



The Journal of  
**HISTORICAL  
NETWORK  
RESEARCH**

CINDARELLA PETZ/RAJI GHAWI/  
JÜRGEN PFEFFER

# Tracking the Evolution of Communities in a Social Network of Intellectual Influences

*Journal of Historical Network Research 7 (2022) 114–154*

**Keywords** longitudinal analysis, social network analysis, history of intellectuals, dynamic networks, community detection, community evolution, big data

**Abstract** The history of intellectuals consists of a long-spanning entangled web of influences, interdependencies, and inspirations. In this paper, we construe the history of intellectuals by means of a formalized network approach, and structurally analyze how communities form and develop based on their intellectual influences. We are working with a unique data set of Linked Open Data, which we critically reflect upon. In this paper we tackle the question of whether community detection can help us to identify schools of thought, as well as patterns in the influence relations of scholars. We provide a detailed description of the process of extracting Linked Open Data, the construction of longitudinal networks, and the methodology of identifying and evaluating intellectual communities in the dataset using a community detection algorithm. Finally, we track the dynamic evolution of these communities over time, and characterize the structural patterns of their evolution, and the mechanisms of their development. We contextualize the changes in selected network structures in order to establish the merit of this method for a new perspective on the history of intellectuals, their influences, and their ideas.

## 1. Introduction

The history of intellectuals consists of a long-spanning and entangled web of influences, interdependencies, and inspirations. In this paper, we work with a unique dataset on the influence relations of intellectuals, and explore their communities' formation, development, and dissolution. The history of intellectuals encompasses an abundance of interdisciplinary research fields, including the history of scientific disciplines and methodologies, the history of ideas and of books, and the origins and anterior social contexts of intellectuals and intellectual thought (Wickberg 2001; Gordon 2013). Research on intellectual history mostly employs a regionally limited perspective within a closed timeframe in order to develop a comprehensive comparative analysis, but this approach lacks an inclusive, global perspective (Haakonssen and Whatmore 2017), as well as focusing on the “usual suspects” from a Eurocentrist perspective (Subrahmanyam 2017). Attempts to rectify this in a *Global Intellectual History*, by Moyn and Sartori (2013), were criticized as focusing on already well-researched intellectuals, despite their transnational approach (Subrahmanyam 2015).

A formalized network approach to the influences of and upon intellectuals allows us to re-frame historical research beyond Lovejoy's “unit-ideas” (1936) and to build on Skinner's “contextual history” of ideas (1969, which Armitage (2012) suggested includes a series of continuous contexts in the history of ideas and concepts), focusing on the personal relations and interdependencies of scholars. This follows the idea of philosophers such as Pierre Bourdieu, Karl Mannheim, or Erwin Panofsky on the relational situatedness of ideas and intellectuals in their historical and cultural context of the time in a social history of ideas (Ringer 1990, pp. 270–4), as well as on the “conditions and modalities of ‘knowledge production’” (Goldman 1994, p. 266). Armitage (2014) and Baring (2016) suggested a transnational focus in a relational intellectual history, which responds to the previously criticized focus on biographical studies (as voiced e.g., by Ringer 1990) limited to specific regions or timespans (Subrahmanyam 2017). This kind of research is facilitated with the methodologies and tools of the Digital Humanities, which Edelstein (2016) considered a “boon for intellectual history”.

In order to respond to the requirements of a global perspective on the history of intellectuals and harness the prospects of formal network analysis on the study of intellectual history, Ghawi et al. (2019) proposed to study this as a network from a global perspective, and identified, among other things, the most influential scholars in time as being those with the longest reaching influences (influence cascades). This analysis was extended in recent work, which introduced a longitudinal perspective on the most central scholars within each period (Petz et al. 2020). In social sciences, longitudinal network analysis is used on panel data to facilitate an understanding of the development of and changes in social structures and node characteristics over time by studying temporal snapshots of social networks (Hennig et al. 2012; Newcomb 1961; Huisman and Snijders 2003;

Snijders et al. 2010; Holme and Saramäki 2019). In historical studies, panel data are usually not available. In the study by Petz et al. (2020), this was solved by constructing temporal snapshots of networks by dividing the timespan of history into periods.

In related work, the epistolary exchanges of Early Modern scholars were mapped as spatial networks (compare, e.g., Edelstein et al. 2017, p. 237). More recent projects on the *Republic of Letters* have also incorporated a temporal perspective to these (Vugt 2017), which provide challenges to the modeling of the discrete and continuous time of fuzzy dating (Kudella 2019, p. 50).

Recently, community detection in historical network research gained traction in the analysis of conflict and coalition politics of Medieval sovereigns using the concept of Georg Simmel's Social Circles (Dahmen et al. 2017; Gramsch-Stehfest 2020), or the identification of communities in the transmission of medieval manuscripts with Gephi, for example by Fernández Riva (2019).

In this paper, we are interested in identifying the formation and evolution of intellectual communities in time. In order to study the patterns and mechanisms of intellectual community evolution over time, we test whether we can computationally identify trends in the history of intellectuals. Can we identify schools of thought? Can we identify hidden patterns in the influence relations of scholars? How do these structures change over time? The perspective on the history of intellectuals as being organized into network communities serves as a starting point for an analysis of the transformation and evolution of thought.

## 2. Data and Method

Our dataset is extracted from YAGO<sup>1</sup>, a large semantic knowledge base developed by the Max Plank Institute for Informatics in Saarbrücken. YAGO is one of the pioneering contributors to Linked Open Data (LOD) and was, alongside DBpedia (Bizer et al. 2009), one of the first to extract semantic knowledge at a large scale from Wikipedia (Suchanek et al. 2007). YAGO compiles information about millions of entities (such as people, cities, countries, and organizations): mining data from Wikipedia's<sup>2</sup> categories, redirects, and infoboxes, covering synsets or hyponymy from WordNet<sup>3</sup>, and matching spatial and topographical entities from GeoNames<sup>4</sup> (Mahdisoltani et al. 2015). This information was compiled with web scraping and text mining techniques, which were employed on Wikipedia's info-

---

1 Yet Another Great Ontology.

2 <https://www.wikipedia.org>.

3 <https://wordnet.princeton.edu>.

4 <https://www.geonames.org>.

boxes and categories, as well as natural language processing applied to, e.g., entity disambiguation and result filtering. In YAGO, the resulting data were merged with the DBpedia ontology<sup>5</sup> and the SUMO ontology<sup>6</sup>. The accuracy of YAGO's data has been manually evaluated to above 95%. We work with the YAGO3 version (released in 2015, which we extracted in 2018), which features over 10 million entities and over 120 million entries within their attributes.

In the following, we will first describe the extraction process of the YAGO influence relation and the pre-processing of the dataset, in which we enriched the dataset with temporal, spatial and disciplinary dimensions, as well as the process of constructing longitudinal networks. Subsequently, we will discuss the peculiarities and possible biases of the dataset presented.

### Mining a social network of intellectuals from YAGO

Most LOD sources, including YAGO, are typically represented using RDF (Resource Description Framework), which is the W3C<sup>7</sup> standard for representing information in the Semantic Web (Manola and Miller 2004). RDF is a data model, where each piece of information (called a statement or fact) is structured in a triple of the form:

(subject, predicate, object)

where subject and object are labeled as noted, connected by an edge labeled predicate. The standard query language for RDF is SPARQL (Prud'hommeaux and Seaborne 2008; Harris and Seaborne 2013), which became a W3C recommendation in 2008. As argued by Ghawi and Pfeffer (2019, 2020), Linked Open Data can be used as a source of information to extract social networks among entities, using various extraction patterns expressed in the SPARQL query language.

YAGO includes a predicate labeled `yago:influences`, which relates intellectuals based on their influence relationships, as recorded in Wikipedia's infoboxes. The accuracy of the `yago:influences` relation was evaluated with a confidence score of 0.96 by YAGO. We are particularly interested in this predicate to extract an influence social network among intellectuals. Table 1 shows a sample of RDF triples from YAGO, depicting the influence relation among several intellectuals. To extract our target influence social network, we used a SPARQL query as shown in Figure 1. The query has been executed over YAGO's SPARQL endpoint.<sup>8</sup>

---

5 <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/linking/>.

6 <http://www.mpi-inf.mpg.de/~gdemelo/yagosumo/>.

7 World Wide Web Consortium, <https://www.w3.org/>.

8 <https://linkeddata1.calcul.u-psud.fr/sparql>, as of July 2019.

|                    |                 |                    |
|--------------------|-----------------|--------------------|
| Ibn_Tufail         | yago:influences | Christiaan_Huygens |
| Ibn_Tufail         | yago:influences | Immanuel_Kant      |
| Ibn_Tufail         | yago:influences | Isaac_Newton       |
| René_Descartes     | yago:influences | Christiaan_Huygens |
| René_Descartes     | yago:influences | Immanuel_Kant      |
| René_Descartes     | yago:influences | Isaac_Newton       |
| Johannes_Kepler    | yago:influences | Isaac_Newton       |
| Maimonides         | yago:influences | Isaac_Newton       |
| Christiaan_Huygens | yago:influences | Isaac_Newton       |
| Francis_Bacon      | yago:influences | Isaac_Newton       |
| Isaac_Newton       | yago:influences | Abraham_de_Moivre  |
| Isaac_Newton       | yago:influences | Immanuel_Kant      |
| Isaac_Newton       | yago:influences | Voltaire           |
| Baruch_Spinoza     | yago:influences | Immanuel_Kant      |

**Tab. 1** Examples of influence relations

This query returns all pairs  $(u, v)$  of entities (scholars in our case) that are connected via the `yago:influences` relation, or in other words, if entity  $u$  influences entity  $v$ . We then use the result of this query as the basic edge-list for the influence network of intellectuals, in which nodes constitute the raw data base of intellectuals available. The raw dataset comprises 12,705 nodes and 22,818 edges, as reported in Ghawi et al. (2019).

### Intellectuals and their influences in YAGO

As the influence relations in YAGO originate from Wikipedia, any findings of this study that use a dataset extracted from YAGO3 perforce reflect the knowledge hosted there. There are several important points to reflect on the type of

```
SELECT ?u ?v
WHERE {
  ?u yago:influences ?v.
}
```

**Fig. 1** SPARQL query used to extract the influence social network

intellectuals and influence recorded. Intellectual is a broad category in YAGO, following its historical dimension that there is no “single definition of the intellectual’s condition that applies universally” as Ringer (1990, p. 281) noted, we might also add, with Wickberg (2001, p. 387), that a generalization of intellectuals as a social type would not be historically correct. YAGO’s intellectuals entail philosophers, writers, and scholars of the natural sciences as well as artists, mathematicians, physicians, polymaths, musicians, and more, among which are an illustrious list of polar explorers. These intellectuals appear in our dataset if there was a known influence from and to other intellectuals; the influences recorded are based on their main influences, and are therefore not exhaustive.<sup>9</sup> The information on the included intellectuals and their main influences originate from Wikipedia, and as such represent a crowd-sourced and semi-popular source of knowledge on intellectuals in history. Wikipedia is an online encyclopedia that provides the “primary source of knowledge for a huge number of people around the world” (Anderka 2013, p. 12) and an “authoritative source of information” also for scientific scholarship (Murgu and Ivings 2019, p. 12),<sup>10</sup> with information reliability ensured by consistent “major peer review activity” (Viseur 2014, p. 3). Due to the professionalization of Wikipedia in the last decade<sup>11</sup>, this collection of scholars closely represents the current state of research<sup>12</sup> and encompasses what can be considered the main intellectuals in history (though the list is not exhaustive).

The intellectuals recorded in YAGO can be considered biased, as they are focused on major figures, and more specifically on men.<sup>13</sup> While scholars like pi-

---

9 For example, the Medieval writer Bernardus Silvestris (1085–c. 1160) is not included in the dataset, whose allegorical philosophical work on the birth of the universe (“Cosmographia”) heavily influenced the “father of English poetry” Geoffrey Chaucer (c. 1340–1400). The latter is included in the dataset, but his main influence is recorded as Ovid.

10 Compare also to Thompson and Hanley’s (2018, p. 1) estimation that “Wikipedia [is] influencing roughly one in every three hundred words in related scientific journal articles”.

11 This entails regular proofreading, peer-reviewing, and facilitated reversals of vandalized articles using the MediaWiki software (Anderka 2013, p. 9).

12 The accuracy of Wikipedia has been examined regularly, such as in a blind comparison of various online encyclopedias including the *Encyclopedia Britannica*. While in 2005, a 30% error difference was reported by e.g. Giles (2005), by 2012 Wikipedia has been evaluated to show “significantly higher [...] accuracy, references, and overall judgment” in comparison to other online encyclopedias (Casebourne et al. 2012, p. 32). Studies on the up-to-dateness of information in Wikipedia, such as by Kousha and Thelwall (2016), showed that roughly 5% of Wikipedia’s references directly cite scientific scholarship. Teplitskiy et al. (2017) noted the importance of Open Access publications in this context (in contrast to publications restricted by paywalls) in order to amplify the diffusion of current scientific insights.

13 WikiProject *Women in Red* by Roger Bamkin (Wikipedia 2020b) and *Project Vox* at Duke University Libraries (2020) are initiatives designed to raise awareness about the “gender gap” in Wikipedia: the absence of female scientist’s entries, and their higher probabil-

oneering psychologist Leta Stetter Hollington (1886–1939) are missing from our dataset<sup>14</sup>, the philosopher Émilie Du Châtelet (1706–1749) is included. Of seven female philosophers from the Early Modern Period highlighted by *Project Vox* (Duke University Libraries 2020), three appear in our dataset.<sup>15</sup> To conclude, the dataset offers an *abstract* form of the history of intellectuals, which records the most important influences of the most important intellectuals, closely reflecting the current state of research as added by the crowd-sourced Linked Open Data community – similar to the broad strokes of a *Meistererzählung* (master narrative). Any findings necessarily iterate the representation on YAGO, and thus on Wikipedia.

### Adding a temporal dimension

We expanded the dataset with the birth and death dates of each scholar in order to incorporate a temporal dimension to the analysis.<sup>16</sup> We used the SPARQL query shown in Figure 2, where the predicates `wasBornOnDate` and `diedOnDate` were used to retrieve birth and death dates for the scholars in our dataset. Since a scholar could be an influencer or be influenced, their entity could appear in the subject or object positions of the triple pattern. Therefore, the query contained a combination of both patterns using a UNION operator. Since the data set may not have information about the birth date or the death date (or both) of some scholars, the triple patterns to retrieve those dates are stated as optional.

The results of this query are as follows:

- 8,073 entities have both dates (119 of these had errors: death dates before birth dates, which had to be manually corrected).
- 4,030 entities have a birth date, but no death date.
- 82 entities have a death date, but no birth date.
- 520 entities have neither dates.

Some entities had no birth or death dates recorded; we corrected such missing information by schematically adding/subtracting 60 years from the birth/death date<sup>17</sup> up to the symbolic year of 2020, in order to get a broad estimation of their

---

ity to be deleted (Krämer 2019). In 2016, 16.72% of English entries in Wikipedia were about women (Stephenson-Goodknight 2016); by 2019, this number was raised to 18% (Krämer 2019). These more recent developments are not included in YAGO3, which was created in 2015.

14 Whose Wikipedia page was introduced later than the YAGO3 database from 2015.

15 These are Mary Astell, Du Châtelet, and Anne Conway.

16 Compare the following data preparation and cleaning procedures to Petz et al. (2020).

17 This process is then followed by another data verification, when introducing a periodization into which the scholars are mapped, as described in the following parts.

```

SELECT ?u ?birthDate ?deathDate
WHERE {
  { SELECT DISTINCT ?u WHERE {
    { ?u yago:influences ?v. }
  UNION
    { ?v yago:influences ?u. }
  } }
  OPTIONAL { ?u yago:wasBornOnDate ?birthDate.}
  OPTIONAL { ?u yago:diedOnDate ?deathDate.}
}

```

**Fig. 2** SPARQL query to extract birth and death dates

lifetime and later periodization. When both dates were missing, we verified them manually. In the course of the data processing, we removed entities that were either conceptual actors, legendary figures, or groups, e.g., the “Megarian school” or “Gilgamesh”. The interim dataset consisted of 12,577 scholars with complete birth and death dates.

### Mapping scholars into a periodization of time

In order to derive a longitudinal perspective from a static network, we compartmentalized the timespan of history manually into consecutive periods (or: eras), into which we embedded the scholars. From this perspective, influences on the micro-level can be studied as influences on macro-level among periods of history. We used the global periodization introduced in Petz et al. (2020) to map scholars into eras that inferred five consecutive eras based on Osterhammel’s global periodization (2006), as seen in Table 2. We decided on a global perspective for the periodization in order to cater to the internationality of intellectual networks and their heterogeneous origins, and to satisfy the criticized lack of international

| Abbrv. | Era                  | Start | End  |
|--------|----------------------|-------|------|
| AN     | Antiquity            |       | 600  |
| MA     | Middle Ages          | 600   | 1350 |
| EM     | Early Modern Period  | 1350  | 1760 |
| TP     | Transitioning Period | 1760  | 1870 |
| MA     | Modern Age           | 1870  | 1945 |

**Tab. 2** Overview on Eras



outlook in intellectual history.<sup>18</sup> We left out the Contemporary period starting in 1945, as for this study we focused on the periods up to the Contemporary age. Of course, any periodization is a construction to facilitate research, and as such depends on the specific caesura for the respective research field (Pot 1999, p. 63; Osterhammel 2006, pp. 50–1).

Every intellectual was then assigned to an unambiguous period. In the process of doing so, we corrected outliers that resulted from different dating (e.g., dates recorded in YAGO in the Hijri calendar instead of the Gregorian, or missing negative signs for BC). We rectified some outliers of impossible influence relations from a later period to an earlier one, which resulted from wrongly switched influence relations and/or when the lifespans of a scholar influencing another were drastically different, thus eliciting chronologically reverse links of eras. Finally, we mapped ambiguous period membership of scholars, who fit more than one era, into a single essential period. The approach of a single period membership avoids redundancy, and offers a more intuitive perspective on the longitudinal structure of the networks in order to grasp macro changes in their influence relations.

### **Adding a spatial dimension, and disciplines**

As we are interested in identifying schools of thought, we manually established the geographic domain of agency for each intellectual in the dataset (compare to Table 3), and surveyed their main discipline as recorded in Wikipedia, which we structured into 14 container categories<sup>19</sup> (see Table 4). These disciplines encompass the main area of work of a scholar.

For the survey of these attributes, we employed a human annotation process, which involved dividing the dataset into ten chunks and manually classifying the main discipline and geo-location for each intellectual. These annotations were then manually verified. We found a further 32 entities to be either non-intellectual inspirations<sup>20</sup> or groups which members already existed in the dataset, which we then removed. The final cleaned dataset consists of 5,287 intellectuals in the network, with 7,803 influence relations. Table 5 shows a snippet of our final dataset of scholars, where each scholar is associated with their birth and death dates, era, geo-category, and main discipline.

---

18 Compare this also to the discussion.

19 These container categories are sometimes anachronistic in nature, e.g., “social studies” as a field developed only in modern times, but a term we nonetheless used to group historians, anthropologists, and social scientists. Also while “writer” as a category could be used for each intellectual in the dataset, we only grouped those who worked as poets, journalists, or essayists, and who did not work more prominently in other fields.

20 Such as the sailor Owen Chase, whose biography inspired the story of *Moby Dick* by Hermann Melville.

| Abbrev. | Geo-Category              | Notes                                                                           |
|---------|---------------------------|---------------------------------------------------------------------------------|
| GR      | Ancient Greeks and Romans |                                                                                 |
| EU      | Europe                    |                                                                                 |
| AR      | Arab world                | including Near and Middle East, "Al-Andalus", Ottoman Empire, and Modern Turkey |
| AS      | Asia                      | e.g., India, China, and Japan                                                   |
| AM      | North America             |                                                                                 |
| OT      | Others                    | Oceania, Africa and South America                                               |

**Tab. 3** Overview on Geo-categories

| Abbrev. | Discipline        | Notes                                     |
|---------|-------------------|-------------------------------------------|
| wrt     | writer            | poets, journalists, essayists             |
| art     | art               | painters, sculptors                       |
| phl     | philosophy        |                                           |
| bio     | bio-sciences      | e.g. biology, physics, chemistry, geology |
| rel     | religious studies | e.g., theology, mystics                   |
| soc     | social studies    | e.g., sociology, anthropology, history    |
| med     | medicine          |                                           |
| pol     | political field   | e.g., politicians, military, statesmen    |
| mat     | mathematics       | including statistics                      |
| eco     | economy           | including businessmen                     |
| leg     | legal studies     | e.g. judges, jurists, lawyers             |
| lan     | language studies  | e.g., linguistics, translation, grammar   |
| eng     | engineering       | including architecture                    |
| ply     | polymath          |                                           |

**Tab. 4** Overview on Disciplines

| actor              | dob  | dod  | era | region | discipline |
|--------------------|------|------|-----|--------|------------|
| Ibn_Tufail         | 1105 | 1185 | MA  | AR     | ply        |
| Maimonides         | 1135 | 1204 | MA  | AR     | ply        |
| Francis_Bacon      | 1561 | 1626 | EM  | EU     | phl        |
| Johannes_Kepler    | 1571 | 1630 | EM  | EU     | bio        |
| René_Descartes     | 1596 | 1650 | EM  | EU     | phl, mat   |
| Christiaan_Huygens | 1629 | 1695 | EM  | EU     | med        |
| Baruch_Spinoza     | 1632 | 1677 | EM  | EU     | phl        |
| Isaac_Newton       | 1642 | 1727 | EM  | EU     | ply        |
| Abraham_de_Moivre  | 1667 | 1754 | EM  | EU     | mat        |
| Voltaire           | 1694 | 1778 | EM  | EU     | phl        |
| Immanuel_Kant      | 1724 | 1804 | TP  | EU     | phl        |

**Tab. 5** A snippet of the final dataset of scholars

### Constructing Longitudinal Networks

We constructed longitudinal network snapshots of the original complete network by subsampling the dataset according to the five consecutive periods (also referred to as eras, compare to Table 2) in order to transform the static final network into a series of time steps. By adding these time-slices of the original network in consecutive order, we derived five progressively *accumulated networks*, which consist of all the influence links of scholars up to and including a target period. For example, the first accumulated network (the Antiquity network) consists of all scholars from the Antiquity era only; whereas the second accumulated network (the Middle Ages network) consists of all scholars from Antiquity and the Middle Ages, and so on. The last accumulated network (the Modern Age network) consists of all scholars of all eras, hence, it is equivalent to the original (complete) network. Table 6 gives an overview of the number of nodes and edges in each accumulated network.

In the following, we describe the network properties of the influence relations of scholars and their time-sliced network projections, and investigate on their communities detected through a community detection algorithm. For analysis, we created directed graphs for each time-sliced network with the *Python::NetworkX* library (Hagberg et al. 2008).

| Era                 | Nodes | Edges |
|---------------------|-------|-------|
| Antiquity           | 209   | 313   |
| Middle Ages         | 5,41  | 786   |
| Early Modern Period | 1,212 | 1,765 |
| Transition Period   | 2,123 | 3,223 |
| Modern Age          | 4,666 | 7,803 |

**Tab. 6** Accumulated-Era Networks

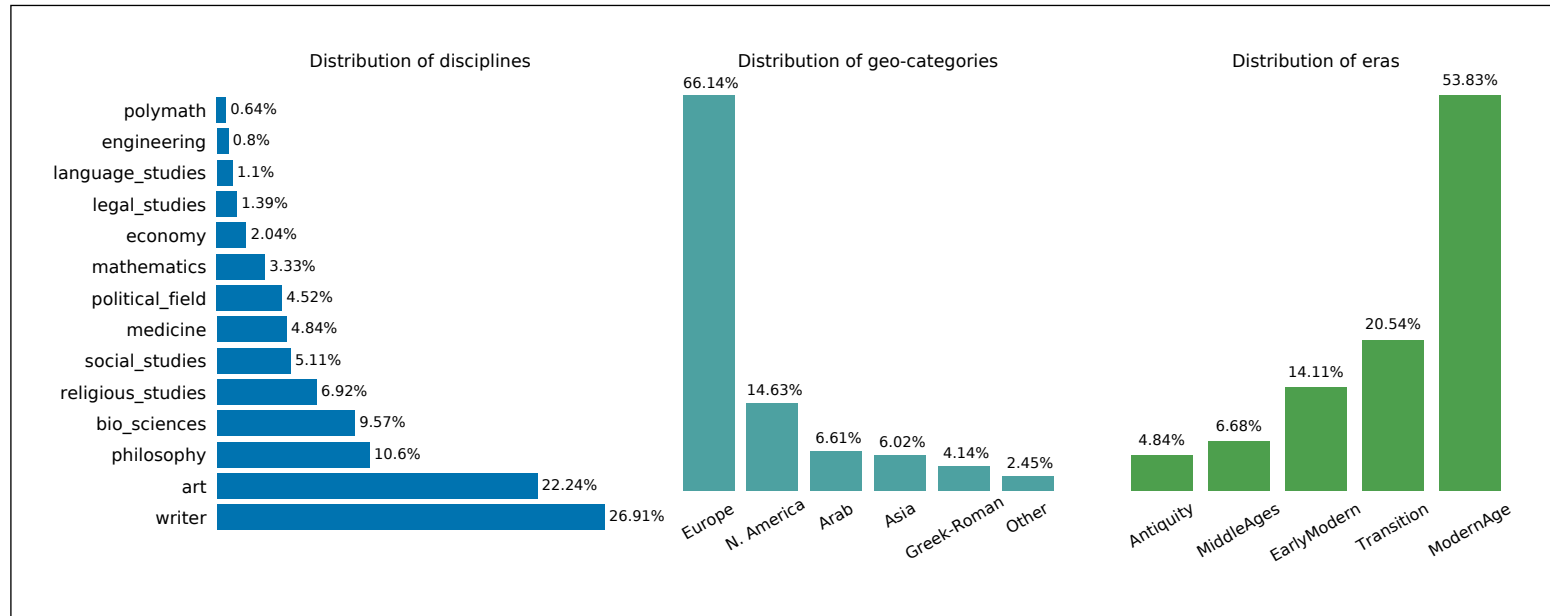
### 3. General Data Exploration

After enriching the dataset in preprocessing, we were able to characterize the dataset of international scholars more thoroughly. In the following we examine the characteristics of scholars in the three dimensions of disciplines, regions, and eras (compare to Figure 3).

The most frequent profession in the database of scholars is *writer* (27%), followed by *arts* (22%), and *philosophy* (11%). The least frequent disciplines are *polymaths* and *engineering* with < 1% each, catering to the relative rarity of polymathy and the recentness of engineering as a discipline. While the dataset takes a global stance, the majority of scholars are from Europe (EU, 66%), followed by North America (AM, 15%) and the Arab world (AR, 7%). The least frequent geo-categories of Oceania, South America, and Africa, which we summarized in the container category OT, together constitute less than 3% of all scholars.

Despite its global representation, we can observe a relative bias of favoring the west in the dataset. For what we defined as the Arab World, intellectuals are relatively well represented for the Medieval era (as a nod to the Islamic Golden Age, in Baghdad as well as in Cordoba), whereas for most of Africa there are almost no scholars, as well as for South America, and only a scrape on the surface of the rich intellectual history of Asia. There is also a prevalence of more recent scholars in the dataset: the percentage of scholars per era continuously increased over time, from 4.8% of all scholars in Antiquity to 53.8% in the Modern Age.

In order to explore the interdependencies among the three dimensions (eras, regions and disciplines), we examine the frequency distribution of scholars over the different combinations of dimensions. Figure 4 shows three 2-dimensional matrices, in which each corresponds to a pair of dimensions: era–discipline, region–era, and region–discipline.



**Fig. 3** Distributions of Disciplines, Regions, and Eras

|  |  |  |  |  |  | Discipline |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|--|--|--|--|--|--|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
|  |  |  |  |  |  | wrt        | art | phl | bio | rel | soc | med | pol | mat | eco | leg | lan | eng | ply |    |
|  |  |  |  |  |  | AN         | 48  | 4   | 134 | 11  | 28  | 14  | 27  |     | 14  |     | 4   | 3   |     | 1  |
|  |  |  |  |  |  | MA         | 61  | 23  | 57  | 41  | 154 | 30  | 19  | 9   | 30  |     | 27  | 17  |     | 17 |
|  |  |  |  |  |  | EM         | 149 | 271 | 101 | 68  | 99  | 26  | 29  | 13  | 42  | 7   | 20  | 7   | 6   | 13 |
|  |  |  |  |  |  | TP         | 319 | 219 | 116 | 204 | 50  | 67  | 78  | 60  | 44  | 26  | 21  | 13  | 12  | 9  |
|  |  |  |  |  |  | MR         | 954 | 688 | 255 | 264 | 117 | 208 | 170 | 212 | 87  | 93  | 34  | 36  | 33  | 10 |

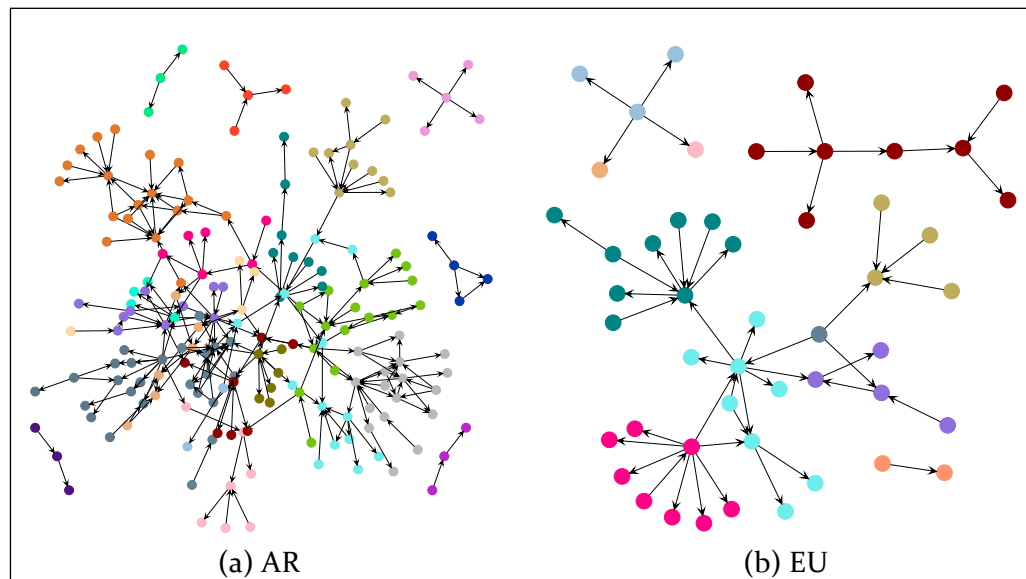
| Region | Era |     |     |     |     | Discipline |     |      |     |     |     |     |     |     |     |     |     |     |     |    |    |
|--------|-----|-----|-----|-----|-----|------------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|
|        | AN  | MA  | EM  | TP  | MR  | wrt        | art | phl  | bio | rel | soc | med | pol | mat | eco | leg | lan | eng | ply |    |    |
|        | GR  | 217 | 3   |     |     | GR         | 44  | 4    | 122 | 9   | 19  | 13  | 18  |     | 14  |     | 4   | 2   |     |    |    |
|        | EU  |     | 75  | 655 | 892 | 1907       | EU  | 1005 | 886 | 422 | 416 | 193 | 224 | 174 | 233 | 161 | 84  | 44  | 47  | 36 | 28 |
|        | AR  | 13  | 232 | 37  | 17  | 54         | AR  | 91   | 4   | 37  | 42  | 159 | 41  | 27  | 9   | 21  |     | 31  | 16  |    | 17 |
|        | AS  | 28  | 44  | 43  | 27  | 179        | AS  | 140  | 44  | 41  | 10  | 60  | 10  | 37  | 3   | 8   | 3   | 4   | 6   |    | 2  |
|        | AM  |     |     | 9   | 139 | 652        | AM  | 186  | 258 | 39  | 113 | 19  | 57  | 56  | 59  | 16  | 39  | 19  | 5   | 14 | 2  |
| OT     |     |     | 2   | 22  | 112 | OT         | 74  | 28   | 8   | 16  | 2   | 7   | 15  | 4   |     | 1   | 4   |     | 2   | 2  |    |

**Fig. 4** Frequency distribution of scholars per Region–Era (left), per Era–Discipline (top), and per Region–Discipline (center). Read like: Frequency of scholars with characteristics *row* and *column*.

The first matrix (top) shows the distribution of scholars per era–discipline combination. We can see that the dominant discipline in Antiquity was *philosophy*, while in the Middle Ages it was *religious studies*, and in the Early Modern period *arts*. In all three eras, the second most prominent profession was *writer*, which became the dominant discipline in the Transitioning period and Modern age.

The second matrix (left) shows the distribution of scholars per region–era combinations. We can see that the majority of scholars in Antiquity were from Greek and Roman Antiquity (GR), while in the Middle Ages they were from the Arab world (AR). This reveals another peculiarity of the dataset: we located Greek and Roman Antique thinkers in this category, rather than counting them as European scholars.

The Arab regional subset has most of its information on Medieval scholars, therefore showing a relative increase during the Medieval period. This could be due to the fact that Medieval research in the Arab world was more centralized than in Europe at the time. In Europe, decentralized Monastic learning was only replaced by the establishment of universities in an intellectual revolution during

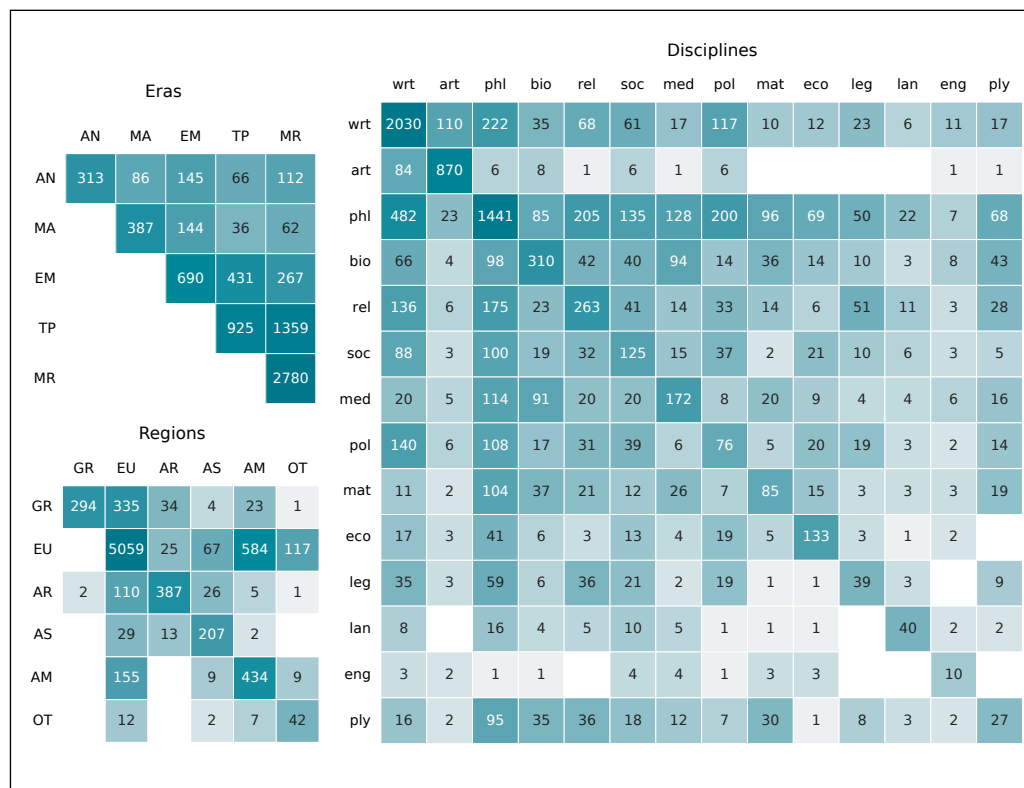


**Fig. 5** Medieval period influence networks of two sub-samples of scholars located in AR and EU, respectively. Communities are indicated in colors; communities of size < 3 are not shown (for the community detection process, please refer to section 4).

the 12<sup>th</sup> century (Burke 2000, p. 36; Lutz-Bachmann 2003, p. 133), thereby resulting in differences in intellectual influence genealogies. In the Arab world, college-like institutions already existed before this, which are thought to have been influential on the creation of the first colleges in Europe, albeit with the difference that they were “fluid system[s]” with an informality towards teaching (Burke 2000, pp. 49–50, Quote on p. 50) and a focus on prominent teachers instead of institutions (Berkey [1992] 2014, p. 16).<sup>21</sup> When comparing the networks of a sub-sample of AR and EU, respectively, the thesis on the greater centralization and interconnection of scholarship in AR during the Middle Ages finds further support: of the AR sub-sample, 79% of scholars (171 out of 215 in total) cluster in the largest weakly connected component – in comparison to the 69% of scholars (34 out of 49 in total) in the EU sub-sample graph (compare to Figure 5). This difference is even more remarkable, when the difference in the number of scholars is taken into account: the AR sub-sample hosted more than 4 times the number of scholars than in the EU sub-sample.

In order to explore the influence relations among aggregated scholars in the different categories of each dimension (era, region, discipline), we look at the

21 It might be noted here that Medieval European universities were also much less formalized before 1800 (Burke 2000, p. 50).



**Fig. 6** Influence relations between eras (top-left), between regions (bottom-left), and between disciplines (right). Read like: Frequency of scholars with characteristic (in row) who influence scholars with characteristic (in column).

distribution of influence relations of a scholar over the different combinations of those categories, i.e. the characteristics of one scholar influencing another. Figure 6 shows three 2-dimensional matrices, each corresponding to one of the dimensions, and iterates the absolute frequency of scholars in each category. These matrices show the aggregated frequency of intellectuals (with characteristic in row) dependent on the three dimensions (era, region, discipline) influencing likewise aggregated intellectuals (with characteristic in column), i.e. the number of scholars from the AR that influence scholars in AS. The first matrix (Figure 6, top-left) shows the influence relations between the different eras. We can see that each period was mostly influenced by (scholars from) that same period, except for the Transitioning period, which had more out-going influences upon the Modern Age. The second matrix (Figure 6, bottom-left) shows the influence relations between the different regions. We can see that the influence received by each region comes from that region itself, except for AM and OT, which received more influences from EU than from themselves. We can also note that each era influenced itself the most, except for GR, which influenced EU more than itself.



The third matrix (Figure 6, right) shows the influence relations between the different disciplines. There, scholars usually influence other scholars from their own disciplines, as well as *writers*.

In the following sections, we develop a method to detect communities and study the patterns of community evolution of intellectuals over time, space and disciplines, as well as the structural dynamics of eras, and identify the mechanisms regarding their development.

#### 4. Community Detection

Most social networks exhibit community structures: their nodes are organized into groups, called communities or clusters, where each group's nodes have a higher probability of being connected to each other than to members of other groups (Fortunato 2010). Pairs of nodes are more likely to be connected if they are both members of the same community, and less likely to be connected if they do not share communities. Identifying communities may offer insight into how the network is organized; it helps to classify nodes based on their role with respect to the communities they belong to. The problem of detecting communities in a network has been extensively studied in the literature, and several methods for community finding have been developed.<sup>22</sup>

In a social network, a community can be considered a set of entities more closely connected to each other than to the rest of the entities in the network (Girvan and Newman 2002), e.g., through more intense interaction with each other (Leskovec et al. 2008). This closeness is based on similarity, and implicitly assumes an underlying structuring principle of homophily (Dakiche et al. 2019, p. 1085; see also McPherson et al. 2001).

As discussed before, we constructed five accumulated networks of scholars over the five eras. In this paper, our goal is to study the evolution of communities of scholars over time. The method of doing so consists of the following steps:

- We first apply a community detection algorithm on these accumulative networks, and
- Then apply an algorithm to track the evolution of communities over time in the five consecutive eras.

In this section, we present the first step of community detection, and then present the second step of tracking the community evolution in the next section.

---

22 We refer to Fortunato (2010) for a comprehensive survey.

### Finding communities with the InfoMap algorithm

We opted to use the *InfoMap community detection algorithm* for directed networks as implemented in the map equation framework by Bohlin et al. (2014). The core of the InfoMap algorithm closely follows the Louvain method (Blondel et al. 2008), where neighboring nodes are joined into modules (clusters), which are subsequently joined into supermodules. The InfoMap algorithm allows us to detect communities in directed networks, which we applied on the accumulated influence networks of intellectuals, consisting of all influence links among scholars who lived up to and including the target era.

Since the InfoMap two-level algorithm is based on random walks, it would provide different results each time it is executed on the same network. Accordingly, we developed an evaluation method to choose the most coherent results: We opted to base such an evaluation on the homogeneity of these communities based on the attributes we surveyed in preprocessing. Thus, we define the following *diversity measure*:

Let  $A$  be a group of items (duplicates allowed) of size  $L$  (number of all items), and let  $N$  be the number of unique items in  $A$ ; we define the diversity of this group as:

$$\text{diversity}(A) = \begin{cases} 0 & \text{if } L = 1 \\ \frac{N-1}{L-1} & \text{otherwise} \end{cases}$$

This measure will equal 1 when the group is completely heterogeneous (e.g., if all items are different:  $N = L$ ), and will equal 0 when it is completely homogeneous (e.g., all items are the same:  $N = 1$ ). We can use this measure to assess the diversity of a group of scholars with respect to any dimension, where the items are the categories of that group's members according to the given dimension. For instance, if a community consists of two scholars from Antiquity, and three scholars from the Middle Ages, its diversity according to this measure is:  $\frac{2-1}{5-1} = 0.25$ .

In the process of detecting communities, we used this diversity measure to evaluate the results obtained. For each era, we executed the community detection algorithm 10 times with a different randomization seed<sup>23</sup> each time (henceforward referred to as a clustering run). For each run, we established the diversity of the clustering results by calculating the diversity of each community with regard to their homogeneity in disciplines, regions, and eras. We combined these

---

23 The randomization seed is the entry configuration of the algorithm.

results with a weighted average, using the weights of 45% for disciplines, 30% for regions, and 25% for eras.<sup>24</sup> The diversity of the clustering run was then calculated as the average diversity of the communities detected in each accumulated network.<sup>25</sup> From this sample of 10 randomized clustering runs, we chose the initial seed of the clustering run that minimized diversity, namely the run with the most homogeneous set of communities.

### Characterizing these communities

When applying the community detection algorithm, we obtained a set of communities for each accumulative-era network. Table 7 shows, for each era, the number of detected communities ( $C$ ), the number of communities with size  $> 3$  ( $C'$ ), the size (number of nodes) of the top three largest communities ( $LC_1$ ,  $LC_2$ ,  $LC_3$ ), and the average number of members per community  $M$ . We clearly observe an increase in the number of communities, which is a result of the increasing number of scholars in the longitudinal accumulation.

For each era and over all the communities from each era, Table 7 shows the average diversity with respect to eras ( $D_e$ ), geo-categories ( $D_g$ ), and disciplines ( $D_d$ ). We observe that the communities of scholars exhibit a very low diversity with respect to eras and regions across all consecutive eras, i.e., in any given community there is one major era and one major region to which most of the members belong. However, communities exhibit a relatively intermediate level of diversity with respect to disciplines over all eras, even though this had the strongest weight in the optimization of community homogeneity in the community detection process.

Table 8 shows the composition of the largest communities per era. These show clear thematic distinctions.

The three largest communities in Antiquity are a group of 19 Classical Philosophers around Aristotle, Plato and Socrates, a cluster of 13 Greek and Roman poets (Virgil, Ovid, Ennius) and Cynic philosophers (Menippos of Gadara), as seen in Figure 8 on top, and a community mirroring the political influences on

---

24 We weighted disciplines the highest, as these we found the main reason for community formation, following the observations in Figure 6. Regional location influences the formation of ties; however, these spatial distances are not unbridgeable, and following the line of thought of Baring (2016), connection should weight more in a truly transnational perspective. We weighted the influence of eras the least as we perceived this is to be a trivial difference between scholars from different time frames.

25 In future work, a robustness analysis could evaluate changes in the diversity measure, if different or multiple disciplines would be recorded for each intellectual (instead of this analysis' focus on only one main discipline), with e.g. Voltaire in the category *writer* as a poet instead of (or including) *philosophy* as an Enlightenment philosopher.

| Era          | <i>N</i> | <i>C</i> | <i>C'</i> | $ LC_1 $ | $ LC_2 $ | $ LC_3 $ | <i>M</i> | <i>D<sub>e</sub></i> | <i>D<sub>g</sub></i> | <i>D<sub>d</sub></i> |
|--------------|----------|----------|-----------|----------|----------|----------|----------|----------------------|----------------------|----------------------|
| Antiquity    | 209      | 31       | 21        | 19       | 13       | 13       | 6.74     | 0.00                 | 0.085                | 0.31                 |
| MiddleAges   | 541      | 80       | 53        | 21       | 19       | 19       | 6.76     | 0.05                 | 0.089                | 0.50                 |
| EarlyModern  | 1212     | 201      | 103       | 41       | 29       | 29       | 6.03     | 0.07                 | 0.072                | 0.37                 |
| TransitionP. | 2123     | 361      | 179       | 51       | 50       | 42       | 5.88     | 0.12                 | 0.063                | 0.36                 |
| ModernAge    | 4666     | 726      | 365       | 66       | 52       | 48       | 6.43     | 0.14                 | 0.103                | 0.33                 |

**Tab. 7** Overview on the sizes of detected communities per era. *N*: No. of nodes, *C*: No. of communities, *C'*: No. of communities with size > 3,  $LC_i$ :  $i^{th}$  largest community, *M*: average members per community, *D<sub>e</sub>*, *D<sub>g</sub>*, *D<sub>d</sub>*: diversity of eras, regions and disciplines, respectively. When the diversity measure is closer to 0, the community is more homogeneous.

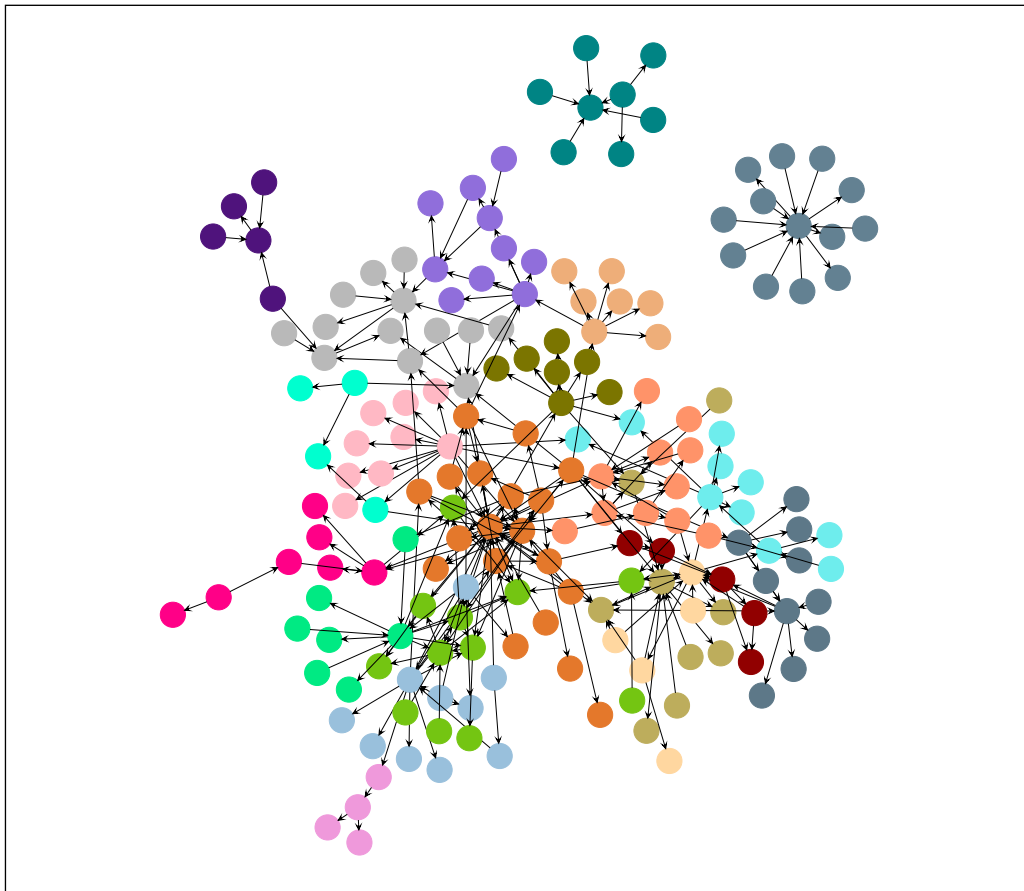
Ashoka the Great of the Mauryan dynasties, and the Seleucid empire. This community is identifiable in the top right of Figure 7, which shows the influence network of intellectuals during the Antiquity era and their communities, distinguished by colors. On the top right, there are two communities not connected to the others: on the right, the community surrounding Ashoka the Great, and on the left, one of the Chinese scholars.

The Classical Philosophers community survived into the Middle Ages as the biggest cluster (21 scholars) influencing Medieval Georgian Neoplatonist philosopher Ional Petrissi, followed by two communities of equal size (19) of Persian poets, Sufi mystics, Sunni poets and philologists dating from the 9<sup>th</sup> to the 13<sup>th</sup> century (Rumi, Saadi Shirazi; see Figure 8, second from top), and a community of Neoplatonic philosophers influencing Christian theologians (Plotinus, Augustine of Hippo, Anselm of Canterbury).

In the Early Modern period, the biggest community (41 scholars) is formed by mostly Medieval Christian theologians and mystics (Thomas Aquinas, John Calvin) and Renaissance humanist Nicholas of Cusa, followed by three communities (29) of Enlightenment scholars of political theory/statesmen (John Locke, Thomas Hobbes, Edmund Burke; compare to Figure 8, second from below), a Persian theosophic poets community (Rumi, Hafez), and a community of Classical philosophers from Antiquity influencing Renaissance Greek philosophers Ioannis Kottounios and surgeon Marco Aurelio Severino. Noteworthy here are also the international and heterogeneous community around Andalusian Medieval scientists/philosophers (Abu al-Quasim al-Zahrawi, Maimonides), and the political theoretical influences of polymath Ibn Tufail on Early Modern scientists, politicians, and religious scholars. The sixth biggest cluster (20) is merged from the Roman Poets with contemporary early Modern Poets of the Renaissance

|                     | # | C  | eras    | regions | disciplines | notable scholars                                  |
|---------------------|---|----|---------|---------|-------------|---------------------------------------------------|
| Antiquity<br>(AN)   | 1 | 19 | AN 100% | GK 95%  | phl 95%     | Plato, Aristotle, Socrates                        |
|                     | 2 | 13 | AN 100% | RO 65%  | wrt 88%     | Virgil, Ovid, Theocritus                          |
|                     | 3 | 13 | AN 100% | IN 77%  | pol 77%     | Buddha, Ashoka,<br>Mahavira                       |
|                     | 4 | 12 | AN 100% | GK 58%  | phl 88%     | Augustine of Hippo,<br>Proclus, Boethius          |
|                     | 5 | 11 | AN 100% | GK 100% | phl 100%    | Arcesilaus, Carneades,<br>Xenocrates              |
| MiddleAges<br>(MA)  | 1 | 21 | AN 95%  | GK 90%  | phl 95%     | Plato, Aristotle, Socrates                        |
|                     | 2 | 19 | MA 100% | AR 95%  | wrt 64%     | Ganjavi, Al-Hallaj, Omar<br>Khayyám               |
|                     | 3 | 19 | AN 89%  | GK 55%  | phl 71%     | Plotinus, Augustine of<br>Hippo, the Areopagite   |
|                     | 4 | 17 | MA 100% | AR 100% | rel 76%     | Ahmad ibn Hanbal, al-<br>Bukhari, Abu Dawood      |
|                     | 5 | 17 | MA 94%  | AR 71%  | phl 30%     | Ibn Tufail, Averroes,<br>Maimonides               |
| EarlyModern<br>(EM) | 1 | 41 | MA 51%  | EU 54%  | phl 48%     | Thomas Aquinas, Ploti-<br>nus, Augustine of Hippo |
|                     | 2 | 29 | EM 93%  | EU 93%  | phl 39%     | John Locke, Francis<br>Bacon, Thomas Hobbes       |
|                     | 3 | 29 | MA 69%  | AR 86%  | wrt 71%     | Ganjavi, Al-Hallaj, Attar<br>of Nishapur          |
|                     | 4 | 29 | AN 90%  | GK 83%  | phl 83%     | Plato, Aristotle, Socrates                        |
|                     | 5 | 29 | MA 62%  | AR 48%  | phl 23%     | Ibn Tufail, Averroes,<br>Maimonides               |
| Transition<br>(TP)  | 1 | 51 | MA 51%  | EU 53%  | rel 47%     | Thomas Aquinas, Ploti-<br>nus, Augustine of Hippo |
|                     | 2 | 50 | EM 66%  | EU 96%  | phl 44%     | René Descartes, John<br>Locke, David Hume         |
|                     | 3 | 42 | AN 90%  | GK 79%  | phl 67%     | Plato, Aristotle, Socrates                        |
|                     | 4 | 33 | TP 55%  | EU 67%  | wrt 77%     | Goethe, Friedrich<br>Schiller, Ganjavi            |
|                     | 5 | 31 | MA 58%  | AR 45%  | phl 31%     | Ibn Tufail, Averroes,<br>Maimonides               |
| ModernAge<br>(MR)   | 1 | 66 | MR. 79% | EU 78%  | art 100%    | Pablo Picasso, Paul<br>Cézanne, Claude Monet      |
|                     | 2 | 52 | MA 52%  | EU 62%  | rel 46%     | Thomas Aquinas, Ploti-<br>nus, Augustine of Hippo |
|                     | 3 | 48 | MR 56%  | EU 83%  | phl 35%     | Hegel, Karl Marx,<br>Friedrich Engels             |
|                     | 4 | 45 | MR 78%  | EU 100% | wrt 87%     | A. Macedonski, B. Fon-<br>dane, Ion Minulescu     |
|                     | 5 | 41 | TP 59%  | EU 98%  | phl 61%     | Immanuel Kant, Baruch<br>Spinoza, Schelling       |

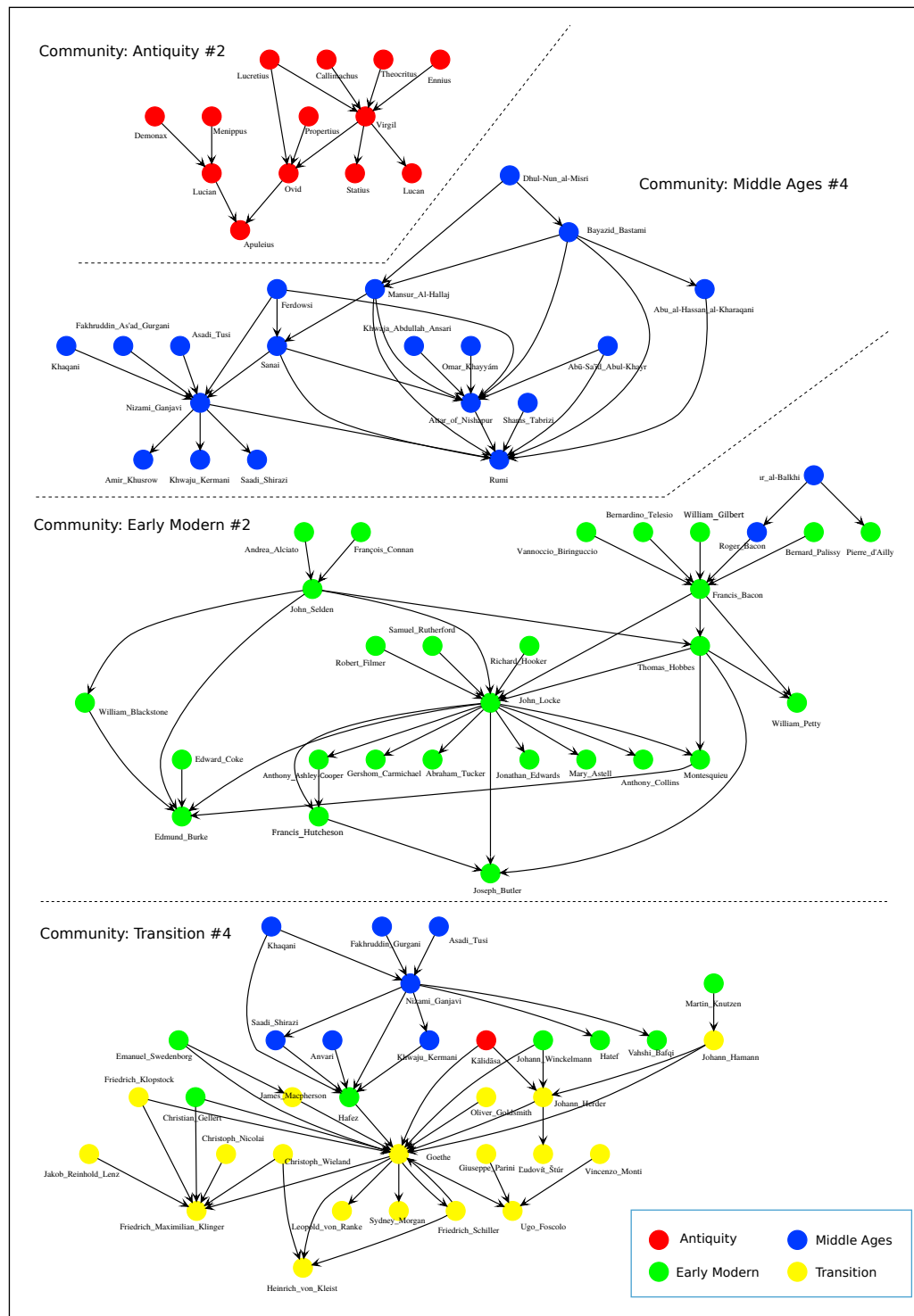
**Tab. 8** Overview on the composition of the largest communities in each era



**Fig. 7** Influence network of intellectuals in the Antiquity period with communities highlighted.

(Ludovico Ariosto), representing their interest in Classical antiquity. An interesting configuration provides the seventh biggest cluster (18), which consists mainly of natural philosophers and mathematicians, and a surprisingly high concentration of polymaths (Gottfried Leibniz, Roger Joseph Boscovich, Benjamin Franklin).

The biggest community from the Transitioning period (51 scholars) consists of Christian Theologians and Mystics, again headed by Thomas Aquinas. The second biggest community (50) consists of Early Modern Enlightenment scholars of political theory and economy (Francis Bacon, John Locke, David Hume, Adam Smith), influencing science philosophers such as August Comte. The Classical Philosophers constitute the third biggest cluster (42) here, in almost unchanged composition. The fourth biggest cluster, of 33, shows an interesting community of the Persian theosophic writer Hafez's influence on German Romantic literature (Goethe, Kleist, Schiller), as seen in Figure 8 below.



**Fig. 8** Tracking influence relation compositions of various exemplary communities.

In the Modern age, a community of 66 Modern artists emerges, ranging from Impressionism (Cézanne, Matisse) to Cubism (Braque) and Social Realism (Ben Shaw). The second biggest community (52) is composed of Antique and Medieval Christian religious philosophers (Plotinus, Thomas Aquinas), with few links to Early Modern and Modern theologians (Nicholas of Cusa, Joseph Maréchal). The third biggest community (48) consists of political philosophers influenced by Hegel and Socialist thinkers formed around Max Stirner, Bruno Bauer and Karl Marx, and Modern influences ranging from Vladimir Lenin, to the political "father of Indonesia" Tan Malaka, and Bertold Brecht.

The communities detected revolve around the main influences of each scholar, and comprise reasonable thematic groups, which allows us to identify various schools of thought.

## 5. Evolution of Communities

In order to track the evolution of these communities computationally and to identify their structural changes, we follow the approach of Greene et al. (2010). They proposed to identify a set of *dynamic communities*, a type of "evolving complex networks" (Qiu et al. 2010; Dakiche et al. 2019, p. 1085) that are present in the network across one or more time steps, whose compositions change according to the behavior of their members, i.e. joining, leaving, or establishing new relations. Communities identified at an individual time step are referred to as *step communities*: these represent specific observations of a dynamic community at a given point in time. Each dynamic community  $D_i$  can be represented by a timeline of its constituent step communities, ordered by time and with at least one step community for each step  $t$ . The most recent observation in a timeline is referred to as the *front* of the dynamic community.

Their model for dynamic community analysis is focused around the life cycle of communities and the key events that characterize the evolution of dynamic communities, such as

- Birth: The emergence of a step community observed at time  $t$  for which there is no corresponding (preceding) dynamic community.
- Death: The dissolution of a dynamic community  $D_i$  occurs when it cannot be observed anymore (i.e. there is no corresponding step community to be observed) for several consecutive time steps.  $D_i$  is subsequently removed from the set  $D$  of dynamic communities.
- Merge: A merge occurs if two distinct dynamic communities ( $D_i, D_j$ ) observed at time  $t - 1$  match to a single step community  $C_{ta}$  at time  $t$ . The pair subsequently shares a common timeline starting from  $C_{ta}$ .
- Split: It may occur that a single dynamic community  $D_i$  present at time  $t - 1$  is matched to two distinct step communities ( $C_{ta}, C_{tb}$ ) at time  $t$ . A branching oc-



curs, with the creation of an additional dynamic community  $D_j$  that shares the timeline of  $D_i$  up to time  $t - 1$ , but has a distinct timeline from time  $t$  onward.

- Continuation: Trivial one-to-one matching where a dynamic community observed at time  $t$  also has an observation at time  $t + 1$ .

This results in the need to track communities over time (Dakiche et al. 2019, p. 1085). The strategy of tracking communities across time steps is a heuristic threshold-based method, which allows for many-to-many mappings between communities across different time steps.<sup>26</sup> The strategy proceeds as follows: Given the first step grouping  $C_1$  (communities of the first time step), a distinct dynamic community is created for each step community. Then, with the next grouping  $C_2$ , an attempt is made to match these step communities with the fronts  $\{F_1, \dots, F_k\}$  (i.e., the step communities from  $C_1$ ). All pairs  $(C_{2a}, F_i)$  are compared, and the dynamic community timelines and fronts are updated based on the key event rules described previously. The process continues until all step groupings have been processed and classified.

### Matching front and step communities

To perform the actual matching between  $C_t$  and the fronts  $\{F_1, \dots, F_k\}$ , we need a measure of similarity between sets. Greene et al. (2010) proposed to use the widely-adopted Jaccard index<sup>27</sup>. The similarity of a united set of communities, however, provides trivial results in our study, as it focuses on the overlap in two communities instead of how much of one community is integrated into the next.

Boujlaleb et al. (2017) proposed using another measure, Quantity Insertion (QI), which reflects the quantity of members of front community  $F_i$  that are inserted into the step community  $C_{ta}$ :

$$\text{sim}_{\text{QI}}(C_{ta}, F_i) = \frac{|C_{ta} \cap F_i|}{|F_i|}$$

26 Many-to-many mappings are a method of system analysis, and refer to the mapping of the relationships of entities' instances. An entity can contain a parent instance for which there are many children instances in another entity, and vice versa. In our context, this means that a community can consist of scholars, who are also present in a subsequent/preceding community of a later/earlier era, whose relationship is observed or tracked by this method. A mapping occurs when the similarity of both entities passes a certain threshold.

27 The Jaccard index or coefficient calculates the similarity of a sample set (in this case, of communities) by dividing the size of overlap of the sample set by the size of the united sample set, i.e. the overlap of two communities divided by the size of the two communities combined.

We opt to use this QI measure in our study, as it is more robust and provides more interpretable results. If the similarity exceeds a matching threshold  $\theta \in [0, 1]$ , the pair is matched and  $C_{ta}$  is added to the timeline for the dynamic community  $D_i$ . The output of the matching process will naturally reveal a series of community evolution events. A step community matching to a single dynamic community indicates a *continuation*, while the case where it matches multiple dynamic communities results in a *merge* event. If no suitable match is found for a step community above the threshold  $\theta$ , a new dynamic community is created for it.

### Characterization of community tracking results

We applied this tracking method to the detected communities over the five eras as time steps. We excluded small communities with 3 or less scholars, hence the number of remaining communities of the five eras are [21, 53, 103, 179, 365]. For matching we used a similarity threshold  $\theta \geq 0.5$ , that is two communities are considered a match when at least 50% of common members belong to the front dynamic community.

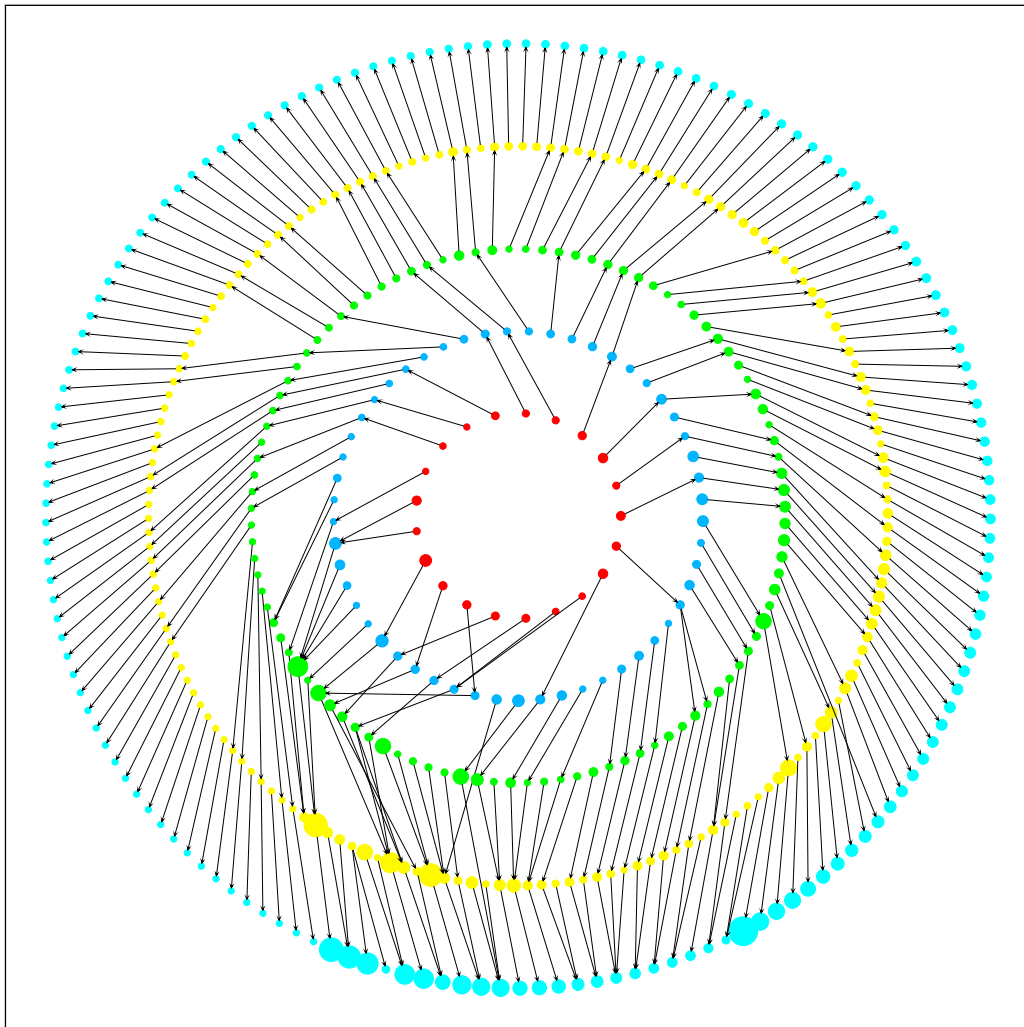
The results of the process of tracking community evolution are 154 dynamic communities over time, among them 132 continued communities (with continuation events only) and 22 merged communities (with merge/split events). Figure 9 gives an overview of the evolution of dynamic communities over all five periods.

We recorded 4 split events (1 at EM, 2 at TP, and 1 at MR), and 36 merge events (2 at MA, 4 at EM, 10 at TP, and 20 at MR). On average, each dynamic community consists of 14.3 members (median: 10 members). For each dynamic community, we calculated the number of constituent step communities (denoted  $N$ ), the number of distinct scholars (across all step communities, denoted  $M_1$ ), and the average number of scholars per step community (denoted  $M_2$ ). Note that in any dynamic community, the constituent step communities are not generally disjointed<sup>28</sup>; thus, a scholar can belong to a different step community in different eras. Therefore,  $M_2$  is not the same as  $M_1$  divided by  $N$ . In fact,  $M_2$  is the sum of the sizes of step communities divided by their member count  $N$ .  $M_1$  is the count of distinct scholars in all step communities of an entire dynamic community over all eras.

The difference  $M_1 - M_2$  provides an indication of the change behavior of a dynamic community. The lower this difference is, the more static the community is; for instance, when this difference is 0 (i.e.,  $M_1 = M_2$ ), the community is self-con-

---

28 They are disjointed only when they are from the same era.



**Fig. 9** Evolution of dynamic communities with more than 3 members over eras, starting with Antiquity from the center outwards. The size of nodes represents the amount of memberships.

tained and does not change its members); conversely, the higher the difference is, the more changes in the community.

In continued dynamic communities, the number of constituent step communities  $N$  is the same as the number of corresponding eras, because in such continued dynamic communities, only one step-community is observed at each era. This is not the case for merged dynamic communities, where multiple step-communities can be observed in a given era (e.g., they are merged in a later era, or split in a former era).

| $N$ | $M_1$ | $M_2$ | $N_{AN}$ | $N_{MA}$ | $N_{EM}$ | $N_{TP}$ | $N_{MR}$ |
|-----|-------|-------|----------|----------|----------|----------|----------|
| 23  | 100   | 15.30 | 7        | 6        | 5        | 3        | 2        |
| 11  | 63    | 19.64 | 2        | 5        | 2        | 1        | 1        |
| 10  | 25    | 8.50  | 0        | 3        | 3        | 3        | 1        |
| 7   | 38    | 13.29 | 0        | 1        | 3        | 2        | 1        |
| 7   | 19    | 10.29 | 1        | 1        | 2        | 2        | 1        |

**Tab. 9** Summary of the top 5 merged dynamic communities.  $N$ : number of step communities,  $M_1$ : number of distinct scholars,  $M_2$ : average number of scholars per step community,  $N_e$ : number of step communities from era  $e$

Table 9 summarizes the 5 largest merged dynamic communities. For each one, it shows the number of constituent step-communities  $N$ , the number of distinct scholars  $M_1$ , and the average number of scholars per step-community  $M_2$ . It also shows the number of step-communities from each era. For instance the largest dynamic community consists of 23 step communities (7 from AN, 6 from MA, 5 from EM, 3 from TP and 2 from MR). This comprises 100 distinct scholars (many of whom belong to multiple step communities in different eras); their step communities contain 15.3 scholars on average. This community is described in detail later.

### Patterns of dynamic communities

The pattern of communities per era is depicted in Table 10, along with some statistics about their occurrence in both continued and merged cases.

For instance, the first pattern represents step communities from all five eras. This pattern occurs in 13 dynamic communities; 9 of these are continued, while 4 are merged. In the 9 that are continued, the average number of step communities is  $\bar{N} = 5$  (in accordance with the five periods we have), the average number of distinct scholars is  $\bar{M}_1 = 10.4$ , and the average scholar-per-community ratio is  $\bar{M}_2 = 8.7$ . In the 4 that are merged, the average number of step communities is  $\bar{N} = 11.75$ , the average number of distinct scholars is  $\bar{M}_1 = 55.5$ , and the average scholar-per-community ratio is  $\bar{M}_2 = 15.3$ .

The most frequent patterns are:

- TP→MR: this pattern occurs 72 times (68 continued, and 4 merged).
- EM→TP→MR: this pattern occurs 41 times (37 continued, and 4 merged).
- MA→EM→TP→MR: this pattern occurs 24 times (16 continued, and 8 merged).

| Eras |     |     |     |    | Continued |           |             |             | Merged |           |             |             |     |
|------|-----|-----|-----|----|-----------|-----------|-------------|-------------|--------|-----------|-------------|-------------|-----|
| AN   | MA  | EM  | TP  | MR | #         | $\bar{N}$ | $\bar{M}_1$ | $\bar{M}_2$ | #      | $\bar{N}$ | $\bar{M}_1$ | $\bar{M}_2$ | #   |
| x →  | x → | x → | x → | x  | 9         | 5         | 10.4        | 8.7         | 4      | 11.75     | 55.5        | 15.3        | 13  |
| x →  | x → | x → | x   |    |           |           |             |             | 1      | 5         | 19.0        | 7.0         | 1   |
|      | x → | x   |     |    | 1         | 2         | 9.0         | 7.5         | 1      | 3         | 12.0        | 7.0         | 2   |
|      | x → | x → | x → | x  | 16        | 4         | 11.0        | 9.8         | 8      | 6.13      | 25.9        | 12.3        | 24  |
|      |     | x → | x → | x  | 37        | 3         | 10.7        | 8.5         | 4      | 4.25      | 32.0        | 15.6        | 41  |
|      |     | x   | x → | x  | 1         | 2         | 22.0        | 15.5        |        |           |             |             | 1   |
|      |     |     | x → | x  | 68        | 2         | 11.0        | 8.5         | 4      | 3         | 42.8        | 19.3        | 72  |
|      |     |     |     |    | 132       |           |             |             | 22     |           |             |             | 154 |

**Tab. 10** Patterns of dynamic communities over eras.  $\bar{N}$ : average number of step communities,  $\bar{M}_1$ : average number of distinct scholars,  $\bar{M}_2$ : average scholar-per-community ratio.

There is another interesting pattern, EM→MR, which consists of a step-community from EM that disappears in TP and reappears again in MR.

In order to better interpret these patterns, we investigated the similarity of communities in each time step, and inspected their loss/gain of members.

### Step-wise similarity of communities

Over all the dynamic communities, we calculated the similarity between matching step communities (inter-period similarity) regarding their common scholars using the Jaccard and the QI measures, and their similarity in terms of the three dimensions: era, geo-categories and disciplines using the cosine similarity<sup>29</sup>). Table II shows the results, where we can observe a very high average similarity of around 92% (based on the QI measure) over all time-steps, revealing that most communities are relatively constant in their composition.

In a dynamic community, a constituent step community does not necessarily contain all the members of its preceding step communities; it also does not necessarily contain all the members of its succeeding step communities. In general, the members of a step community  $X$  in a certain era will be members of different communities  $Y_i$  in the next era. If the similarity between  $X$  and  $Y$  is above a

29 For each step community, each dimension is described as a vector; for example, a community with 6 scholars from EU and 4 scholars from AM, is represented as (0,6,0,0,4,0).

| step    | #   | common scholars |       | dimensions |        |            |
|---------|-----|-----------------|-------|------------|--------|------------|
|         |     | Jaccard         | QI    | era        | region | discipline |
| AN → MA | 21  | 0.795           | 0.921 | 0.995      | 0.994  | 0.990      |
| MA → EM | 53  | 0.725           | 0.894 | 0.954      | 0.963  | 0.962      |
| EM → TP | 97  | 0.754           | 0.916 | 0.955      | 0.975  | 0.969      |
| TP → MR | 166 | 0.718           | 0.911 | 0.918      | 0.988  | 0.975      |
| MA → TP | 1   | 0.450           | 0.692 | 0.874      | 0.949  | 0.965      |
| EM → MR | 2   | 0.604           | 0.892 | 0.882      | 0.999  | 0.968      |
| All     |     | 0.746           | 0.920 | 0.916      | 0.987  | 0.976      |

**Tab. II** Similarity between step communities

threshold, it matches, and the two communities will be identified as belonging to the same dynamic community. However, if the similarity is below the threshold (and  $\neq 0$ ), this means there are some of X members who moved to Y without having an observed connection between X and Y. Thus, there is an unobserved exchange of members, which means a loss of members for X, and a gain/introduction of new members for Y.

In order to analyze this behavior of gaining/losing scholars within communities, we calculate several measures for a step community:

- Loss number ( $Loss_N(x)$ ): the number of lost members, i.e., the number of X members who are not present in its successors. We also calculate the Loss ratio ( $Loss_r(x)$ ) by dividing  $Loss_N(x)$  by the size of X.
- Gain number ( $Gain_N(x)$ ): the number of newly-introduced members, i.e., the number of X members who are not present in its predecessors. We also calculate the Gain ratio ( $Gain_r(x)$ ) by dividing  $Gain_N(x)$  by the size of X.
- Forward Stability ( $FS(X)$ ): the ratio of X members who are present in its successors.
- Backward Stability ( $BS(X)$ ): the ratio of X members who were present in its predecessors.

Note that  $Loss_r(x) + FS(X) = 1$ , and  $Gain_r(x) + BS(X) = 1$ .

The average loss per dynamic community is 2.2 members, where 90 out of the 154 (58%) communities have 0 loss; the maximum gain is 44. The average gain per dynamic community is 3.3 members, where 61 out of the 154 (40%) communities have 0 gain, and the maximum is 60. The average forward stability is 0.92, while the average backward stability is 0.85; this means we can expect

| Era | $\text{loss}_N$ |     |      |     | $\overline{FS}$ | $\text{gain}_N$ |     |      |     | $\overline{BS}$ |
|-----|-----------------|-----|------|-----|-----------------|-----------------|-----|------|-----|-----------------|
|     | #               | sum | avg  | max |                 | #               | sum | avg  | max |                 |
| AN  | 21              | 16  | 0.76 | 5   | 0.921           |                 |     |      |     |                 |
| MA  | 53              | 50  | 0.94 | 5   | 0.907           | 19              | 19  | 1.00 | 6   | 0.925           |
| EM  | 98              | 87  | 0.89 | 9   | 0.925           | 48              | 97  | 2.02 | 12  | 0.858           |
| TP  | 165             | 186 | 1.13 | 24  | 0.917           | 86              | 141 | 1.64 | 26  | 0.888           |
| MR  |                 |     |      |     |                 | 153             | 538 | 3.52 | 60  | 0.819           |

**Tab. 12** Summary of loss and gain of scholars per dynamic community

that 92% of the members of a step community will be observed in its successor communities, while 85% of the members were observed in its predecessor communities. Table 12 provides a summary of these measures over all eras. These figures again suggest little structural change in communities on average, and relatively stable communities over time.

Given a dynamic community, there is a very strong correlation between the difference between the distinct scholars and the average number of scholars in a step community  $M_1 - M_2$  on the one hand, and the sum of  $\text{Loss}_N$  as well as the sum of  $\text{Gain}_N$  over its constituent step communities on the other:

$$\text{corr}(M_1 - M_2, \sum \text{Loss}_N) = 0.868$$

$$\text{corr}(M_1 - M_2, \sum \text{Gain}_N) = 0.844$$

A low difference in the number of distinct scholars and the average number of scholars in a step community ( $M_1 - M_2$ ) means little loss and/or little gain, therefore communities have a high degree of stability, and can be considered almost static, whereas a high  $M_1 - M_2$  signifies more loss and/or more gain, characterizing low stability in a highly changing community.

An interesting sub-class of dynamic communities are those who are *self-contained*. We say a dynamic community is self-contained if its constituent step communities consistently consist of the same set of scholars. This means that there is no exchange of scholars with other communities at all. Based on this definition, a self-contained community has the following characteristics:

- It is necessarily a *continued* community.
- $M_1 = M_2$ , i.e., the number of distinct scholars equals the ratio of scholars per step-community.

| Eras |     |     |     |    | #  | $M_1 = M_2$ |     |     |
|------|-----|-----|-----|----|----|-------------|-----|-----|
| AN   | MA  | EM  | TP  | MR |    | avg         | min | max |
| x →  | x → | x → | x → | x  | 2  | 5.50        | 4   | 7   |
|      | x → | x → | x → | x  | 8  | 7.50        | 4   | 17  |
|      |     | x → | x → | x  | 16 | 5.81        | 4   | 11  |
|      |     |     | x → | x  | 26 | 5.58        | 4   | 10  |
|      |     |     |     |    | 52 | 5.94        |     |     |

**Tab. 13** Self-contained communities

- The similarity of step-wise communities is 100% in terms of scholars and in terms of the three dimensions.
- Both the loss number,  $\text{Loss}_N(x)$ , and the gain number,  $\text{Gain}_N(x)$ , are 0.
- Both the forward- and backward-stability (the ratio of common scholars) is 100%.

We found that there are 52 such self-contained communities (2 start in Antiquity, 8 in MA, 16 in EM, and 26 in TP). All of them survive intact until the Modern Age. Table 13 shows the patterns of self-contained communities, along with their average sizes.

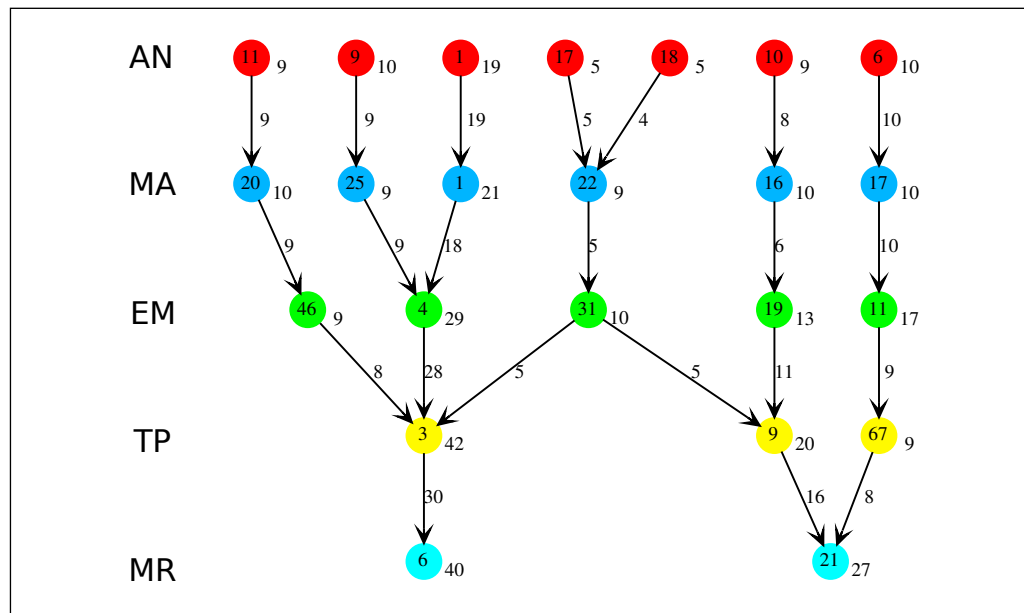
### Description of the largest dynamic community

The largest merging dynamic community consists of two clusters on the left and right, and one intermediate branch merging with each of those two.

The cluster of the left side of Figure 10 consists of two sub-branches that meet during the Transitioning period, the first of which consists of a community of the influence of the Greek poet Aesop (6<sup>th</sup> BC) on later writers of fables (Avianus, Babrius), tragedians (Sophocles), historians (Herodotus), rhetoricians (Himerius), and grammarians (Dositheus Magister), and as such continues into the Middle Ages and the Early Modern period.

The left cluster's second sub-branch consists of two separate branches, which merge in the Early Modern period. One part of this (sub-)sub-branch consists of a community of philosophers and mathematicians around Ionian philosopher Pythagoras, which continues into the Medieval era. The other part of this (sub-)sub-branch consists of the group of Classical Philosophers of Socrates, Plato, and Aristotle. This community continues, self-contained, into the Medieval period, and finally merges with the Antique mathematical philosophers community of Pythagoras in the Early Modern period, joined through Parmenides. This joined





**Fig. 10** Largest dynamic community with merge/split events ( $\theta = 0.5$ ).

community influences various Early Modern Neoplatonian philosophers and the Greek scholar Ionnais Kottounios (17<sup>th</sup> century).

Finally, both sub-branches of the left cluster join in the Transitioning period, which include the poetic influences on Socrates, and merge with another Sophist community from the intermediate branch into a combined community of Sophist, poetic, mathematical, and Stoic influence network on Plato and Aristotle, which then continues into the Modern period, losing the poet's branch, but incorporating the Aristotelian influence on Francis Bacon.<sup>30</sup>

The intermediate branch consists of the community of Socrates' student Antisthenes's influences on Cynicism (Diogenes of Sinope, Crates of Thebes) merging with a community of Stoic philosophers surrounding Zeno of Citium into a joined community of Cynics' influence on Stoicism through the Megalarian School. These continue into a community, which also incorporates early Sophist influences on Cynicism. A split event leads the Sophist part of the community to merge with the aforementioned Classical philosopher's community around

<sup>30</sup> Bacon's work on natural philosophy drew heavily from ancient sources, and as Pesic (2014, p. 79) argued, his terminology – such as the contested usage of “violence” of nature and the dominion of man (compare to Merchant 2008) – can only be understood “depend[ent] on their Aristotelian context”, though he also departed from his ancient sources (see also Cushing 1998, pp. 15–28).

Socrates, and the Cynics part to merge with the community of Cynics around Zeno of Citium and the Stoics around Epictetus. The Stoics in turn influence a cluster of Jansenists around Early Modern scientist Blaise Pascal from the second cluster on the right who, after his religious epiphany in the 1650s, was influenced by Epiktet and turned to Jansenism, a heretic branch of Catholic Popedom fighting against the Jesuits.

The second cluster on the right consists of two continuing communities, a Stoic community from the school of Athens around Cleanthes and Epictetus, and another Stoic community in the Roman Republic around the teacher-student pair of Panaetius and Posidonius that continues until the Modern Age, with the exception of the merging event with the intermediate branch of Cynics, Stoics and Jansenists. This group then merges into a combination of both branches of Stoicism of Panaetius and Cleanthes via Zeno of Citium, losing the Early Modern Jansenists group around Pascal.

In order to infer the question as to why these groups show these evolution dynamics in communities, it is equally important to look for who is part of a community as for who is not part of it. The scholarly part of a community shows a greater homogeneity than other possible communities with regards to their disciplines, regions, and periods, as this was optimized in the clustering algorithm. We can clearly identify schools of thought and reasonable thematic clusters. These communities are again structured by a sub-group of cores – members of communities that stay together even though the group changes communities over time, such as the Cynics genealogy of Diogenes of Sinope, Crates of Thebes, his wife Hipparchia of Maroneia, and her brother Metrocles, which closely resemble schools and fields within the communities. The changing composition of communities, exchanging core sub-groups and floating members, leads to these core-groups integrating into other groups that provide a stronger homogeneity relative to the other possible groups of that era.

## 6. Conclusion

This study is founded on a database encompassing the influence relations of intellectuals similar to the broad strokes of a “*Meistererzählung*”: an abstract view on the main influences of the main intellectuals as collected in the Linked Open Data base YAGO3; it is thus sourced from Wikipedia, the biggest and most accessible encyclopedia of crowd-sourced origin<sup>31</sup>. This view on the history of intellectuals closely iterates the state-of-knowledge compiled in the knowledge base YAGO3, and consequently represents a crowd-curated, contemporary view on the

---

31 Wikipedia grows by approx. 1,500 articles per day and offers an unparalleled rapid potential for correctional prowess, with an average of 1.9 edits per second (Wikipedia 2020a).

history of intellectuals and their main influences on one another. Despite the focus of the database on main influences, and biases in representation that favor male European intellectuals, and a general global stance of the dataset (as described in Sections 2 and 3), this unique dataset nonetheless constitutes the most complete available on the history of intellectuals, albeit from an abstract perspective, similar to the broad strokes of a master narrative.

In this study, we offered a network methodology to analyze the history of intellectuals. We provided a detailed description of the process of creating, enriching, and preparing an extracted dataset from YAGO with SPARQL, and how to create longitudinal networks of such data based on global periodization. We investigated the community formation processes of scholars over time, and developed a method to evaluate the quality of the resulting communities by taking their diversity into account. This community detection helps to understand the genealogy of scholars, and the variety of relational influence, and provides a means to computationally identify schools of thought. We traced the evolution of these communities as a dynamic process over time and differentiated between 154 dynamic communities of size 4 or greater, and tracked the continuation and merging of communities throughout their evolution, as well as their similarity in each time step. We described the exemplary mechanisms and characteristics of their development based on the largest merging community in the dataset, exemplifying the change in core groups and floating members.

These approaches helped to bring more quantitative/computational evidence for certain assumptions derived from qualitative research, and offer the potential for further falsification. In order to achieve this potential, a more “fine-grained” database would be necessary. As the above analyses iterate the abstract “broad stroke” representation in YAGO/Wikipedia, we would ideally like to broaden the database to include a representative and global outlook, and apply the established methodology of analyzing the YAGO network to a more “fine-grained” influence network that takes into account more than just the most important influences of the most important scholars. This could be based on a selection of primary sources from within intellectual history, which we would like to evaluate based on differences and insights of the dynamics of intellectual influences, and to then compare those with the results of the extracted YAGO3 dataset. A community analysis based on a more fluid interpretation of the main discipline of each scholar, taking multiple heterogeneous disciplines into account, would elevate the robustness of the formation of computationally detected communities. We would also like to add to this study with an extended in-depth analysis of the various interrelations of the core groups these communities consist of, and their interactions (and exchanges) with other communities.

## References

- Anderka, Maik. 2013. "Analyzing and predicting quality flaws in user-generated content: the case of Wikipedia." PhD Thesis, Professor Content Management/Web-Technologien. Faculty of Media. Bauhaus University Weimar. DOI: <https://doi.org/10.25643/bauhaus-universitaet.1977>.
- Armitage, David. 2012. "What's the Big Idea? Intellectual History and the Longue Durée." *History of European Ideas* 38 (4): 493–507. <http://www.the-tls.co.uk/tls/public/article1129685.ece>.
- Armitage, David. 2014. "The International Turn in Intellectual History." In *Rethinking Modern European Intellectual History*, edited by Darrin M. McMahon and Samuel Moyn, 232–252. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199769230.003.0012>.
- Baring, Edward. 2016. "Ideas on the Move: Context in Transnational Intellectual History." *Journal of the History of Ideas* 77 (4): 567–587. DOI: <https://doi.org/10.1353/jhi.2016.0031>. <https://muse.jhu.edu/article/635628>.
- Berkey, Jonathan P. (1992) 2014. *The Transmission of Knowledge in Medieval Cairo: A Social History of Islamic Education*. Princeton University Press.
- Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. "DBpedia – A Crystallization Point for the Web of Data." *Web Semant. (Amsterdam, The Netherlands)* 7, no. 3 (): 154–165. DOI: <https://doi.org/10.1016/j.websem.2009.07.002>.
- Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 10 (): P10008. DOI: <https://doi.org/10.1088/1742-5468/2008/10/p10008>.
- Bohlin, Ludvig, Daniel Edler, Andrea Lancichinetti, and Martin Rosval. 2014. "Community Detection and Visualization of Networks with the Map Equation Framework." In *Measuring Scholarly Impact: Methods and Practice*, 3–34. DOI: <https://doi.org/10.1007/978-3-319-10377-8>.
- Boujlaleb, Loubna, Ali Idarrou, and Driss Mammass. 2017. "Tracking Community Evolution in Social Networks." *Journal of Theoretical and Applied Information Technology* 95, no. 22 (): 6127–6137.
- Burke, Peter. 2000. *A social history of knowledge. From Gutenberg to Diderot*. Cambridge: Polity Press.
- Casebourne, Imogen, Chris Davies, Michelle Fernandes, and Naomi Norman. 2012. "Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopedias: A comparative preliminary study across disciplines in English, Spanish and Arabic." (Brighton, UK). Accessed 2021-12-16. [http://commons.wikimedia.org/wiki/File:EPIC\\_Oxford\\_report.pdf](http://commons.wikimedia.org/wiki/File:EPIC_Oxford_report.pdf).
- Cushing, James. 1998. *Philosophical concepts in physics: the historical relation between philosophy and scientific theories*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139171106>.

- Dahmen, Silvio, Ana Bazzan, and Robert Gramsch. 2017. "Community Detection in the Network of German Princes in 1225: A Case Study." In *Complex Networks VIII*, edited by Bruno Gonçalves, Ronaldo Menezes, Roberta Sinatra, and Vinko Zlatić, 193–200. Cham: Springer International Publishing.
- Dakiche, Narimene, Fatima Benbouzid-Si Tayeb, Yahya Slimani, and Karima Benatchba. 2019. "Tracking community evolution in social networks: A survey." *Information Processing & Management* 56 (3): 1084–1102.
- Duke University Libraries. 2020. "Project Vox." Duke University Libraries. <https://projectvox.library.duke.edu/>.
- Edelstein, Dan. 2016. "Intellectual History and Digital Humanities." *Modern Intellectual History* 13 (1): 237–246.
- Edelstein, Dan, Paula Findlen, Giovanna Ceserani, Caroline Winterer, and Nicole Coleman. 2017. "Historical Research in a Digital Age: Reflections from the Mapping the Republic of Letters Project." *The American Historical Review* 122, no. 2 (): 400–424. DOI: <https://doi.org/10.1093/ahr/122.2.400>.
- Fernández Riva, Gustavo. 2019. "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods." *Journal of Historical Network Research* 3: 30–49.
- Fortunato, Santo. 2010. "Community Detection in Graphs." *Physics Reports* 486 (3): 75–174. DOI: <https://doi.org/10.1016/j.physrep.2009.11.002>.
- Ghawi, Raji, Cindarella Petz, and Jürgen Pfeffer. 2019. "On the Shoulders of Giants' – Analysis of a Social Network of Intellectual Influence." In *Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, 248–255*. DOI: <https://doi.org/10.1109/SNAMS.2019.8931821>.
- Ghawi, Raji, and Jürgen Pfeffer. 2019. "Mining Social Networks from Linked Open Data." In *Graph-Based Representation and Reasoning*, edited by Dominik Endres, Mehwish Alam, and Diana Sotropa, 221–229. Cham: Springer International Publishing.
- Ghawi, Raji, and Jürgen Pfeffer. 2020. "Extraction Patterns to Derive Social Networks from Linked Open Data Using SPARQL." *Information* 11 (7). <https://www.mdpi.com/2078-2489/11/7/361>.
- Giles, James. 2005. "Special Report. Internet encyclopaedias go head to head." *Nature* 438 (). Accessed 2021-12-16. <https://www.nature.com/articles/438900a.pdf>.
- Girvan, Michelle, and Mark Newman. 2002. "Community Structure in Social and Biological Networks." *Proceedings of the National Academy of Sciences* 99 (12): 7821–7826. DOI: <https://doi.org/10.1073/pnas.122653799>.
- Goldman, Harvey. 1994. "From social theory to sociology of knowledge and back: Karl Mannheim and the sociology of intellectual knowledge production." *Sociological Theory*: 266–278.

- Gordon, Peter E. 2013. "What is Intellectual History? A Frankly Partisan Introduction to a Frequently Misunderstood Field." [http://projects.iq.harvard.edu/files/history/files/what\\_is\\_intell\\_history\\_pgordon\\_mar2012.pdf](http://projects.iq.harvard.edu/files/history/files/what_is_intell_history_pgordon_mar2012.pdf)
- Gramsch-Stehfest, Robert. 2020. "Community detection and structural balance. Network analytical modelling of political structures and actions in the Middle Ages." In *The Power of Networks. Prospects of Historical Network Research*, edited by Florian Kerschbaumer, Linda von Keyserlingk-Rehbein, Martin Stark, and Marten Düring, 37–55. London New York: Routledge.
- Greene, Derek, Donal Doyle, and Padraig Cunningham. 2010. "Tracking the Evolution of Communities in Dynamic Social Networks." In *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining*, 176–183. ASONAM '10. USA: IEEE Computer Society. DOI: <https://doi.org/10.1109/ASONAM.2010.17>.
- Haakonssen, Knud, and Richard Whatmore. 2017. "Global possibilities in intellectual history: a note on practice." *Global Intellectual History* 2 (1): 18–29. DOI: <https://doi.org/10.1080/23801883.2017.1370248>.
- Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart. 2008. "Exploring Network Structure, Dynamics, and Function using NetworkX." In *Proceedings of the 7th Python in Science Conference*, edited by Gaël Varoquaux, Travis Vaught, and Jarrod Millman, 11–15. Pasadena, CA.
- Harris, Steve, and Andy Seaborne. 2013. "SPARQL 1.1 Query Language." *W3C Recommendation*. <http://www.w3.org/TR/sparql11-query>.
- Hennig, Marina, Ulrik Brandes, Jürgen Pfeffer, and Ines Mergel. 2012. *Studying Social Networks: A Guide to Empirical Research*. Frankfurt a. Main: Campus Verlag.
- Holme, Petter, and Jari Saramäki. 2019. "A Map of Approaches to Temporal Networks." In *Temporal Network Theory*, edited by Petter Holme and Jari Saramäki, 1–24. Computational Social Sciences. Cham: Springer International Publishing. DOI: [https://doi.org/10.1007/978-3-030-23495-9\\_1](https://doi.org/10.1007/978-3-030-23495-9_1).
- Huisman, Mark, and Tom A. B. Snijders. 2003. "Statistical Analysis of Longitudinal Network Data With Changing Composition." *Sociological Methods & Research* 32 (2): 253–287. DOI: <https://doi.org/10.1177/0049124103256096>.
- Kousha, Kayvan, and Mike Thelwall. 2016. "Are Wikipedia Citations Important Evidence of the Impact of Scholarly Articles and Books?" *Journal of the Association for Information Science and Technology*. DOI: <https://doi.org/10.1002/asi.23694>.
- Krämer, Katrina. 2019. "Female scientists' pages keep disappearing from Wikipedia – what's going on?" *Chemistry World* (newspaper). <https://www.chemistryworld.com/news/female-scientists-pages-keep-disappearing-from-wikipedia-whats-going-on/3010664.article>.

- Kudella, Christoph. 2019. "How do we model the Republic of Letters? With Contributions from Neil Jefferies." In *Reassembling the Republic of Letters in the Digital Age. Standards, Systems, Scholarships*, edited by Howard Hotson and Thomas Wallnig, 41–53. Göttingen University Press. DOI: <https://doi.org/10.17875/gup2019-1146>.
- Leskovec, Jure, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. 2008. *Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters*. arXiv: 0810.1355 [cs.DS].
- Lovejoy, Arthur O. 1936. *The Great Chain of Being: A Study of the History of an Idea*. Harvard University Press.
- Lutz-Bachmann, Matthias. 2003. "Die Vielheit des Wissens und das neue Konzept von Wissenschaft. Zu den epistemologischen Voraussetzungen der Wissensrevolution im 12. Jahrhundert." In *Wissenskulturen. Beiträge zu einem forschungsstrategischem Konzept*, edited by Johannes Fried and Thomas Kailer, 133–140. Berlin: Akademie Verlag.
- Mahdisoltani, Farzaneh, Joanna Biega, and Fabian M. Suchanek. 2015. "YAGO3: A Knowledge Base from Multilingual Wikipedias." In *CIDR 2015, Seventh Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, Online Proceedings*. <http://cidrdb.org/cidr2015/Papers/CIDR15%5C%5C Paper1.pdf>.
- Manola, Frank, and Eric Miller. 2004. "RDF Primer." *W3C Recommendation*.
- McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27 (1): 415–444. DOI: <https://doi.org/10.1146/annurev.soc.27.1.415>.
- Merchant, Carolyn. 2008. "'The Violence of Impediments': Francis Bacon and the Origins of Experimentation." *Isis* 99 (4): 731–760.
- Moyn, Samuel, and Andrew Sartori, eds. 2013. *Global Intellectual History*. New York: Columbia University Press.
- Murgu, Cal, and Krisandra Iivings. 2019. "'Blind Trust Is Not Enough': Considering Practical Verifiability and Open Referencing in Wikipedia." *Journal of Critical Library and Information Studies* 2 (2: Information/Control: Control in the Age of Post-Truth). DOI: <https://doi.org/10.24242/jclis.v2i2.62>.
- Newcomb, Theodore M. 1961. "The Acquaintance Process as a prototype of human interaction." (New York). DOI: <https://doi.org/10.1037/13156-015>.
- Osterhammel, Jürgen. 2006. "Über die Periodisierung der neueren Geschichte (Vortrag in der Geisteswissenschaftlichen Klasse am 29. November 2002)," edited by Berlin-Brandenburgische Akademie der Wissenschaften, 10: 45–64. *Berichte und Abhandlungen*. Berlin: Akademie Verlag.
- Pesic, Peter. 2014. "Francis Bacon, violence, and the motion of liberty: The Aristotelian background." *Journal of the History of Ideas* 75 (1): 69–90.
- Petz, Cindarella, Raji Ghawi, and Jürgen Pfeffer. 2020. "A Longitudinal Analysis of a Social Network of Intellectual History." *IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM)*: 340–347. DOI: <https://doi.org/10.1109/ASONAM49781.2020.9381318>.

- Pot, Johan Hendrik Jacob van der. 1999. *Sinneutung und Periodisierung der Geschichte. Eine systematische Übersicht der Theorien und Auffassungen*. Leiden Boston Cologne: Brill.
- Prud'hommeaux, Eric, and Andy Seaborne. 2008. "SPARQL Query Language for RDF." *W3C Recommendation*. <http://www.w3.org/TR/sparql11-query>.
- Qiu, Baojun, Kristinka Ivanova, John Yen, and Peng Liu. 2010. "Behavior evolution and event-driven growth dynamics in social networks." In *2010 IEEE Second International Conference on Social Computing*, 217–224. Washington, DC: IEEE, IEEE Computer Society.
- Ringer, Fritz. 1990. "The intellectual field, intellectual history, and the sociology of knowledge." *Theory and society* 19 (3): 269–294.
- Skinner, Quentin. 1969. "Meaning and Understanding in the History of Ideas." *History & Theory* 8.
- Snijders, Tom A. B., Gerhard G. Bunt, and Christian Steglich. 2010. "Introduction to Actor-Based Models for Network Dynamics." *Social Networks* 32: 44–60. DOI: <https://doi.org/10.1016/j.socnet.2009.02.004>.
- Stephenson-Goodknight, Rosie. 2016. "Viewpoint: How I tackle Wiki gender gap one article at a time – BBC News." *BBC* (newspaper). Accessed 2020-09-12. <https://www.bbc.com/news/world-38238312>.
- Subrahmanyam, Sanjay. 2015. "Global Intellectual History Beyond Hegel and Marx." *History and Theory* 54 (1): 126–137. DOI: <https://doi.org/10.1111/hith.10746>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/hith.10746>.
- Subrahmanyam, Sanjay. 2017. "Beyond the usual suspects: on intellectual networks in the early modern world." *Global Intellectual History* 2 (1): 30–48.
- Suchanek, Fabian M., Gjergji Kasneci, and Gerhard Weikum. 2007. "Yago: A Core of Semantic Knowledge." In *Proceedings of the 16th International Conference on World Wide Web*, 697–706. WWW '07. Banff, Alberta, Canada: ACM. DOI: <https://doi.org/10.1145/1242572.1242667>.
- Teplitskiy, Misha, Grace Lu, and Eamon Duede. 2017. "Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science." PUBLISHED ONLINE 2016-10-13 IN WILEY ONLINE LIBRARY, *Journal of the Association for Information Science and Technology* 68 (9). DOI: <https://doi.org/10.1002/asi.23687>.
- Thompson, Neil, and Douglas Hanley. 2018. "Science is Shaped by Wikipedia: Evidence From a Randomized Control Trial." MIT SLOAN RESEARCH PAPER NO. 5328-17 (). DOI: <https://doi.org/10.2139/ssrn.3039505>.
- Viseur, Robert. 2014. "Reliability of user-generated data: The case of biographical data in Wikipedia." In *Proceedings of The International Symposium on Open Collaboration*, 1–3. New York, NY: Association for Computing Machinery. DOI: <https://doi.org/10.1145/2641580.2641581>.
- Vugt, Ingeborg van. 2017. "Using Multi-Layered Networks to Disclose Books in the Republic of Letters." *Journal of Historical Network Research* 1, no. 1: 25–51. Accessed 2020-04-15. [//jhnr.uni.lu/index.php/jhnr/article/view/7](http://jhnr.uni.lu/index.php/jhnr/article/view/7).



- Wickberg, Daniel. 2001. "Intellectual History vs. the Social History of Intellectuals." *Rethinking History* 5 (3): 383–395. DOI: <https://doi.org/10.1080/13642520110078505>.
- Wikipedia. 2020a. "Wikipedia: Statistics." Wikipedia. Page Version ID: 977514767. <https://en.wikipedia.org/w/index.php?title=Wikipedia:Statistics%5C&oldid=977514767>.
- Wikipedia. 2020b. "Wikipedia: WikiProject Women in Red." Wikipedia. Page Version ID: 977332878. [https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Women\\_in\\_Red](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red).