JOURNAL OF INTERDISCIPLINARY HISTORY OF IDEAS

2019

Volume 8 Issue 16
Item 5

– Section 3: Notes –

Exploring Knowledge Dynamics in the Humanities
Two Science Mapping Experiments

by
Eugenio Petrovich, Emiliano Tolusso

JIHI 2019

Volume 8 Issue 16

. . . .

Exploring Knowledge Dynamics in the Humanities
Two Science Mapping Experiments

Eugenio Petrovich, Emiliano Tolusso *

This paper reports and briefly discusses the results of two science mapping experiments we conducted in two humanities fields: analytic philosophy and human geography. In the first section, we provide a non-technical introduction to science mapping techniques, presenting the steps required to produce distance-based science maps. The two following sections present the datasets of our experiments and the maps we produced. Lastly, we discuss the main limitations of science mapping when it is applied to areas in the humanities.
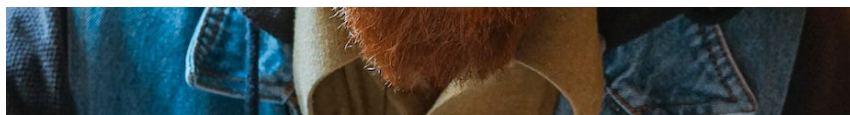
Science mapping is a flourishing research topic at the crossroad of scientometrics, information visualization, network analysis, and sociology of science (Börner, Chen, & Boyack, 2005; Börner, Theriault, & Boyack, 2015; Chen, 2013). It aims to display the structure and dynamics of scientific knowledge by using 2- or 3-d visualizations, known as "science maps" (Chen, 2017). Among the different kinds of science maps that can be found[1], *bibliometric maps* are the most developed. Bibliometric maps are based on the idea that scientific knowledge can be represented by a *network*, in which scientific publications (e.g., articles

---

* University of Siena (*eugenio.petrovich @ unisi.it*); University of Milan and PoliS-Lombardia (*emiliano.tolusso @ unimi.it*).
[1] See (Börner, 2010) for an excellent and aesthetically stunning overview.

published in academic journals) are represented by the nodes or vertices of the network and the citations between them by the links or edges (Waltman & van Eck, 2014).

Until recently, science mapping has been mainly used to investigate the structure of the sciences, with little interest in the humanities. However, in the last years, scientometricians and digital humanists have started to map humanistic fields too (see for instance (Colavizza, 2018; Kreuzman, 2001)). In this working paper, we aim to contribute to this strand of research, by displaying and shortly discussing the results of two mapping experiments that we originally presented at the first DR2 conference, held in Turin in January 2017.[1] Our experiments targeted analytic philosophy and human geography, respectively.

The rest of the paper is organized as follows. First, we provide a short introduction to bibliometric science maps to better understand the methodology and the tool we used to produce the maps (a software called VOSviewer). In the second section, we present and comment on our maps. Lastly, we discuss some limitations of bibliometric mapping applied to the humanities.

1. A short introduction to bibliometric mapping

As we said above, the bibliometric approach to science mapping is based on the analysis of the links between scientific publications, i.e., on the analysis of *citation networks*. The classic workflow to produce a bibliometric map includes the following steps: data selection and retrieval, data cleaning, network extraction, normalization of the co-occurrence matrix, mapping and visualization, and, lastly, interpretation (Chen, 2017; Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011).

The first step consists of selecting a set of scientific publications that reasonably represents the research area we want to map, a process known as field

---

[1] This paper is a revised version of the post we published on the DR2 open peer review blog on February 2018. In particular, we tried to integrate the several and very useful comments we got during the process of open peer review happened on the blog (`https://tinyurl.com/dr2oprblog-exploring`). We thank anyone who contributed to it by sharing with us helpful suggestions. Note that improved versions of the mapping experiments can be found in (Petrovich & Buonomo, 2018), (Petrovich, 2019), and (Tolusso, 2019).

Eugenio Petrovich, Emiliano Tolusso

delineation (Laurens, Zitt, & Bassecoulard, 2010; Zitt & Bassecoulard, 2006). Second, the data for the mapping is retrieved from a citation database, i.e., a database that collects the meta-data of scientific publications (authors, title, publication year, etc.) along with their citation links with other publications. Clarivate's Web of Science (WoS) and Elsevier's Scopus are the main resources for this kind of data.[1] After the pre-processing phase in which the downloaded records are cleaned and standardized (e.g., by correcting the misspelling of author names or disambiguating homonyms), the citation network is extracted from publications and the relative *occurrence matrix* is thus obtained. An example of such a matrix is displayed in Table 1. Each row of the matrix represents a *citing* publication, whereas each column represents a reference, i.e., a *cited* publication. If the citing paper cites one of the references, the corresponding cell is marked with 1, 0 otherwise. In the example below, paper A cites the references (a, b, c), whereas paper B cites (b, d), and so on. Thus, each row-vector associated with a capital letter represents the *bibliography* of that publication, whether each column-vector relative to a lowercase letter represents the *citations* obtained by that publication (publication *a* is cited by A and B, publication *b* by A and B, and so on). The sum of the values on the row-vectors is equal to the length of the bibliography of the citing publication (for publication A, it amounts to 3), whereas the sum of the values on the column-vector is equal to the total number of citations gathered by the cited publication (for publication *a*, it amounts to 2).

	a	b	c	d	e
A	1	1	1	0	0
B	0	1	0	1	0
C	0	0	1	1	0
D	1	0	0	0	1
E	0	0	1	1	1
F	0	0	0	0	1

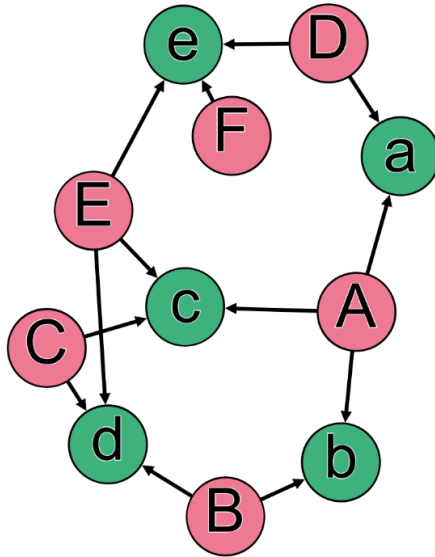Table 1: Example of occurrence matrix

Fig. 1: Example of a citation network.[1]

The same information contained in the occurrence matrix can be represented by a *directed graph* in which the nodes represent the publications and the arrows the citation ties (Fig. 1).

The citation network represented in the occurrence matrix can be used to produce a *direct-linkage bibliometric map*, such as the one shown in Fig. 1. However, since the citation occurrence matrix is usually very sparse (i.e., most of its cells contain zeros), this technique is little used in science mapping (see however (Waltman & van Eck, 2012)).

In fact, the two most common techniques of bibliometric mapping, namely *bibliographic coupling* (Kessler, 1963) and *co-citation analysis* (Small, 1973), do

---

[1] Citing publications (capital letters) are in pink, cited publications (lowercase letters) in light green. Citations are represented by directed arrows.

Eugenio Petrovich, Emiliano Tolusso

not work directly on the occurrence matrix itself but on a matrix derived from it, containing *co-occurrence data* (van Eck & Waltman, 2009).

1.1. Bibliographic coupling and co-citation analysis

Both bibliographic coupling and co-citation analysis use citations to calculate the *similarity* between publications. In bibliographic coupling, publications sharing many references, i.e., publications the bibliographies of which largely overlap, are considered as more similar than publications sharing few or no references at all. In co-citation analysis, on the other hand, the similarity of publications is measured based on the number of times they are *cited together* in the bibliographies of a set of other publications. Once again, the idea is that, if publications are frequently cited together, they are likely to be more similar than if they are seldom or never cited together.

Now, both the strength of the bibliographic coupling and the strength of the co-citation link between two publications can be calculated by multiplying the original occurrence matrix with its transposed. In network analysis, this operation amounts to *project* a bi-modal network on one of its two one-mode networks. In the case of the example in Fig. 1, the resulting matrices are shown in Table 2 and Table 3.

	A	B	C	D	E	F
A	-	1	1	1	1	0
B		-	1	0	1	0
C	1	1	-	0	2	0
D	1	0	0	-	1	1
E	1	1	2	1	-	1
F	0	0	0	1	1	-

Table 2: Bibliographic coupling matrix derived by the occurrence matrix in Fig. 1.

	a	b	c	d	e
a	-	1	1	0	1
b	1	-	1	1	0
c	1	1	-	2	1
d	0	1	2	-	1
e	1	0	1	1	-

Table 3: Co-citation matrix derived by the occurrence matrix in Fig. 1.

Note that both the co-occurrence matrices are squared and symmetrical, while the original occurrence matrix is rectangular and asymmetrical.[1] In the bibliographic coupling matrix in Table 2, the *citing* publications C and E have the highest bibliographic coupling, whereas in the co-citation matrix in Table 3, the *cited* publications *e* and *d* have the highest co-citation strength.

1.2. Normalization

Once all the co-occurrence values have been computed, the next step in the production of the science map is to *normalize* the raw values, i.e., to replace the raw co-occurrences with statistical processing. As we saw above, the bibliographic coupling and the co-occurrence strengths are used as proxies of the similarity between publications. What we want from them, thus, is to represent similarity accurately. The problem with raw co-occurrences is that they are not very good at doing that, since they are affected by a "size effect", which produces a distorted picture of the real similarities (van Eck & Waltman, 2009).

To see this, suppose that article A and article B are very similar in content. Suppose also that article A contains ten times as many references as article B. Other things being equal, one would expect article A to share more references with other articles in the same research field as article B. Article A, thus, would have more raw co-occurrences with other research articles than B. However, this does not indicate that article A is more similar to other articles than article

[1] Note also that, in both cases, it is difficult to interpret the meaning of the values lying on the diagonal (Ahlgren, Jarneving, & Rousseau, 2003).

Eugenio Petrovich, Emiliano Tolusso

B. It only indicates that article A has more references than B. Therefore, the raw co-occurrences are not a good indicator of similarity, since they reflect not only the similarity, but the *size* of the items as well.

There exist different options to avoid such distortion by *normalizing* the raw values. In science mapping, the most frequently used include the cosine (the most popular one), the association strength (used in the VOSviewer tool, see below), the inclusion index, and the Jaccard index (van Eck & Waltman, 2009).[1] Clearly, depending on the chosen normalization, the final maps will be different.

Once the normalized version of the co-occurrence matrix is obtained, the next, crucial step is to *visualize* the information contained in it, using a representation that is more intelligible by humans than matrices of numbers, i.e., the *science map.*

1.3. Mapping and visualization

The easiest way to do so is to produce the network version of the matrix, using some graphical strategy to represent the strength of the links. A classic solution is to set the width of the link proportional to the strength (see Fig. 2 and Fig. 3) or to report the values as an index close to the link. This resulting visualization is known as a *graph-based science map* (Waltman & van Eck, 2014).

When interpreting this type of map, the most important thing to keep in mind is that the *distance* between the nodes is, in general, *not* significant. The similarity between the nodes is expressed by the width of the link, not by their distance on the plane. In software for network analysis and visualization, such as Pajek,[2] which produce graph-based visualizations, the distance between the nodes on the map does not reflect similarities. Rather, the nodes are placed in

---

[1] Cosine similarity: $cos = \frac{c_{ij}}{\sqrt{s_i s_j}}$. Association strength: $as = \frac{c_{ij}}{s_i s_j}$. Inclusion index: $inc = \frac{c_{ij}}{min(s_i, s_j)}$. Jaccard index: $J = \frac{c_{ij}}{s_i + s_j - c_{ij}}$. Where c_{ij} is the number of co-occurrences of the publications i and j, s_i is the total number of co-occurrences of publication i, and s_j is the total number of co-occurrences of publication j. In the past, also the Pearson's r, a common statistical measure of correlation, was frequently used, but it was demonstrated that, as a *similarity* measure, it is flawed (Ahlgren et al., 2003).

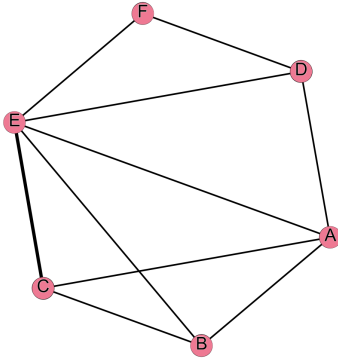[2] http://mrvar.fdv.uni-lj.si/pajek/.

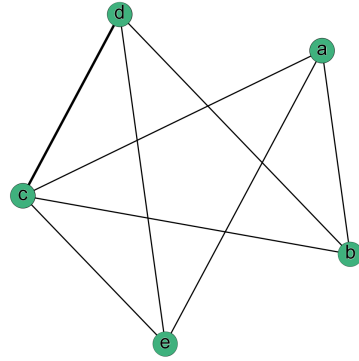Fig. 2: Graph-based visualization of the bibliographic coupling matrix. Nodes are citing publications.

Fig. 3: Graph-based visualization of the co-citation matrix. Nodes are cited publications.

order to enhance the readability of the network (for instance, by avoiding the overlap between the nodes' labels).[1]

A more interesting kind of science map, however, attempts to follow more closely the metaphor of the geographical map, by placing the nodes on the plane according to their similarity. This kind of map is called *distance-based* (Waltman & van Eck, 2014). Ideally, the distance of the nodes on the map (i.e., their Euclidean distance) should be inversely proportional to their similarity, so that similar nodes are placed close in the space and dissimilar nodes distant from each other.

However, as cartographers know well, it is impossible to fulfill this task when the object we want to represent has a higher number of *dimensions* than our representation. Think of the Earth and a geographical map: since the Earth is a 3-dimensional object while the map is in two dimensions only, it is impossible to represent our planet on a map without introducing some degree of *distortion* (e.g., the exaggerated size of Greenland in maps based on Mercator projection).

---

[1] The most common algorithms for doing so are known as Kamada-Kawai and Fruchterman-Reingold.

Eugenio Petrovich, Emiliano Tolusso

The same, but even worse, happens for science maps. In the case of the Earth, which is a 3-dimensional object, any point on its surface can be fully described by a vector of three coordinates. However, in the case of our matrices, any publication is described by a vector of coordinates, the length of which is equal to the number of other publications in our set. For instance, to fully individuate the "position" of the publication A of our example in relation to the others, we needed 5 coordinates! This means that each publication can be conceived as a *point in a multi-dimensional* structure, which is strictly *impossible* to represent perfectly on a 2-d map. A certain degree of distortion will always be present.

Fortunately, statisticians have elaborated several techniques to address this problem, technically known as 'dimensionality reduction'. The main ones are a family of techniques called MDS, which stands for Multi-Dimensional Scaling (Borg & Groenen, 2010).[1] Such techniques allow to find the 2-d representation of an n-dimensional space which minimizes the degree of distortion between the real n-dimensional object and its 2-d representation:

The goal is to find a new representation for the N objects as k-dimensional vectors, where $k < d$ such that the interim proximity nearly matches the original similarities or dissimilarities. Stress is the most common measure of how well a particular configuration reproduces the observed distance matrix. (Chen, 2013, p. 114)

It is important to note that, depending on the MDS technique chose, the resulting 2-d visualization will be different, since the distances will be calculated in different ways. Moreover, certain techniques tend to produce "artifacts", such as quasi-circular layout, or to place the most important items in the center of the map (van Eck, Waltman, Dekker, & van den Berg, 2010).

1.4. Clustering

After the 2-d visualization of the similarity matrix is obtained, a common step in science mapping is to detect the presence of *communities* in the network, i.e., groups of similar nodes. This is done by using statistical techniques of *clustering*,

---

[1] Factor analysis and Principal Component Analysis (PCA) are sometimes used for the same aim of MDS.

a family of various techniques that are able to sort objects (i.e., data-points) into disjoint groups, called clusters. In network science, the most popular technique for clustering is based on the modularity function, but there exist several other clustering methods, such as hierarchical clustering (Hennig, Meilă, Murtagh, & Rocci, 2016). In science maps, clusters are usually represented by colors, so that the information about clusters can be superimposed on the network layout, enhancing the readability and interpretation of the map.

Once again, different clustering methods or variations in the parameters of the clustering functions (e.g., resolution) will result in different clusters and, thus, different science maps. In fact, the main goal of this relatively technical *excursus* on science mapping was to show that science maps are not a simple "photograph" of the "real structure" of a citation network. Like any other map, they are the result of two elements: the features of the mapped object on the one hand, and the methodological choices of the mapper, on the other.

When we proceed to the final step of science mapping, namely the interpretation of the map based on our expert knowledge of the field, we should always remember these *caveats* and clarify our methodological choices to the readers.

1.5. Higher-level maps

The same methodology we described above can be extended to other bibliometric units. For instance, co-citations can be calculated between authors (author co-citation analysis (White & Griffith, 1981)) or between journals (journal co-citation analysis, (McCain, 1991)). The corresponding maps will display authors and journals, respectively, as nodes of the network. The higher-level maps based on relationships between journals are used to investigate science at the level of fields, which contain millions of publications and thousands of journals, or even to map the entire scientific literature (Boyack, Klavans, & Börner, 2005; Leydesdorff & Rafols, 2009).

1.6. Term maps

As we saw above, bibliometric maps use the citations between publications as input. Another common type of science map uses instead the *textual meta-data* of scientific publications, namely their titles, abstracts, and keywords. Clearly, these textual properties of scientific publications are a very rich source of information about their content. Thus, they can be used to obtain fine-grained pictures of the structure and dynamics of scientific knowledge (Callon, Courtial, Turner, & Baui, 1983).

The workflow for producing a term map is not very different from what we saw for bibliometric mapping. Citation indexes store also the textual meta-data of publications. Once the target publications are singled out, the titles, abstracts, and keywords can be easily retrieved. Their processing then follows the standard steps of NLP (Natural Language Processing).[1] Irrelevant words (called 'stop words') such as 'and', 'or', and so on, are removed, and noun-phrases (i.e., sequences of nouns plus adjectives) are extracted.

Usually, not all the extracted noun-phrases are equally relevant. Compare, in the scientific literature, the different relevance of a specific term like 'cardiovascular' with that of a generic term like 'paper'. Even if the former is likely to occur *fewer* times than the latter, it carries more information about the topic of a research field than the other. To retain only the 'specific' terms and remove the 'generic' ones, different statistical techniques are available. The most popular is to calculate the *tf-idf* (which stands for 'term frequency—inverse document frequency') of each term, but other approaches are possible.[2]

Once the relevant noun-phrases are selected, the occurrence matrix (publications × terms) is constructed (see Table 4) and transposed to the co-occurrence matrix of terms (Table 5). Each cell of the matrix tells in how many publications the two noun-phrases occur together.

---

[1] An introduction to text mining techniques can be found in (Taheo, 2018).
[2] VOSviewer, the tool we used to produce the term maps of human geography, employs a different technique for selecting the most relevant noun-phrases, based on the comparison of co-occurrence distributions (van Eck & Waltman, 2011).

	"abcd"	"efgh"	"ilmn"
A	0	0	1
B	1	0	1
C	1	0	0
D	0	1	0

Table 4: Example of term occurrence matrix. Rows are publications, columns are relevant noun-phrases.

	"abcd"	"efgh"	"ilmn"
"abcd"	-	0	1
"efgh"	0	-	0
"ilmn"	1	0	-

Table 5: Example of term co-occurrence matrix derived from Table 4.

Once again, the raw values are statistically processed to correct for differences in sizes between the items, to avoid that publications with longer abstracts or titles distort the similarity measures.

From the normalized co-occurrence matrix onward, the processing pipeline is basically the same as in bibliometric mapping: the n-dimensional information contained in the similarity matrix is turned to a 2-dimensional representation by dimensionality reduction techniques and clustering is applied to detect groups of similar terms. Lastly, the interpretation of the maps is developed based on the expert knowledge of the field under study.

After this relatively technical *excursus*, we hope that the results of our two mapping experiments, that we present in the following sections, will be easier to assess.

2. First mapping experiment: Late Analytic Philosophy

Even if there are good historiographical arguments to refuse the label "analytic philosophy" and to contest any definition of analytic philosophy that has been provided in the literature (see, for instance, (Beaney, 2013; Glock, 2008)), the existence of a social group of contemporary philosophers sharing a common "analytic" tradition can be hardly denied. Social practices such as job offers, journal choices, editorial policies, and publication styles attest to the presence of such a group (Preston, 2010). Therefore, trying to map analytic philosophy starting from its main "social traces", namely the publications in analytic journals, seems to be a reasonable endeavor.[1]

In this first mapping experiment, our goal is to map the structure and evolution of recent analytic philosophy by mapping the research articles published in five journals, that are deemed by analytic philosophers themselves as the "best" generalist journals in their field:[2]

- *Philosophical Review* (PhR)
- *Journal of Philosophy* (JPh)
- *Mind*
- *Noûs*
- *Philosophy and Phenomenological Research* (PPR)

The time span chosen for the analysis is 1985-2015, which falls within what Tripodi and Bonino call "late analytic philosophy", that is the most recent phase of analytic philosophy (Bonino & Tripodi, 2018; Tripodi, 2015). We retrieved on Web of Science Core Collection *all the publications* published by these journals in that period, obtaining a corpus of 11167 records. Table 1 reports some descriptive statistics of the dataset.

---

[1] For a science mapping experiment investigating a field the existence of which is frequently contested, i.e., Integrated History and Philosophy of Science, see (Weingart, 2015): "The best we can do to empirically show that a certain social structure exists is to study their institutional traces. In the case of academia, that means looking at publications, citations, institutional affiliations, mentorship relationships, conference attendance, and so on" (Weingart, 2015, p. 203).
[2] See (Petrovich & Buonomo, 2018) for a discussion of the reasons behind the choice of these journals.

Rank	Journal	Publications (1985-1994)	Publications (1995-2004)	Publications (2005-2015)	Tot	Avg pubs per year
1	PhR	821 [7.4%]	722 [6.5%]	440 [3.9%]	1983 [17.8%]	66,1
2	Noûs	755 [6.8%]	606 [5.4%]	554 [5.0%]	1915 [17.1%]	63,8
3	JPh	593 [5.3%]	416 [3.7%]	387 [3.5%]	1396 [12.5%]	46,5
4	Mind	750 [6.7%]	1064 [9.5%]	1268 [11.4%]	3082 [27.6%]	102,7
5	PPR	752 [6.7%]	1062 [6.7%]	977 [8.7%]	2791 [25.0%]	93,0
	Total	**3671 [32.9%]**	**3870 [34.7%]**	**3626 [32.5%]**	**11167 [100%]**	372,2

Table 6. Analytic Philosophy dataset. In brackets percentage values.

The data were divided into three timespans for the historical analysis (see below). Each of the three sub-timespan accounts for around one-third of the publications. The journals, however, are not equally represented, due to differences in publication frequencies (the number of issues per year, the average number of articles per issue, etc.). In particular, the most prolific journals *Mind* and *Philosophy and Phenomenological Research* account alone for more than half of the total publications.

Note that in the follow-up study by Buonomo and Petrovich, we restricted the publication type to *research articles*, excluding book reviews, editorials, etc., that are instead included in this dataset. The choice of focusing or not on research articles only produces significant differences in the dimension of the datasets, especially in the case of the *Philosophical Review*, which publishes mainly book reviews. It is very interesting to compare the different maps of analytic philosophy which result from different choices in the dataset. In particular, in the inclusive dataset we examine here, the cluster of philosophy of language is placed at the center of the map, at the crossroads of the other philosophical sub-disciplines. In the restricted map, it is in between the metaphysics and philosophy of mind clusters (Petrovich & Buonomo, 2018, fig. 1).

2.1. Results

The publications retrieved were mapped with the technique of *co-citation analysis* using the software VOSviewer, a freely available science mapping tool

Eugenio Petrovich, Emiliano Tolusso

developed by Ludo Waltman and Nees Jan van Eck at the Center for Science and Technology Studies (CWTS) in Leiden (van Eck & Waltman, 2010). VOSviewer extracts the citation network from the Web of Science records, calculates the co-occurrence matrix and then normalizes raw co-occurrences data using the association strength measure (see above).[1] It produces distance-based visualizations, in which the relative distance between the nodes on the 2-d map reflects the similarity of the publications (with the lowest stress possible), using a special technique (the VOS technique) which does not produce the visual artifacts that are sometimes introduced by standard MDS methods (van Eck et al., 2010). VOSviewer also groups publications with a clustering technique based on a modified version of the modularity function (Waltman, van Eck, & Noyons, 2010).

Remember that in co-citation analysis, the items on the map are the *cited items*, i.e., the cited references of our target publications. The map thus represents the structure of the literature cited in our corpus, not the structure of the corpus itself. To investigate the latter, i.e., the *citing* publications, bibliographic coupling would be the right technique.

In the map in Fig. 4, the size of the nodes is proportional to the number of citations they receive in the corpus, i.e., the number of times they appear in the bibliographies of the corpus publications. Note that only the top-100 most cited references are shown on the map.[2] The width of the link between two nodes is proportional to the number of co-citations of the publications connected by the link. Only links with 10 co-citations or more are shown on the map. Lastly, the color of the nodes corresponds to the cluster.[3]

The map shows some interesting features of the dataset. First, some clear clusters can be recognized, showing that late analytic philosophy literature is grouped into different areas. Indeed, each cluster can be easily mapped to a sub-specialty of analytic philosophy. The red cluster in the northern part of the map relates to the philosophy of mind, the yellow eastern cluster to moral and political philosophy, the southern green cluster to metaphysics, and the western

---

[1] Association strength is the default option for normalization, but other are possible. We tested other methods and the overall structure of the maps remained almost the same, providing a certain degree of robustness to our analyses.

[2] Therefore, the citation threshold for being included on the map was 44 citations.

[3] The resolution clustering parameter was set to 1.00.

light blue to epistemology. Interestingly, at the center of the map, lies a violet cluster containing several works in the field to which analytic philosophy was originally closely associated, namely the philosophy of language.
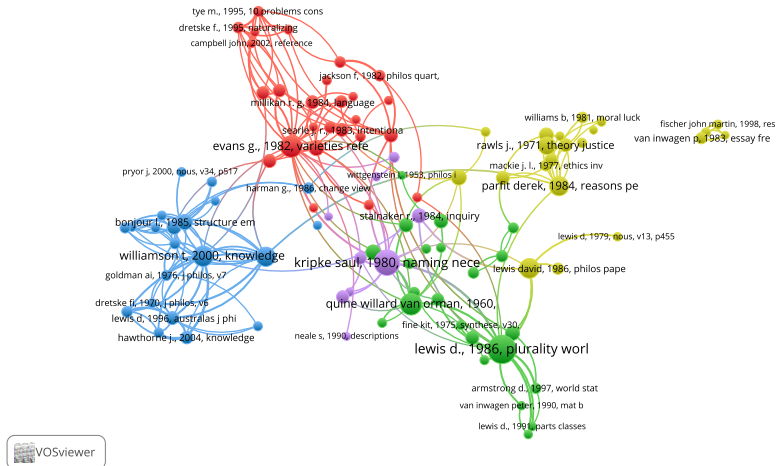

Fig. 4: Overall map of analytic philosophy dataset (timespan: 1985-2015).

It seems, therefore, that the *organizing principle* of the late analytic literature lies in *sub-disciplinary differences*, rather than, say, major authors, philosophical theories or meta-philosophical principles.

A further interesting feature of the map is its center-periphery structure. Remember that in distance-based visualizations, relative distances are meaningful, therefore they can be interpreted. In our case, it seems that the center of the map hosts the 'paradigms' of the analytic tradition, such as Quine's *Word and Object* and Kripke's *Name and Necessity* (Levy, 2003). The periphery, on the other hand, is populated by specialized sub-disciplines. Apparently, the farther a document appears on the map, the more specialized its content is.

Considering the top-100 documents represented by the nodes of the map, we note that no 'Continental' author or publication is present (Table 2). The isolation between the two 'camps' of contemporary philosophy seems therefore

Eugenio Petrovich, Emiliano Tolusso

still alive, notwithstanding the efforts, from both sides of the divide, to overcome the gap (Biletzki, 2001).

David Lewis has the highest citation score and is the author of two out of ten publications in the top-10, *On the Plurality of Worlds* and *Counterfactuals.* By knowing the reaction of the analytic community to the former, it is easy to understand why high citation counts should not easily be interpreted as a sign of *agreement* or *success* of a philosophical theory. In fact, the main philosophic thesis advanced in *On the Plurality of Worlds* (modal realism) has been *criticized* by most analytic philosophers, not accepted as an uncontested philosophical "achievement". From this point of view, citation scores in philosophy should be interpreted very differently than in the sciences, when they are frequently equated with scientific quality or scientific consensus on a certain claim.[1]

Rank	Author	Title	Year	Links	Co-cits	Cits
1	Lewis, David	On the Plurality of Worlds	1986	80	595	260
2	Kripke, Saul	Naming and Necessity	1980	89	552	223
3	Evans, Gareth	Varieties of Reference	1982	82	558	176
4	Quine, W.V.O.	Word and Object	1960	84	472	172
5	Williamson, Timothy	Knowledge and Its Limits	2000	82	467	163
6	Lewis, David	Counterfactuals	1973	78	327	156
7	Parfit, Derek	Reasons and Persons	1984	64	293	151
8	Nozick, Robert	Philosophical Explanations	1981	86	463	150
9	Rawls, John	Theory of Justice	1971	61	179	128
10	Davidson, Donald	Essays on Actions and Events	1980	82	308	110

Table 7. Top ten most cited publications (timespan: 1985-2015).

Interestingly, no journal article appears in the top-10 of the most cited publications. Indeed, of the 100 most cited documents, only 18 (less than one on five) are journal articles. Thus, although analytic philosophers usually present the journal as their favorite publication outlet—sometimes highlighting this publication choice as a sign of the fact that analytic philosophy is more 'scientific'

[1] To be sure, this is a naïve interpretation of citation scores *also in the sciences.* See (Aksnes, Langfeldt, & Wouters, 2019) for an introduction to the debate on the limits of evaluative bibliometrics.

than other philosophical traditions—in their cited references they are quite loyal to the standard humanities practices. Consider, however, that some of the most-cited documents are collections of articles previously published as journal articles (e.g., the *Essays on Actions and Events* by Donald Davidson), so that the underrepresentation of articles could be overstated.

Until now, we investigated the structure of the entire dataset, producing a map of analytic philosophy, which aggregates all the corpus publications. In the next three maps, we divided the corpus into three sub-periods: 1985-1994, 1995-2004, 2005-2015. This allows us to carry out a *longitudinal mapping exercise*, which uncovers the temporal evolution of the field. The maps are shown in Fig. 5 (a, b, c). They are all co-citation networks, with a citation threshold of 20 citations, a link strength threshold of 4 co-citations, and the clustering resolution parameter set to 1.0.

The three maps present a *clear pattern*: the quite sparse and unstructured network of the first decade becomes, in the last decade, an organized network with three definite sub-areas. The clusters become more and more compact, highlighting that the sub-disciplinary literatures become more and more delineated.[1]

We interpret this pattern as the empirical evidence of a process of *specialization* that occurred within analytic philosophy during the last thirty years, confirming the qualitative perception of analytic philosophers and historians of analytic philosophy (Marconi, 2014; Soames, 2005; Tripodi, 2015). The maps show also the birth of the new sub-specialty of analytic epistemology in 1994-2004, and its consolidation in the next decade.

---

[1] Maximilian Noichl provided a very nice quantitative follow-up of this qualitative assessment of specialization in a post published on the DR2 blog (`https://dr2blog.hcommons.org/2019/11/04/a-short-and-informal-replication-of-petrovich-and-buonomo-2018/`). Noichl is also the author of a detailed map of philosophy made with bibliographic coupling (the pre-print is available at `http://philsci-archive.pitt.edu/16470/1/modeling_recent_philosophy2.pdf`).
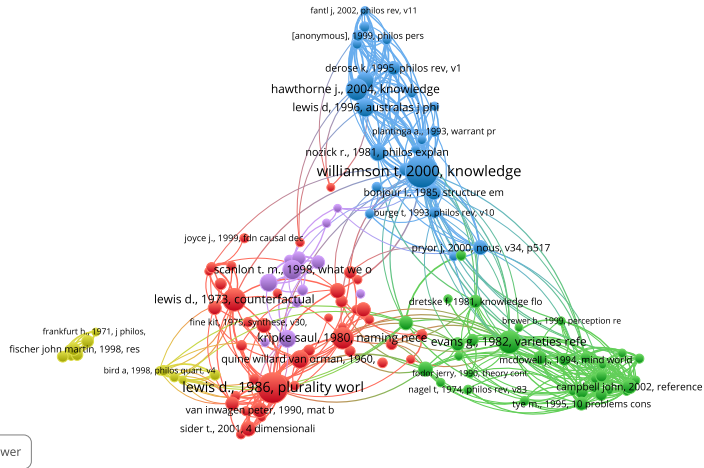
Eugenio Petrovich, Emiliano Tolusso
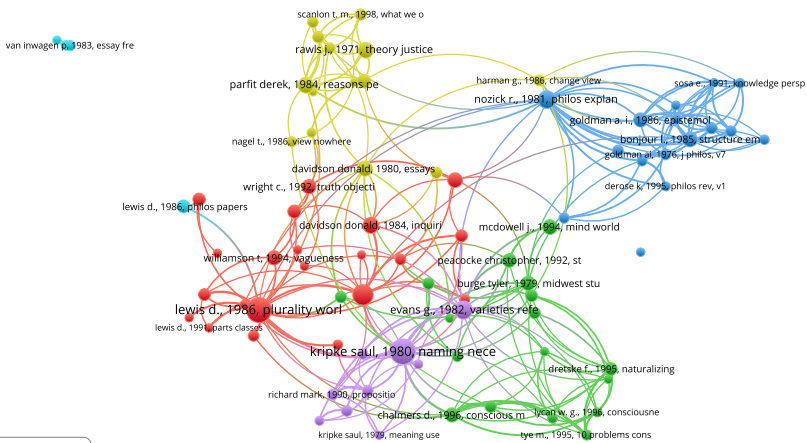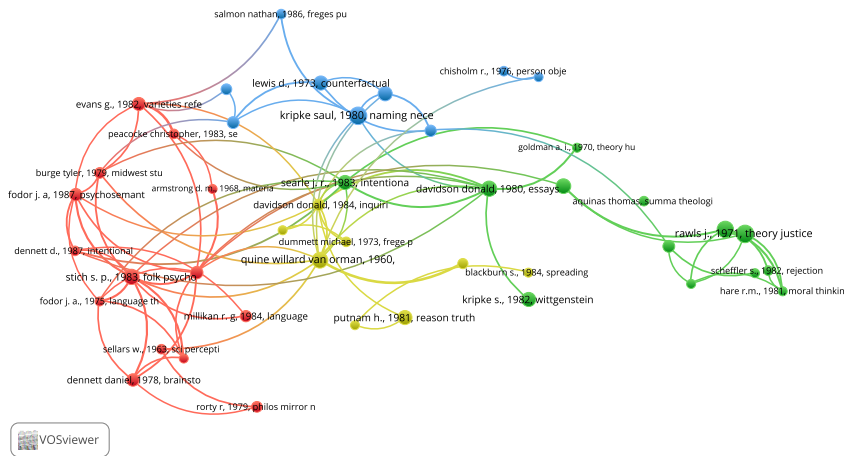
Fig. 5 (a, b, c). Longitudinal mapping of analytic philosophy in three timespans. A: 1985-1994, B: 1995-2004, C: 2005-2015.

3. Second mapping experiment: Human Geography

For our second experiment, we focused on a different task. Our main interest was to explore how science mapping can be actively used to make sense of the complex cognitive content of a broad research field. Thus, we turned to term-based science maps.

The case of human geography constitutes in our eyes a very fitting example. Human geography is a complex and heterogeneous field of research, usually dealings with different research topics, employing a vast array of research methodologies and a varied lexicon. Given these specific conditions, an analysis of the textual meta-data seems very promising in terms of experimenting the power of term-mapping.

Once again, we based our analysis on Web of Science Core Collection. We used the Scimago Journal Ranking[1] to individuate the top journals in the field,

---

[1] https://www.scimagojr.com/

Eugenio Petrovich, Emiliano Tolusso

selecting the five journals in the category "Social Science—Geography" with the highest h-index in the year 2015:

- *Progress in Human Geography* (PHG)
- *Global Environmental Change: Human and Policy Dimensions* (GEC)
- *Transactions of the Institute of British Geographers* (TIBG)
- *Journal of Economic Geography* (JEC)
- *Economic Geography* (EG)

Documents were divided into three different ten-year timespans to capture the development of the field over time. Descriptive statistics of the dataset are displayed in Table 7.

Rank	Journal	Publications (1985-1994)	Publications (1995-2004)	Publications (2005-2015)	Tot	Avg pubs per year
1	PHG	1136 [13.2%]	1132 [13.2%]	1080 [12.6%]	3348 [39%]	108
2	GEC	154 [1.8%]	361 [4.2%]	1028 [12%]	1543 [17.9%]	49.8
3	TIBG	667 [7.8%]	538 [6.2%]	480 [5.6%]	1685 [19.6%]	54.3
4	JEG	0 [0%]	99 [1.1%]	523 [6.1%]	622 [7.3%]	20.1
5	EG	518 [6%]	443 [5.1%]	428 [5%]	1389 [16.2%]	44.8
	Total	**2475 [28.8%]**	**2573 [30%]**	**3539 [41.2%]**	**8587 [100%]**	**286.2**

Table 7: Human geography dataset. In brackets percentage values.

3.1. Results

The whole dataset (1985-2015) was converted into the first map (Fig. 5) . Terms (noun-phrases) serve as the node of the network, while the link width between terms represents the number of their co-occurrences. Terms appearing closer in the map have high co-occurrence values, meaning that they are frequently coupled in titles and abstracts of our corpus.
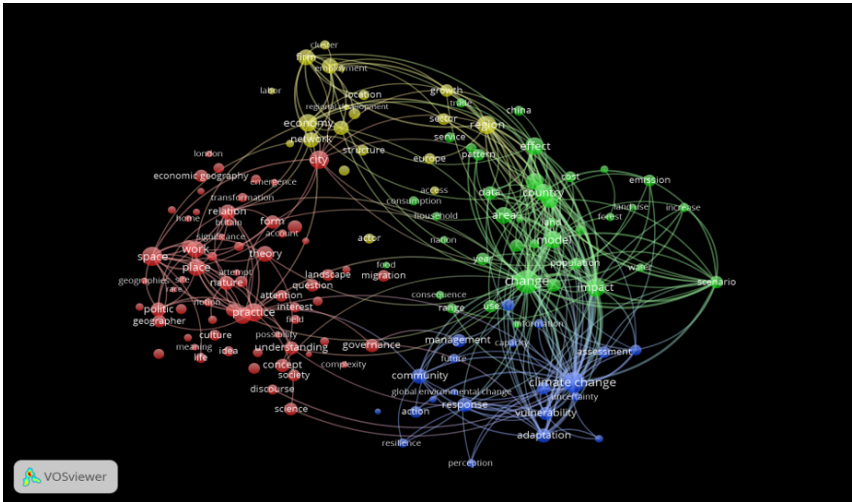
Fig. 5: Overall map of human geography dataset (timespan: 1985-2015).

The *thematic structure* of contemporary research in human geography is easily readable in the map. Specifically, four main clusters are recognizable. The yellow cluster in the northern part of the map represents the sub-discipline of economic geography, the red cluster in the western part stands out as a compendium for social geography, while the eastern cluster colored in green is representative of the environmental geography field. The last one, colored in blue, represents a specific subfield dealing with climate change, a topic faced so often in our sample to be recognizable as a cluster on its own, telling explicitly of the emerging structures and patterns in the literature.

A compelling feature of the map is the absence of a real center. The four clusters gravitate around autonomous centroids, with different levels of integration with each other. The resulting structure is donut-shaped, showing the lack of a real thematic focal point in the broad discipline and its very distinctive specializations. However, the widespread presence of inter-cluster connections highlights, to a certain extent, the presence of a coherent lexicon shared by the different branches of human geography. Table 8 shows the 10 most common terms in the corpus.

Rank	Term	Cluster	Links	Occurrences	Avg. Pub. Year
1	change	2	145	1591	2005
2	climate change	3	143	1307	2008
3	space	1	143	1063	2005
4	impact	2	145	1032	2007
5	city	1	143	1009	2004
6	practice	1	145	975	2007
7	region	4	145	954	2006
8	economy	4	145	926	2004
9	model	2	145	922	2005
10	country	2	145	902	2006

Table 8. Top ten most recurring terms (timespan: 1985-2015).

As for the historical reconstruction, three maps were constructed based on three subperiods (Fig. 6, a, b, c). The main aim of these maps is to show the morphological evolution of the term network overtime.
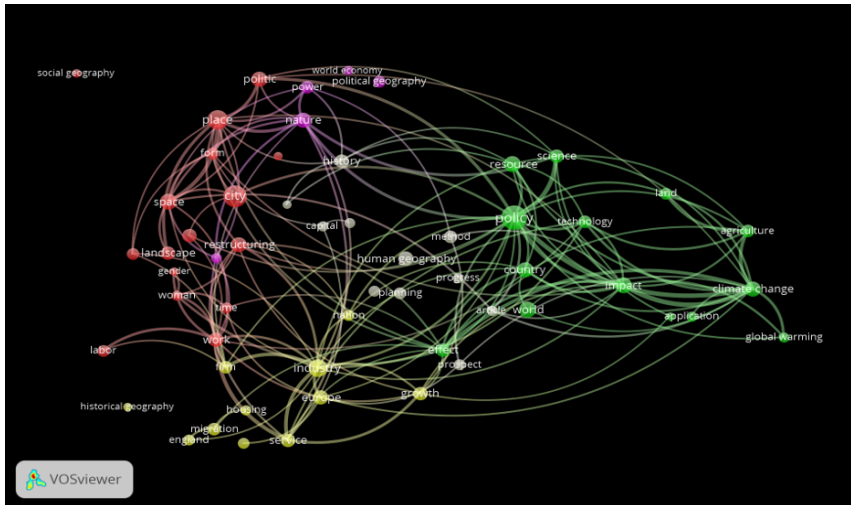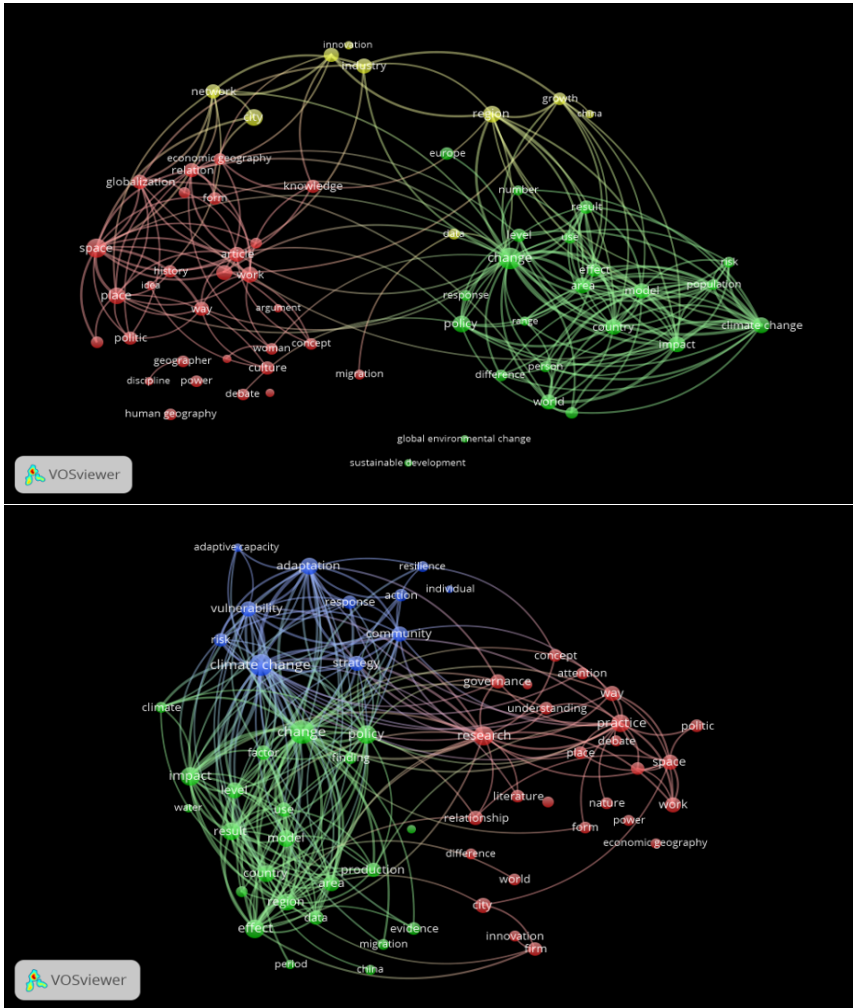
Fig. 6: (a, b, c). Longitudinal mapping of human geography in three timespans. A: 1985-1994, B: 1995-2004, C: 2005-2015.

Eugenio Petrovich, Emiliano Tolusso

The longitudinal mapping highlights the unfolding of the specialization process: the maps display a clear pattern of clusterization, with the gradual definition of subclusters and specialized terms increasingly occurring.

Particularly notable is the gradual establishment of the cluster of climate change, emerging clearly in the last time span as a strong branch of environmental geography. The two subfields are still highly interconnected, as the latter deals with the everyday conundrums of global environmental change, but the salience of climate change as a topic proves to be extraordinarily relevant for human geography, detaching from the fold.

Specialization, however, might be matched also by the opposite process of *unification*: the cluster of social geography, for instance, seemingly coalesces with the cluster of economic geography, which stands as an independent family of terms only in the first map (Fig. 6 a).

Besides the dynamics of specialization and unification, the maps show also examples of clusters perduring overtime. The main terms representing social and environmental geography are stable through the different reconstructions, suggesting the existence of two affirmed strains of research that are hardly affected by the passing of three decades.

Clusters, therefore, are to be intended as very dynamic entities, the presence and detectability of which is not necessarily constant, as they evolve in different directions before our eyes. In general, we believe that such trends are coherent with the recent dendrogram of geography's subfields, but a more in-depth, qualitative evaluation is needed, and still to be performed.

4. Main limits of science mapping as applied to the humanities

Science mapping allows us to visualize the structure and dynamics of knowledge produced in the humanities from a new and different perspective, but it is subject to some limitations that must be considered (Nederhof, 2006). At least four of them deserve attention.

First, the Web of Science database (like Scopus) does not include monographs, which are still one of the most important publication outlets in the humanities. However, it is crucial not to misunderstand this point. The lack of monographs

in WoS means that references cited in monographs are not indexed, not that monographs do not appear at all. Indeed, monographs do appear in so far as the citing articles contain citations pointing to them. Therefore, monographs are part of the set of *cited* items but are not part of the set of *citing* items of WoS.

Secondly, the scope of the data that can be accessed from Web of Science (and Scopus) depends on the terms of the agreement of the subscription. In our case, for instance, we could not access Web of Science before 1980 due to the conditions of the University of Milan's agreement. The fact that bibliographic and bibliometric data lie behind a paywall poses strong limitations to the reproducibility of citation analyses. Hopefully, the launch of the free citation database Dimensions will change the situation for the better.

Thirdly, Web of Science's subject categories are often inaccurate for the humanities. Sometimes, journals are misplaced in the wrong category. Hence, they should be used with great care when data are selected during the field delineation step.

Lastly, Web of Science indexes mainly journals in English. Non-English journals are under-represented. Scopus' linguistic coverage is larger, but still, for the humanities, it is far for being complete. These limitations significantly reduce the possibility of studying national scholarly traditions in the humanities.[1]

Despite these limitations, however, we think that science mapping is a valuable tool for exploring knowledge dynamics in the humanities, especially for recent periods, characterized by a massive production of scholarly literature that would be hard to investigate by traditional close reading methods. Therefore, we would recommend historians of ideas to include it in their methodological toolbox.

---
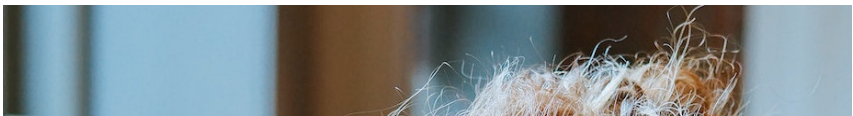
[1] For instance, the *Rivista di storia della filosofia*, an Italian journal publishing high-quality studies in the history of philosophy, is not covered by Web of Science Core Collection.

Eugenio Petrovich, Emiliano Tolusso

References

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, *54*(6), 550-560. `https://doi.org/10.1002/asi.10242`

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*, *9*(1), 215824401982957. `https://doi.org/10.1177/2158244019829575`

Beaney, M. (Ed.). (2013). *The Oxford handbook of the history of analytic philosophy*. Oxford: Oxford University Press.

Biletzki, A. (2001). Introduction: Bridging the Analytic-Continental Divide. *International Journal of Philosophical Studies*, *9*(3), 291-294. `https://doi.org/10.1080/09672550110058867`

Bonino, G., & Tripodi, P. (2018). Introduction to 'History of Late Analytic Philosophy'. *Philosophical Inquiries*, *6*(1), 7-14. `https://doi.org/10.4454/philinq.v6i1.206`

Borg, I., & Groenen, P. J. F. (2010). *Modern multidimensional scaling: Theory and applications*. New York; London: Springer.

Börner, K. (2010). *Atlas of science: Visualizing what we know*. Cambridge, Mass.: MIT Press.

Börner, K., Chen, C., & Boyack, K. W. (2005). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, *37*(1), 179-255. `https://doi.org/10.1002/aris.1440370106`

Börner, K., Theriault, T. N., & Boyack, K. W. (2015). Mapping science introduction: Past, present and future: Mapping Science Introduction: Past, Present and Future. *Bulletin of the Association for Information Science and Technology*, *41*(2), 12-16. `https://doi.org/10.1002/bult.2015.1720410205`

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351-374. `https://doi.org/10.1007/s11192-005-0255-6`

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191-235. `https://doi.org/10.1177/053901883022002003`

Chen, C. (2013). *Mapping scientific frontiers: The quest for knowledge visualization* (Second Edition). London: Springer.

Chen, C. (2017). Science Mapping: A Systematic Review of the Literature. *Journal of Data and Information Science*, *2*(2), 1-40. `https://doi.org/10.1515/jdis-2017-0006`

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Jour-*

nal of the American Society for Information Science and Technology, *62*(7), 1382-1402. `https://doi.org/10.1002/asi.21525`

Colavizza, G. (2018). A diachronic study of historiography. *Scientometrics*, *117*(3), 2117-2131. `https://doi.org/10.1007/s11192-018-2934-0`

Glock, H.-J. (2008). *What is analytic philosophy?* Cambridge, UK; New York: Cambridge University Press.

Hennig, C. M., Meilă, M., Murtagh, F., & Rocci, R. (Eds.). (2016). *Handbook of Cluster Analysis*. Boca Rato: CRC Press, Taylor & Francis Group.

Kessler, M. M. (1963). Bibliographic coupling extended in time: Ten case histories. *Information Storage and Retrieval*, *1*(4), 169-187. `https://doi.org/10.1016/0020-0271(63)90016-0`

Kreuzman, H. (2001). A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science. *Scientometrics*, *51*(3), 525-539.

Laurens, P., Zitt, M., & Bassecoulard, E. (2010). Delineation of the genomics field by hybrid citation-lexical methods: Interaction with experts and validation process. *Scientometrics*, *82*(3), 647-662. `https://doi.org/10.1007/s11192-010-0177-9`

Levy, N. (2003). Analytic and Continental philosophy: Explaining the differences. *Metaphilosophy*, *34*(3), 284-304. `https://doi.org/10.1111/1467-9973.00274`

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, *60*(2), 348-362. `https://doi.org/10.1002/asi.20967`

Marconi, D. (2014). *Il mestiere di pensare: La filosofia nell'epoca del professionismo*. Torino: Giulio Einaudi editore.

McCain, K. W. (1991). Mapping economics through the journal literature: An experiment in journal cocitation analysis. *Journal of the American Society for Information Science*, *42*(4), 290-296. `https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<290::AID-ASI5>3.0.CO;2-9`

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, *66*(1), 81-100. `https://doi.org/10.1007/s11192-006-0007-2`

Petrovich, E. (2019). *The Fabric of Knowledge. Towards a Documental History of Late Analytic Philosophy* (Ph.D. Dissertation, University of Milan). Retrieved from `https://air.unimi.it/handle/2434/613334`

Petrovich, E., & Buonomo, V. (2018). Reconstructing late Analytic philosophy: a quantitative approach. *Philosophical Inquiries*, *6*(1). `https://doi.org/10.4454/philinq.v6i1.184`

Preston, A. (2010). *Analytic philosophy: The history of an illusion*. London; New York: Continuum.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265-269. https://doi.org/10.1002/asi.4630240406

Soames, S. (2005). *Philosophical analysis in the twentieth century. Vol. 2: The age of meaning* (Repr., 1. paperback print). Princeto, NJ: Princeton Univ. Press.

Taheo, J. (2018). *Text mining: Concepts, implementation, and big data challenge*. New York, NY: Springer Science+Business Media.

Tolusso, E. (2019). *Alpine Protected Areas in a Changing Climate. A Geography of Science Perspective* (Ph.D. Dissertation, University of Milan). Retrieved from https://tinyurl.com/naturalsciences-ch-tolusso-phd

Tripodi, P. (2015). *Storia della filosofia analitica: Dalle origini ai giorni nostri* (1a edizione). Roma: Carocci editore.

van Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, *60*(8), 1635-1651. https://doi.org/10.1002/asi.21075

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ArXiv:1109.2058 [Cs]*. Retrieved from http://arxiv.org/abs/1109.2058

van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, *61*(12), 2405-2416. https://doi.org/10.1002/asi.21421

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378-2392. https://doi.org/10.1002/asi.22748

Waltman, L., & van Eck, N. J. (2014). Visualizing bibliometric networks. In *Measuring scholarly impact: Methods and practice* (pp. 285-320). Springer.

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*(4), 629-635. https://doi.org/10.1016/j.joi.2010.07.002

Weingart, S. B. (2015). Finding the History and Philosophy of Science. *Erkenntnis*, *80*(1), 201-213. https://doi.org/10.1007/s10670-014-9621-1

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, *32*(3), 163-171. `https://doi.org/10.1002/asi.4630320302`

Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, *42*(6), 1513-1531. `https://doi.org/10.1016/j.ipm.2006.03.016`

Bearded young man consulting a map in a Belfast café. Photo by Toa Heftiba on Unsplash (`https://unsplash.com/photos/HGhIH73WTSE`).