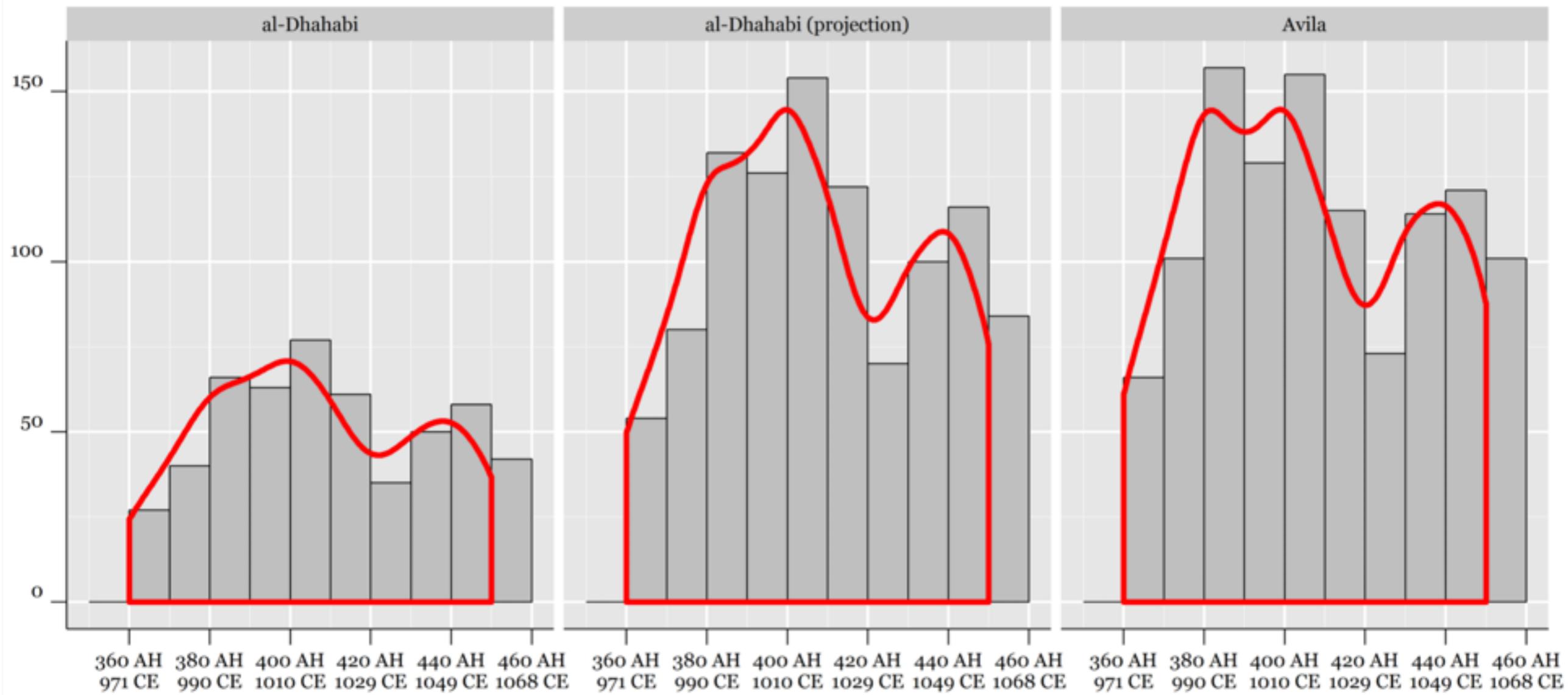


# Geo-Chronological Coverage: Comparative Perspective

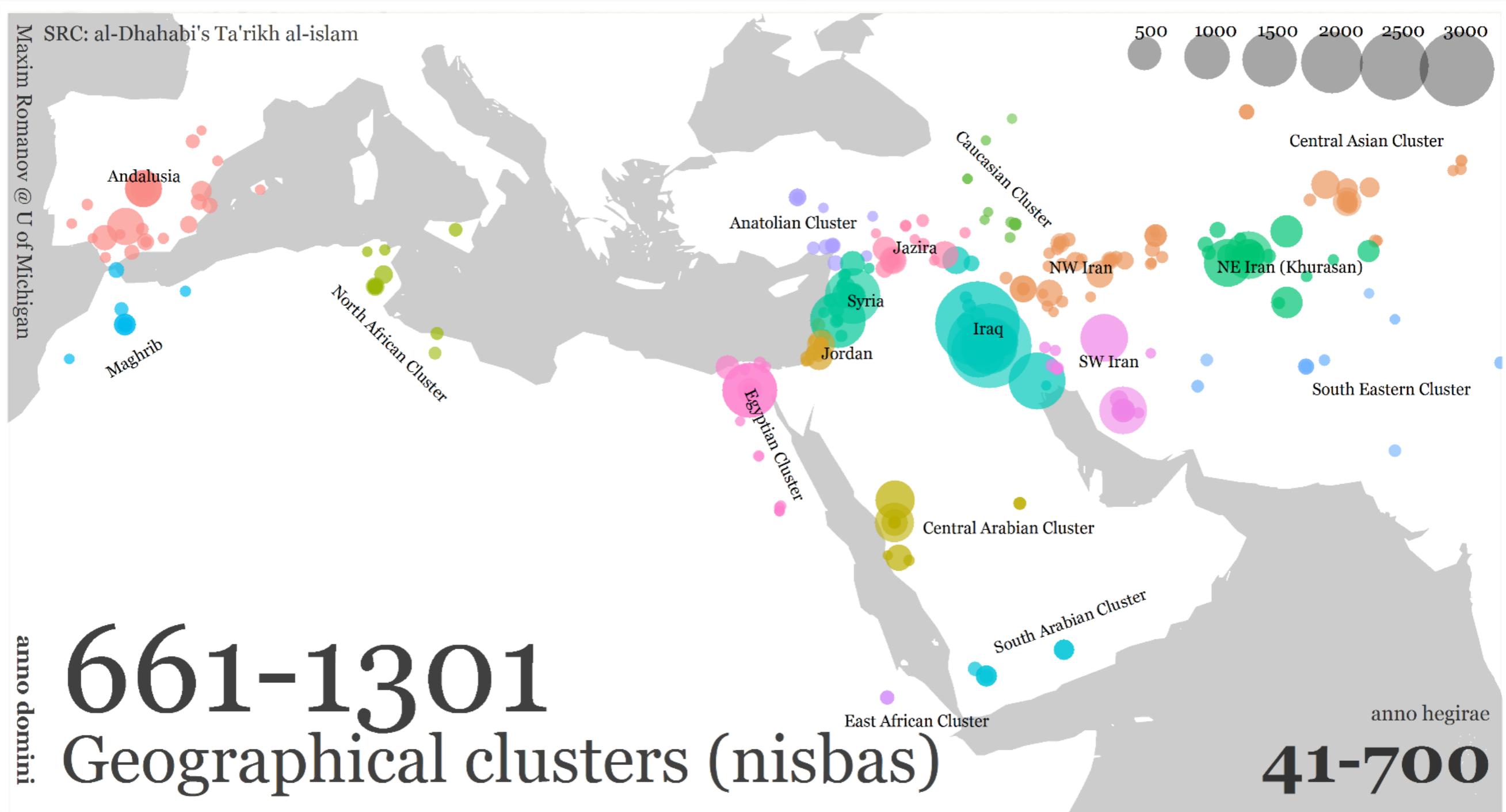


# Geo-Chronological Coverage: Comparative Perspective

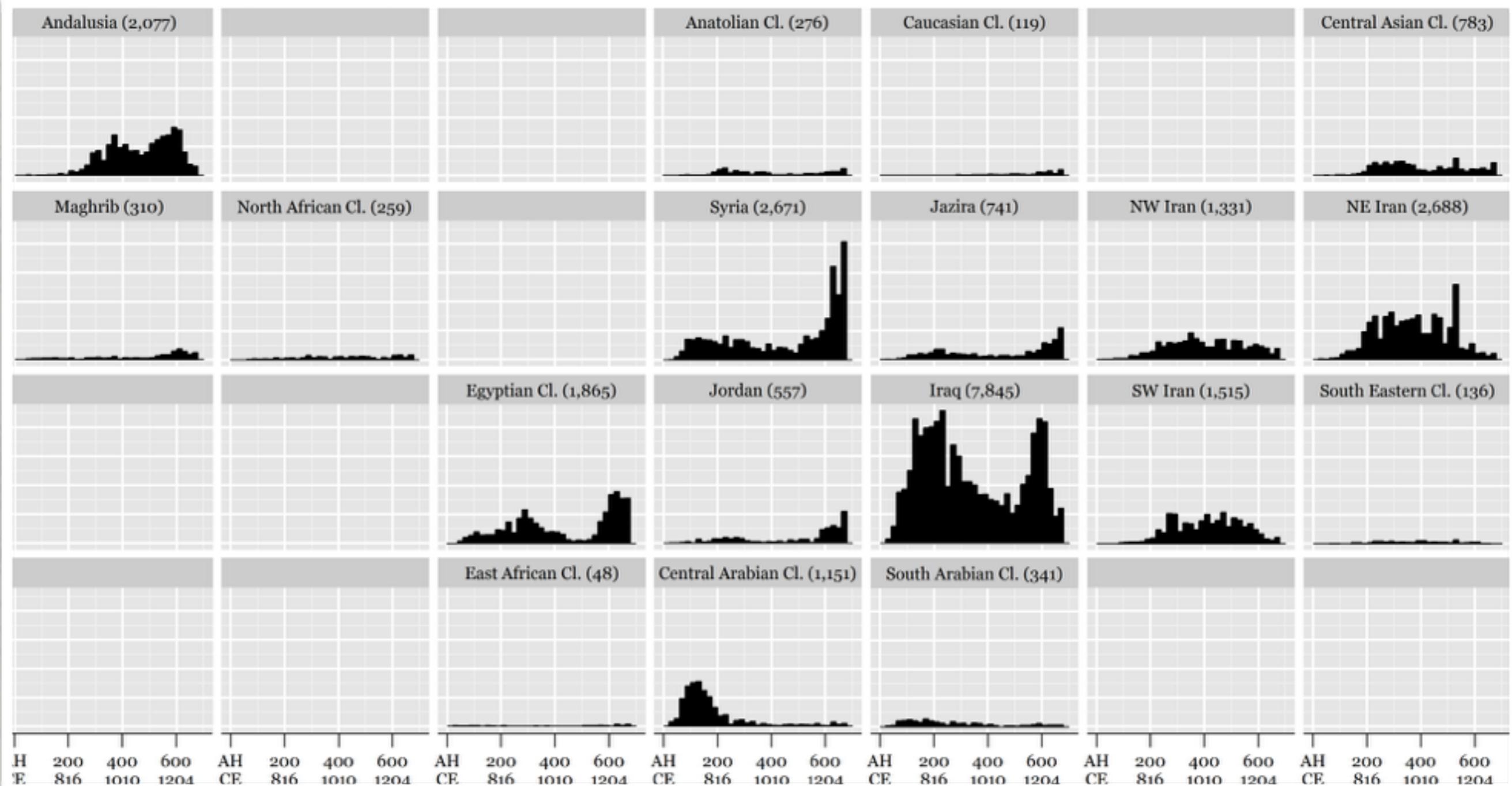
AH	370	380	390	400	410	420	430	440	450	460
in TI	42%	40%	42%	49%	49%	53%	49%	44%	48%	41%

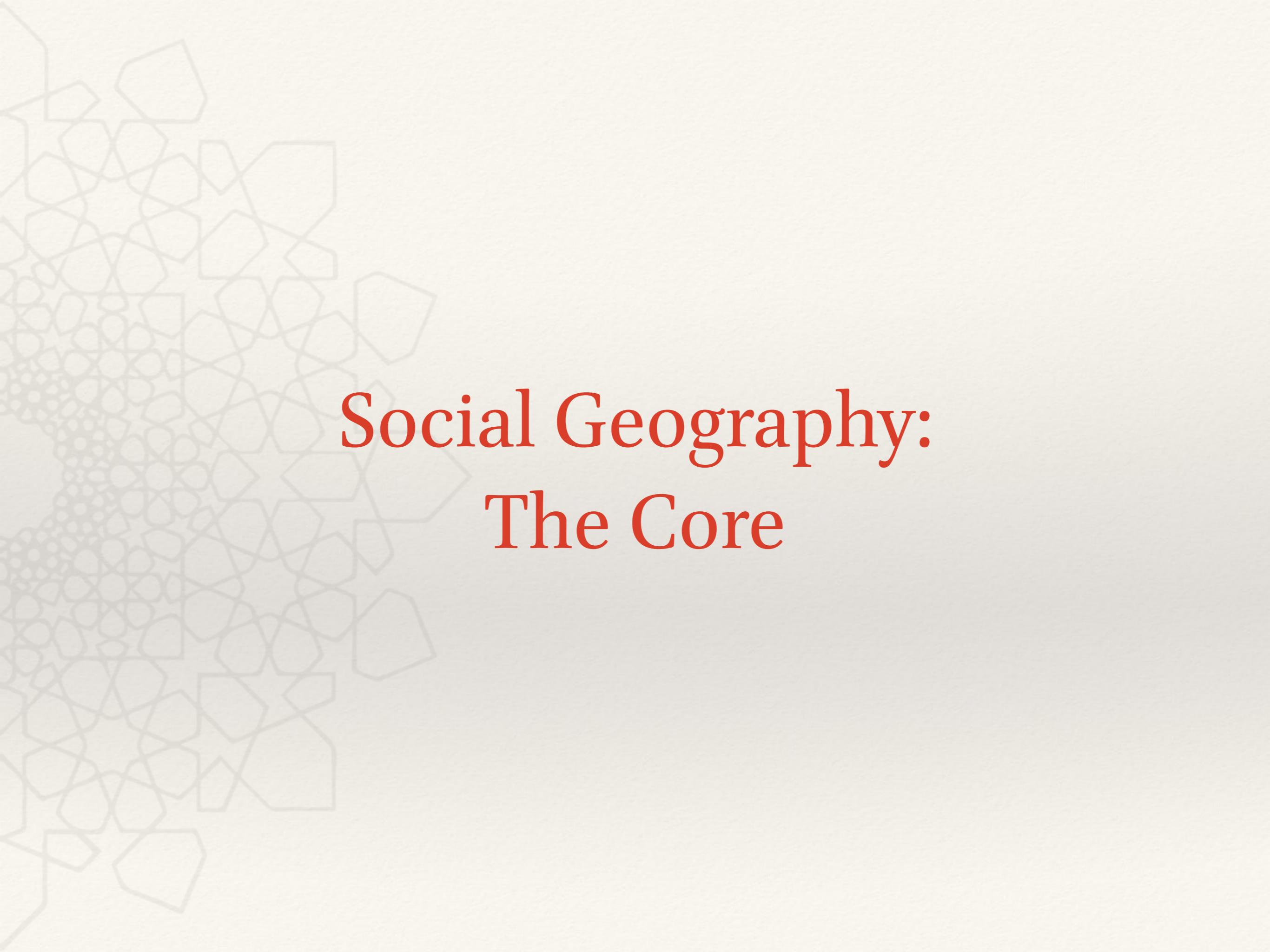


# Geographical Coverage



# Geo-Chronological Coverage





# Social Geography: The Core

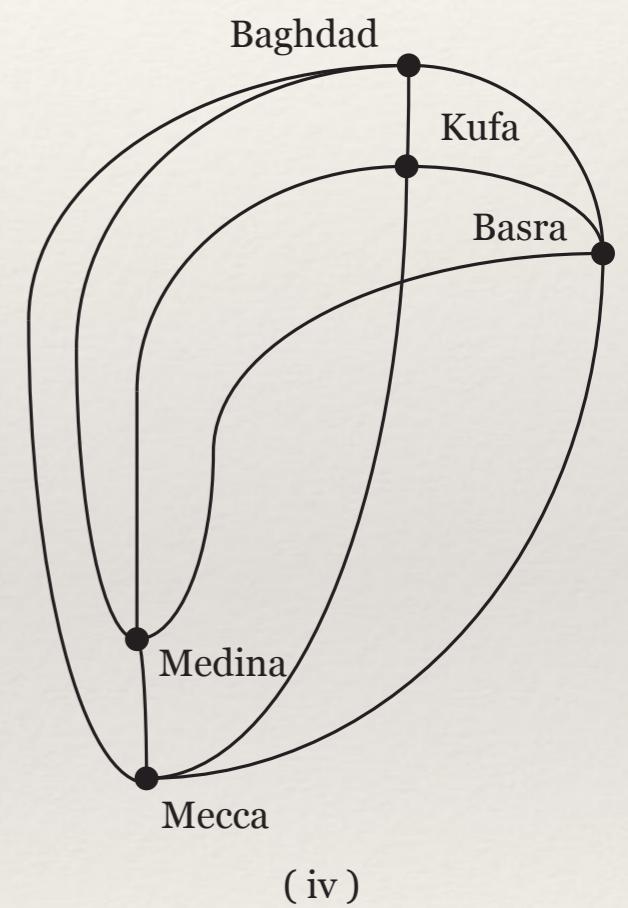
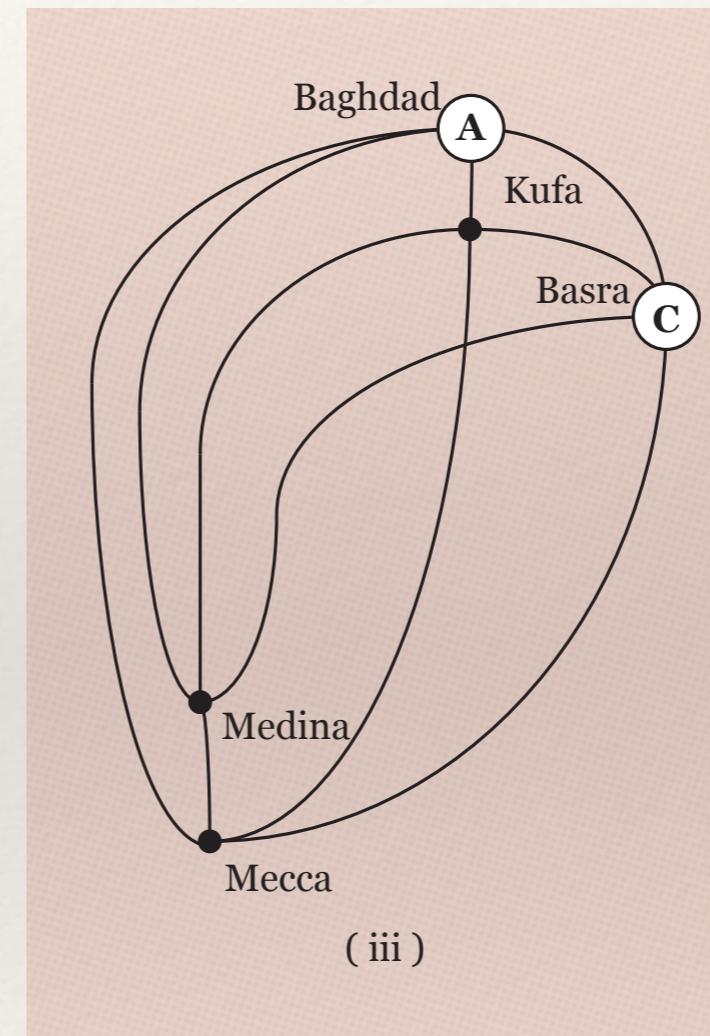
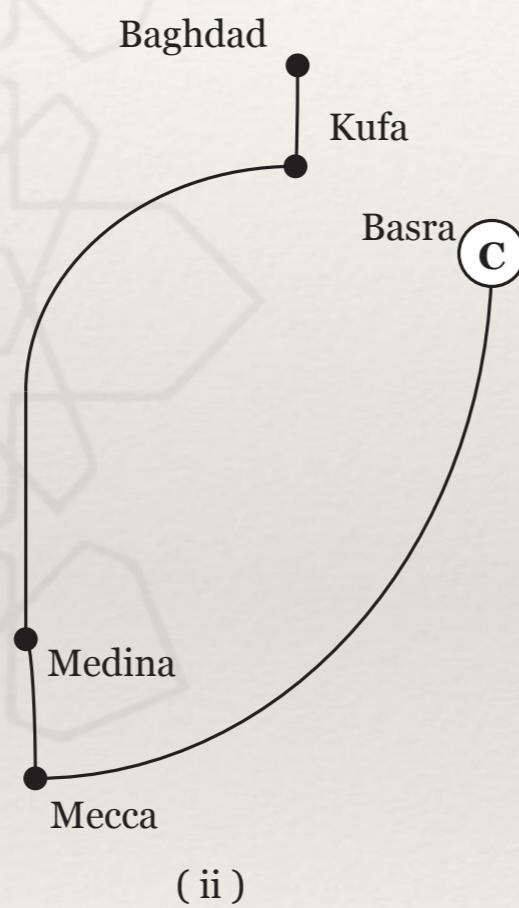
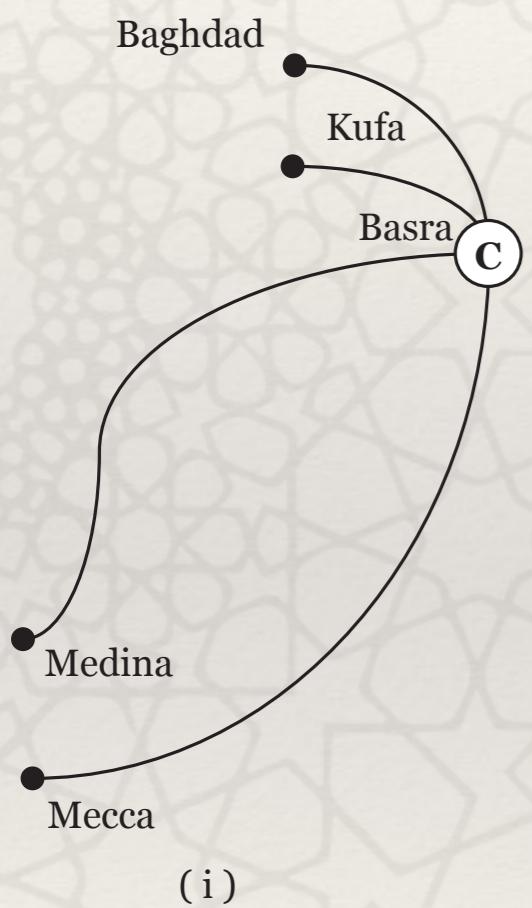
# Modeling the Core



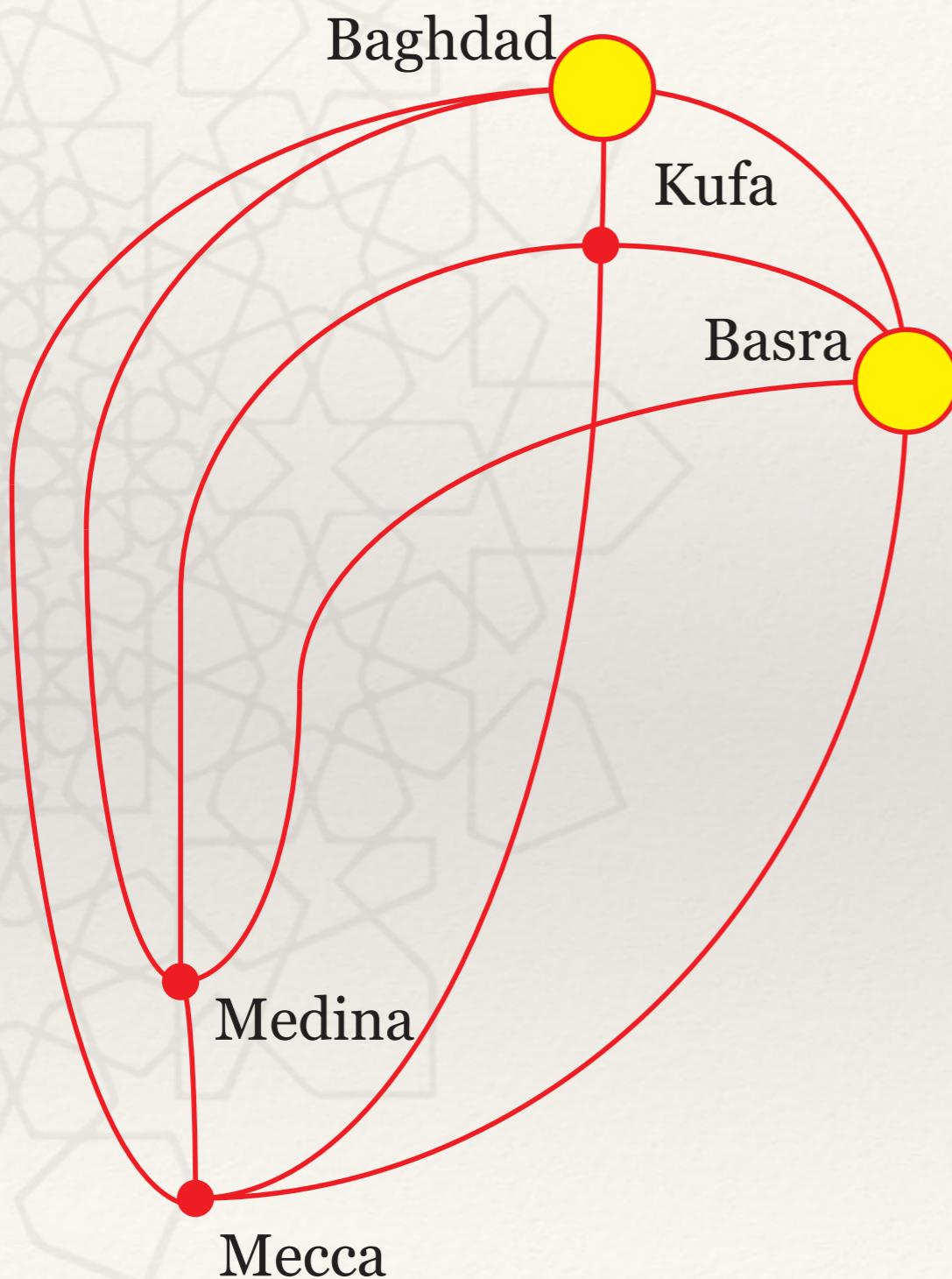
facebook

SOURCE

# Modeling Geographical Connections

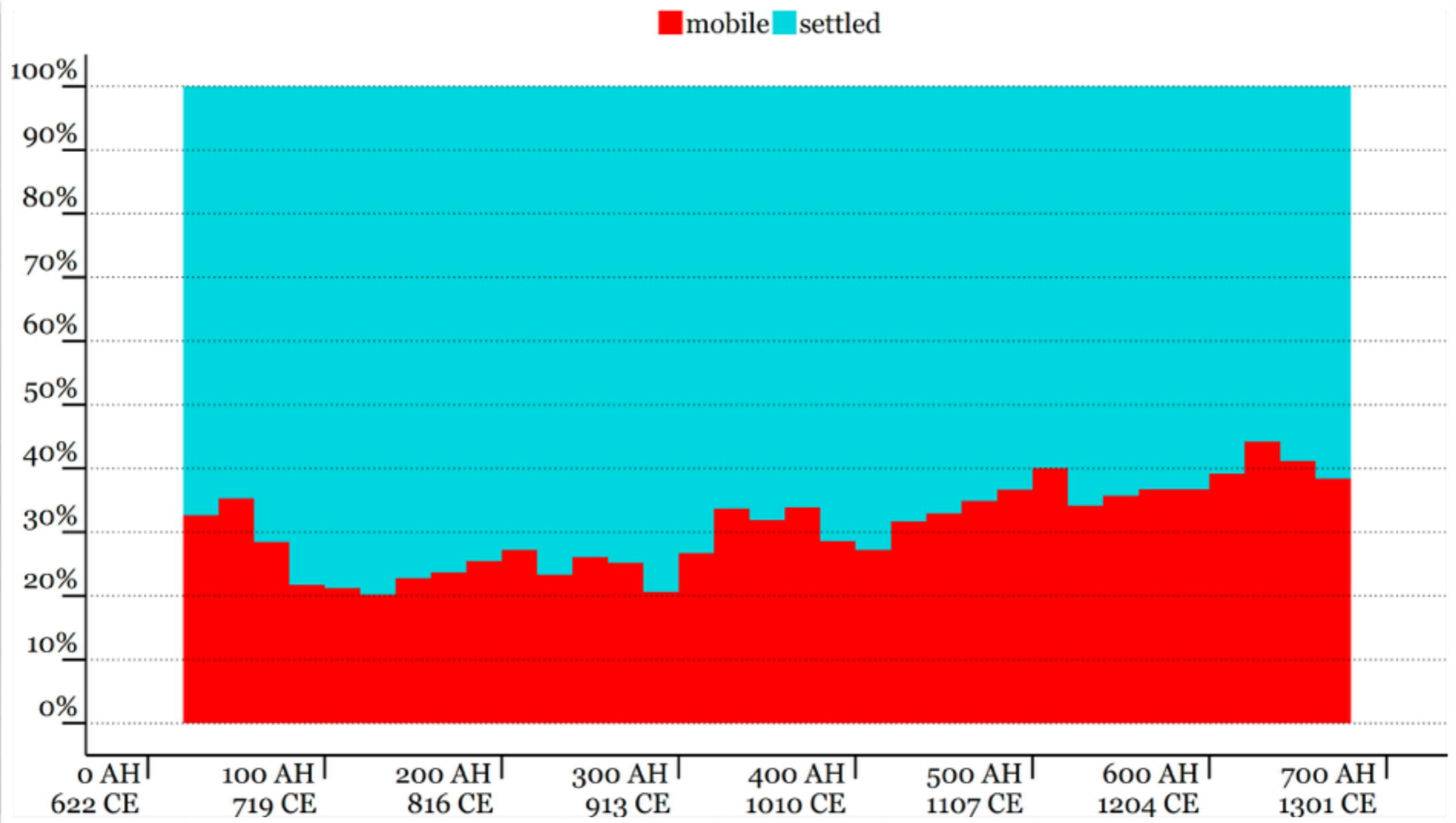


# Modeling Geographical Connections



- ❖ 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

# Transregional Mobility



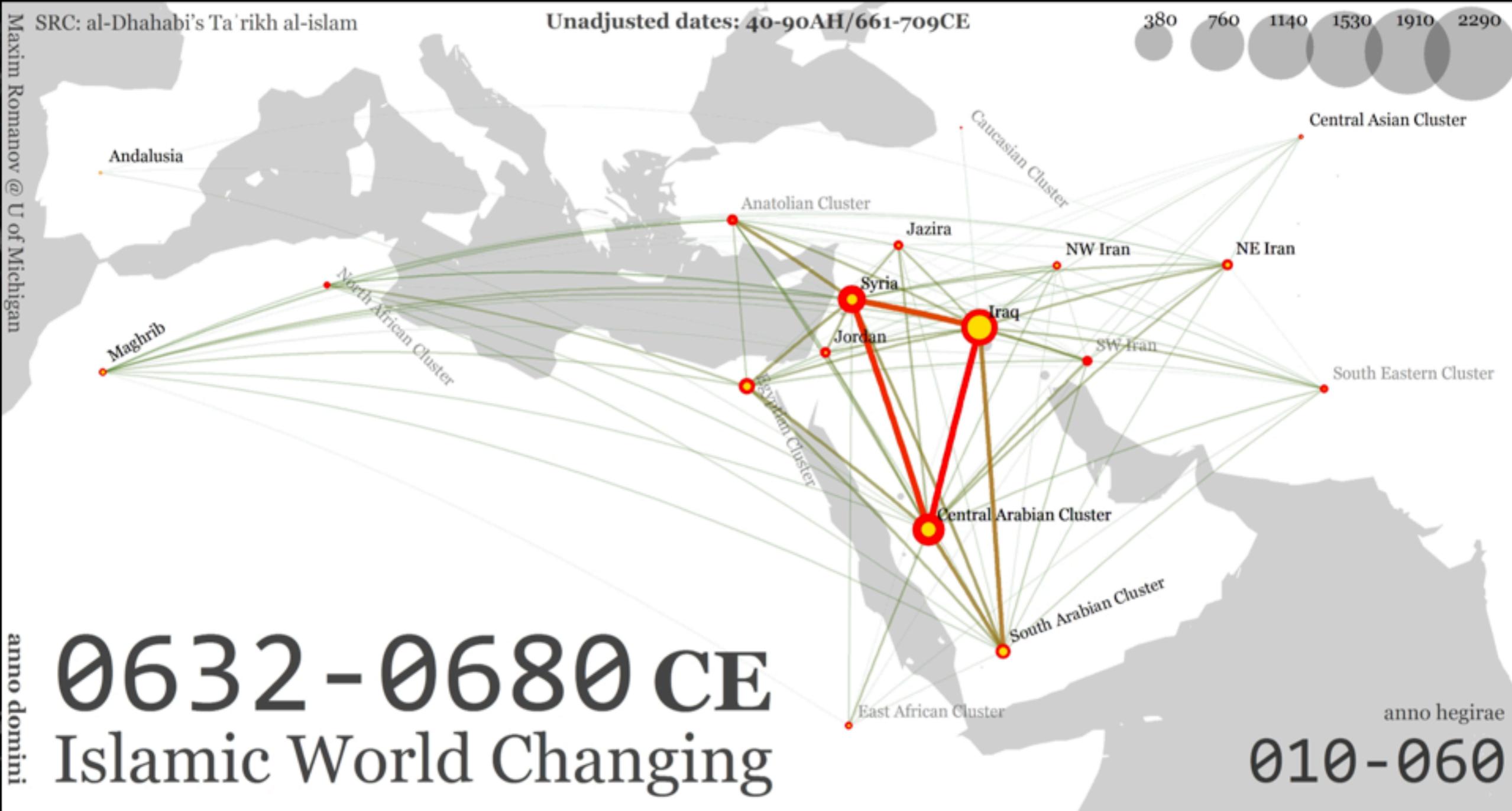
Note: 33% (9,700) with transregional connections, fluctuating within the range of 22-46%

# 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



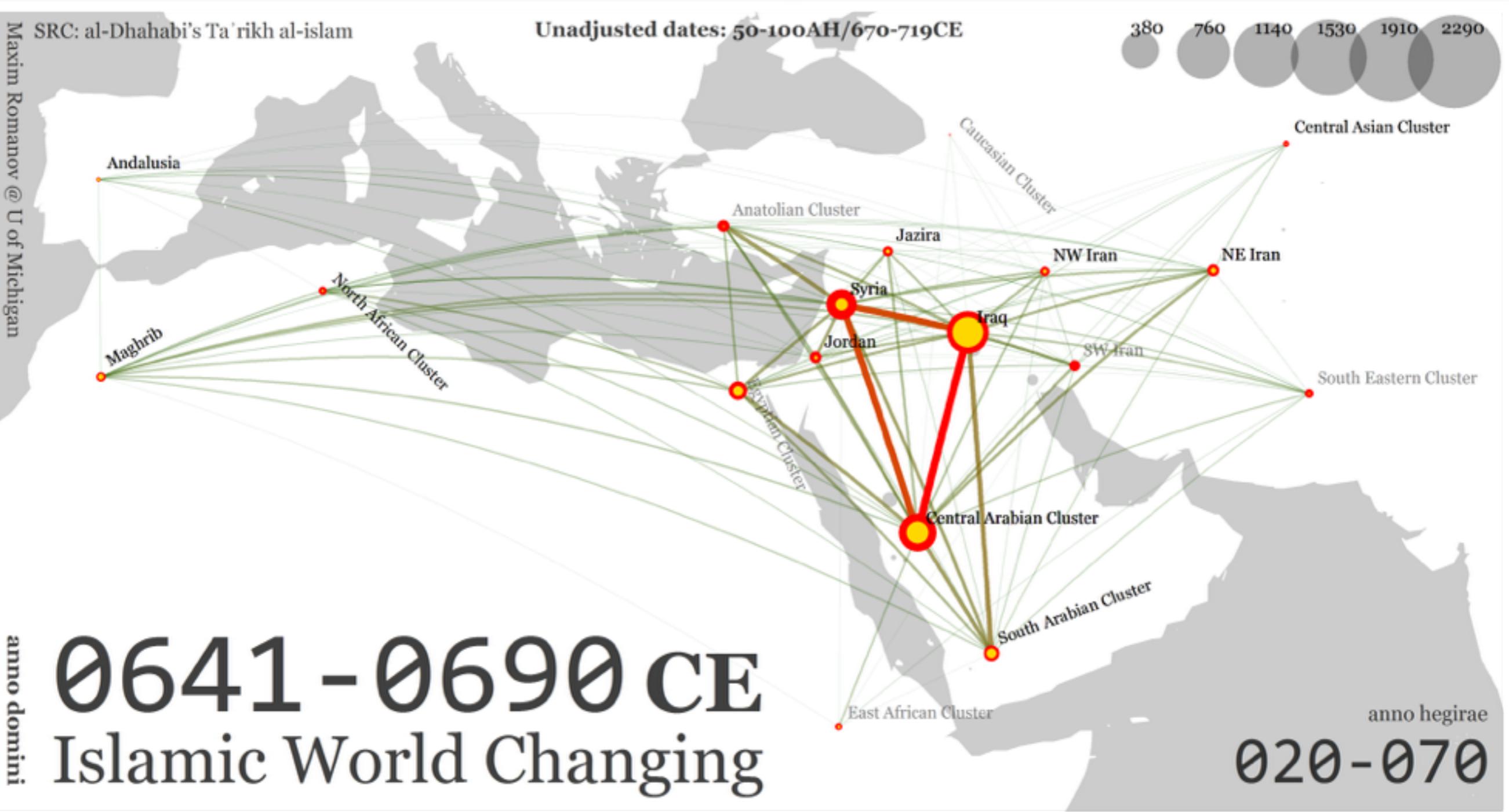
anno hegirae

010-060

# 0641 - 0690 CE Islamic World Changing

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

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

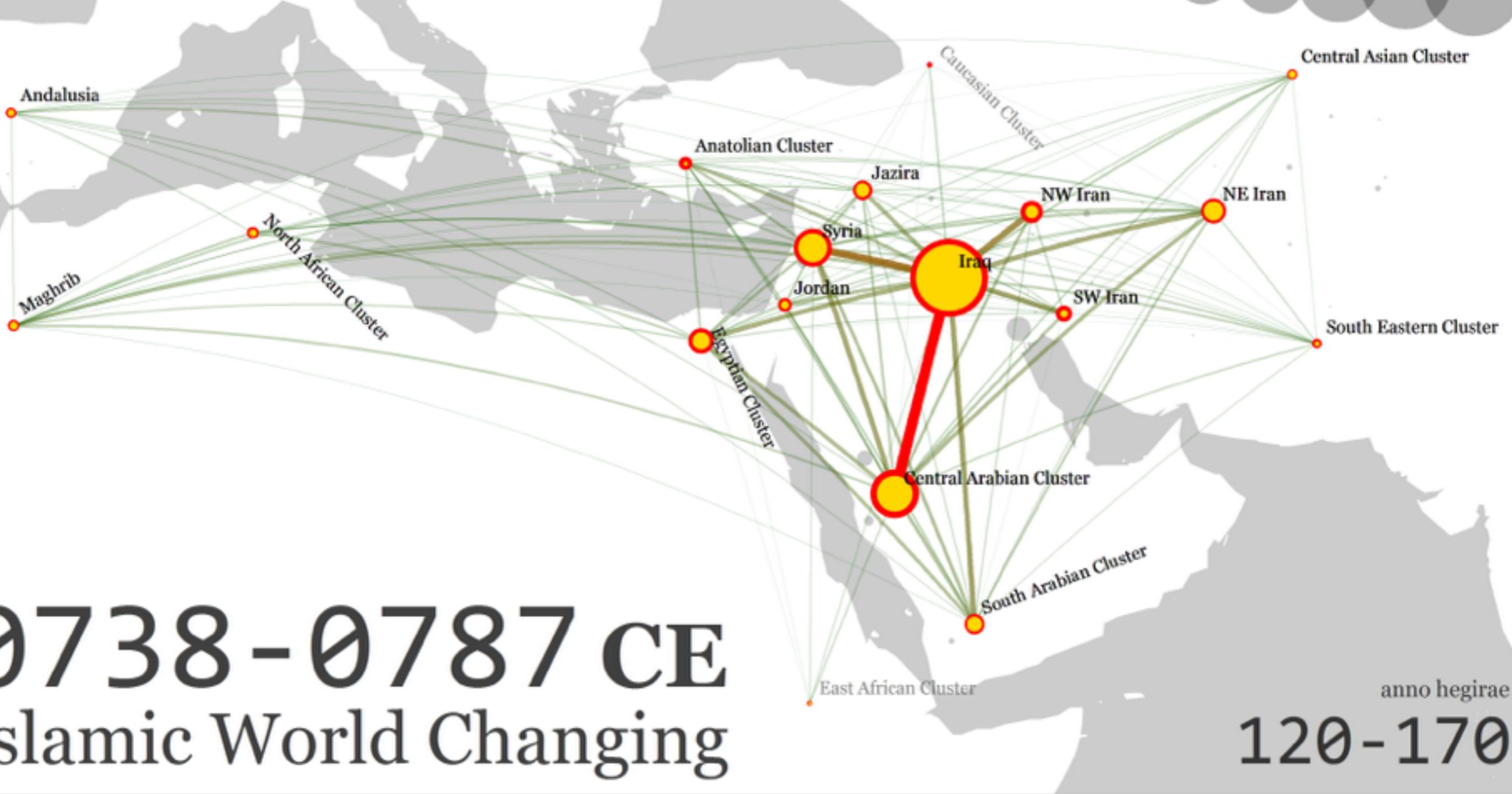


0738-0787 CE  
Islamic World Changing

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

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

380 760 1140 1530 1910 2290



anno hegirae

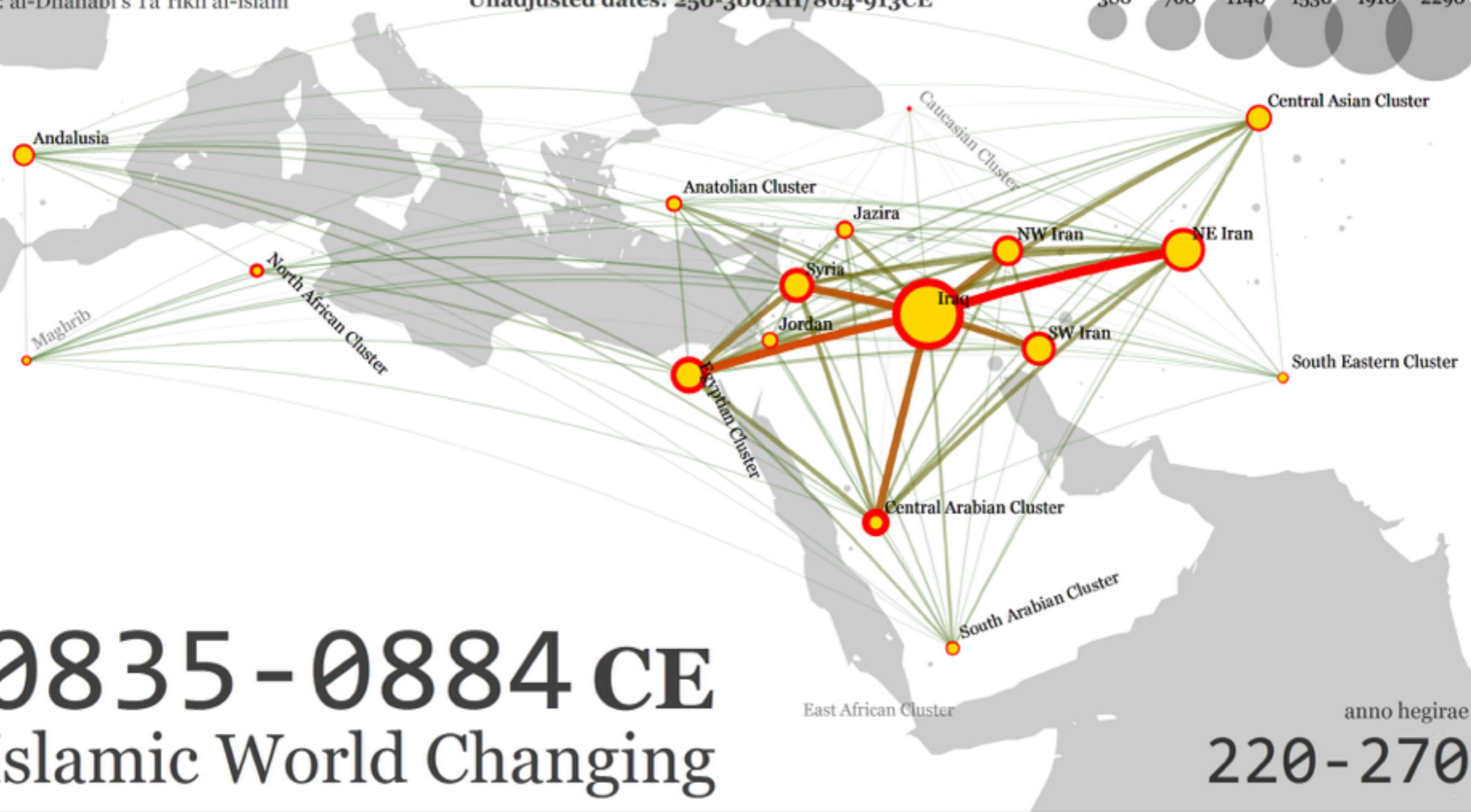
120-170

# 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

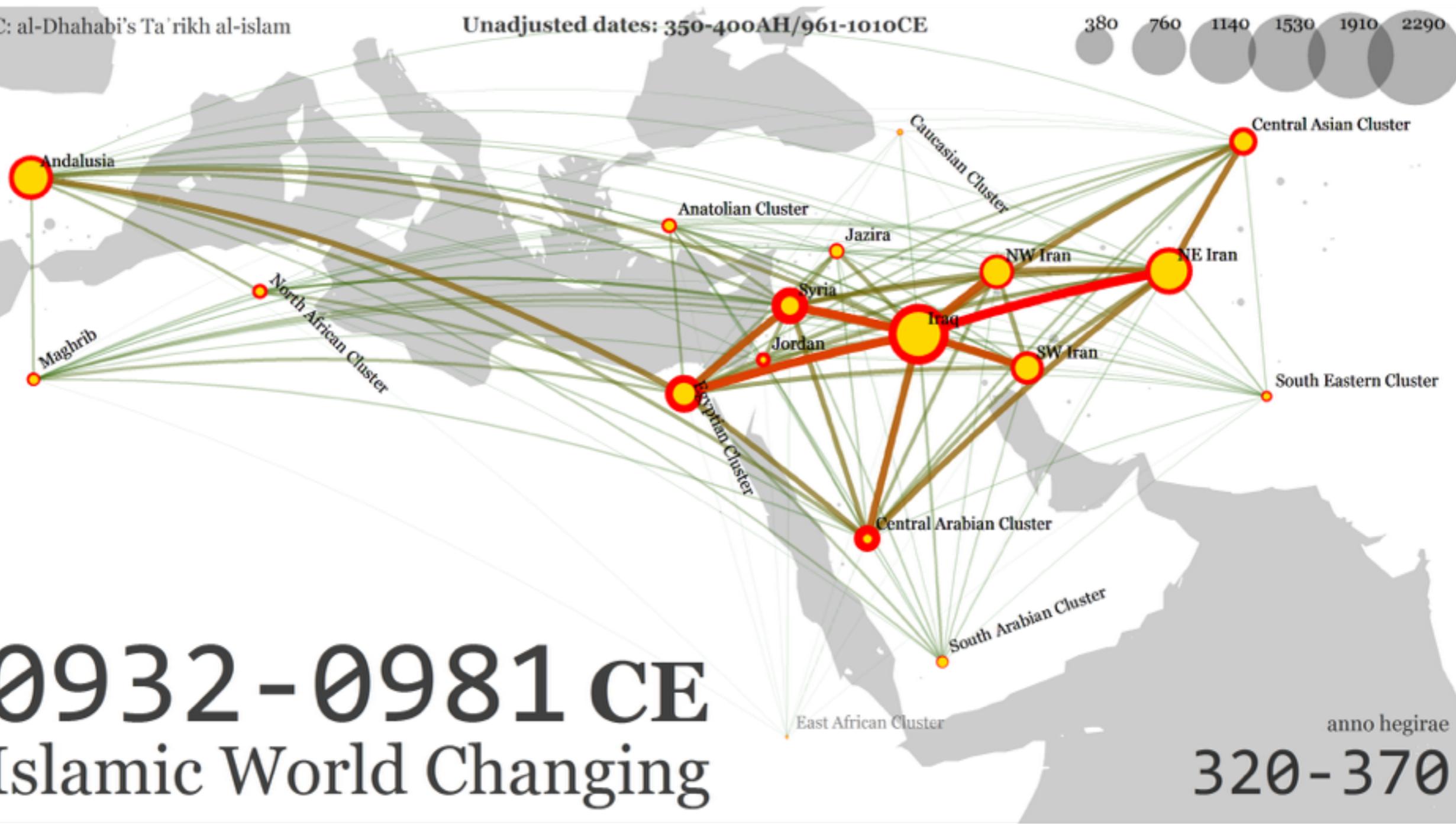
220-270

# 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

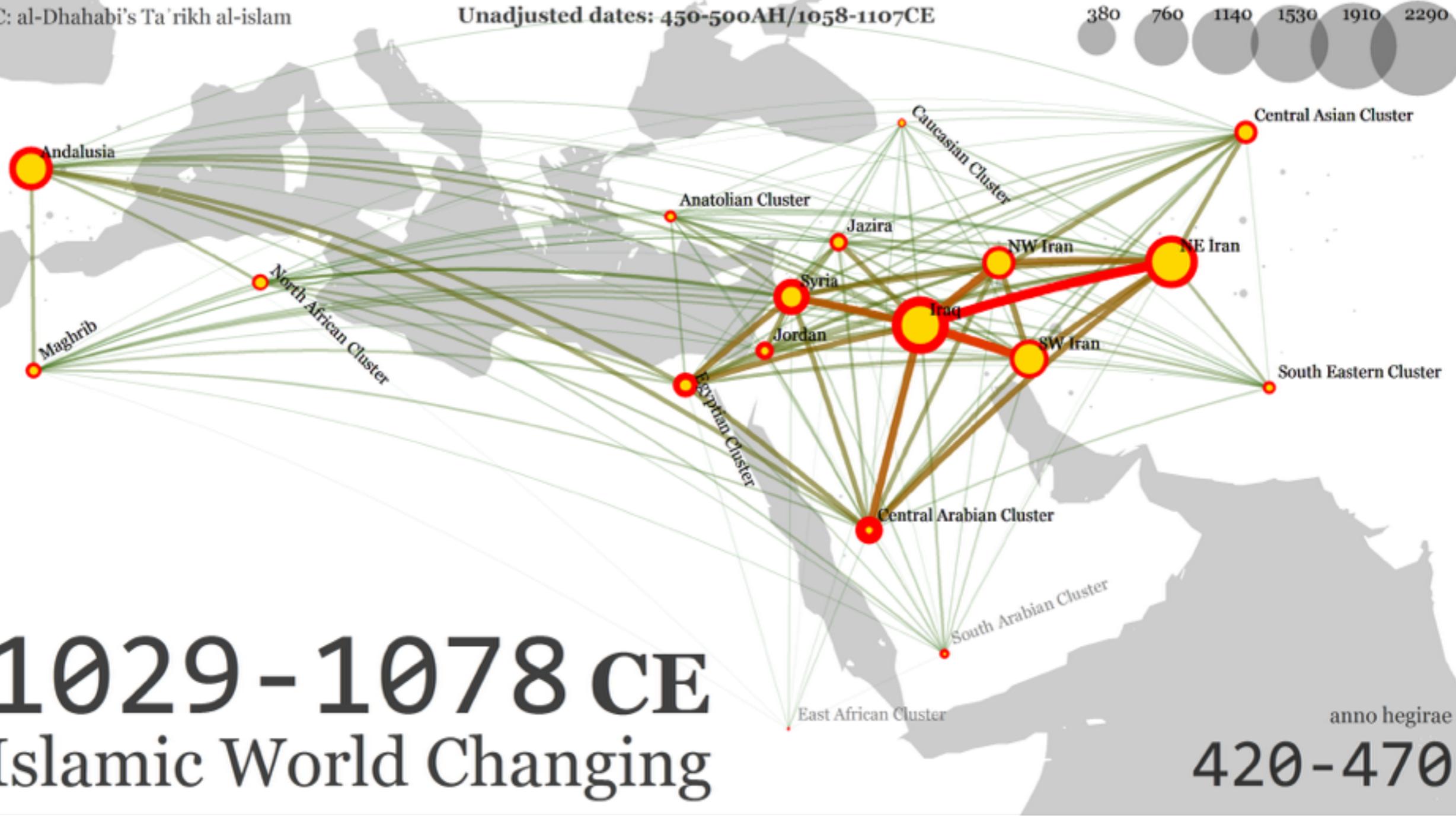
320-370

# 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

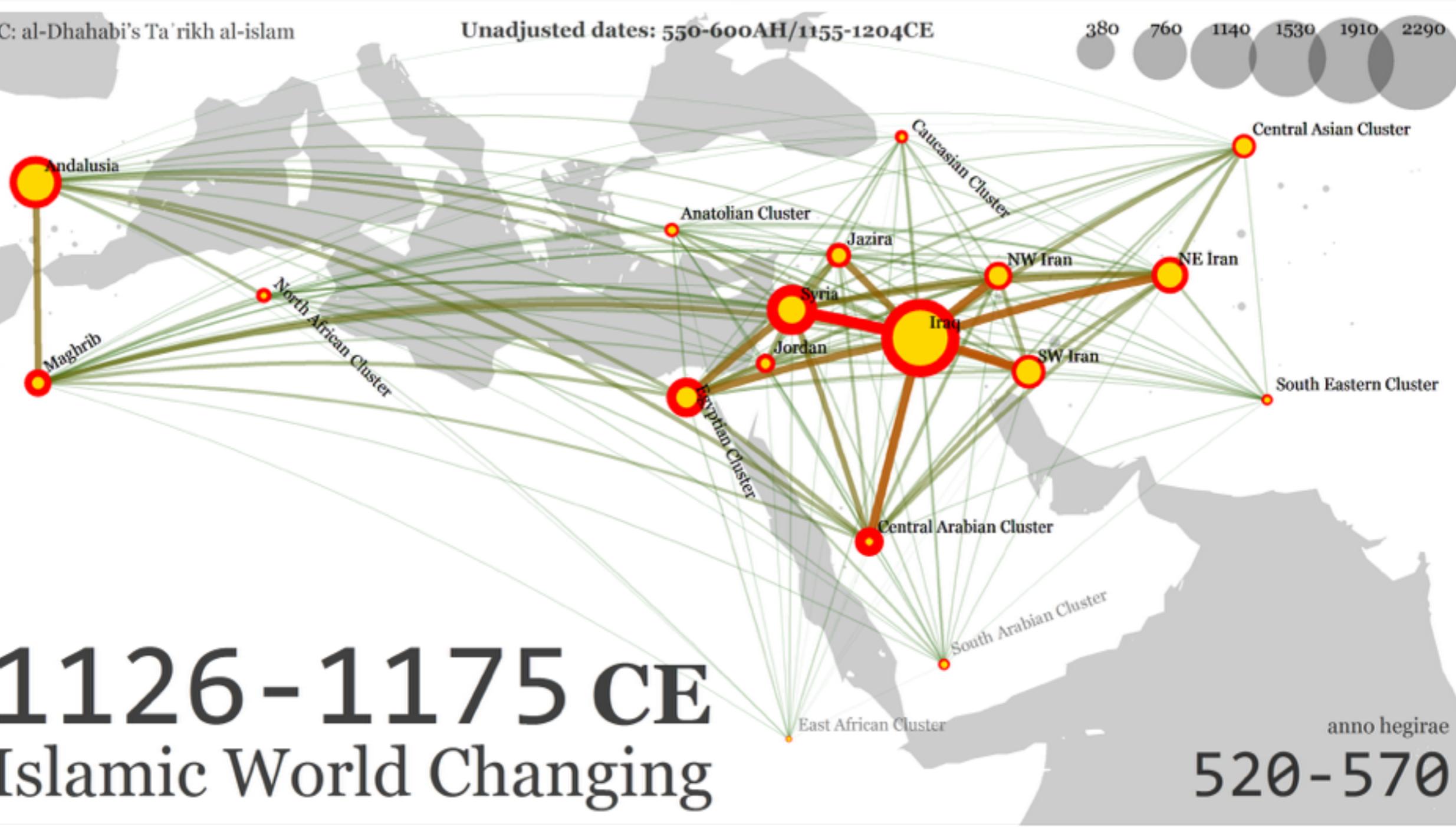
420-470

# 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

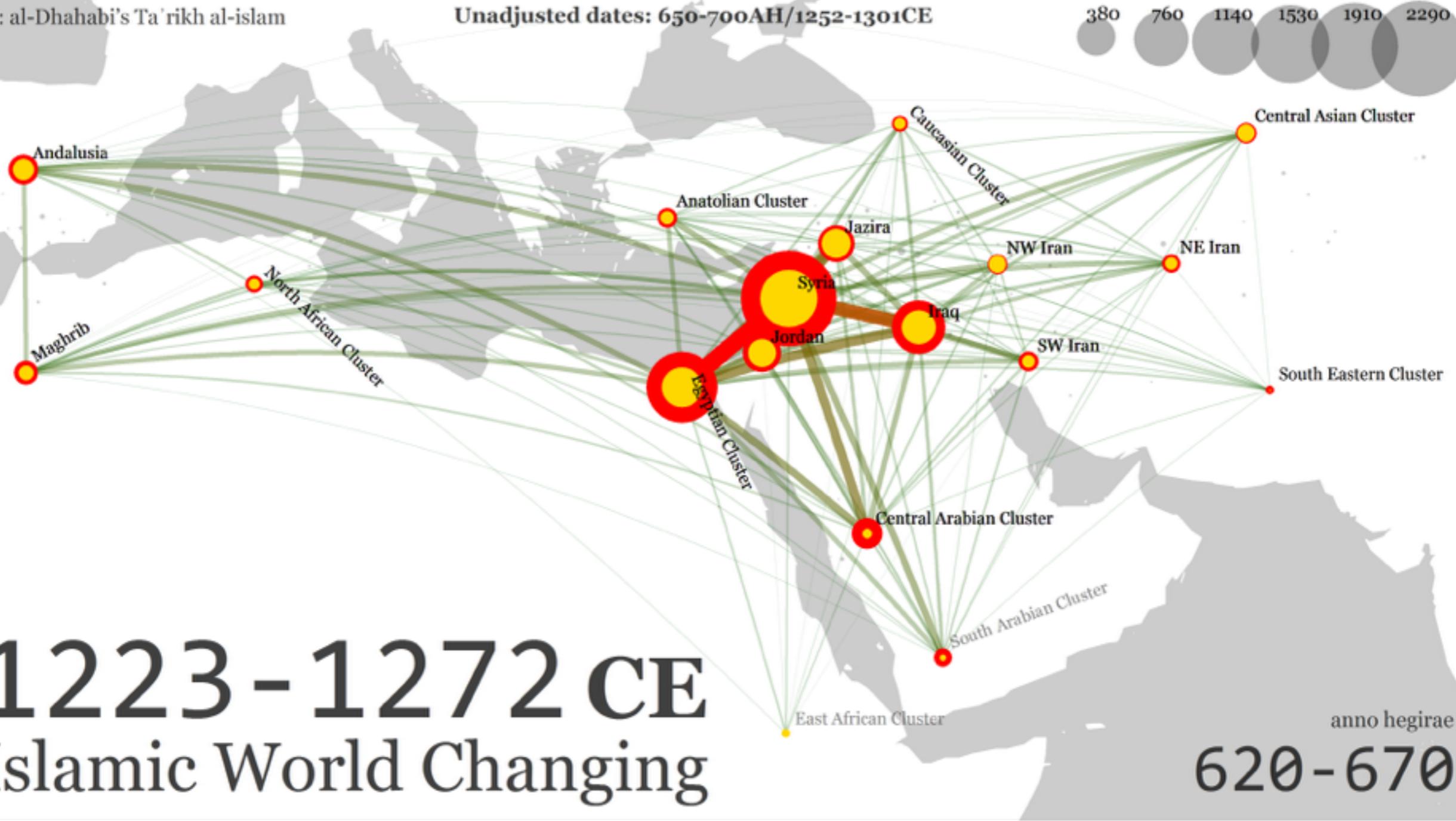
520-570

# 1223-1272 CE Islamic World Changing

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

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

380 760 1140 1530 1910 2290



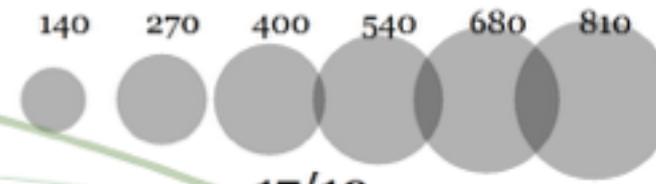
anno hegirae

620-670

# Regional Integration

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

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



802/

190/22

138/32

31/7

259/23

27/3

154/33

135/6

28/4

44/9

137/22

38/19

62/44

33/4

8/1

1/

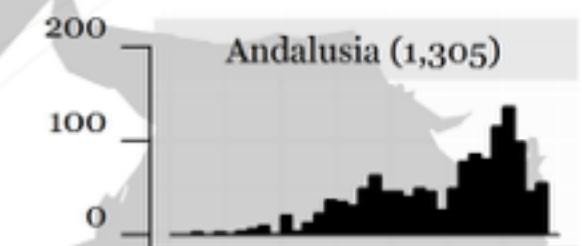
**2,048** Visitors**241** Migrants**2,360** Natives/Residents:

802 - with local connections

1,241 - with transregional connections

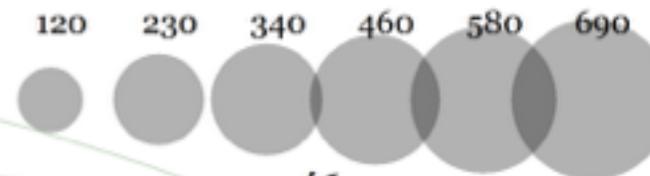
**1,305** - total number of transregional connections**0641-1272 CE**

Connections: Andalusia

**020-670**

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

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



14/4

8/3

47/5

20/2

38/7

39/10

143/15

249/81

54/10

181/42

452/

684/192

138/42

49/13

47/2

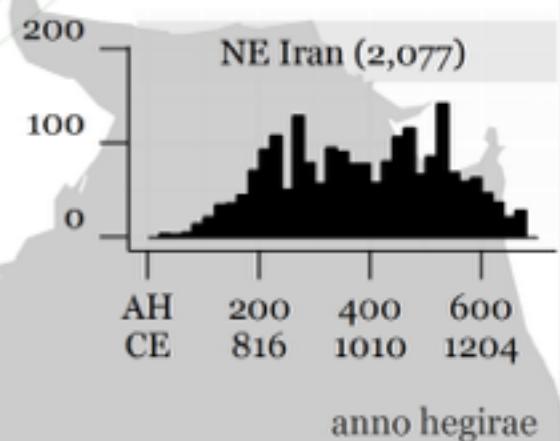
214/30

2/1

- 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

**0641-1272 CE**

Connections: NE Iran

**020-670**





# Urban centers

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

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

130

260

390

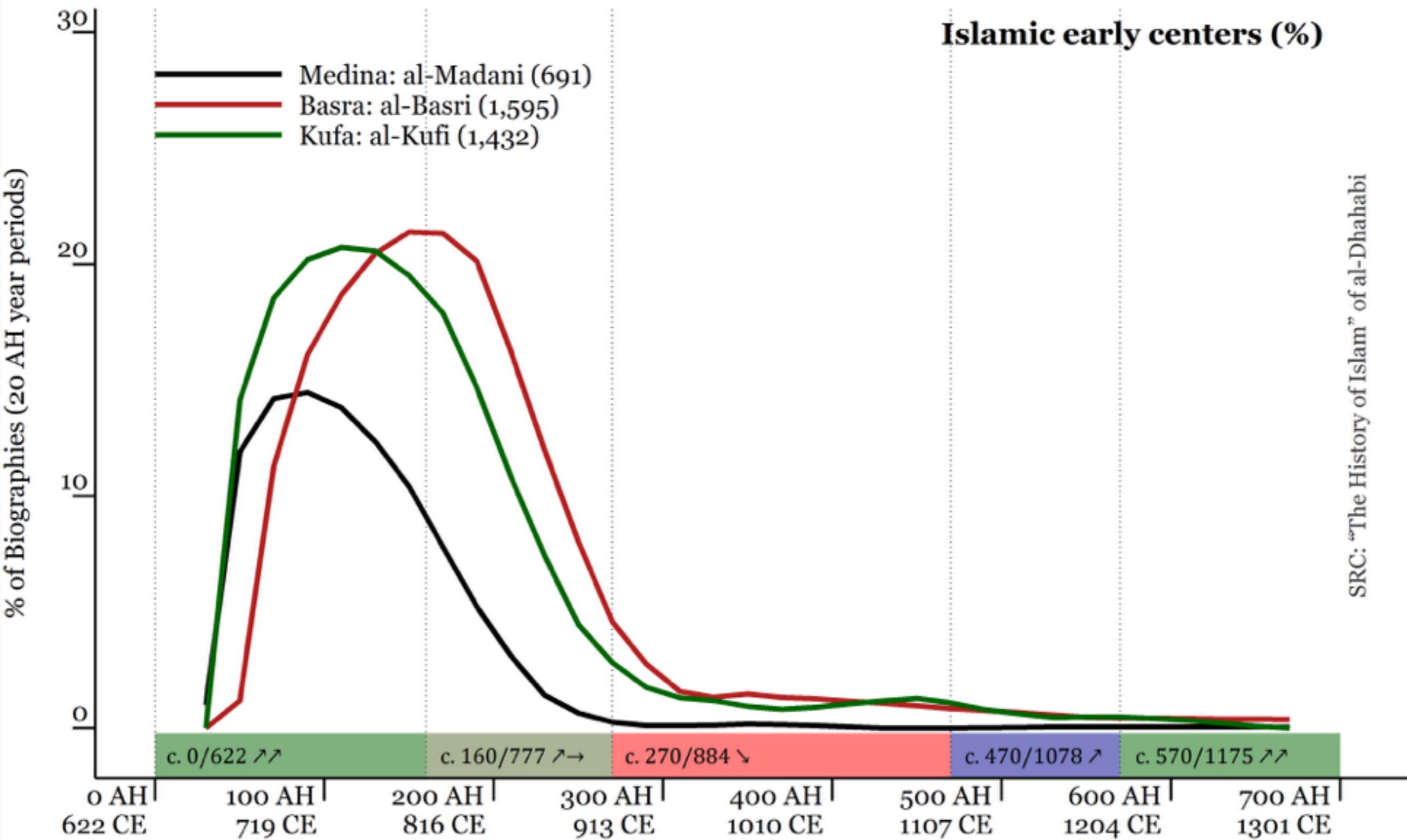
520

650

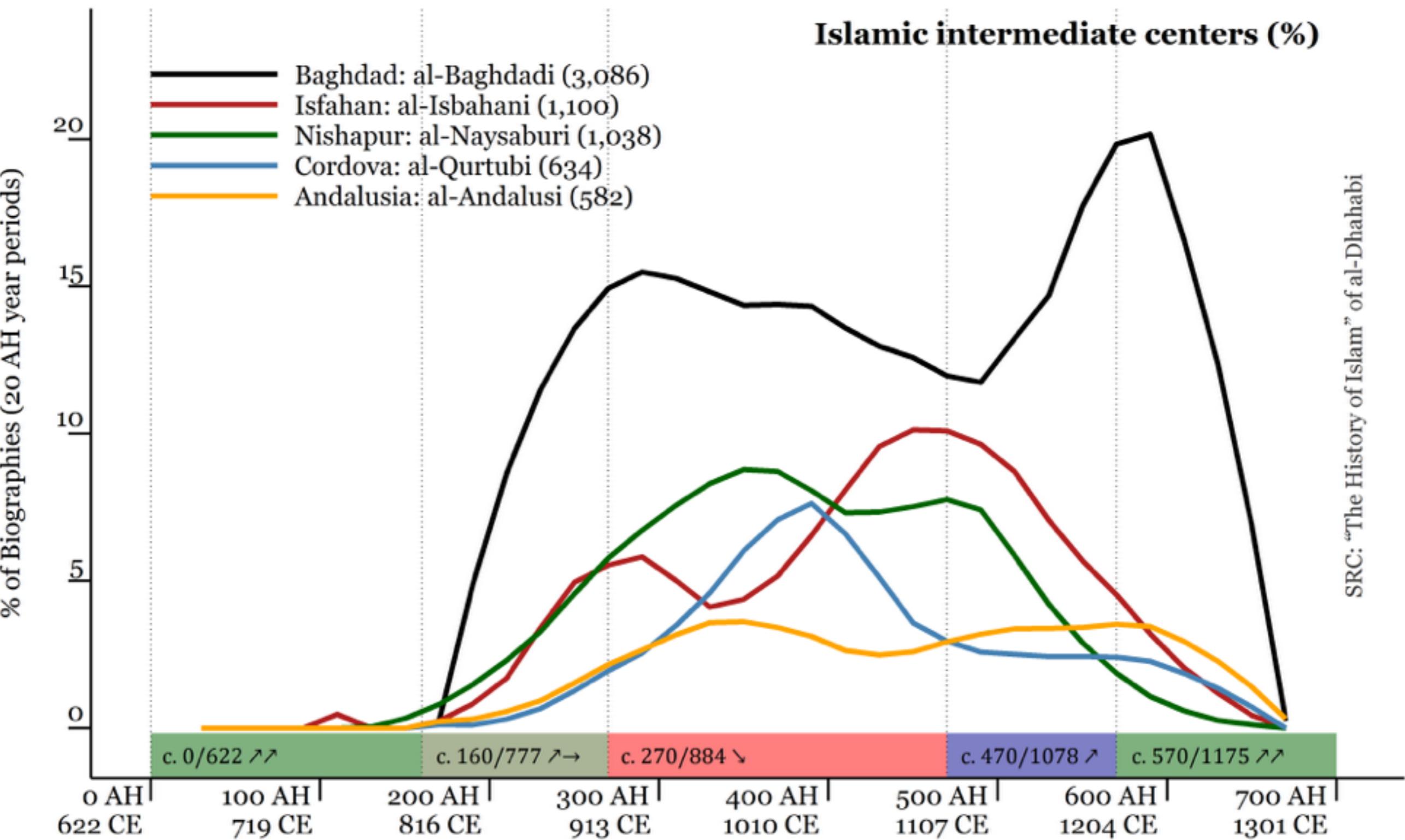
780



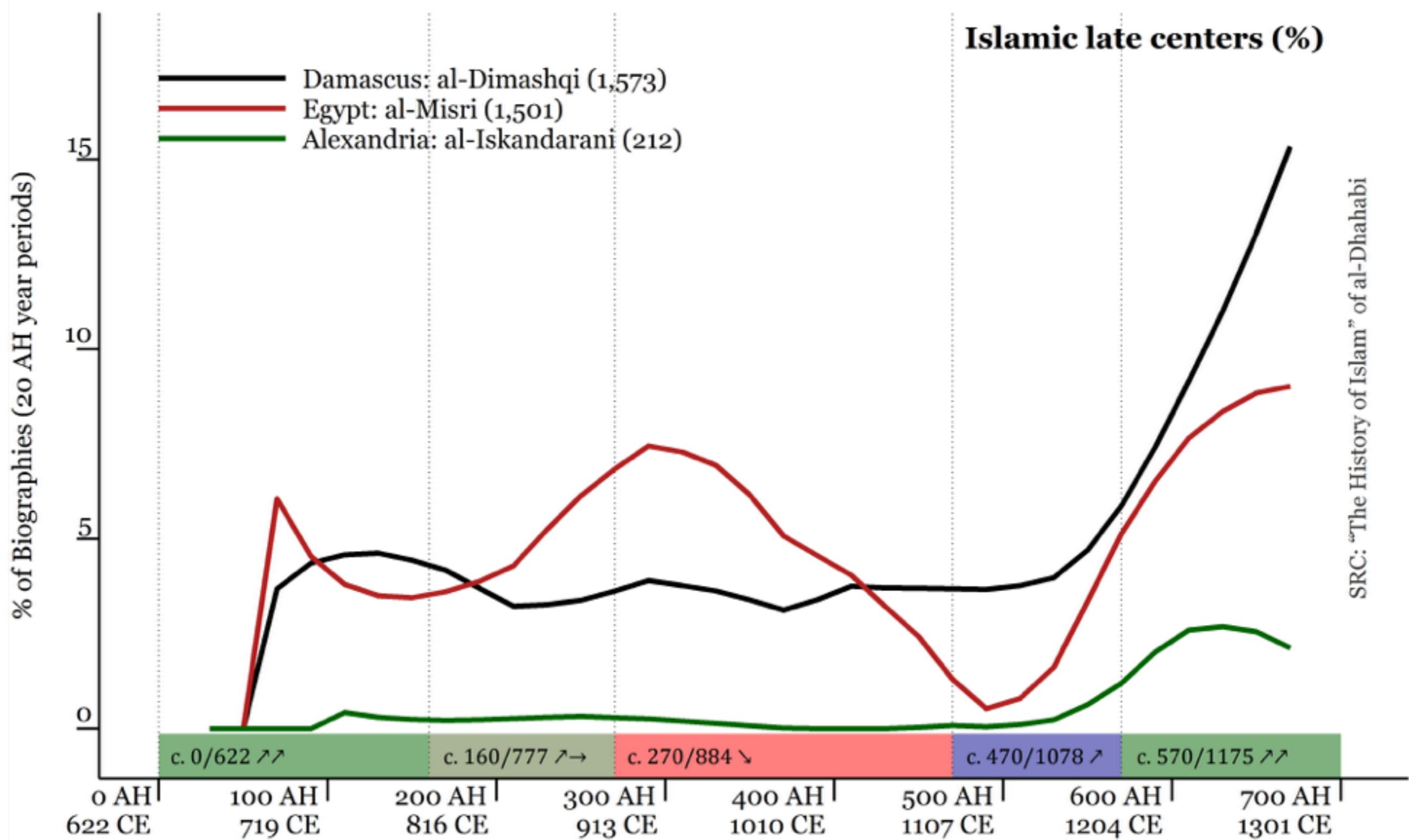
## Islamic early centers (%)

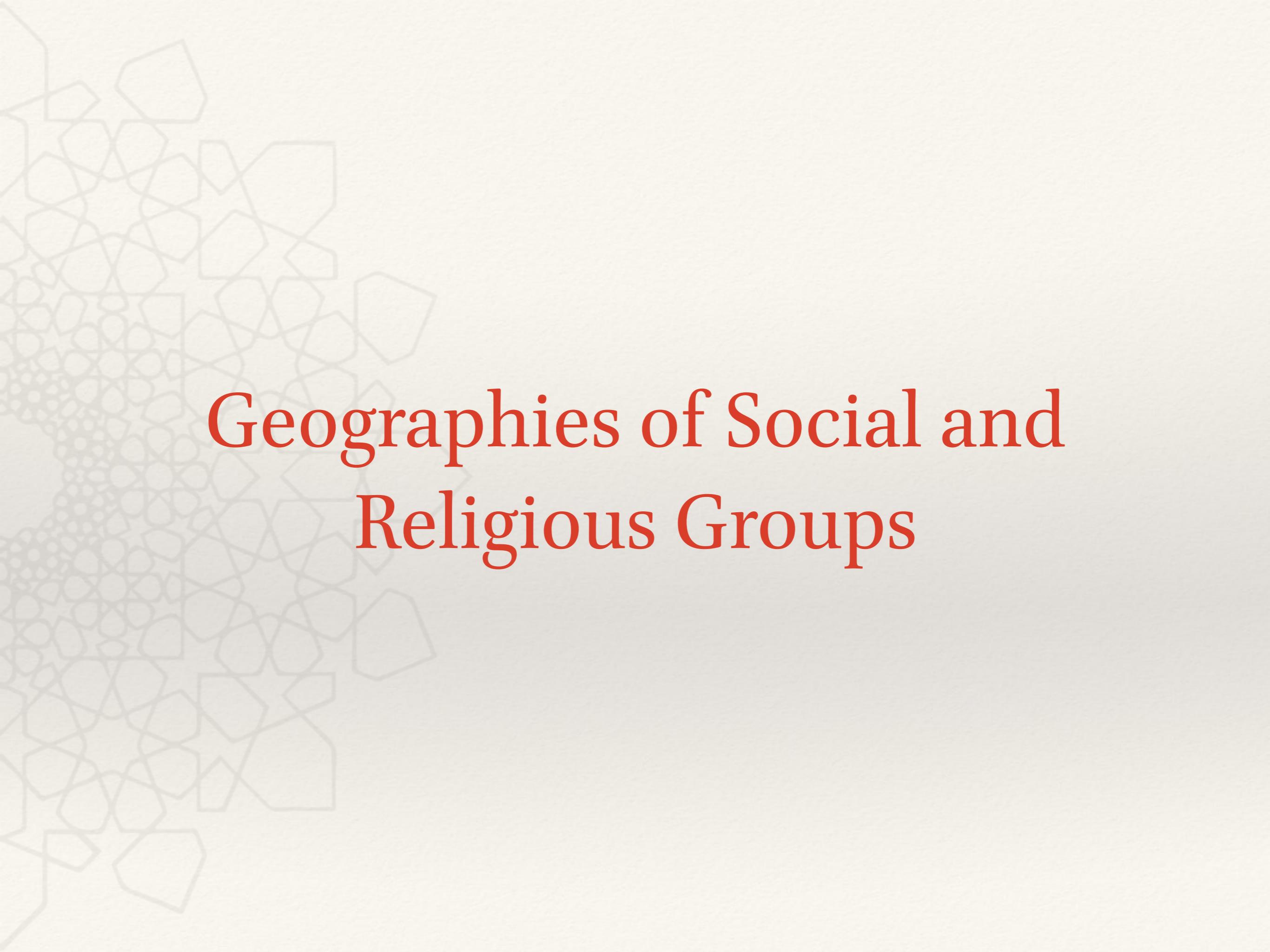


## Islamic intermediate centers (%)

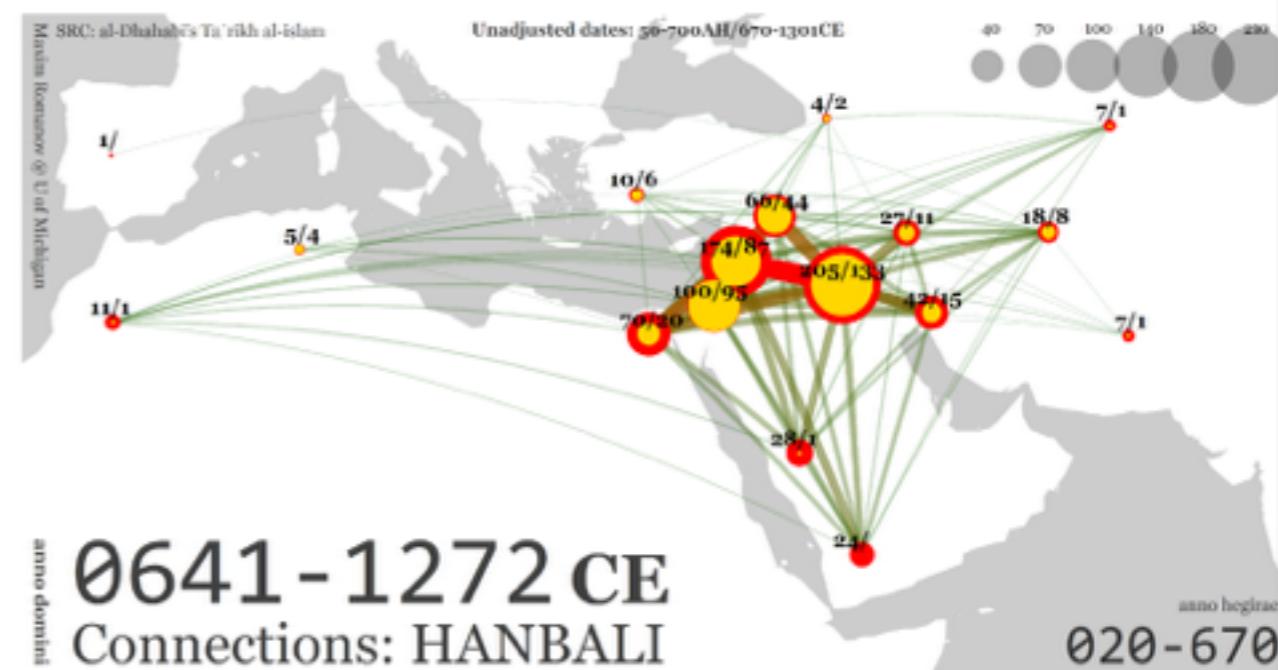
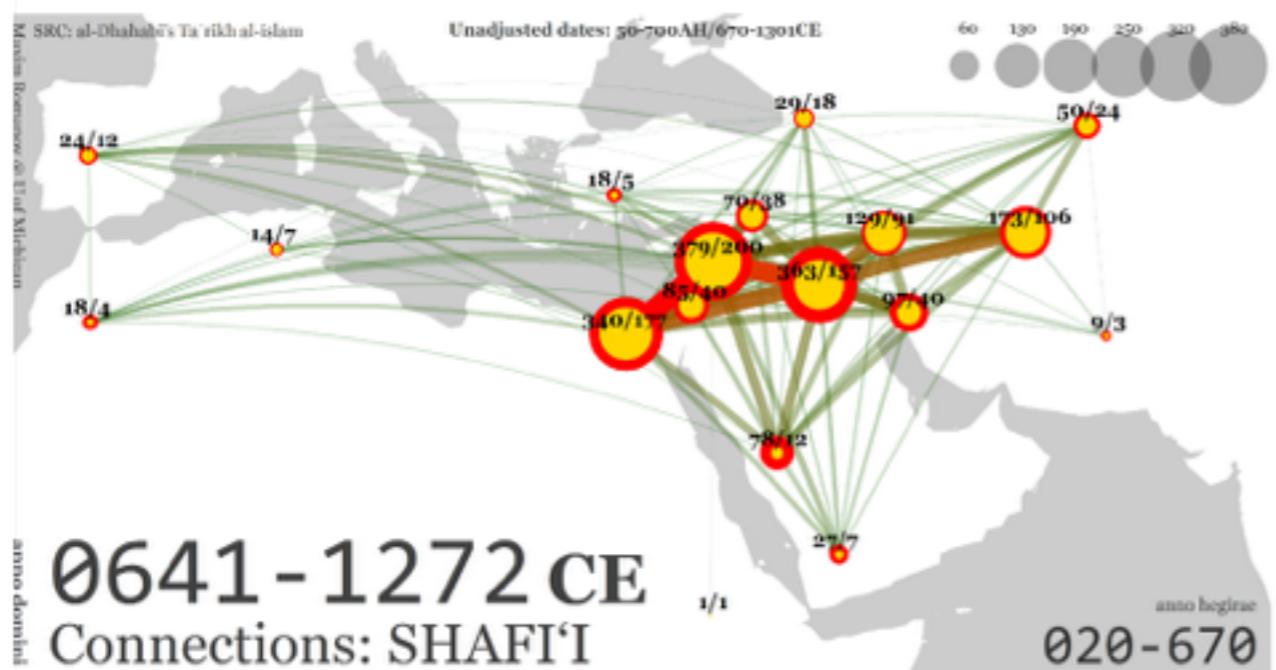
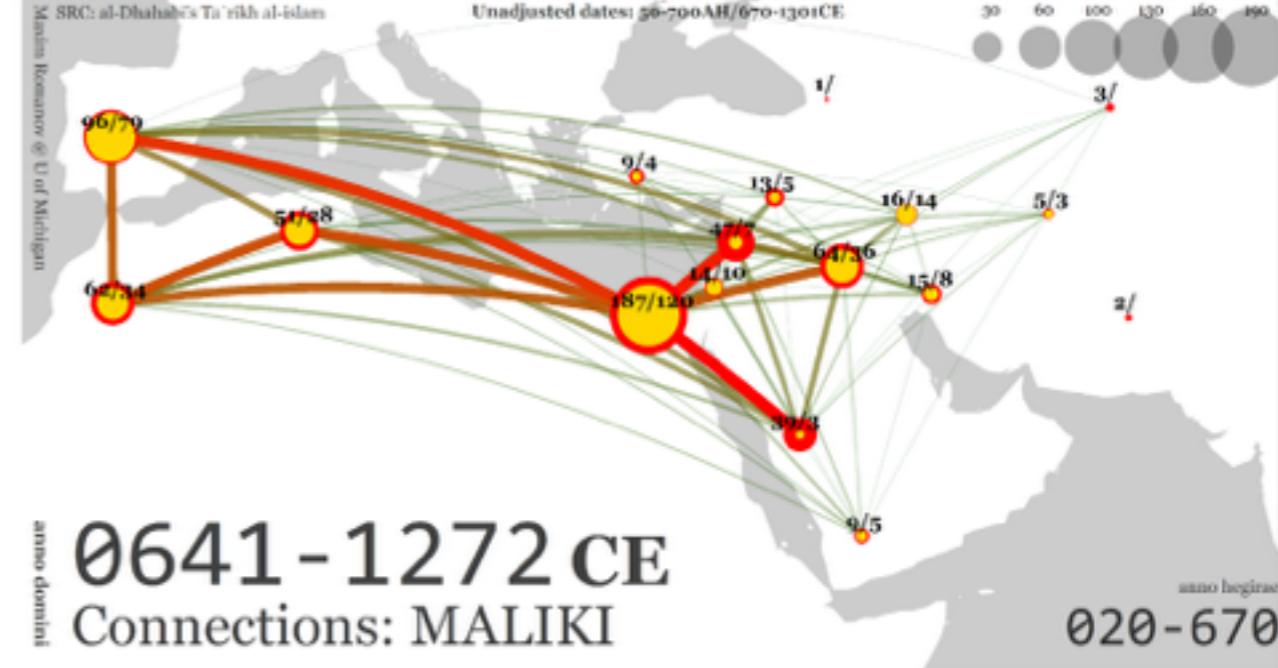
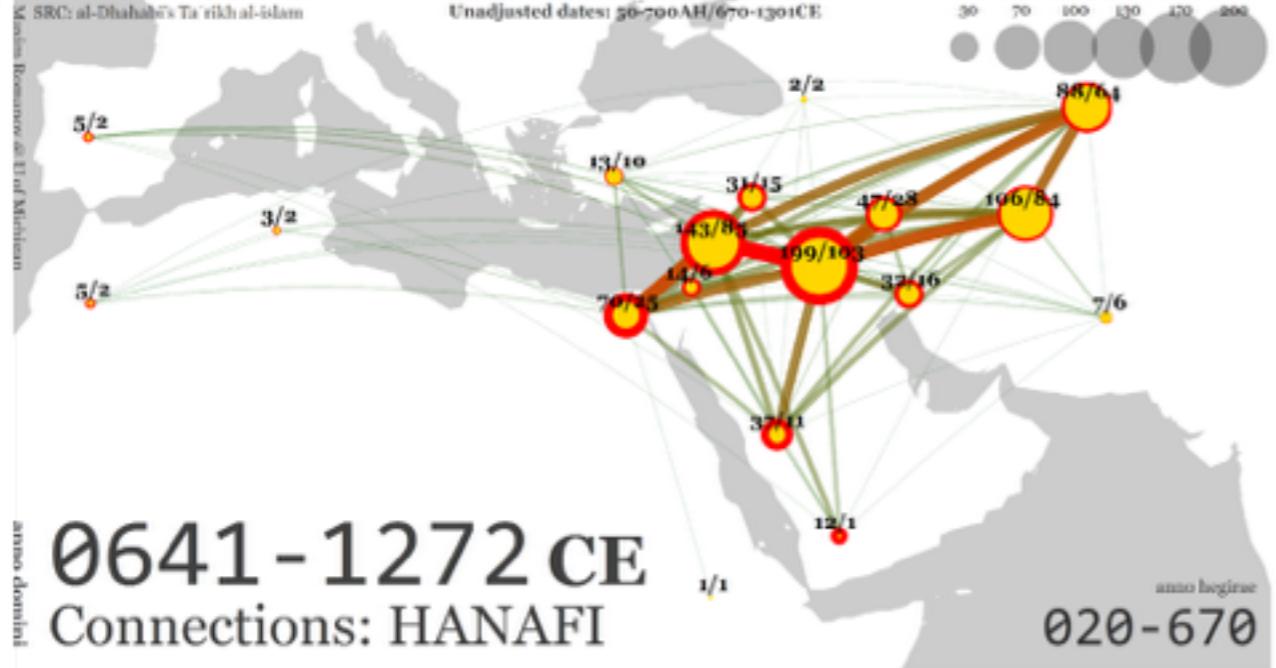


## Islamic late centers (%)





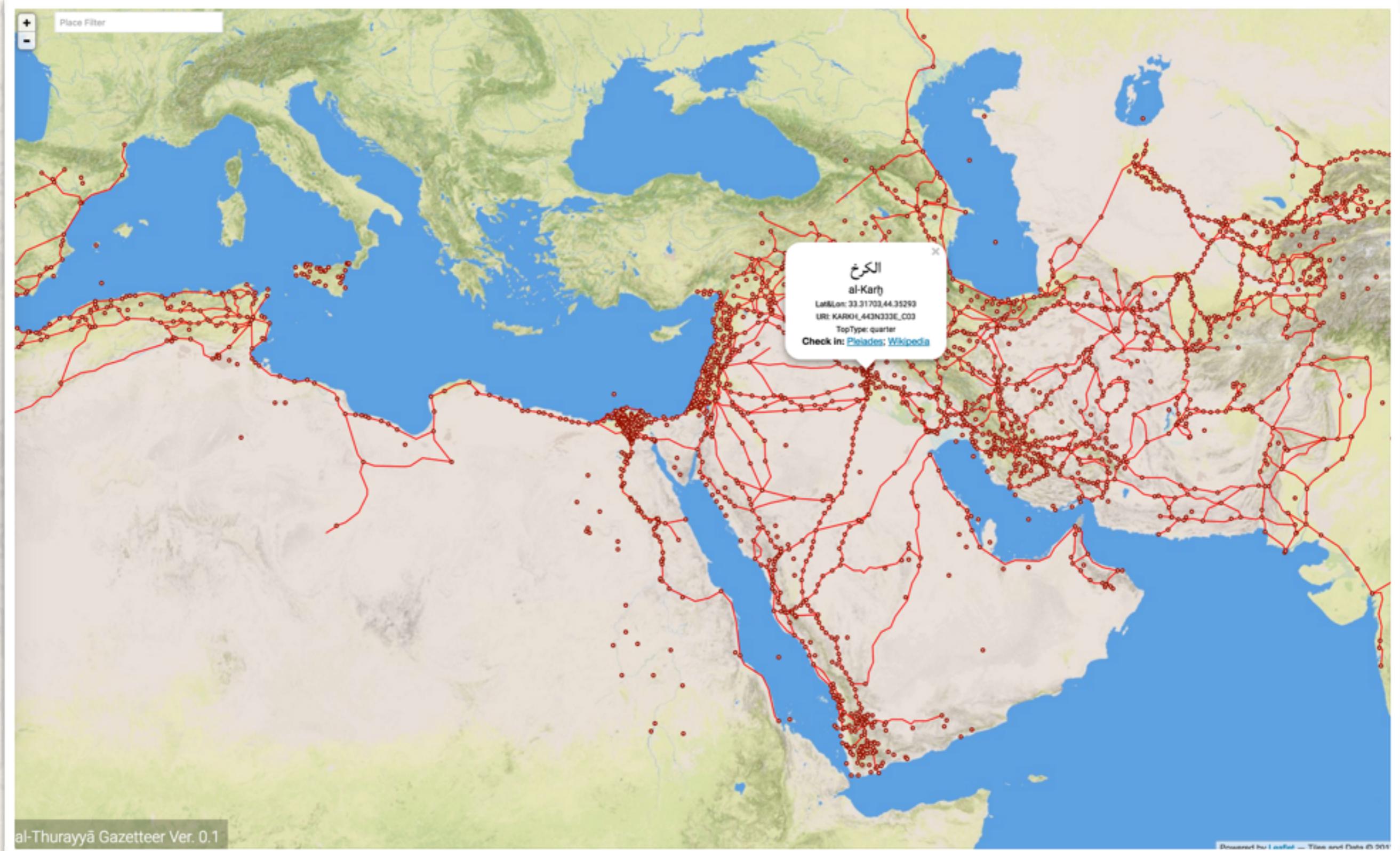
# **Geographies of Social and Religious Groups**



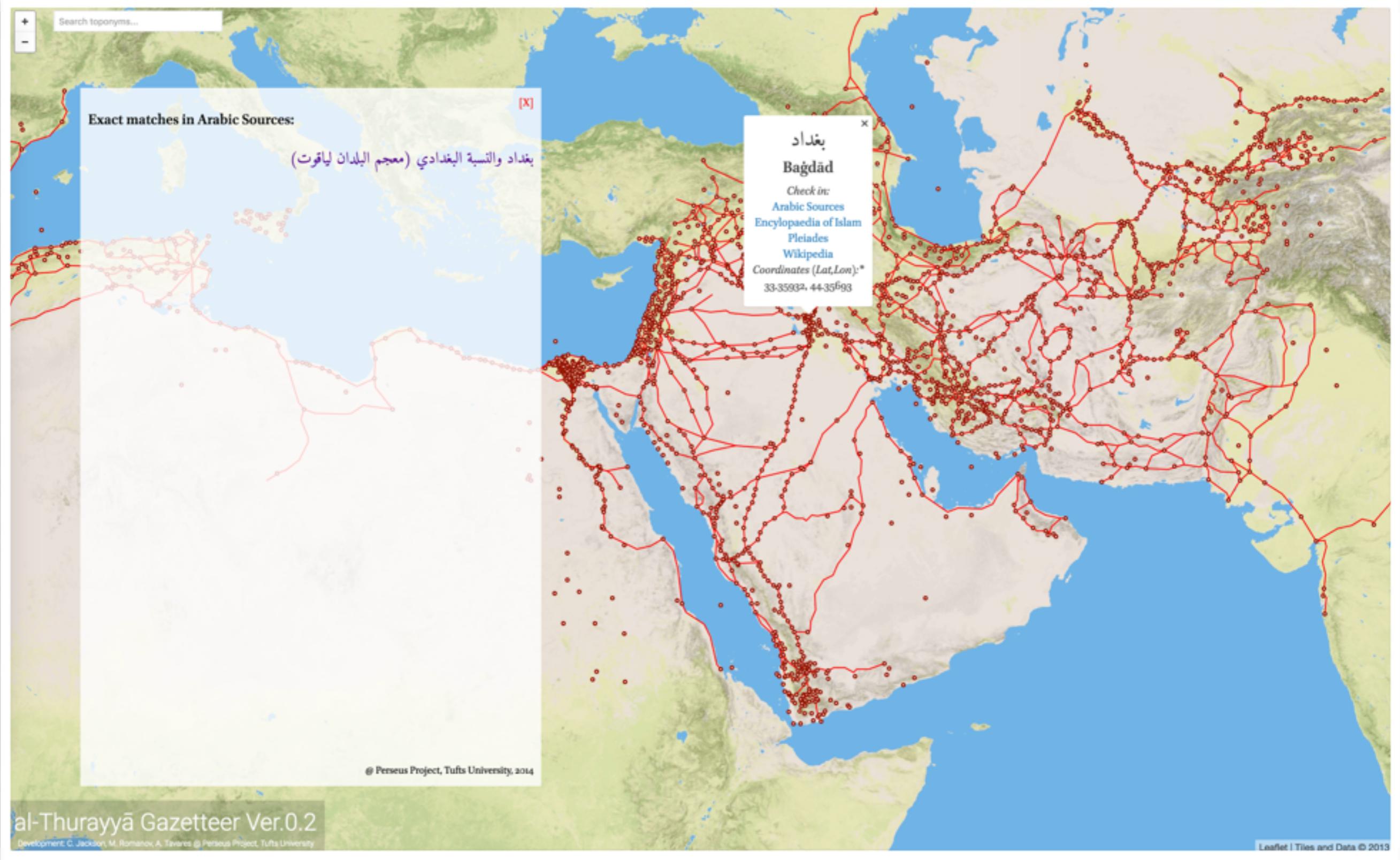
# Overview

- ❖ Collections of Arabic texts
- ❖ Analysis of a biographical collection (focusing on cultural geography)
- ❖ **Recent work on Islamic geography:**
  - ❖ **Developing a Gazetteer**
  - ❖ **Modeling geography from primary texts**

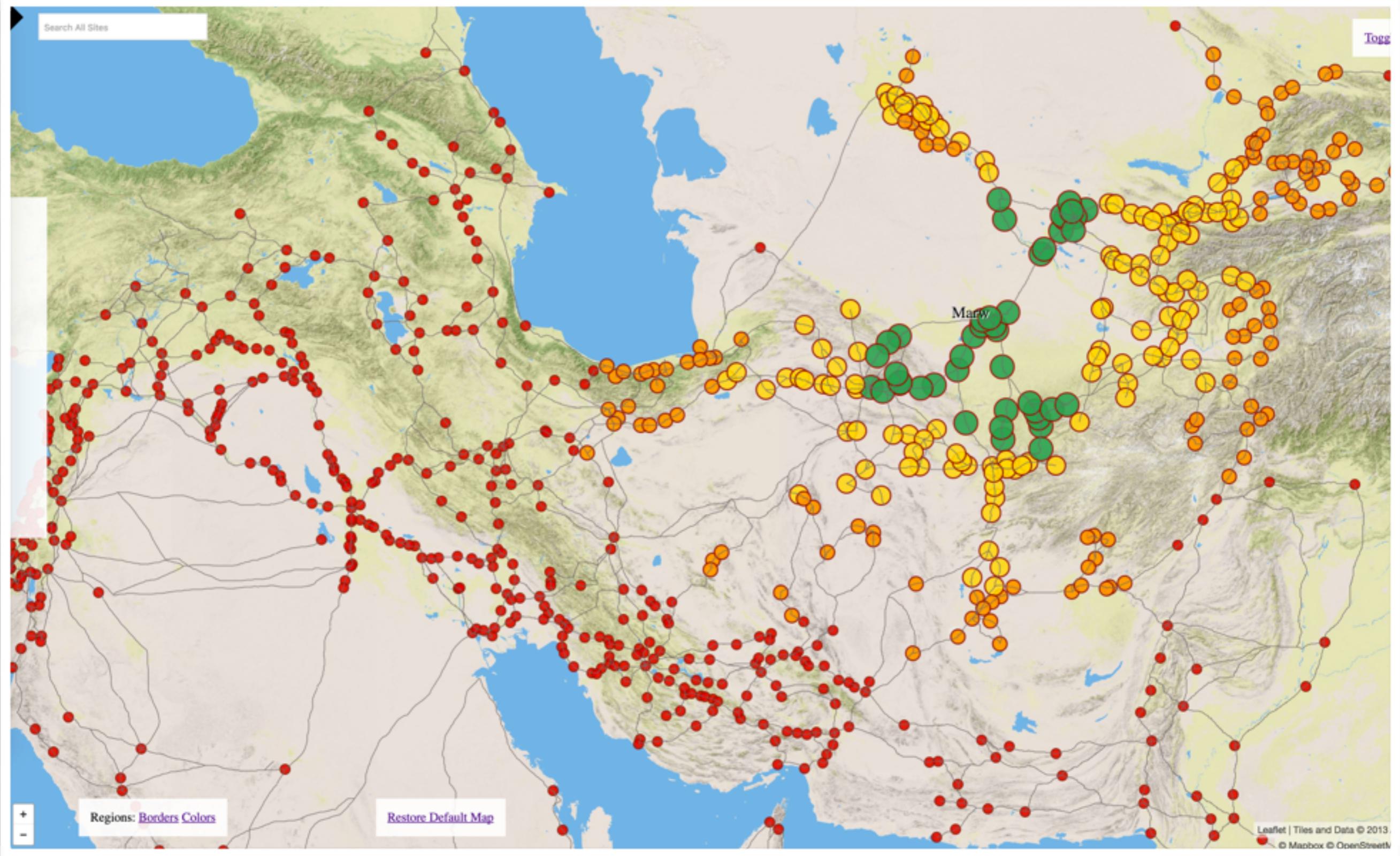
# Gazetteer of the Classical Islamic World



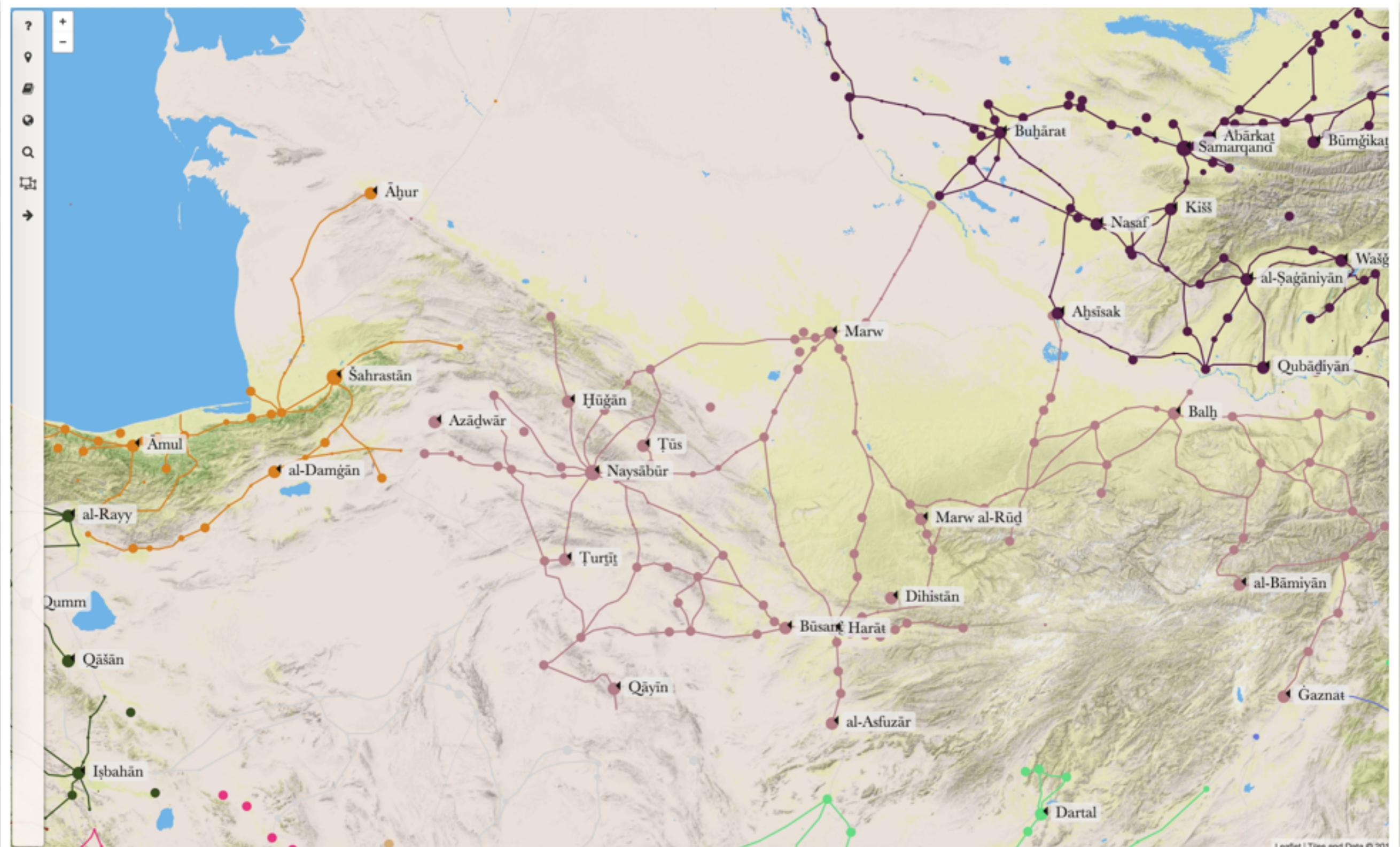
# Gazetteer of the Classical Islamic World



# Geospatial Model: Flooding Network



# al-Turayyā: Modeling Provinces



# al-Turayyā: *the latest iteration*

**?**

**al-Turayyā Gazetteer**

This is a new working version of *al-Turayyā Gazetteer* (or *al-Thuryyā Gazetteer*). Currently it includes over 2,000 toponyms and almost as many route sections georeferenced from Georgette Cornu's *Atlas du monde arabo-islamique à l'époque classique: IXe-Xe siècles* (Leiden: Brill, 1983). The functionality is still under development. You can use an earlier version of *al-Turayyā*, where you can browse the Gazetteer by clicking on any toponym marker. The popup will show the toponym both in Arabic script and transliterated. We are using a slightly modified transliteration system that facilitates conversion between fully transliterated, transliterated, and Arabic forms of toponyms. It should be easily understandable. There may be typos, because of the nature of how the data has been generated, so please, let us know if something should be corrected. The popup also offers a selection of possible sources on a toponym in question. You can check Arabic Sources: currently, al-Sam'āni's *Kitāb al-ansāb* and Yāqūt's *Mujam al-buldān*. Currently, the Gazetteer will only check for exact matches, which means that in some cases there will not be any entry at all, while in other cases there may be more than one and they may refer to other places with the same name. Improving the precision of this lookup is on our to-do list. You can also check if there is information on a toponym in question in Brill's *Encyclopaedia of Islam*, *Pleiades*, and *Wikipedia*. It can be found [here](#).

*Note on transliteration:* The website uses a somewhat unconventional transliteration system, which was developed to facilitate computational analysis. Unlike more traditional transliteration schemes the current one uses one-to-one letter representation, with every Arabic letter transcribed distinctively, which allows for an automatic conversion between transliteration and the Arabic script. The overall scheme should be easily recognizable to Arabists (new letters are as follows: *t* for *til'* marbutat; *d* for dagger *alif*; and *đ* for *alif magrūfah*).

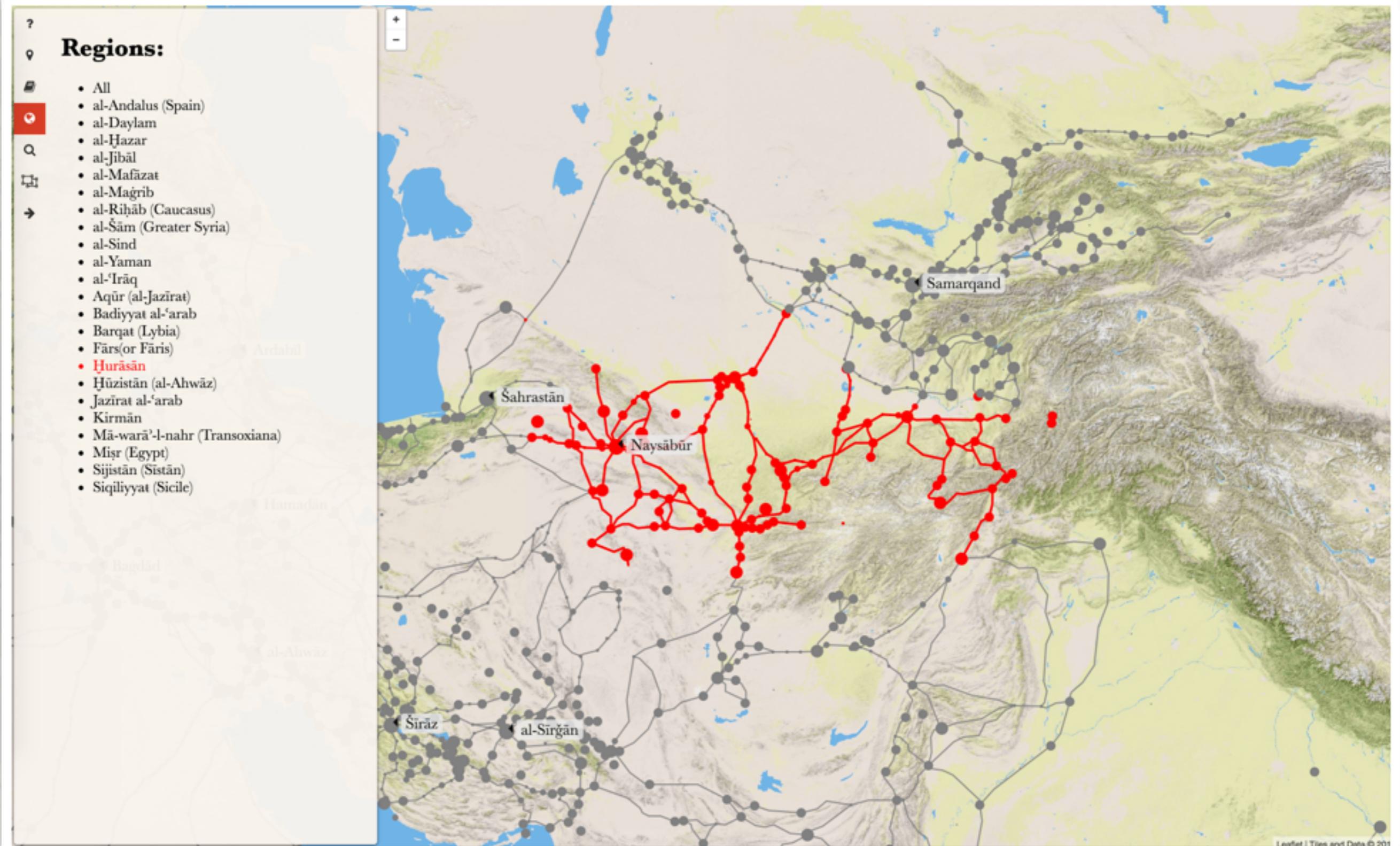
**Credits and Acknowledgments:**

**Current team:** Masoumeh Seydi and Maxim Romanov @ U Leipzig. **Former contributors:** 2013–2014: Cameron Jackson (class of 2014, double-major in Arabic and Computer Science, Tufts)—technical and conceptual development; 2013: Adam Tavares, programmer @ Perseus Project, Tufts—technical development. **Special thanks to:** 2013–2014: Vickie Sullivan (Chair, Classics Department, Tufts U), 2013—Gregory Crane and the Perseus DL and the U Leipzig teams for support and inspiration.

**Open Data:**

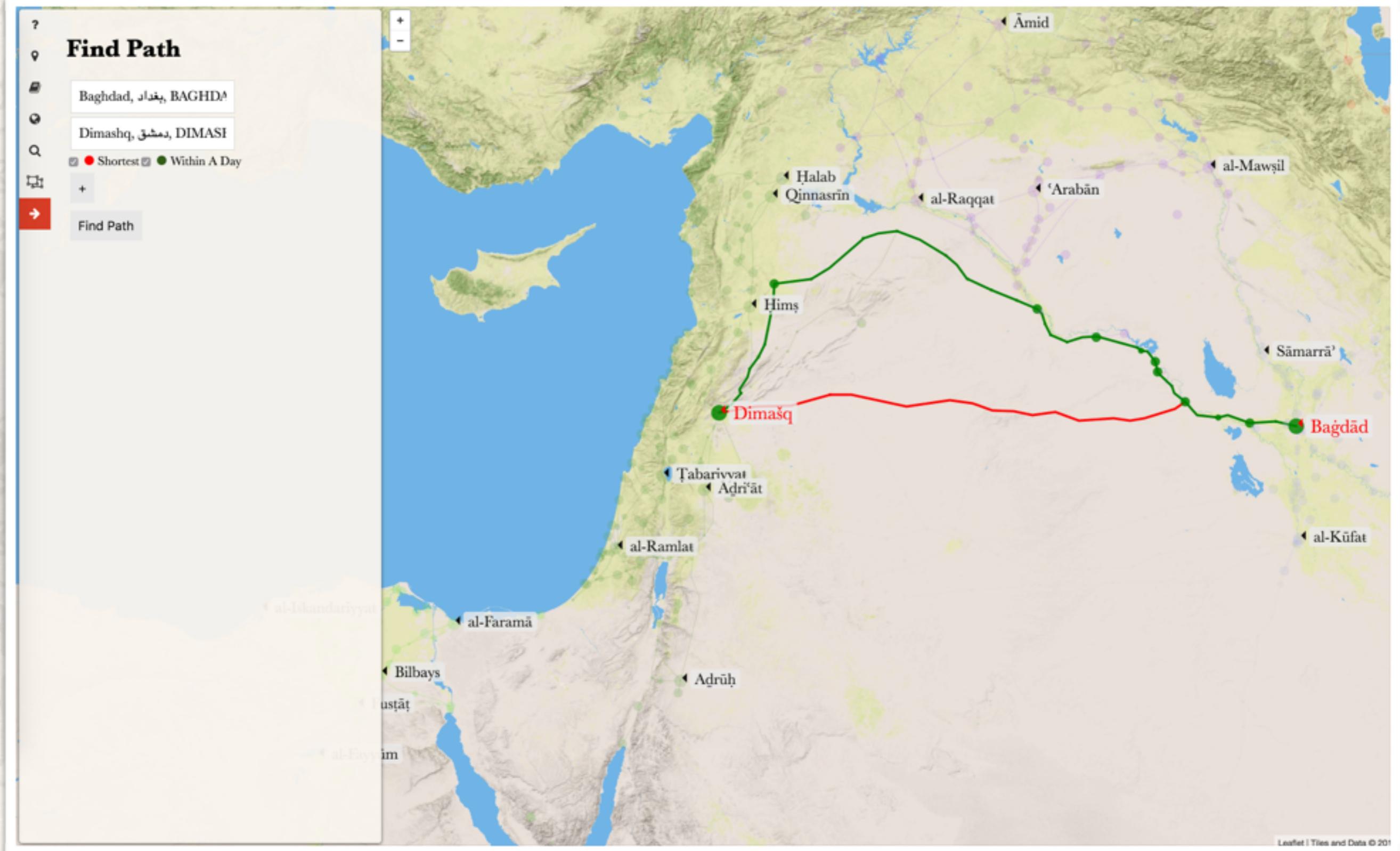
Leaflet | Tiles and Data © 2019

# al-Turayyā: Modeling Regions



With Masoumeh Seydi (Phd Candidate in DH, U Leipzig)  
Building on Cameron Jackson's work (Class 2015, Tufts)

# al-Turayyā: *Finding Routes*



# al-Turayyā: Consulting Primary Sources

**Sources on: بغداد**

- al-Himyari's *Rawd al-mi'ār*: بغداد (100% match)

بغداد

دار مملكة خلقاء بني العباس، وفيها أربع لغات: بغداد بدالين مهملتين، وبغداد معجمة الأخيرة، وبغدان بالتون، ومقدان باليم بدلا من الباء، وذكر وترت، قالوا: وبغدان بالفارسية خطبة الصنم لأن يخ صنم وذاد عطية، ولذلك [v.1, p.109] كره الأصحاب هذه التسمية، وكانت قرية من قرى القدس فأخذها أبو جعفر عصبا طرق فيها مدينة وقال الحرجاني: ياخ بالفارسية هو البستان الكبير الشجر، وذاد: معطي، فعناء معطي البستان، قال أبو عثمان النبدي: كاسير مع جرير بن عبد الله الجل حق انتهى إلى موضع فقال: أي موضع هذا؟ قالوا: قطريل، فترك دابه ثم قال: حمت التي حل الله عليه وسلم يقول: يجئ مدينة بين دجلة والدجل والصرارة وقطriel يعني إليها خراج كل أرض وتجمع إليها جبارية الأرض، وفي رواية يختلف بها، كما أحسب، سميت بغداد لأنه أهدي إلى كسرى خصي من الشرق وكان له صنم يقال له يخ فقال الخصي: بغدادي أي أعطائي إله يعني الصنم، وهذا كان المترجون يكرهون أن يسموا بغداد بهذا الاسم ويقولون بغداد بالدل المهملة.

وكان أبو جعفر المنصور بعث رجالاً سنة تسع وأربعين ومائة يطلبون له موضعها يعني فيه مدينة فطلبوا لهم يرضاوها موضعها حتى جاء موضعها بالصراة وقال: هذا موضع أرضاء تأثيره الميرة من القرارات وجبلة والصرارة، وكان أبو جعفر هذا وهو عبد الله بن محمد بن علي بن عبد الله بن العباس يعني مدينة بين الكوفة والجزيرة جحشاً لها خاصية فأقام بها مدة إلى أن عزم على توجيه أبي محمد النبدي لفزو الصائفة في سنة أربعين ومائة فصار إلى بغداد فوقف بها وقال: ما اسم هذا الموضع؟ فقبل: بغداد، فقال: هذه والله المدينة التي أعلمني أبي محمد بن علي أبي أبيها وأزتها ويزتها ولدي من بعدي، ولقد غفلت عنها الملوك في الجاهلية والإسلام حتى يتم تدبر الله تعالى وحكمه في وتصح الروايات وبين الدلالات والعلامات تأثيرها الميرة في الدجلة والقرارات من واسط والأبلة والأهواز وفارس وعمان والجامعة وما يصل بذلك، وكذلك ما يأتي من الموصل وديار ربيعة وأذربيجان وأرمانيا والرقة والشام والغور ومصر والمغرب وأصفهان وكبور خراسان فالحمد لله الذي ذكرها لي وأفضل عنها كل من تقدمي والله لأنينا ثم أسكننا أيام حياني ويسكتها ولدي من بعدي ثم تكون أغير مدينة في الدنيا ثم لأبين بعدها أربع مدن لا تقرب واحدة منها أنها فربناها وبين الرافقة ولم يستكناها وبين مملكتها والمصيحة والنصرة، فوجه في حشر الصناع والقعلاة من الشام والموصل

# *Modeling geography from primary texts*

- ❖ Extracting and modeling data from primary sources
  - ❖ route networks
  - ❖ locations of extinct settlements
  - ❖ visualization of regions
  - ❖ dynamic travel networks (based on biographical data)



map  
~~There's a graph~~  
for that