

Mapping India in Pliny the Elder's *Natural History*

Dawn, Lizao Zhuang (r0914937)

Abstract

this is an abstract

Table of Contents

1	Introduction	1
1.1	<i>Natural History</i> and its complexity	1
1.2	Spatial perspective in <i>Natural History</i>	2
1.3	Text source for the study	4
2	Research Question	4
2.1	Prominent mentioned places in <i>Natural History</i>	4
2.2	Why India?	6
2.3	India-related text as a case study	7
3	Methodology	7
3.1	Workflow	7
3.2	Data preparation	9
3.2.1	HTML scraping from TOPOSText	9
3.2.2	Filtered dataset of “India-related text”	11
3.2.3	Data completeness check	12
3.2.4	Preprocessing of texts	14
4	Data Analysis	14
4.1	Place name distribution in India-related text	14
4.2	Word frequency and collocating bi-grams	16
4.3	Topic modeling	20
4.4	Network analysis for Named Entity	29
4.4.1	Person name annotation/tag retrieve	29
4.4.2	Network graph generation	33
5	Conclusions	38
5.1	Comprehension of “India” in the narrative	38
5.2	Distant reading as a method	39
5.3	Reflection and limitation	40
References		41

List of Figures

1	Normalized distribution of place names in <i>Natural History</i>	4
2	Place name distribution map	6
3	Occurrence count for all place names and place names of Indian subcontinent in each book	15
4	Occurrence count for all place names and place names of Indian subcontinent in each book_different y-axis scales	15
5	Top 1% frequent words in Indian subcontinent related text as tree map .	17
6	Top 1% frequent words in Indian subcontinent related text as word cloud	17
7	Coherence distribution of different topic numbers when passes=40 . .	21
8	Topic cluster (overall)	23
9	Topic cluster (highlighting on Topic_0)	24
10	Topic cluster (highlighting on Topic_1)	24
11	Topic cluster (highlighting on Topic_2)	25
12	Context type distribution in books containing Indian place names . .	28
13	Occurrence frequency of different context type in India-related text .	28
14	Place/person/book number network of India-related content in <i>Natural History</i>	35
15	Example of edged groups in the network graph of India-related text in <i>Natural History</i>	36
16	Clusters of Book 6 & Book 7 in India-related content in <i>Natural History</i> .	36
17	Clusters of Book 37 in India-related content in <i>Natural History</i>	37

List of Tables

1	Normalized distribution of place names in <i>Natural History</i>	3
2	Top 20 mentioned place names in Natural History	5
3	Example for the reference dataset containing the plain text in paragraphs of <i>Natural History</i>	9
4	Example for the <i>Natural History</i> geographical-related text dataset . .	10
5	Example for the India-related dataset	11
6	Example for supplement annotation to Indian places in <i>Natural History</i> .	13
7	Example of retrieved and validated person name annotations in <i>Natural History</i>	30
8	Example for person name annotation merged into India-related text dataset	30
9	Example of person name tags retrieved with WikiNEuRal from India-related text in <i>Natural History</i>	31
10	Example of person name tags retrieved with Flair from India-related text in <i>Natural History</i>	32
11	Quantity of person name entities retrieved with different methods . .	32
12	First 15 distinct person names retrieved with different methods . . .	33

1 Introduction

1.1 *Natural History* and its complexity

Pliny the Elder's *Natural History* is widely recognized as the earliest encyclopedia in the world, manifesting a pioneering effort in comprehensively cataloging the vast array of human knowledge from that era.

The work is thematically divided into 37 books, covering a diverse range of subjects including astronomy, geography, zoology, botany, medicine, and more. Pliny meticulously consulted a wide range of Greek and Roman references, totaling approximately 2,000 volumes¹, and interwove his own literary interpretation or comments to the narratives.

Despite the carefully designed knowledge-ordering framework (Lao 2016), scholars have observed a paradoxical complexity in *Natural History*, evident in its linguistic style, narrative approach, and use of references. The work compiles inconsistent toponyms from Greek and Latin, includes digressions in descriptions (Roller 2022), and exhibits changes in vocabularies and sentence structures (Pinkster 2005). However, it is precisely this complexity that makes the work more fascinating and not only a valuable source of the knowledge and worldview of the ancient world, but also a gateway into Pliny's conceptualization, imagination, and even the prevailing imperial ideology of that time.

The complexity and interconnectivity of the general structure of *Natural History* are further highlighted in different aspects by refreshing approaches. For the content organisation, Healy (1999) advocated Pliny's original contribution in revealing the technology and science engagement of the Rome Empire. Taking the historical, political and linguistic context into consideration, he found that in addition to providing informative descriptions of natural phenomena and scientific experiments, Pliny also developed the scientific language in Latin through his work. And Naas (2002) discussed how Pliny formulated the diverse materials into his encyclopaedic structure, revealing the work's multifaceted nature as an epistemological, ideological, and moral project. Analysing Pliny's employment of the historical exemplum in the work, Schultze (2011) argued how that literary device directed and teased the readers as well as established a profound connection between human beings and the entire spectrum of nature in *Natural History*.

In addition to the close reading methods used in the above analyses of the context and references in *Natural History*, Rydberg-Cox (2021) employed network analysis method with different metrics to map the interrelationships between Pliny's sources and the topics discussed in the work. Furthermore, Fantoli (2022) presented a comparative study

¹ *Natural History* 1.5.1 (<https://topostext.org/work/148>)

of book 2 of *Natural History* and book 7 of Seneca's work *Natural Questions*, both of which are astronomy sections in the Latin classics about nature. Her study utilized statistical method to analyse variations in their discourse distribution, and identified Pliny's unique stylistic features. The encyclopaedic authorial intent shown in *Natural History* is also vindicated through correspondence and tree analysis. These two studies demonstrate how distant reading methodologies offer novel insights into our understanding of the ancient treatises.

In addition to the contextual and referential analyses about *Natural History*, Rydberg-Cox (2021) employed network analysis techniques, applying various metrics to illustrate the connections between Pliny's sources and the topics discussed in the work. Furthermore, Fantoli (2022) conducted a comparative examination of book 2 of *Natural History* and book 7 of Seneca's *Natural Questions*, both focused on astronomy as part of the Latin classics about nature. Statistical methods were applied to investigate variations in their discursive distribution, effectively identifying Pliny's distinctive stylistic attributes. Moreover, the encyclopaedic authorial intent evident in *Natural History* is validated through correspondence and tree analysis. These studies collectively showcase how distant reading methodologies offer novel insights into our understanding of ancient treatises.

1.2 Spatial perspective in *Natural History*

As pointed out by Beagon (2011), differentiating from his predecessors, Pliny showed a “terrestrial curiosity” in *Natural History*, emphasising recognition of the physical, material world. In this regard, the vision of geography plays a pivotal role in distributing information, knowledge, and events throughout *Natural History*.

Drawing from the long-established topographical and ethnographic traditions, Pliny seamlessly connects volumes dedicated to geography (books 3-6) with broader elements, activities, and cultural, historical, and societal contexts(Roller 2022), exemplified in his portrayal of exotic plants, communities' habitats, imperial expeditions, and trade products. In other words, geographical names that occurred in each book of *Natural History* served as signposts guiding readers through diverse lands, shedding light on how Pliny and his contemporaries perceived and conceptualized the world around them.

A normalized frequency of place name occurrence in the work is calculated as the ratio of counts of the occurrences of place names in each book to the word lengths of the book (Table 1). The bar chart (Figure 1) depicts the comparison of the distribution of place names in the books of *Natural History*. The observation aligns with the content structure of *Natural History*, that books 3-6 centred around the themes of “geography and ethnography”, contain the most mentions of location names, and place names are also frequently mentioned in books about agriculture and horticulture (book 12-14),

aquatic life (book 31), and mining and mineralogy (book 34-37).

Table 1: Normalized distribution of place names in *Natural History*

Book	Total_length	Place_count	Place_freq
1	2778	1	0.000360
2	30570	406	0.013281
3	18037	1007	0.055830
4	15434	1309	0.084813
5	18872	1112	0.058923
6	27891	1012	0.036284
7	21204	225	0.010611
8	24176	185	0.007652
9	19197	140	0.007293
10	20816	121	0.005813
11	27345	77	0.002816
12	13906	188	0.013519
13	13243	164	0.012384
14	15277	189	0.012372
15	14552	135	0.009277
16	25442	180	0.007075
17	29387	82	0.002790
18	35850	222	0.006192
19	18822	146	0.007757
20	22743	21	0.000923
21	17896	95	0.005308
22	16491	24	0.001455
23	15764	17	0.001078
24	17491	56	0.003202
25	16734	85	0.005079
26	15448	35	0.002266
27	12444	40	0.003214
28	26476	28	0.001058
29	13976	31	0.002218
30	14395	23	0.001598
31	12204	222	0.018191
32	14635	76	0.005193
33	17946	113	0.006297
34	18972	193	0.010173
35	21283	277	0.013015
36	21295	357	0.016764

	Total_length	Place_count	Place_freq
Book			
37	22255	282	0.012671

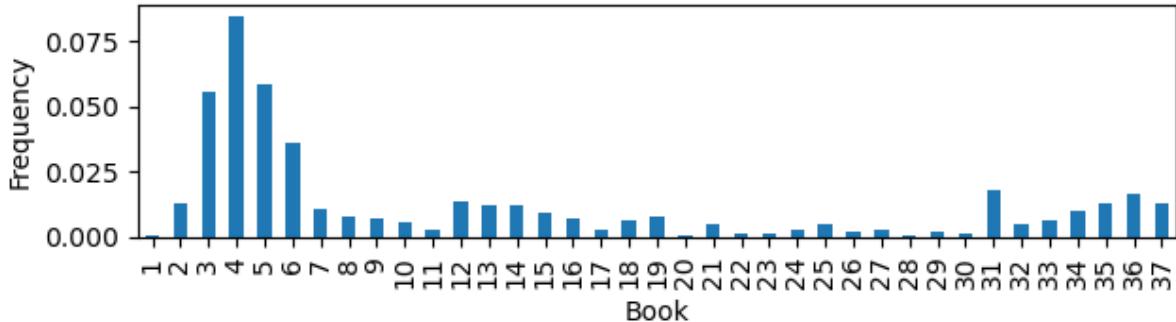


Figure 1: Normalized distribution of place names in *Natural History*

1.3 Text source for the study

Natural History is originally in Latin. For the purpose of this study, an English translation conducted by Henry T. Riley (1816-1878) and John Bostock (1773-1846), first published in 1855, is utilized. The translated text is obtained in a digitized version from the [TOPOSText project](#), having been sourced from the Perseus Project and governed by a Creative Commons Attribution-Share-Alike 3.0 U.S. License.

Annotations of proper names, place names and their corresponding coordinates are available together with the text of *Natural History* ([Book1-11](#), [Book12-37](#)) on [TOPOS-Text project](#). This invaluable resource allows for creating a dataset that includes both the textual contents and geographical annotations, which can be utilized to investigate the distribution of place names in the entire work and examine the frequencies and patterns of geography-related content.

The extension of the extracted corpora and the workflow of the extraction will be further explained in the Methodology chapter (Section 3).

2 Research Question

2.1 Prominent mentioned places in *Natural History*

Based on the geographical annotations in *Natural History* provided by the TOPOSText, there are 2052 unique places mentioned in *Natural History*.

The top 20 most frequently mentioned place names (as 1% of the total) in *Natural History* is shown in Table 2.

Table 2: Top 20 mentioned place names in Natural History

ToposText_ID	Place_Name	Lat	Long	Count
1687 https://topostext.org/place/406163RIta	Italy	40.6000	16.30000	292
2034 https://topostext.org/place/419125PRom	Rome	41.8910	12.48600	269
52 https://topostext.org/place/271307REgy	Egypt	27.1000	30.70000	261
82 https://topostext.org/place/300740RInd	India	30.0000	74.00000	167
57 https://topostext.org/place/280400RAra	Arabia	28.0000	40.00000	123
320 https://topostext.org/place/355390RSyr	Syria	35.5000	39.00000	109
255 https://topostext.org/place/350330RCyp	Cyprus	35.0000	33.00000	85
109 https://topostext.org/place/312301WNil	Nile	30.0918	31.23130	85
2282 https://topostext.org/place/441073LAlp	Alps	44.1420	7.34300	82
766 https://topostext.org/place/376145RSic	Sicily	37.6000	14.50000	71
275 https://topostext.org/place/352252IKre	Crete	35.2052	25.18360	64
7 https://topostext.org/place/130350REth	Ethiopia	13.0100	35.01000	58
417 https://topostext.org/place/364282IRho	Rhodes	36.4408	28.22440	56
966 https://topostext.org/place/380237PAth	Athens	37.9718	23.72793	56
2043 https://topostext.org/place/419125SCap	Capitol	41.8933	12.48300	52
298 https://topostext.org/place/353403WEup	Euphrates	35.2791	40.27080	47
2241 https://topostext.org/place/435335WPon	Pontus	43.5000	33.50000	47
1839 https://topostext.org/place/411146RCam	Campania	41.1000	14.60000	46
1480 https://topostext.org/place/397443RArm	Armenia	39.7020	44.29800	45
17 https://topostext.org/place/195390WEry	Red Sea	19.5000	39.00000	42
545 https://topostext.org/place/369103PCar	Carthage	36.8500	10.32000	42
602 https://topostext.org/place/370340RCil	Cilicia	37.0100	34.01000	42

The place names referenced in *Natural History* are geographically mapped, with each location marked on the map using its corresponding coordinates. A dot is assigned to represent a place, with the size and color of the dot reflecting the frequency of its mention in the work. The larger and darker the dot, the more frequently the place is referred to within the context of *Natural History*.

An intriguing observation from the output, as depicted in Figure 2, is the prominence of India, a region outside the Mediterranean, despite its high frequency of mentions.

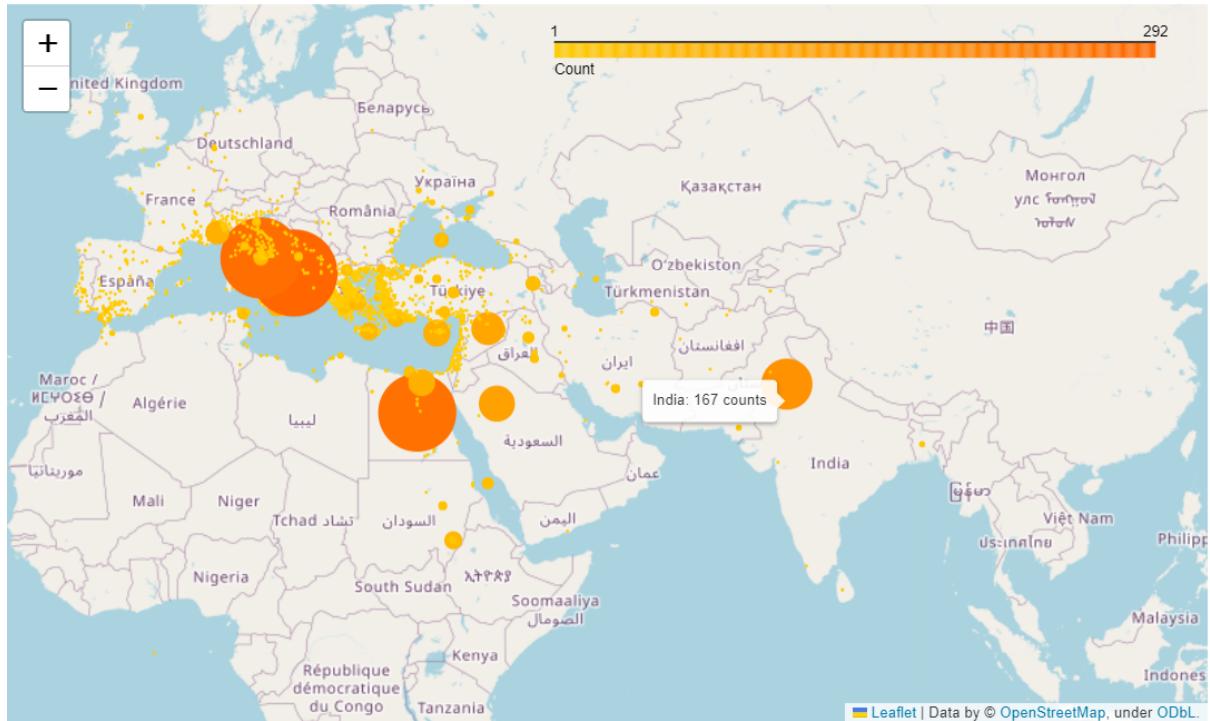


Figure 2: Place name distribution map

2.2 Why India?

Geographically, India seems a distant and disconnected territory from the Roman Empire, lacking any direct aquatic or land routes with the Mediterranean region. However, the exotic curiosity Pliny attempted to integrate within his work, and the history of the Indo-roman trade relations, can be regarded as a broader context for the prominent mentioning of India in *Natural History*.

As suggested by (Murphy 2003), the *mirabilia*, encompassing accounts of extraordinary landscapes, people, plants, and animals, assumes a substantial proportion within the books of *Natural History*. Pliny's inclusion of such exotic elements not only catered to the prevailing curiosity of his Roman readers but also fostered a comparative perspective between distant locales, exemplified by his references to India, and their natural counterparts within Rome (Naas 2011). Within a research framework of Roman Imperialism, the detailed portrayal of foreign lands, such as India, holds significant importance in shaping both Pliny's and his contemporary Roman readers' perception of their place within the global landscape (Pollard 2009).

In addition, *Natural History* serves as a valuable reference for tracking the Indo-Mediterranean network of exchange (Pollard 2009). Through the depiction of cities, ports, and rivers along the trade routes, the work provides substantive evidence of the flourishing trade relations between the Roman Empire and the Indian subcontinent (Neelis 2011). The extensive exemplification of diverse commodities, such as

gemstones, glass, spices, textiles, plants, and wine, along with the accounts of the currency *sestertii* involved in the merchandise exchange in the work shed lights to the compelling details and social and cultural implications of this long-distance trade (Székely 2006; Pollard 2009). Furthermore, the direct criticisms regarding the high cost of the luxury items imported from India imply both the magnitude of the trade volume and Pliny's stance towards this commercial interaction (Neelis 2011).

2.3 India-related text as a case study

In light of the observations and foundational research mentioned above, the present study centres its investigation on the spatial perspective within Pliny's *Natural History*, focusing on the texts about India, seeking to delve into the discourse surrounding this region. To achieve this goal, distant reading methodologies, including statistical analysis, topic modelling, and social network analysis, are employed in the study.

The main aim of this study is to explore how India is described, and how the information about India is structured in *Natural History*, which may also contribute to a more profound comprehension of the inherent complexity and interconnectivity that permeates this monumental work.

3 Methodology

3.1 Workflow

The workflow for this study involves the following key stages:

Data Collection:

As mentioned in the Introduction chapter (Section 1), the text employed for this study is obtained from the digitized English translation (by Henry T. Riley (1816-1878) and John Bostock (1773-1846)) of Pliny's *Natural History* available on [TOPOSText project](#).

The two parts of *Natural History* ([Book1-11](#), [Book12-37](#)) are scraped for their the textual contents together with the geographical annotation of the mentioned ancient places, and their book, chapter and paragraph affiliations with the function provided in [Beautiful Soup](#) library of Python.

Data Preprocessing:

The information extracted from the HTML is structured into separate columns as [Pandas](#) dataframe, a dataframe for plain text of the entire work, and a dataframe for geographical-related text in *Natural History* with the geographical annotations are generated and stored in CSV format.

After a preliminary exploration, the research focus is narrowed down to India-related text in *Natural History*. Referring to the geographical territories in the scope of ancient Greek and Roman world (Talbert 2000b), a dataframe for India-related text is filtered from the abovementioned dataframe for geographical-related text with the range of corresponding coordinates of Indian subcontinent in the era of *Natural History*. The filtered India-related text dataframe is also stored in CSV format.

The location names mentioned in the India-related text were checked manually for its completeness. The identified location names without annotation in TOPOSText were appended to the India-related text dataset.

Additionally, the textual contents in the datasets were preprocessed with tokenization, lemmatization and the exclusion of stop-words processes.

Data Analysis:

Statistical analysis is conducted in the preliminary exploration of the extracted dataframes. A normalized frequency of geographical name occurrence in each book is calculated for an overview of the place name distribution in *Natural History*. And the top 1% prominently mentioned place names in the entire work are sorted out with the time of their occurrences. The specific attention on India-related text as a case study is drawn from this initial observation.

In the analysis of the India-related text (target corpus) in *Natural History*, three analysis methods are employed:

1. Word frequency and collocation analysis: single word frequency and bi-gram collocation of the target corpus are measured with the functions in [NLTK](#) package for an overview of the common word patterns relating to India in *Natural History*.
2. Topic modelling: [Genism](#) library is used for semantic vectorization and implementation of Latent Dirichlet Allocation (LDA) model for the topic modelling of the target corpus, and the library of [pyLDAvis](#) is utilized for an interactive visualisation. The output of this method shows the potential topics in the India-related text in *Natural History*.
3. Network analysis for named entities: Person names mentioned in the India-related text are retrieved from annotations on TOPOSText. Person and place names occurred in the target corpus are extracted as nodes together with their affiliated book numbers, and the co-occurrence between the nodes are counted as edges for network analysis. The output of this method is a graph showing the clusters of the nodes in the target corpus, indicating the structure of the content related to India in *Natural History*.

Interpretation and Conclusion:

The workflow and parameter setting of each research method are explained at the beginning of each analysis section. The results acquired from each method are interpreted with a dialogue to the broader literature and close reading of the related text.

In the Conclusion chapter (Section 5), the findings are illustrated comprehensively as a response to the research questions. And the limitations of the methods are discussed and evaluated.

3.2 Data preparation

This section provides an overview of the data preparation process, encompassing four sub-sections: HTML scraping from TOPOSText, creation of a filtered dataset of “India-related text,” completeness check and preprocessing of textual data. The tools and procedures employed in data collection and dataset generation for the study are introduced in the subsequent content.

3.2.1 HTML scraping from TOPOSText

As previously stated, the textual contents of Pliny’s *Natural History* are available on the [TOPOSText project](#), presented in two distinct parts: Book 1-11, Book 12-37. Both parts are provided in HTML format, offering separate sections of the complete work.

To extract the relevant data, the [Beautiful Soup](#) tool, a Python library renowned for parsing HTML and XML documents, was employed.

The text in the HTML documents is organised into paragraphs, each uniquely identified by an “id” attribute that specifies its corresponding book, chapter, and paragraph number. For instance, a typical paragraph has an “id” tag as follows:

```
<p id='urn:cts:latinLit:phi0978.phi001:3.9.7'>
```

The textual contents are retrieved with the information from these “id” attributes and structured as a reference dataset, comprising the plain text of *Natural History* divided into paragraphs, with each paragraph assigned a unique identifier, and separate columns indicating its affiliated book, chapter, and paragraph number. An illustrative example of the dataset’s structure is shown as Table 3.

Table 3: Example for the reference dataset containing the plain text in paragraphs of *Natural History*

UUID4	Reference	Book	Chapter	Paragraph	Text
0 e9e67565-bb...	urn:cts:lat...	1	1	1.0	PREFACE IN ...

UUID4	Reference	Book	Chapter	Paragraph	Text
1 010b853d-b8...	urn:cts:lat...	1	2	1.0	But who cou...
2 2d10e332-9c...	urn:cts:lat...	1	3	1.0	But if Luci...
3 113e0b4c-5b...	urn:cts:lat...	1	4	1.0	My own pres...
4 19115032-9f...	urn:cts:lat...	1	5	1.0	For my own ...

There are a total of 3493 paragraphs in the English-translated version of *Natural History* used in this study. The extracted text contains 343096 tokens and 28606 types after preprocessing. This reference dataset has been saved in CSV format for record.

Moreover, the geographical annotations concerning the ancient places mentioned in the text are labelled with a class attribute denoted as “place”, including a URI for the location, the place name and its corresponding coordinates, exemplified by the following HTML code snippet:

```
<a about="https://topostext.org/place/419125LPal" class="place" lat="41.8896"
long="12.4884">Palatine</a>
```

All annotations under the “place” class are extracted, and the URI, name, geographical coordinates, textual content and the book/chapter/paragraph number of each place are structured into the dataset for geographical-related text in *Natural History*. An example of this dataset is shown in Table 4 for reference.

Table 4: Example for the *Natural History* geographical-related text dataset

UUID4	ToposText_ID	Place_Name	Reference	Lat	Long	Book	Chapter	Paragraph	Text
0 bf12...	http...	Academy	urn:...	37.9920	23.7070	1	8	1.0	For ...
1 f782...	http...	Pala...	urn:...	41.8896	12.4884	2	5	1.0	For ...
2 a0f9...	http...	Esqu...	urn:...	41.8950	12.4960	2	5	1.0	For ...
3 b8d8...	http...	Capitol	urn:...	41.8933	12.4830	2	5	1.0	For ...
4 f81b...	http...	Rome	urn:...	41.8910	12.4860	2	6	3.0	Belo...

According to the geographical annotations of the ancient places occurred in *Natural History*, there are 5595 occurrences of place names in book 1-11 and 3281 in book 12-37, adding up to a combined total of 8876 annotated places throughout the work. The geographical-related text in *Natural History* contains 199507 tokens and 23937 types after preprocessing. This dataset including place names and their textual context in *Natural History* is saved in CSV format for record.

3.2.2 Filtered dataset of “India-related text”

As outlined in the Research Question chapter (Section 2), this thesis examines texts concerning the Indian region in Pliny’s *Natural History* as a case study. The objective is to explore how India is described, portrayed, and imagined within this extensive work, providing valuable insights into its complexity.

To ensure a comprehensive contextual analysis, the dataset creation considers not only instances where the word “India” is directly mentioned but also text related to the Indian region. This broader approach aims to encompass a wider scope of relevant information. Drawing from the research and mapping of the Indian region in the perception of the ancient Greek and Roman world, as explained and manifested in the *Barrington Atlas of the Greek and Roman World* (Talbert 2000a, 2000b), the approximate coordinates defining the target region are as follows²:

- Latitude: 5-35 degrees North
- Longitude: 65-95 degrees East

Utilizing the aforementioned dataset of geographical-related text in *Natural History*, the records with corresponding coordinates falling within the specified range are extracted to construct a dataset relevant to the discourse about the Indian region in the work. The filtering process ensures not only the text explicitly mentioning “India” but also those including other place names situated within the defined boundaries of the Indian region were retained.

The new dataset comprises the textual content and the geographical coordinates of the mentioned Indian place in *Natural History*. An example of the structure of the dataset of India-related text is shown as Table 5.

Table 5

UUID4	ToposText_ID	Place_Nam
85 fd43dc48-2410-4edd-8131-ce6cc6a3c809	https://topostext.org/place/300740RInd	India
92 30fe041c-3577-4375-a1f9-a7cc8c9212dc	https://topostext.org/place/300740RInd	India
93 0d786fbe-aed1-4f9a-9fbf-fa7c228a57cb	https://topostext.org/place/300740RInd	India
218 b28b62fc-8505-4fa9-aace-bd7b3df792c8	https://topostext.org/place/254683WInd	Indus
343 b010b6b5-0608-4c97-a980-171813b8b06a	https://topostext.org/place/300740RInd	India

There are 229 occurrences of paragraphs mentioning the places in Indian region with geographical annotation. And the distinct places mentioned are [‘India’ ‘Indus’ ‘Ganges’ ‘Acesinus’ ‘Hydaspes’ ‘Taprobane’ ‘Arachosia’ ‘Muziris’ ‘Baragaza’ ‘Ceylon’]. The tex-

²As indicated in the map-by-map directory, the range spans territories of “modern states of India (minus the Punjab), Bangladesh, Bhutan, Burma, Nepal, and Sri Lanka”.

tual content about the Indian region compiles 18029 tokens and 5384 types after pre-processing. The dataset for India-related text in *Natural History* are saved respectively in CSV format for further reference.

3.2.3 Data completeness check

The paragraphs extracted from the India-related text dataset undergo manual verification for the completeness of Indian place name annotations. Each distinct paragraph in the dataset is individually extracted and stored in TXT format as separate files within a corpus folder. The file names contain information about the affiliating book, chapter, and paragraph numbers.

There are in total 146 distinct paragraphs mentioning Indian places in *Natural History* according to the annotations on TOPOSText.

An example of the exported file name can be referred as follows:

Exported india_corpus\37.77.1_text.txt

The text files are uploaded to [Recogito](#) platform, which offers a semantic annotation tool and automatic geographical annotation suggestions from its supported gazetteers. This process aims to find Indian place names mentioned in the target corpus that were not annotated in TOPOSText. These unidentified place names are then marked on the [Recogito](#) workspace with the available geographical coordinates information as additional annotations. And the identified annotations are exported in CSV format for a supplement to the dataset of India-related text in *Natural History*.

As shown in Table 6, the supplement annotations are organized in the following manner:

FILE: This column contains the name of the file indicating the book, chapter, and paragraph number where the mentioned place name appears.

QUOTE_TRANSCRIPTION: This column contains the textual name of the place as mentioned in the text.

URI: The URI column contains the unique identifier for the place obtained from the gazetteers available on the [Recogito](#) platform.

VOCAB_LABEL: This column contains the confirmed automatically matched geographical name with the corresponding place name mentioned in the text.

LAT&LNG: The LAT and LNG columns contain the corresponding coordinates (latitude and longitude) associated with the place. Note that some places identified may not have matching coordinates.

PLACE_TYPE: This column contains the automatically matched geographical role provided by the gazetteers. It describes the type of the place.

VERIFICATION_STATUS: This column indicates whether the place names have been “verified” with confirmed coordinates that match the gazetteers’ information.

COMMENTS: This column includes manual remarks for the place names that do not have matching coordinates but are believed to indicate Indian place names based on the context.

Table 6: Example of place names

FILE	QUOTE_TRANSSCRIPTION	TYPE	URI	VOCAB
0 2.75.1_text.txt	hypasis	PLACE	http://pleiades.stoa.org/places/60110	Zadar
3 6.21.4_text.txt	sydrus	PLACE	http://pleiades.stoa.org/places/60110	Zadar
4 6.21.4_text.txt	rhodapha	PLACE	http://pleiades.stoa.org/places/60019	Rhodop
5 6.21.4_text.txt	palibothra	PLACE	http://pleiades.stoa.org/places/59978	Paliboth
6 6.21.5_text.txt	prinas	PLACE	http://pleiades.stoa.org/places/60008	Prinas (

PLACE_TYPE	
river	14
settlement	13
island	6
unknown	4
cape	3
mountain	2
people	2
lake	1
unlocated	1
unlocated,river	1
unlocated,settlement	1
Name: UUID, dtype: int64	

After the manual annotation check, 56 Indian place names were identified and added as supplementary annotations to the existing dataset, most of which are names of rivers, settlements, and islands. Among these, 45 place names have confirmed corresponding coordinates based on the reference in Recogito. For the other 11 place names, though have no matching coordinates on Recogito, there are contextual clues indicating that they are probably Indian location names.

The supplemented place name annotations were added to the India-related text dataset. The updated dataset contains 285 occurrences of paragraphs mentioning Indian places. And the distinct place names mentioned are ['India' 'Indus' 'Ganges' 'Acesinus' 'Hydaspes' 'Taprobane' 'Arachosia' 'Muziris' 'Baragaza' 'Ceylon' 'Hypasis'

‘Sydrus’ ‘Rhodapha’ ‘Palibothra’ ‘Prinas’ ‘Cainas’ ‘Condochates’ ‘Erannoboas’ ‘Cosogagus’ ‘Sonus’ ‘Protalis’ ‘Peucolaitis’ ‘Taxilla’ ‘Modogalinga’ ‘Andarae’ ‘Dardae’ ‘Methora’ ‘Chrysobora’ ‘Dandaguda’ ‘Tropina’ ‘Patala’ ‘Capitalia’ ‘Automula’ ‘Amenda’ ‘Cantaba’ ‘Prasiane’ ‘Argyre’ ‘Crocalia’ ‘Bibraga’ ‘Toralliba’ ‘Hippuros’ ‘Palaesimundus’ ‘Megisba’ ‘Palesimundus’ ‘Cydara’ ‘Coliacum’ ‘Emodian mountains’ ‘Capisa’ ‘Parabeste’ ‘Cartana’ ‘Tonberos’ ‘Arosapes’ ‘Gedrusi’ ‘Arbis’ ‘Sigerus’ ‘Catarcludi’ ‘Meros’ ‘Perimula’ ‘Chenab’ ‘Oratae’].

And the supplemented Indian place names are updated to the dataset of geographically-related text, which contains all place name occurrences in *Natural History*. The supplement expanded the dataset from 8876 occurrences of geographical names to 8932.

3.2.4 Preprocessing of texts

The textual contents stored in the “Text” column of the mentioned datasets are utilized as corpora for different analyses with three distinct scales: the text of the entire work, the text specifically related to geographical content, and the text related to Indian content. To prepare the data for analysis, a preprocessing process is applied using a defined function, which employs tools from the [NLTK](#) package.

During the preprocessing, the texts are tokenized, preserving punctuation marks, and lemmatized to their base forms. Furthermore, common English stopwords are excluded from the corpus, considering the text applied for this study is an English-translated version. To reduce the noise of short strings, tokens with lengths lower than two will not be appended to the output token list. The output of this preprocessing is a refined corpus presented as a nested list structure, with paragraphs forming the smallest nesting unit.

The extensions computed for each corpus mentioned earlier are based on this preprocessing procedure. The preprocessing steps help optimally organise the corpora, ensuring that they are conducive to meaningful analyses and facilitating the extraction of valuable insights from the text at varying scales.

4 Data Analysis

4.1 Place name distribution in India-related text

The comparison between the total number of place names and the place names specifically related to the Indian subcontinent mentioned in each book, is depicted in Figure 3. The difference in numbers between the two categories is significant, as indicated by the large disparity.

To facilitate a more effective comparison of the referencing trends across different books, Figure 4 presents subplots with varying y-axis scales. This approach allows for a clearer visualization of the trends and patterns in place name references throughout the various books.

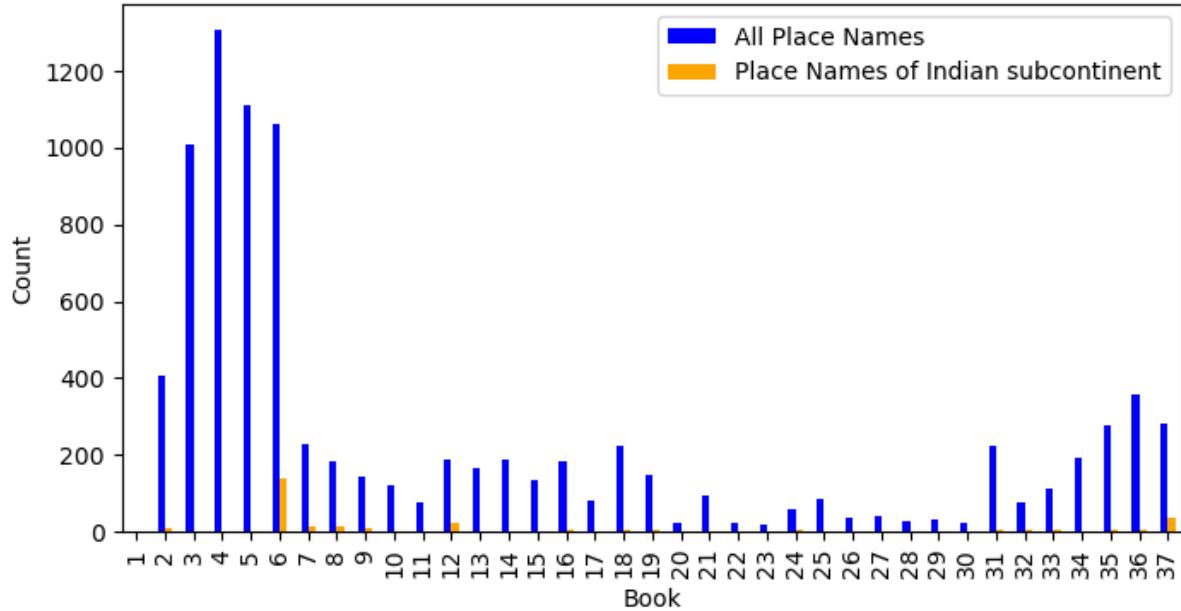


Figure 3: Occurrence count for all place names and place names of Indian subcontinent in each book

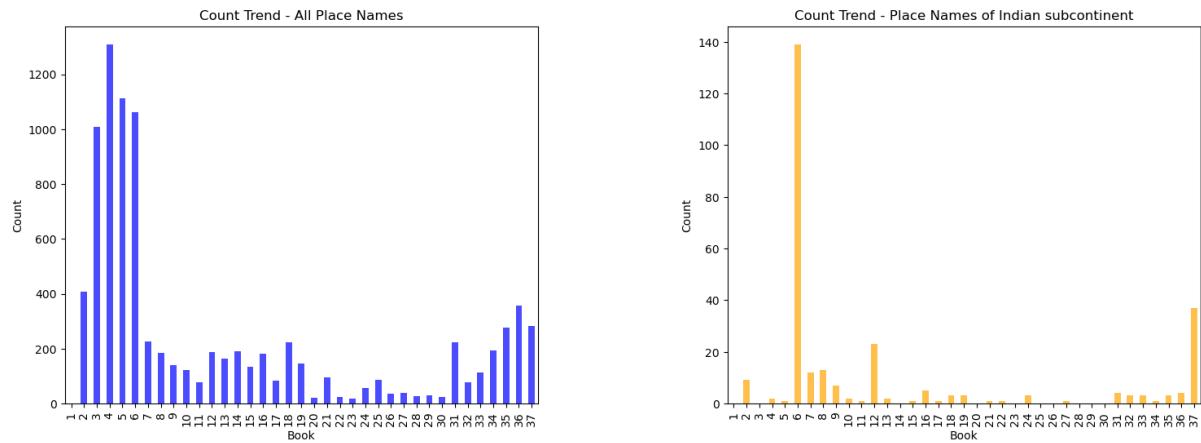


Figure 4: Occurrence count for all place names and place names of Indian subcontinent in each book_different y-axis scales

The figures reveal a distinct difference between the occurrence trends of place names related to the Indian subcontinent and all place names collectively. Specifically, the referencing of the Indian subcontinent is highly concentrated in books 6, 12, and 37 of Pliny's narrative. This discrepancy indicates that the mentioning of place names from the Indian subcontinent is closely tied to specific themes and topics within Pliny's work.

In this regard, three methodologies have been employed to analyze the texts pertaining to the Indian subcontinent in *Natural History*, including word frequency and collocation analysis, topic modeling, and network analysis. The objective of these analyses is to delve deeper into the textual content, unraveling the intricate relationships and uncovering the underlying themes and connections associated with the place names of the Indian subcontinent.

Through word frequency and collocation analysis, the aim is to identify keyword and significant word combinations co-occur in the textual content about India in *Natural History*. This analysis provides insights into the specific linguistic patterns and contextual associations surrounding the Indian places mentioned in the work, providing an overview of the keyword in the discourse.

Topic modeling allows for a broader exploration of the thematic landscape within which the Indian subcontinent place names are embedded. By clustering related words and identifying prevalent topics, this methodology helps to discern the major themes and subject matters that emerge from Pliny's narrative, providing a comprehensive understanding of the broader context in which these place names are mentioned.

Furthermore, network analysis offers a visual representation of the interconnections among the place names of the Indian subcontinent and other entities in Pliny's work. By examining the relationships between different locations and named entities, this analysis uncovers the geographical and conceptual networks that exist within the text, revealing how the Indian subcontinent place names contribute to the overall structure and narrative flow of *Natural History*.

Together, these methodologies aim to provide a nuanced and comprehensive exploration of the texts related to the Indian subcontinent in *Natural History*. By delving into the linguistic, thematic, and network aspects of these place names, a deeper understanding of their role in shaping Pliny's narrative can be achieved.

4.2 Word frequency and collocating bi-grams

By utilizing the measurements available in the [NLTK](#) package, a word frequency list and collocating bi-grams were generated from the text associated with Indian place names in *Natural History*. These outputs provide an overview of the prevalent words and word patterns, as potential keywords in the text.

In the initial observation, the words “India” and “one” ranked high in the frequency list. However, it is apparent that the passages would include the word “India” when discussing about India, making it less informative as a keyword. Likewise, the word “one” appeared as a generic descriptor for bringing up a type of tribe, plant, or attributes like distance, volume, or range, offering limited insight as a keyword. To enhance the

relevance and descriptive nature of the frequency list, these two common but less informative words, “India” and “one”, are further excluded from the token list.

Among 17729 tokens excluding “India” and “one”, 201 (the top 1%) frequently occurring words in the India-related text in *Natural History* are shown in Figure 5 and Figure 6.

		well	said	give	length	whose	large	gulf	wine	small	native	spot	vast	men	dry
	like	country	sun	according	given	still	greater	man	root	except	purpose	rock	different	head	grows
	name	arabia	red	long	land	egypt	away	rock-crystal	ethiopia	horse	would	night	take	yet	fish
	colour	among	region	although	thence	certain	gem	scent	held	sail	six	named	east	round	general
	hundred	another	may	nature	ganges	weight	sometimes	appearance	human	flow	woman	fruit	resembling	south	light
	called	king	another	body	bear	seen	resembles	fifty	rest	price	pepper	shadow	produced	set	promontory
	river	black	state	shore	le	wild	numerous	though	mentioned	famous	lake	greek	territory		
		many	number	spring	desert	case	since	grow	earth	ocean	world	far	look		
		water	whole	every	without	indeed	size	live	bird	tribe	thing	always	mouth	side	
		upon	thousand	find	plant	glass	purple	beyond	five	produce	lie	four	eye		
	stone	part	say	great	others	three	writer	way	fact	town	stated	near	indus		
		white	mountain	time	elephant	however	salt	gold	foot	distance	used	variety			
	also	known	come	day	tree	two	made	indian	nation	place					
		found	sea	island	people	even	city	mile	kind						

Figure 5: Top 1% frequent words in Indian subcontinent related text as tree map

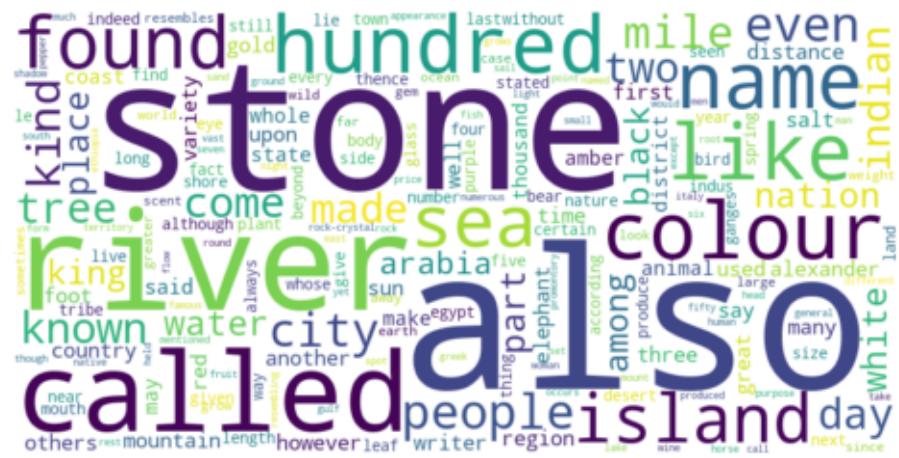


Figure 6: Top 1% frequent words in Indian subcontinent related text as word cloud

An intriguing observation from the word frequency sorting is the prominence of the word “also” in the given text. The word “also” appears frequently, which can be attributed to the encyclopaedic nature of the work, where it is often used to draw comparisons in introductions about species and natural phenomena.

As shown in the following examples where “also” appears in the India-related text in *Natural History*, it is indeed used when comparing the counterparts in India after introducing a natural phenomenon, plant or human activity. In this regard, the common use of “also” may imply that India holds significance as a contrast in the broader narrative.

2.75.1 “**Also** in India at the well-known port of Patala the sun rises.....”

12.10.1 “In India there is **also** a thorn the wood of which resembles ebony.....”

12.15.1 “There is **also** in India a grain resembling that of pepper, but larger and more brittle.....”

12.17.1 “Arabia **also** produces cane-sugar, but that grown in India is more esteemed.”

The hypothesis is further confirmed towards the end of the work in book 37, where Pliny concludes his comprehensive discourse on “Nature”. In 37.77.1, Pliny bestows the highest praise upon Italy, considering it to have earned “Nature’s crown”. And in this context, when expressing his preference and overall judgment, Pliny makes one final mention of “India”. He indicates that, “if we leave aside the fabulous marvels of India”³, Spain can be appreciated as a significant and attractive destination, second only to Italy. This unintentional highlight of India suggests that it holds considerable importance as a distant contrast to the Mediterranean area, where the focal point locates in the *Natural History*’s world scope.

And following by “also”, the words “stone”, “river”, “called” and “colour” notably stand out in the word frequency sorting. These frequent occurrences suggest some potential themes related to India in the content of *Natural History*, which aligns with the distribution of Indian place names as depicted in Figure 4.

Looking into the text in the three books pertaining the most mentions of Indian place names, book 6 includes specific topics on Nations of India, the Ganges and Indus (two main rivers in India) and routes of voyages to India, and book 12 contains introductions to trees and the economic values of their roots and leaves, as well as plants and their medical and flavouring effects. While book 37 focuses on descriptions of different types of gemstones, where the principle types are introduced in a sequence/category of colours, alongside critics about the luxury trade they represent.

The frequent use of the word “river” in India-related text may be related to the mention of voyage and trading routes concerning India in *Natural History*. On the other hand, “stone” and “color” clearly connect to the content in book 37, which deals with gemstones.

³*Natural History* 37.77.1 (<https://topostext.org/work/153>)

As indicated by Roller (2022), “Rivers, especially, were of significance to the Romans: Rome owed its origin to its control of a major crossing of the Tiber that affected travel throughout the Italian peninsula, and river cults were a major part of Roman ideology”.

These two potential themes observed from the frequent occurring words indicate that the geographical location and routes toward Indian subcontinent, and its role as an origin of many plants, animals and gemstones, processes a significance in the content about India in *Natural History*.

In addition to word frequency observation, collocation analysis is utilized to explore the common word patterns within the India-related text in *Natural History*.

The top 0.1% of the most likely collocating bi-grams are extracted using the likelihood ratio measurement. This selection process yields 18 out of 17728 bi-grams that are most significant and likely to co-occur together in the target corpus.

However, during the initial observation, it was noted that approximately one-third of the extracted bi-grams contained the word “hundred”, such as in ('hundred', 'fifty') and ('six', 'hundred'). These bi-grams typically denoted measurements for distance, object length, or quantity, offering limited descriptive information about the content of the text. Consequently, the word “hundred” was excluded from further bi-gram extraction to focus on more informative and relevant co-occurring words.

And the updated output bi-grams are listed as follows:

```
[('alexander', 'great'),
 ('father', 'liber'),
 ('caspian', 'gate'),
 ('fifty', 'mile'),
 ('denarii', 'pound'),
 ('gold', 'silver'),
 ('precious', 'stone'),
 ('river', 'indus'),
 ('fourteen', 'equinoctial'),
 ('olive', 'oil'),
 ('asia', 'minor'),
 ('equinoctial', 'hour'),
 ('lapis', 'lazuli'),
 ('red', 'sea'),
 ('mile', 'breadth'),
 ('emperor', 'nero'),
 ('already', 'mentioned'),
```

('ft.', 'long')]

The extracted bi-grams can be broadly categorized into four types:

Historical figures: ('alexander', 'great'), ('father', 'liber'), ('emperor', 'nero')

Geographical locations and features: ('caspian', 'gate'), ('river', 'indus'), ('asia', 'minor'), ('red', 'sea')

Meseuarments (distance, currency, length, time): ('fifty', 'mile'), ('denarii', 'pound'), ('fourteen', 'equinoctial'), ('equinoctial', 'hour'), ('mile', 'breadth'), ('ft.', 'long')

Trading goods: ('gold', 'silver'), ('precious', 'stone'), ('olive', 'oil'), ('lapis', 'lazuli')

On the one hand, the presence of bi-grams associated with geographical locations, distance, and time measurements in the India-related text reaffirms India's position as a geographic reference, consistent with the earlier findings from the word frequency list and literature review. On the other hand, within the context of the Indo-Mediterranean network of exchange, the occurrence of bi-grams related to geographical locations, currency measurements, and trading goods underscores the importance of India's role in merchandise trade within the narratives of *Natural History*.

Furthermore, the occurrence of historical figures such as "Alexander III, the Great (king of Macedon)", "Nero (Roman emperor)", and "Father Liber (referring to Dionysus, the Greek god of winemaking and wine)" suggests their connections with India in the history of expeditions or mythical tales (Dionysus is believed to have conquered India in Greek epic). This observation opens up a perspective for clustering the human names mentioned in the text to reveal the content structure about India in *Natural History*, which will be further explored in the Network Analysis section.

In conclusion, the analysis of word frequency and collocation in the India-related text within *Natural History* reveals noteworthy word patterns. These patterns suggest that India holds a significant role as a geographical contrast, being compared in terms of distance, natural phenomena, and origin of products with other regions introduced in the narrative. Furthermore, it is highlighted for its prominent role in merchandise trade in the portrait of India in *Natural History*.

4.3 Topic modeling

Furthermore, topic modeling approach is applied to delve further into the underlying topics about India in the work. Topic modeling is a widely used method for text analysis that infers the latent topics in a collection of documents(Bail n.d.; Underwood 2012). Latent Dirichlet Allocation (LDA), as its most commonly employed algorithm, operates under an assumption that each document contains a mixture of different topics, and

each topic is defined as a collection of words with varying probabilities of appearance in the passages(Underwood 2012; Kapadia 2022).

In this study, the collection of India-related text in *Natural History*, segmented into paragraphs, are considered as different “documents”. And the [Genism](#) library in Python is utilized for semantic vectorization and the implementation of the LDA model on the groups of words within the text⁴.

Since the corpus size for text pertaining to Indian place names is relatively small, after several attempts with a consideration of the coherence score distribution depicted in Figure 7, the number of topics is determined to be 3, with 40 passes to obtain the most optimal and non-overlapping topic clusters.

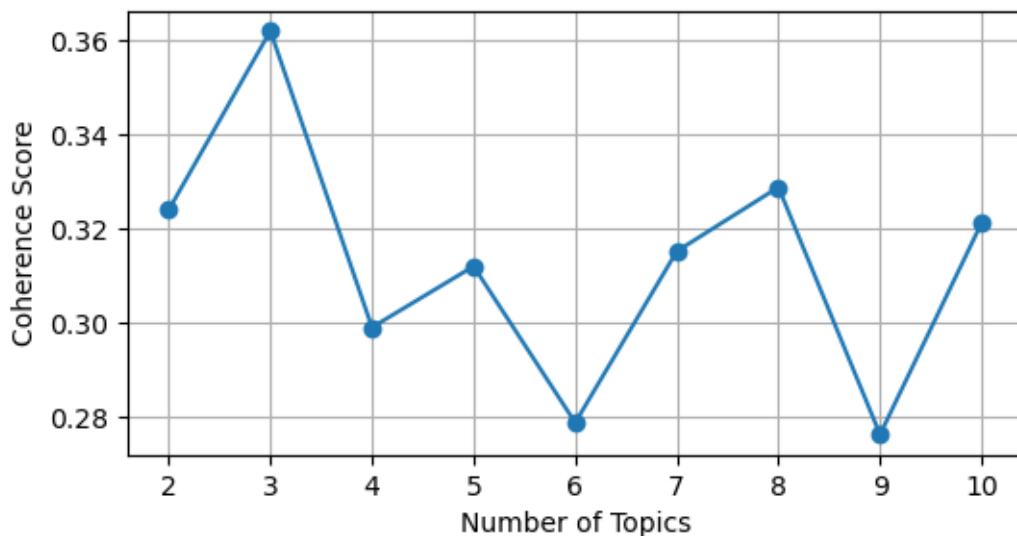


Figure 7: Conherence distribution of different topic numbers when passes=40

And the list of 30 keywords, grouped by the 3 assigned topics, is presented below.

```
[0,
'0.014*"stone" + 0.011*"also" + 0.007*"colour" + 0.007*"india" + '
'0.007*"found" + 0.006*"like" + 0.004*"one" + 0.004*"black" + 0.004*"tree" + '
'0.004*"amber" + 0.004*"name" + 0.003*"known" + 0.003*"white" + 0.003*"kind" +
'+ 0.003*"gold" + 0.003*"made" + 0.003*"even" + 0.003*"called" +
'0.003*"indian" + 0.003*"glass" + 0.002*"part" + 0.002*"people" +
'0.002*"used" + 0.002*"variety" + 0.002*"many" + 0.002*"river" +
'0.002*"make" + 0.002*"rock-crystal" + 0.002*"island" + 0.002*"red"""),
(1,
'0.009*"stone" + 0.008*"also" + 0.007*"india" + 0.006*"like" + 0.006*"kind" +
'+ 0.006*"colour" + 0.005*"name" + 0.005*"one" + 0.004*"called" +
'
```

⁴The code for LDA implementation was referred to the tutorial of Barber (n.d.)

```

'0.004*"even" + 0.004*"white" + 0.003*"pepper" + 0.003*"variety" + '
'0.003*"another" + 0.003*"known" + 0.003*"tree" + 0.003*"black" + '
'0.003*"weight" + 0.003*"leaf" + 0.003*"purple" + 0.002*"grain" + '
'0.002*"people" + 0.002*"denarii" + 0.002*"used" + 0.002*"nard" + '
'0.002*"resembles" + 0.002*"pound" + 0.002*"taste" + 0.002*"found" + '
'0.002*"small"'),
(2,
'0.011*"river" + 0.010*"hundred" + 0.008*"one" + 0.007*"india" + '
'0.007*"also" + 0.007*"city" + 0.007*"mile" + 0.007*"sea" + 0.006*"island" + '
'0.005*"nation" + 0.005*"called" + 0.005*"people" + 0.004*"day" + '
'0.004*"come" + 0.004*"two" + 0.004*"distance" + 0.004*"name" + '
'0.004*"place" + 0.004*"salt" + 0.004*"king" + 0.003*"elephant" + '
'0.003*"water" + 0.003*"country" + 0.003*"foot" + 0.003*"alexander" + '
'0.003*"thousand" + 0.003*"upon" + 0.003*"mountain" + 0.003*"even" + '
'0.003*"part")]

```

Based on the prominent categories of the keywords and the possible interconnection within each group, a interpretation of the topic emerged from each keyword clusters can be drawn as follows.

Group 0: This group consists of keywords related to precious stones, such as “stone”, “amber”, “rock-crystal”and “glass”, with colour references like “black”, “white”, and “red”. Therefore, the underlying topic appears to be the description of precious stones.

Group 1: This group comprises various natural products, such as “tree”, “leaf”, “pepper”, “grain” and “nard”, along with descriptive words like “weight”, “pound” and “taste”. In addition, the ancient Roman currency “denarii” appears in the group, suggesting a possible topic related to merchandise trade with Indian subcontinent.

Group 2: This group contains different geographical features, such as “river”, “island”, “sea” and “mountain”. It also includes terms related to cities, nations, kings, and distances, and the name “Alexander”, referring to Alexander the Great, who has undertook an expedition to Indian subcontinent. “Elephant”, as an important property of the King in India during Pliny’s era, also represents the power and size of the Indian kingdoms. In this regard, the underlying topic for this group likely pertains to geography and society in India.

The interactive visualisation of the 3 topic clusters, manifesting the intertopic distance map and the most salient/relevant terms within the given textand their contributing weights for each topic, can be accessed on the html version of this [thesis](#).

The static demonstration of the visualisation can be referred in Figure 8, Figure 9, Figure 10 and Figure 11.

The intertopic distance map is shown on the left panel of the interactive chart. Each bubble represents a topic, and the size of the bubble indicates the percentage of the texts in the corpus contributing to the topic. The distance between the bubbles implies the extent of difference between them. And a good topic model is expected to have big and non-overlapping bubbles scattered throughout the chart (Tran 2022).

And the most salient/relevant keyword is shown on the right panel. The blue bars represent the overall frequency of each word in the corpus. If no topic is selected, the blue bars of the most frequently used words will be displayed. The contribution of the frequent word to each topic will be shown in size difference when hovering. And when hovering on the bubbles in the left panel, there will be red bars in the right panel giving the estimated number of times a given term was generated by a given topic. The word with the longest red bar is estimated to be used the most in the texts belonging to that topic.

<IPython.core.display.HTML object>

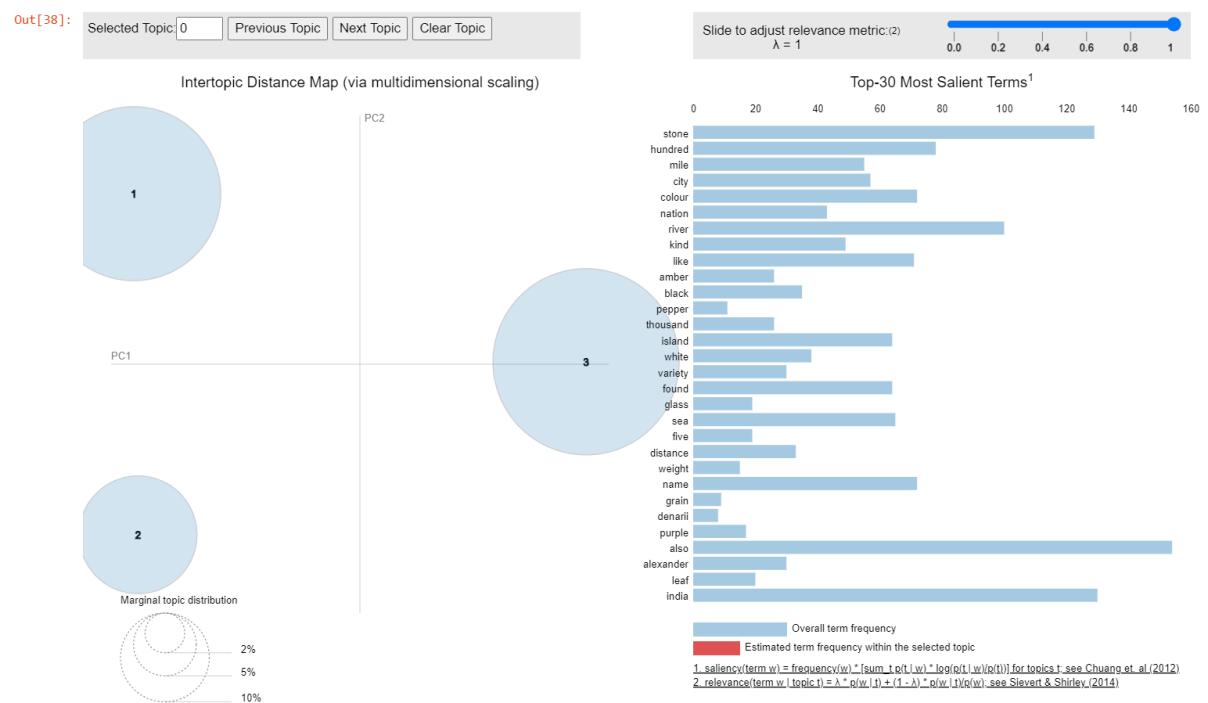


Figure 8: Topic cluster (overall)

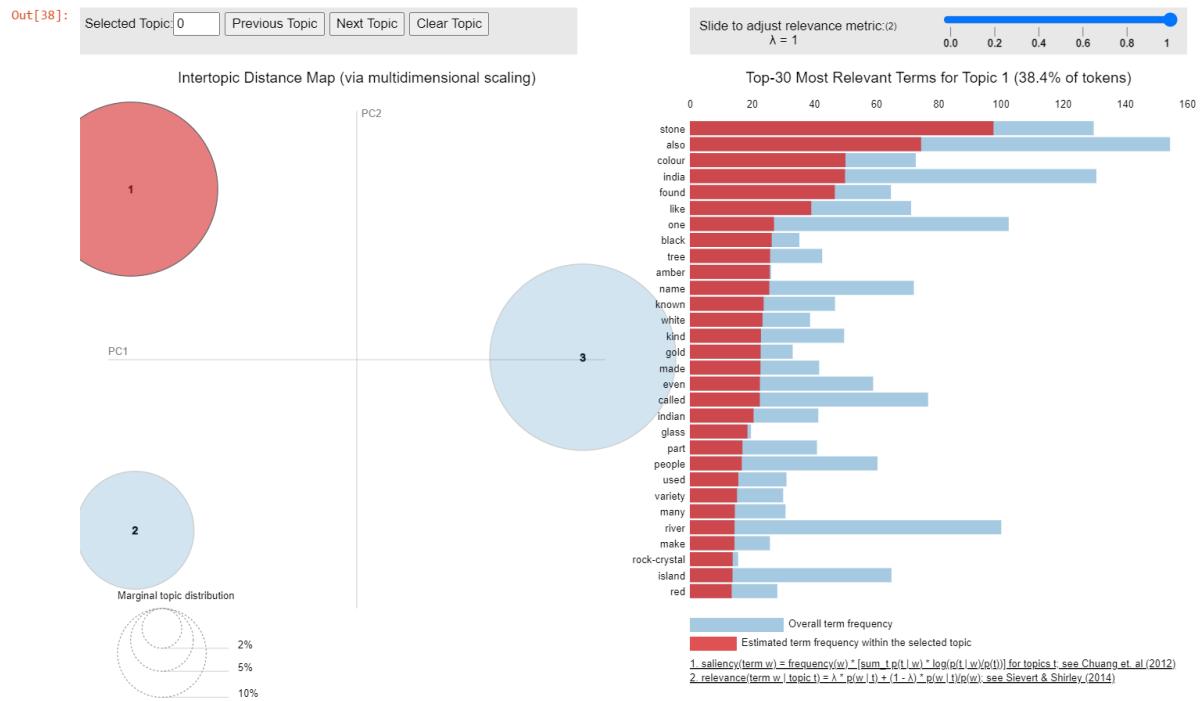


Figure 9: Topic cluster (highlighting on Topic_0)

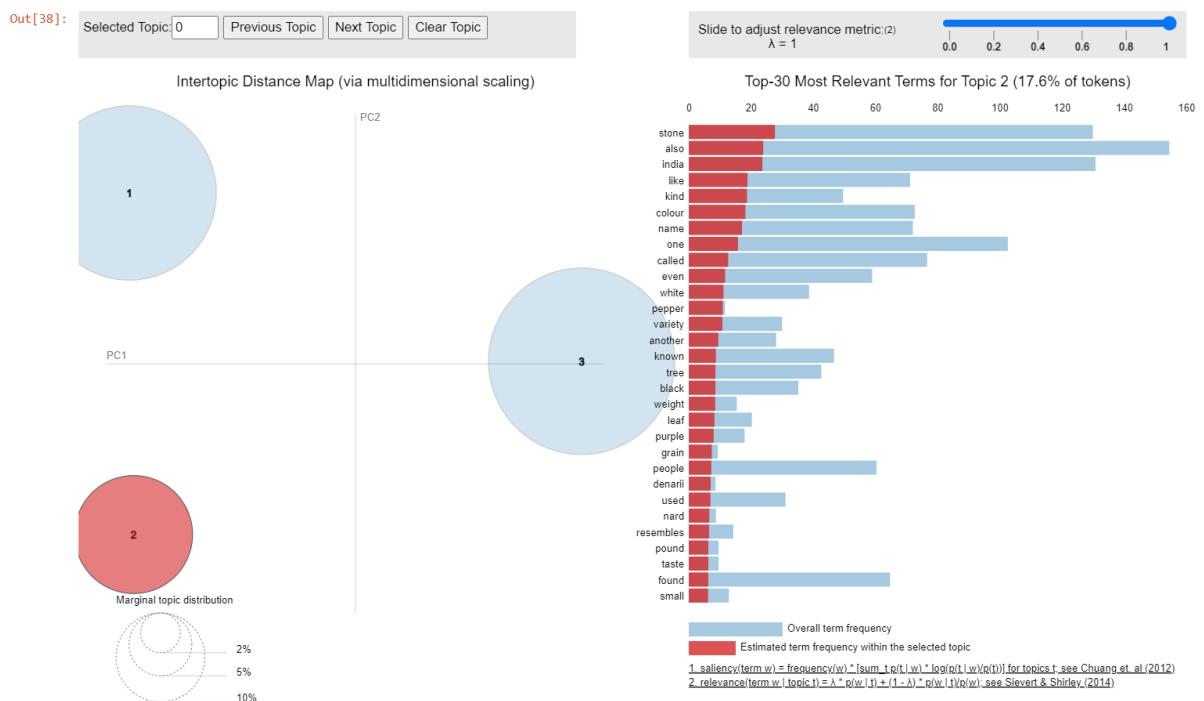


Figure 10: Topic cluster (highlighting on Topic_1)

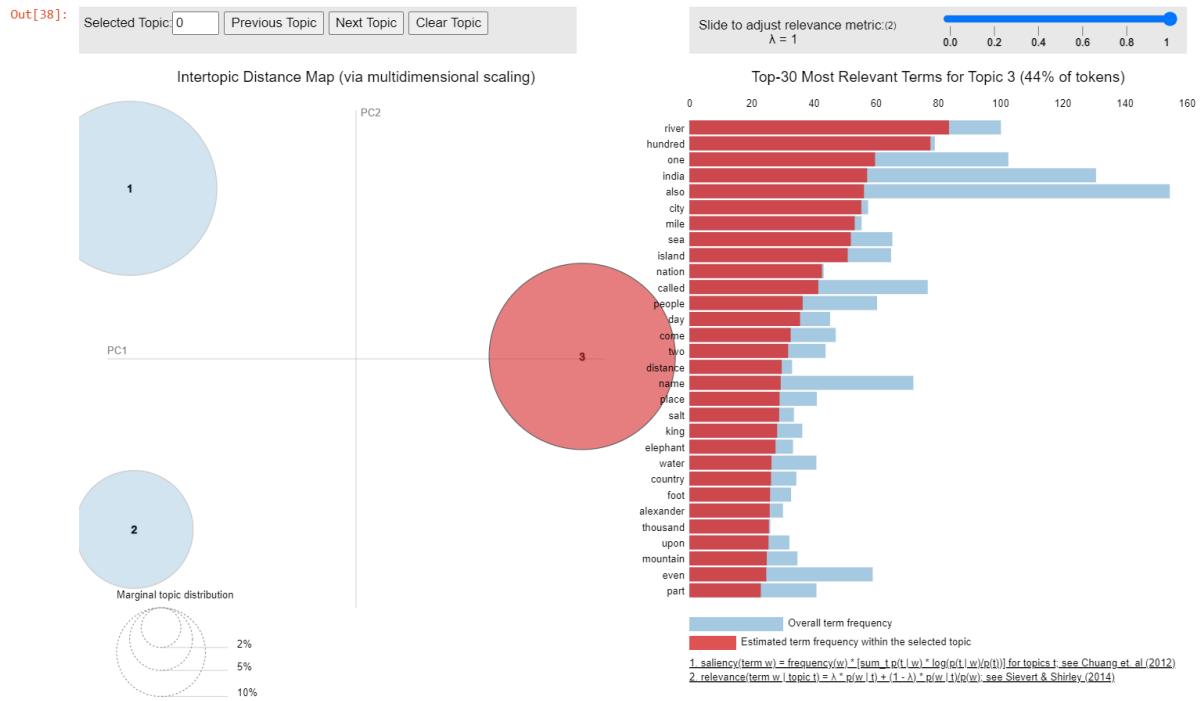


Figure 11: Topic cluster (highlighting on Topic_2)

In the above visualisation, the three keyword clusters are distanced from each other, indicating that they formed distinct potential topics within the given text. And the first (about stones) and the last (about Indian geography and society) topics took up a more significant portion comparing to the second topic (about merchandise trade).

Additionally, the specific context of the Indian places mentioned in *Natural History* is manually summarized and categorized into broader types to serve as a comparison and extension to the topics generated by the model. The initial comments about the context were drawn upon a close reading of the related text, with consideration of the book and chapter theme indicated in the text. The comments and summary are stored in CSV format and imported as pandas data frame.

And the summarized comments were further categorized into seven types, namely ['geographical reference', 'conquest history', 'passing mention', 'general introduction', 'criticism', 'prominent features', 'goods/animals/plants origin', 'producing activity', 'product/knowledge exchange'].

Geographical reference: refers to the occurrence of Indian place names as geographical reference in the narrative, for example:

2.112.1 "...from the river **Ganges** and its mouth where it flows into the Eastern Ocean, through **India** and Parthyene to the Syrian city of Myriandrus situated on the Issic gulf 5,215..."

Passing mention: refers to the condition that the names of Indian place included as a side note, for example:

5.11.1 "...Coptos, which from its proximity to the Nile, forms its nearest emporium for the merchandise of **India** and Arabia..."

Conquest history: refers to the mentions of Indian place names in the context of recalling the conquest history of Alexander the Great, for example:

8.61.2 "...When Alexander the Great was on his way to **India**, the king of Albania had presented him with one dog of unusually large size..."

General introduction: refers to the focused introduction about the general situation of places in Indian subcontinent, including transportation in the work, it is normally directly indicated at the beginning of the affiliated chapter. For example, the whole chapter 21 of book 6 has a leading topic as "The nations of India", the whole chapter 22 of book 6 has a leading topic as "The Ganges".

Prominent features: in *Natural History*, the plants and animals from India are often noted for their large size, and Pliny often highlight the good quality of Indian products in his discussion, these genre of context is categorized as "prominent features" of India. For example:

8.14.1 "...Megasthenes writes that in **India** snakes grow so large as to be able to swallow stags and bulls whole..." 37.21.1 "...**India**, likewise, is the sole producer of these stones and combining, as they do, the brilliant qualities of the most valuable gems, they above all others description..."

Goods/animals/plants origin: there are also many descriptions about India as the origin of different goods, animals and plants, this genre refers to those without concrete comments on their size or quality, just simply mentioned the object is originated in India, for example:

8.25.2 "Hyrcaenia and **India** produce the tiger, an animal of terrific speed..."

Producing activity: this genre refers to the cases that introduces about the human activity or specific producing process about natural creatures or trading products, such as:

8.8.1 "The method of capturing them in **India** is for a mahout riding one of the domesticated elephants..." 37.20.1 "...The **Indians** have found a way of counterfeiting various precious stones, and beryls in particular by staining rock-crystal."

Criticism: woven in the description about merchandise trade with Indian subcontinent,

Pliny had drawn direct criticism about the human greed and unnecessary interference on nature it represents, hence these narratives are specifically grouped together, for example:

12.14.2 “To think that its only pleasing quality is pungency and that we go all the way to India to get this! Who was the first person who was willing to try it on his viands, or in his greed for an appetite was not content merely to be hungry?”

22.56.1 “I myself shall not touch upon drugs imported from India and Arabia or from the outer world. Ingredients that grow so far away are unsatisfactory for remedies...Let them be bought if you like to make perfumes, unguents and luxuries, or even in the name of religion, for we worship the gods with frankincense and costmary. But health I shall prove to be independent of such drugs, if only to make luxury all the more ashamed of itself.”

33.2.1 “It came to be deemed the proof of wealth, the true glory of luxury, to possess something that might be absolutely destroyed in a moment. Nor was this enough: we drink out of a crowd of precious stones, and set our cups with emeralds, we take delight in holding India for the purpose of tippling, and gold is now a mere accessory.”

Product/knowledge exchange: in some circumstances, Pliny also mentioned about the knowledge and product exchange during the trade, such as:

34.48.3 “India possesses neither copper nor lead, and procures them in exchange for her precious stones and pearls.”

37.23.2 “...Indeed, as is generally known, in India the stone is exposed to view by the mountain streams...Later we persuaded the Indians to share our appreciation of it.”

And the distribution of context type concluded from close reading is shown in Figure 12 and Figure 13.

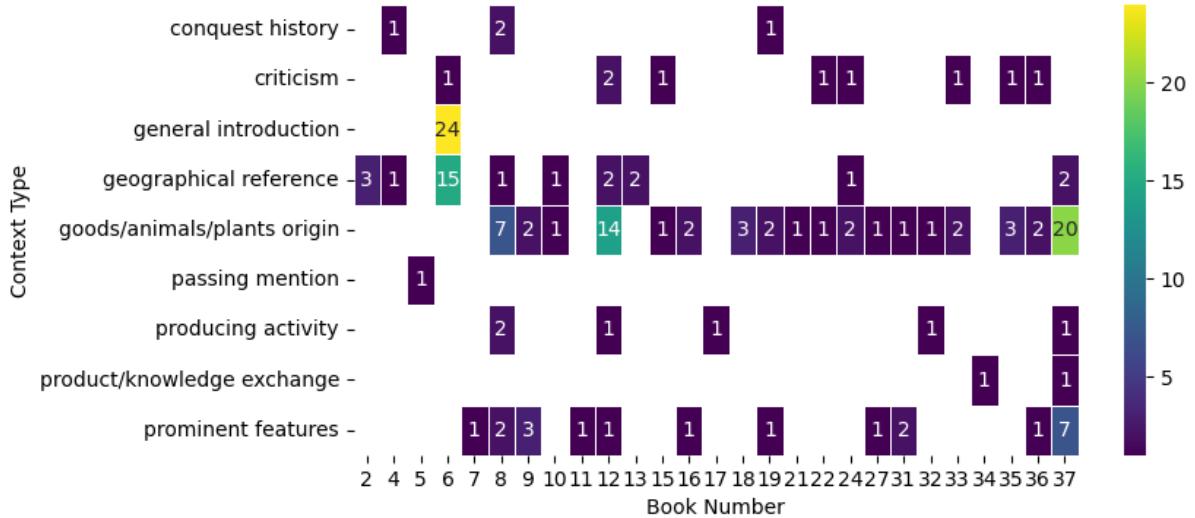


Figure 12: Context type distribution in books containing Indian place names

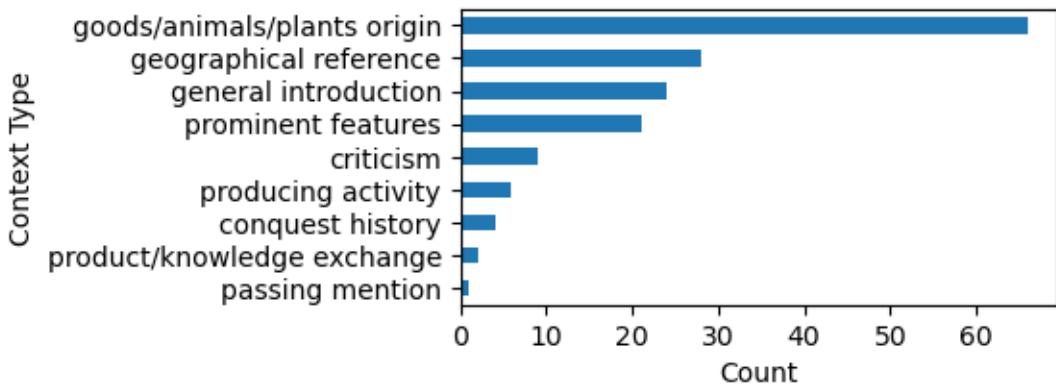


Figure 13: Occurrence frequency of different context type in India-related text

In general, the contexts concluded from the close reading can relate to the latent topics generated from the distant reading method and added more details to the keywords clusters. That is, if we understand the content about India in *Natural History* in three main topics, namely stones, trade, and contrasting nation, under the “stone” topic, except for describing the colour and texture of different types of gemstone, there also include the description about India as an origin of gemstones in a high quality. And under the large proportion of Indian places as geographical reference and focused introduction mentioned in the narrative, contributes to the formation of “nation” condition as a major topic. Additionally, there were conquest history of the Roman Empire to Indian subcontinent referred in the description. And for the topic about “trade”, though it seem inferior to the previous two in terms of quantity, but it is this part that reflects more sentimental judgement from Pliny out of his stoicism standpoint that worshiping nature as a devine and against luxurious desire (Beagon 1996).

On top of the observation from the previous section, that India is seen as an important

geographical contrast, as well as a significant trading partner in the narrative, a more nuanced portrait of India, linking the conquest story/history of the Roman Empire and criticism about human greed from the stoicism with abundant natural wonders and products of the best quality in the world can be drawn from the comprehension of topic modelling and close reading.

4.4 Network analysis for Named Entity

In addition to the previous methods partly answered on the question about “how India is described?” in *Natural History*, a network analysis of the named entities in the text may shed a light on the structure of the India-related content in the work. As inspired by the person name collocations observed from the collocation analysis, the initial idea for this section is to extract person names mentioned in the India-related text, and generate a entity cluster for the person names, place names, and book number, to explore how the content is structured in the context.

Employing the dataset with supplemented Indian place name annotations, distinct place names (including annotated places outside Indian subcontinent), person names and book numbers are considered as nodes, and the co-occurrences of person name and place names, as well as that of two different place names in the same paragraph, and the book number the person and place name occurred, are counted as edges.

To identify the impact of different entities in the discourse, the betweenness centrality of the nodes are measured and represented by the size of the node. And the weight of edge between two nodes represents the time the two entities appeared in the same context. Gone through a Force Atlas 2 layout algorithm, the graph also demonstrates the rough cluster of entities which tend to be mentioned together.

Indexing with the book-chapter-paragraph number of the India-related text, the place names outside Indian subcontinent mentioned in the same paragraph are also included for the network analysis. And according to the original TOPOSText and supplemented place name annotations, there are total of 908 occurrences of place names in paragraphs of Indian-related text in *Natural History*.

4.4.1 Person name annotation/tag retrieve

In order to extract the people names within the captioned text, approaches of retrieving from the TOPOSText annotation and from the tagging of text given by the pretrained multilingual named entity recognition model [WikiNEuRal](#) (Tedeschi et al. 2021) and [Flair](#) (Akbik, Blythe, and Vollgraf 2018) are compared for adopting the most completed output.

1. Annotation retrieving from TOPOSText

On TOPOSText, other than place names, a list of proper names (including people, gods, festivals, animals and artworks) are also annotated to the text of the classics. Each proper name has a unique URI as identifier in the html structure, such as: **Muses**.

Utilizing the tools of [Beautiful Soup](#), the textual proper names, their URLs along with the book-chapter-paragraph number they occurred in, can be retrieved with the URLs of both parts of *Natural History* on TOPOSText.

In the annotated text of *Natural History* on TOPOSText, there are 5109 such proper name annotations in book 1-11 and 7038 in book 12-37, adding up to a combined total of 12147 throughout the work.

The specific annotation type as “person name” can be further validated with the [api](#) on TOPOText. In the api file, all the entities of people/gods, has been valued as “male” or “female” as its “gender” key. Based on this criteria, those with other values in their “gender” key, such as “gender”:“animal” or “gender”:“other”, are not considered as “people” and filtered out from the retrieved output. Moreover, as there is a “concat” value indicating a detailed name of the signified person and shared with all variants of the same URI, the “concat” value is also fetched as a more accurate information of person name entity.

Table 7: Example of retrieved and validated person name annotations in *Naturl History*

FILE_ID	Person_name	ToposText_ID	Reference	Person_name_concat
0 1.1.1	Muses	/people/54	urn:cts:latinLit:phi0978.phi001:1.1.1	Muses (goddesses)
1 1.1.1	Catullus	/people/1881	urn:cts:latinLit:phi0978.phi001:1.1.1	Catullus
2 1.2.1	Cicero	/people/2300	urn:cts:latinLit:phi0978.phi001:1.2.1	Marcus Tullius Cicero
3 1.2.1	Cicero	/people/2300	urn:cts:latinLit:phi0978.phi001:1.2.1	Marcus Tullius Cicero
4 1.2.1	Manius	/people/382	urn:cts:latinLit:phi0978.phi001:1.2.1	Manius

An example of the data structure of the retrieved output of person name annotations in *Natural History* is shown in Table 7. There are total 6568 occurrences of person names annotated in the whole work, **330** of which are within the India-related text.

Table 8: Example for person name annotation merged into India-related text dataset

FILE_ID	Place_Name	Book	Text	Person_name_concat
0 2.75.1	Syene	2.0	Similarly it is reported that at the town of S... Onesicritus	
1 2.75.1	Syene	2.0	Similarly it is reported that at the town of S... Alexander III, the Great	
2 2.75.1	Syene	2.0	Similarly it is reported that at the town of S... Alexander III, the Great	

FILE_ID	Place_Name	Book	Text	Person_name_concat
3 2.75.1	Syene	2.0	Similarly it is reported that at the town of S...	Onesicritus
4 2.75.1	India	2.0	Similarly it is reported that at the town of S...	Onesicritus

As shown in Table 8, the validated annotated person names are added to the India-related text dataset. With the person name annotation retrieving, there found total **3583** co-occurrences of single place name and person name in the same paragraph within the India-related text in *Natural History*.

2. Named entity recognition with [WikiNEuRal](#)

To evaluate the quality and completeness of the extracted person name annotations, two widely practiced pretrained machine learning models for named entity recognition, [WikiNEuRal](#) (Tedeschi et al. 2021) and [Flair](#) (Akbik, Blythe, and Vollgraf 2018) are utilized for tagging person names in the India-related text.

Table 9: Example of person name tags retrieved with WikiNEuRal from India-related text in *Natural History*

FILE_ID	Text	Text_ner
0 2.75.1	Similarly it is reported that at the town of S...	[{'entity_group': 'LOC', 'score': 0.99636}...]
0 2.75.1	Similarly it is reported that at the town of S...	[{'entity_group': 'LOC', 'score': 0.99636}...]
0 2.75.1	Similarly it is reported that at the town of S...	[{'entity_group': 'LOC', 'score': 0.99636}...]
0 2.75.1	Similarly it is reported that at the town of S...	[{'entity_group': 'LOC', 'score': 0.99636}...]
21 2.112.1	Our own portion of the earth, which is my subj...	[{'entity_group': 'LOC', 'score': 0.74122}...]

An example of the person name tags retrieved from WikiNEuRal of the given text are shown in Table 9. There are total **225** occurrences retrieved within the India-related text.

And after merging the person name tags with the dataset of place name occurrences, there found total **1881** co-occurrences of single place name and person name in the same paragraph within the India-related text in *Natural History*.

3. Named entity recognition with [Flair](#)

As a comparison, another well-esteemed NLP pretrained model, [Flair](#) (Akbik, Blythe, and Vollgraf 2018) is also applied for the named entity recognition to the same text.

Table 10: Example of person name tags retrieved with Flair from India-related *Natural History*

FILE_ID	Text	Text_ner
0 2.75.1	Similarly it is reported that at the town of S...	[{"entity_group": "LOC", "score": 0.99}
0 2.75.1	Similarly it is reported that at the town of S...	[{"entity_group": "LOC", "score": 0.99}
0 2.75.1	Similarly it is reported that at the town of S...	[{"entity_group": "LOC", "score": 0.99}
0 2.75.1	Similarly it is reported that at the town of S...	[{"entity_group": "LOC", "score": 0.99}
66 4.17.4	Such is Macedonia, which was once the mistress...	[{"entity_group": "LOC", "score": 0.95}

An example of the person name tags retrieved from WikiNEuRal of the given text are shown in Table 10. There are total **102** occurrences retrieved within the India-related text.

And after merging the person name tags with the dataset of place name occurrences, there found total **938** co-occurrences of single place name and person name in the same paragraph within the India-related text in *Natural History* with the NER tool of Flair.

As shown in Table 11, the TOPOSText annotation provide more person name entities comparing to the other two methods.

Table 11: Quantity of person name entities retrieved with different methods

	Distinct person name	Person name occurrence	Place-person co-occurrence
TOPOSText Annotation	153	330	3583
WikiNEuRal NER	121	225	1881
Flair NER	71	102	938

And the first fifteen distinct retrieved person names of the three methods can be referred in Table 12. For the case of “Alexander III, the Great (general)”, since the annotation retrieved from TOPOSText is validated with the URI, both “Alexander” and “Alexander the Great” will be recognized as the same entity. While from the output of WikiNEuRal and Flair, the two variants are recognized as two different entities. And for the case “Leo” retrieved from Flair NER, it actually refers to the zodiac constellation, rather than a person/god name, and the same entity has a “gender”:“astronomic” value in the TOPOSText annotation, which will prevent it being added to the dataset as a person name.

Table 12: First 15 distinct person names retrieved with different methods

	TOPOSText Annotation	WikiNEuRal NER	Flair NER
0	Onesicritus	Onesicritus	Leo
1	Alexander III, the Great (general)	Alexander	Alexander
2	Heracles (s. of Zeus), Hercules	Artemidorus	Onesicritus
3	Artemidorus Ephesius	Isidore	Paulus Aemilius
4	Isidoros of Charax	Father Liber	Cyrus
5	Dionysus (s. of Zeus), Bacchus	Hercules	SCYTHIA
6	Paulus	Paulus Aemilius	Aramii
7	Aemilius	Amasis	M. Varro
8	Amasis	Alexander the Great	Pompey
9	Memnon (s. of Tithonus)	Antiochus	Laros
10	Osiris	Seleucus	Hecataeus
11	Calliope (d. of Zeus)	M	Patrocles
12	Antiochus, Antiochos (misc)	Varro	Alexander the Great
13	Seleucus I Nicator s. Antiochus	Amometus	Baeton
14	Margos	Hecataeus	Protalis

Both in terms of quantity and quality, the person names retrieved from the TOPOSText is more reliable for this study. Therefore, the output of TOPOSText annotation retrieving is appended to the dataset for node/edge generation for the network analysis.

4.4.2 Network graph generation

As mentioned, there will be three types of nodes for the expected network graph, namely **place name**, **person name** and **book number**.

And there are four types of edges taken into consideration, including the co-occurrence of:

1. **place name** and **person name** in the same paragraph
2. **person name** and **book number** it appears in
3. **place name** and **book number** it appears in
4. **place name** and **place name** in the same paragraph

There are total 519 nodes and 14177 edges for the network of place name/person name/book number in India-related text of *Natural History*.

The network graph generated from the captioned nodes and edges can be explored as Figure 14. The size of the nodes represents the betweenness centrality of the entity in the context, which means the bigger the node, the more central it is in the discourse.

The types of nodes (place name, person name, book number) are differentiated by colour.

And the curved lines linking different nodes indicate the co-occurrences between single place name and person name, two place names, and the book number of place name and person name occurs. The thickness of the line represent the counts of the co-occurrence edge, which means, the thicker the line linking, the more frequent the two nodes are mentioned together.

Apparently, “India” is the absolute center of the discourse. And drawing from both the node size and edge thickness, “Arabia” is the place that often mentioned together with India. Actually, the two countries are often mentioned in the same time, for example:

13.28.1 “Ethiopia, which is on the borders of Egypt, has virtually no remarkable trees except the wool-tree, like the one described among the trees of **India** and **Arabia**.”

22.56.1 “I myself shall not touch upon drugs imported from **India** and **Arabia** or from the outer world.”

One possible reason for this pairing occurrence is that Arabia locates in the middle of the route from Rome to India, and the two places similarly are significant importers for the exotic and luxurious goods trade.

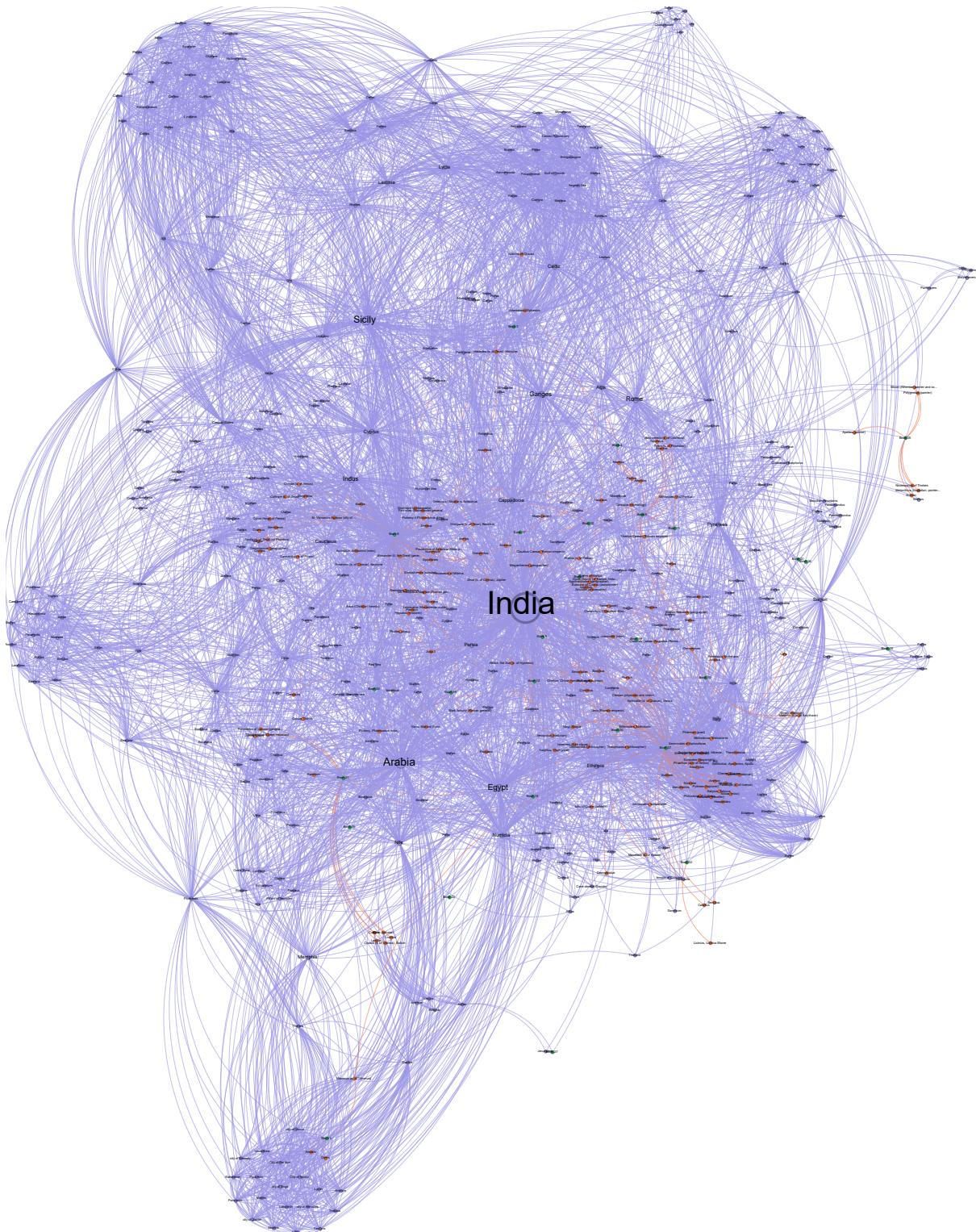


Figure 14: Place/person/book number network of India-related content in *Natural History*

Besides, there are several groups of place names located the edge of the graph, as shown in Figure 15, indicating they tend to be mentioned together, but relatively distanced from the discussion of India. In most cases India is just a passing mention in the enumeration context.

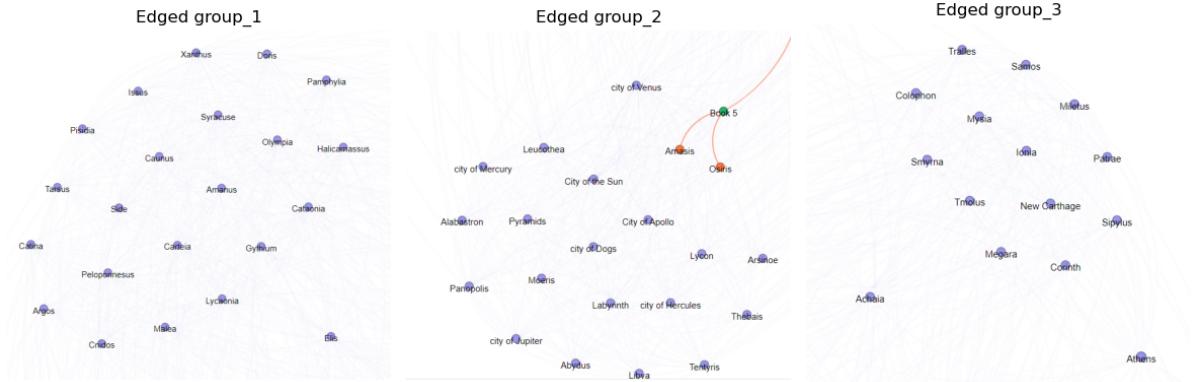


Figure 15: Example of edged groups in the network graph of India-related text in *Natural History*

And there are three obvious person name clusters centering on Book 6, Book 7 and Book 37, the zoomed in captures are shown in Figure 16 (for Book 6 and Book 7) and Figure 17 (for Book 37).

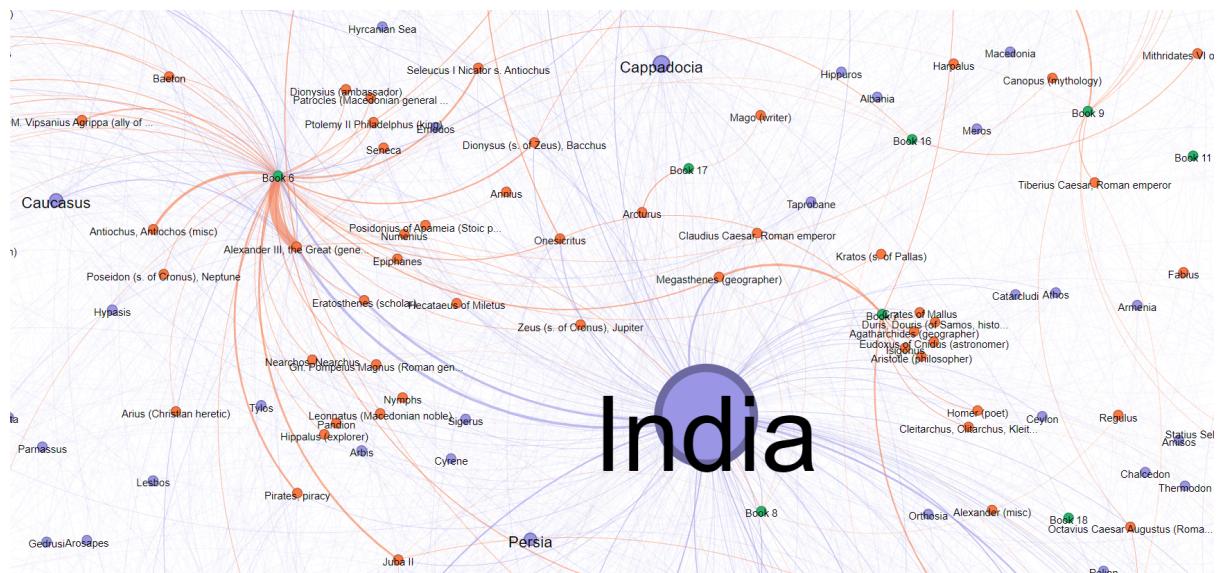


Figure 16: Clusters of Book 6 & Book 7 in India-related content in *Natural History*

From the cluster of Book 6, the most frequent mentioned historical figures are Alexander the Great and Dionysus (often with a variant name as “Father Liber”), which has been mentioned to be related to the epic and history of conquest from Roman Empire to Indian subcontinent. Also other governors/generals/navigators have significant connectness within the discourse, such as “Juba II”, “Ptolemy II Philadelphus”, “Hippalus”, “Patrocles”, “Leonnatus” etc. And it also show the frequent referred scholars of Pliny in the India-related narratives, for example, “Seneca”, “Eratosthenes”, “Posidonius of Apameia”.

And in the cluster of Book 7, where the discussion focus is on anthropology (human

races, tribes, human behaviors), a group of scholars are significantly referred, including “Aristotle”, “Eudoxus of Cnidus”, “Duris”, “Agatharchides”.

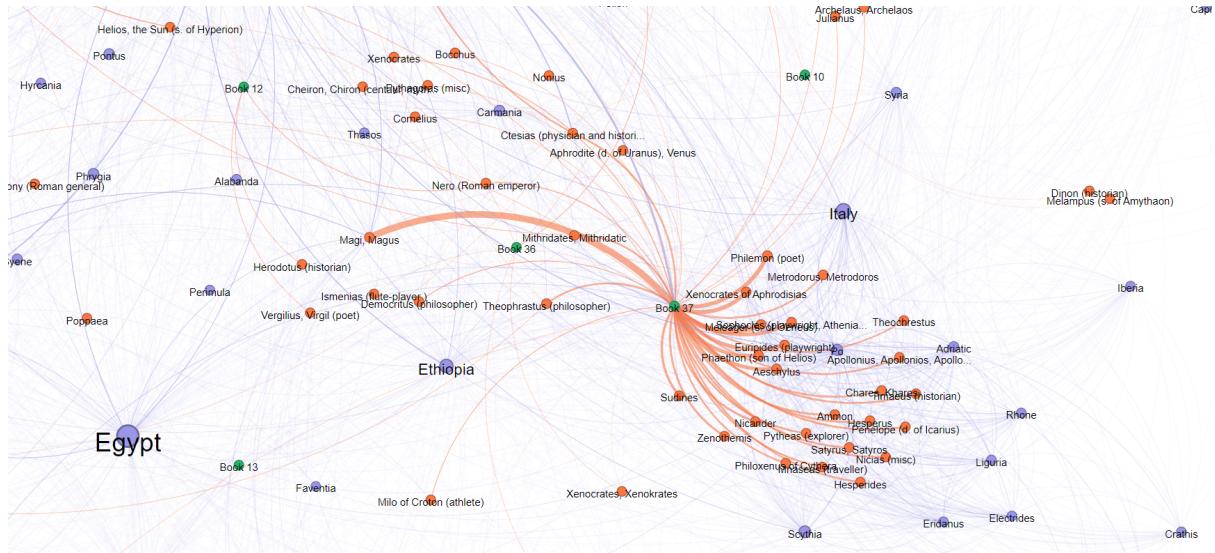


Figure 17: Clusters of Book 37 in India-related content in *Natural History*

While in the cluster of Book 37, strong edges are connected to Greek scholars, such as “Xenocrates”, “Aeschylus”, “Sudines” and “Zenothemis”. Also there is a significant connectness with “Magi”, referring to the priests in Zoroastrianism, a traditional Persian religion mentioned in the discussion about tales and myths related to the precious stones.

An interesting finding in close reading is that in Book 37, Pliny had addressed direct disapproval to these two groups, with a strong critical attitude. As quoted in the following paragraphs, the Greek scholars and Magus are often mentioned as debatable subjects in Book 37, as a strategy for him to express his own worship of nature and the material world.

37.11.2 “Here is an opportunity for exposing the falsehoods of the Greeks.”

37.14.1 “Now I shall discuss those kinds of gemstones that are acknowledged as such, beginning with the finest. And this shall not be my only aim, but to the greater profit of mankind I shall incidentally confute the abominable falsehoods of the Magi, since in very many of their statements about gems they have gone far beyond providing an alluring substitute for medical science into the realms of the supernatural.”

To sum up, the network graph visualizes the connectedness between different types of clustered entities and the central entities within the narrative. In this study, taking the relationship of the place names, person names and book numbers in the target text into

consideration, it shows clearly that India is the center of the narrative, and Arabia and Egypt are closely connected to this center. Also, it depicted how the person names of history figures, mythological gods, explorers and scholars are located in the narrative, shedding a light on the structure of the India-related content in *Natural History*.

If including more types of entities, such as animals, events, ethnic groups, into the network, it will contribute to a more nuanced distance map and entity clustering, which shows more dimensions of the content structure.

5 Conclusions

5.1 Comprehension of “India” in the narrative

Leading by the research questions of “How is India described, and how is the information about India structured in the work?”, the above analysis of different methods, including word frequency and collocation pattern observation, topic modeling and network analysis, integrating with close reading of the text, comprise a comprehension of “India” in the narrative of *Natural History*.

Generally, India is regarded as a significant geographical contrast to the Roman Empire and the Mediterranean. It is frequently mentioned for comparison in the description of landscape, expedition, navigating or trading routes and origins of natural creatures and elements of other locations. And the name of rivers in Indian subcontinent, including some tributaries, are given a highlight as references. Another important role of India in the narrative is a major trading partner who frequently import rare and precious treasures into the Mediterranean.

More specifically, the content about India is potentially themed in three topics, integrating with a close reading of the text, these three topics can be illustrated in details. The more prominent topics are the geographical and social condition about Indian nations and stones originated and produced in India. In the former topic, inspite of addressing India as a origin for numerous marvels, Pliny incidentally mentioned the epic tales or history related to the conquest of India, which suggest an imperial perspective of the work. And in the less prominent topic, which focused on the merchandise trade, Pliny expressed his criticism about the human greed and unnecessary interference on nature it reflects, which echos his Stoic claim of nature and human world.

And a network of place names, person names and book numbers of the India-related text has been created to depict the focus and co-occurrence clusters in the content structure. It shows that Arabia and Ethopia are the two places that close to the mentioning of India in the text, and for different books, there are specific history figures and referencing scholars clustering around. And an interesting observation is that in the person name cluster of Book 37, which discussed about the gemstones, Mangus and

Geek scholars played an important role as disproving subjects, for Pliny to express his advocate worshiping the genuine nature and material world as a Stoic.

In summary, in Pliny's unintentional portrait, India showed its complexity as a geographical contrast, an origin of marvels and treasures, a major trading partner, and a context of unnecessary luxurious pursuit of Pliny's contemporary that he debate against in the scope of *Natural History*.

5.2 Distant reading as a method

The attempt of utilizing several distant reading methods for a research of classical text in this study has, to a certain extent, achieved a response to the research question.

From pinning down the research question to conducting a meticulous analysis of the designated text, the application of distant reading methods yielded insightful perspectives and enriched information. However, as shown in the topic modeling section of Data analysis, an integration of close reading is essential in interpreting the outputs of distant reading.

Furthermore, when implementing these methods, data completeness check and validation are crucial steps to be taken before running the analysis. Since the sophisticated textual contents are simplified and restructured as data values in tablular format, it will make big difference for the outcome of missing or mistaken inputs.

Another benefit of adopting the distant reading approaches is the output dataset can be repurposed for subsequent investigations. Together with this thesis, the following dataset and outputs are available on the [GitHub repository](#) for potentially re-uses:

1. Dataset of Indian place names mentioned in *Natural Hisoty* with the coordinating textual passages (scraped from TOPOSText with supplemented mannually checked annotations)
2. Dataset of geographical names mentionned in *Natural History* with the coordinating textual passages (scraped from TOPOSText)
3. Corpus files of textual passages from the *Natural History* related to India
4. Network graph of place names, person names and book numbers about India-related contents in *Natural History*

Beyond that, the presentation of this thesis is practiced in an integration of Quarto with jupyter notebook. Embedding the YAML metadata into the notebook, it helped mekke a decent presentation both in HTML and PDF format. With rendering the output in HTML, the interactive maps and graphs can be published online.

5.3 Reflection and limitation

Though making the best effort to conduct a complete and consistent study for this thesis, there are still some limitation in the data preprocessing and parameter tuning sections that could be enhanced.

As introduced in the Data preparation section, the Indian place names scraped from TOPOSText is found incomplete and gone through a manual check. About 24% of the amount of original retrieved dataset were appended after the manual check. Though the supplemented Indian place annotation has been updated to the dataset compelling all geographical annotations, there might be missing identifications of places outside the specified coordinates range for India, which is better to also go through a manual check and supplement. However, as it requires more effort and time for the scale of the current study, the step has not been taken. It will make the dataset more complete and the context for this study more consistent if the same supplementing process of all place names could be done.

And for the topic modelling section, when deciding the topic number and passed for assigning the topics, though there is a validation process for topic number designation, and it aligns with the intuitive observation about different outputs, it is better to go through validation other parameters and general comparison for a more validating fine tuning.

On the other hand, with time limitation, it only added person name as an additional entity node for the network analysis. The current network graph shows the distance relationship between person names, place names and book numbers. It may give more dimensions for the whole content structure when including more types of entities, as animals, events, and ethnic groups, which are available on the TOPOSText annotations. In this regard, if the study will continue in the future, this aspect can be certainly enhanced.

And for the presentation of network graph, alternative platforms should be explored for a more interactive output. The optimal scheme is to make the graph easily zooming in to different clusters, and filtering with different types of nodes.

References

- Akbik, Alan, Duncan Blythe, and Roland Vollgraf. 2018. “Contextual String Embeddings for Sequence Labeling.” In *COLING 2018, 27th International Conference on Computational Linguistics*, 1638–49.
- Bail, Christopher A. n.d. “Topic Modeling.” Accessed August 3, 2023. https://cbail.github.io/textasdata/topic-modeling/rmarkdown/Topic_Modeling.html.
- Barber, Jordan. n.d. “Latent Dirichlet Allocation (LDA) with Python.” Accessed March 15, 2023. https://rstudio-pubs-static.s3.amazonaws.com/79360_850b2a69980c4488b1db95987a24867a.html.
- Beagon, Mary. 1996. “Nature and Views of Her Landscapes in Pliny the Elder.” In *Human Landscapes in Classical Antiquity*, 285–309. Routledge.
- . 2011. “Chapter Five. The Curious Eye Of The Elder Pliny.” In *Pliny the Elder: Themes and Contexts*, 71–88. Brill. https://brill.com/display/book/edcoll/9789004210073/Bej.9789004202344.i-248_006.xml.
- Fantoli, Margherita. 2022. “Statistics and Linguistics: Can We Tell Something More about Pliny the Elder?” <https://classics-at.chs.harvard.edu/statistics-and-linguistics-can-we-tell-something-more-about-pliny-the-elder/>.
- Healy, John F. 1999. *Pliny the Elder on Science and Technology*. Oxford: university press.
- Kapadia, Shashank. 2022. “Topic Modeling in Python: Latent Dirichlet Allocation (LDA).” *Medium*. <https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0>.
- Lao, Eugenia. 2016. “Taxonomic Organization in Pliny’s Natural History.” In *Greek and Roman Poetry, the Elder Pliny*, edited by Francis Cairns and Roy Gibson, 209–46. Papers of the Langford Latin Seminar 16. Prenton: Francis Cairns Publications.
- Murphy, Trevor. 2003. “11. Pliny’s Naturalis Historia: The Prodigal Text.” In, 301–22. BRILL. https://doi.org/10.1163/9789004217157_012.
- Naas, Valérie. 2002. *Le Projet Encyclopédique de Pline l’Ancien*. Collection de l’école Française de Rome 303. Rome: Ecole française de Rome.
- . 2011. “Chapter Four. Imperialism, Mirabilia, And Knowledge: Some Paradoxes In The Naturalis Historia.” In *Pliny the Elder: Themes and Contexts*, 57–70. Brill. https://brill.com/display/book/edcoll/9789004210073/Bej.9789004202344.i-248_005.xml.
- Neelis, J. 2011. “Chapter Three. Trade Networks In Ancient South Asia.” In *Early Buddhist Transmission and Trade Networks*, 183–228. Brill. https://brill.com/display/book/9789004194588/Bej.9789004181595.i-372_004.xml.
- Pinkster, Harm. 2005. “The Language of Pliny the Elder.” *Journal of Asthma - J ASTHMA* 129 (November): 239–56. <https://doi.org/10.5871/bacad/9780197263327.003.0011>.
- Pollard, Elizabeth Ann. 2009. “Pliny’s Natural History and the Flavian Templum Pacis: Botanical Imperialism in First-Century C. E. Rome.” *Journal of World History* 20 (3): 309–38. <https://www.jstor.org/stable/40542802>.

- Roller, D. W. 2022. "Introduction." In *A Guide to the Geography of Pliny the Elder*, 1–14. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108693660.003>.
- Rydberg-Cox, Jeff. 2021. "Modeling the Sources and Topics of Pliny's Natural History." *Umanistica Digitale*, no. 11: 217–29. <https://doi.org/10.6092/issn.2532-8816/12521>.
- Schultze, Clemence. 2011. "Chapter Ten. Encyclopaedic Exemplarity In Pliny The Elder." In *Pliny the Elder: Themes and Contexts*, 167–86. Brill. https://brill.com/display/book/edcoll/9789004210073/Bej.9789004202344.i-248_011.xml.
- Székely, Melinda. 2006. "Eastern Trade of the Roman Empire Based on Pliny the Elder's Natural History." *Chronica* 6 (January): 199–206. <https://www.proquest.com/docview/2379648941/citation/93A42D142D614235PQ/1>.
- Talbert, Richard J. A. 2000a. *Barrington Atlas of the Greek and Roman World: Map-by-Map Directory*. Princeton (N.J.): Princeton university press.
- . 2000b. *Barrington Atlas of the Greek and Roman World*. Princeton (N.J.): Princeton university press.
- Tedeschi, Simone, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. 2021. "WikiNEuRal: Combined Neural and Knowledge-Based Silver Data Creation for Multilingual NER." In, 25212533. Punta Cana, Dominican Republic: Association for Computational Linguistics. <https://aclanthology.org/2021.findings-emnlp.215>.
- Tran, Khuyen. 2022. "pyLDAvis: Topic Modelling Exploration Tool That Every NLP Data Scientist Should Know." <https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know>.
- Underwood, Ted. 2012. "Topic Modeling Made Just Simple Enough." *The Stone and the Shell*. <https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/>.