



Modeling the structure of recent philosophy

Maximilian Noichl¹ 

Received: 1 March 2019 / Accepted: 4 September 2019 / Published online: 26 October 2019
© The Author(s) 2019

Abstract

This paper presents an approach of unsupervised learning of clusters from a citation database, and applies it to a large corpus of articles in philosophy to give an account of the structure of the discipline. Following a list of journals from the PhilPapers-archive, 68,152 records were downloaded from the Reuters Web of Science-Database. Their citation data was processed using dimensionality reduction and clustering. The resulting clusters were identified, and the results are graphically represented. They suggest that the division of analytic and Continental philosophy in the considered timespan is overstated; that analytical, in contrast to Continental philosophy does not form a coherent group in recent philosophy; and that metaphors about the disciplinary structure should focus on the coherence and interconnectedness of a multitude of smaller and larger subfields.

Keywords Bibliometrics · Metaphilosophy · Analytic philosophy · Continental philosophy · UMAP · hDBSCAN

1 Introduction

It is clear that philosophers are fond of strong metaphors when describing the structure of their discipline. In particular, talk about the analytical/Continental divide is a common subject of mostly informal interactions between philosophers. This is quite interesting when taken together with the result that “[...] philosophers have substantially inaccurate sociological beliefs about the views of their peers.” (Bourget and Chalmers 2014, p. 489), as it suggests that these metaphors can not necessarily be taken at face value.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11229-019-02390-8>) contains supplementary material, which is available to authorized users.

✉ Maximilian Noichl
noichlmax@hotmail.co.uk

¹ University of Vienna, Vienna, Austria

Therefore it seems sensible to try and accompany them with more data-driven pictures. This article will borrow some young machine learning techniques to give a visual representation of recent philosophy as a whole. This representation is hopefully interesting for various reasons, and is supposed to demonstrate the viability of the employed method as much as anything else, but it will be considered in the context of this paper only with regard to its bearing on the analytic-Continental question.

Imaging investigation into the citation-structure of philosophy is surprisingly sparse. Notable exceptions include Kreuzman (2001) who conducted a co-citation analysis of 62 authors to give a mapping of the division between philosophy of science and epistemology (a result that will be replicated in this study). Healy (2013), has given an account of the citation structure in four highly cited Journals of philosophy, including 2200 articles, and has constructed a network visualization which shows relations between the major publications in the field. Similar work has been done by Higgins and Smith (2013) who have given a citation based view of the ontology-community, based on 2260 articles. And Weingart (2015) has used the methodology of bibliographic coupling on 12,510 articles to identify the connection between history and philosophy of science. There are also several projects that try to give an account of the structure of philosophy by community driven methods, either in the form of surveys (Bourget 2010), or in combination of surveys with text-mining applications (Buckner et al. 2011). Both are potentially very useful to practitioners.

This project differs from community-driven approaches in that it tries to avoid the potential biases of conscious classification by experts in their respective fields through the deployment of unsupervised machine-learning techniques. It is very much in the line of the former, which apply bibliometric techniques to philosophical datasets, and in the general tradition of mapping approaches to science, like Boyack et al. (2005). The standard technique that lies at the heart of this approach is bibliographic coupling, which tries to grasp the relation between two texts by assessing how many sources they have in common. It has been considered the best method for this purposes for a while (Boyack and Klavans 2010). Recent evidence has called that primacy into question (Boyack and Klavans 2017), but it clearly remains among the best methods for the clustering of scientific literature (Waltman et al. 2019).

The main novelty of the method presented in this essay is, that it considers bibliographic-couplings not as singular links between documents, which have to be cut off after a certain threshold in order to be sensibly visualized in a network-structure, but treats them like word-vectors in a text-classification problem: When two texts cite the same third text, this increases the similarity between documents, and the resulting similarity-matrix can then be used to perform dimensionality-reduction for visualization and classification purposes.

2 Methods

2.1 Sampling strategy

As the establishment of the disciplinary boundaries of philosophy is a notoriously hard problem, the adopted sampling strategy was as expansive and inclusive as possible.

As there is some evidence that the highest tiers of common journal rankings do not represent all of philosophy adequately—they seem to favour for example analytical philosophy journals over Continental ones (Schwitzgebel 2012; Katzav 2018)—they could not be used for our sample. PhilPapers, an expansive index of recent philosophy, has compiled a list of 1349 journals of philosophy, which links, at the time of writing to 1,782,816 indexed articles (Bourget 2019). This list has been used in other publications assessing the composition of philosophical literature (Schwitzgebel et al. 2018). The list was downloaded and compiled into a Web of Science (WoS) query, to get the citation data for the individual papers.¹ Journals were removed from the query if they were clearly from the core of another discipline (e.g. *Experimental Psychology*, *Historia*) and had more than a thousand entries in the result of the query, as those journals could be expected to strongly change the results of further analysis.² The result of the query was then ordered by the Web of Science estimates of their citation-counts, and the 87,820 records which had been cited more than three times were kept. It is important to note that the sample that was drawn deviates from what is archived in PhilPapers in important respects. Not only is it comprised near exclusively from anglophone philosophy, but also gravitates towards higher impact journals, as those will have proportionally more articles cited more than three times. A comparison of the distribution of articles in PhilPapers and in our sample can be found in *Online Resource 1*.

2.2 Processing

These records were then processed with the python package *metaknowledge* (McLevey and McIlroy-Young 2017).³ To get a representation of the citation-structure, a standard text-mining approach was employed: Like words in a text-classification-problem, the citations were vectorized using *scikit-learns* *CountVectorizer* (Pedregosa et al. 2011). Then all records were removed which shared less than three citations with any of the other records in the dataset, a step which was necessary to ensure the connectedness of the manifold, which was to be estimated with UMAP (Uniform Manifold Approximation and Projection), yielding a total of 68,152 records, with 1,194,451 citations events to 159,647 unique sources. All further steps are illustrated in Fig. 1. As preliminary dimensionality-reduction has been shown to be beneficial to text-classification-tasks (Shafiei et al. 2007; Underhill et al. 2007), truncated SVD (Singular Value Decomposition) was conducted on the citation vectors for visualisation, a process that reduces noise in the dataset.⁴ If, for example, we find that citation to a certain work by Gareth Evans is strongly correlated with citation to a work by Saul Kripke, we could merge these two into one combined factor without losing a lot of important information, while reducing our dataset by one column. By keeping only some of the principal

¹ The corresponding code is provided as: *Online Resource 2.ipynb*. All online resources are available as a GitHub-Repository: https://github.com/struct-phil/struct_phil.

² A list of the 32 removed journals is provided at the end of *Online Resource 2.ipynb*.

³ The corresponding code for the downsampling is provided in *Online Resource 3.ipynb*. Due to restriction under the WoS-user-agreement, the raw data can not be shared. Direct replication of the whole process should nonetheless be easily accomplishable for anyone with WoS-access.

⁴ The corresponding code for the dimensionality reduction, clustering and generation of the clustering-report is provided in *Online Resource 4.ipynb*.

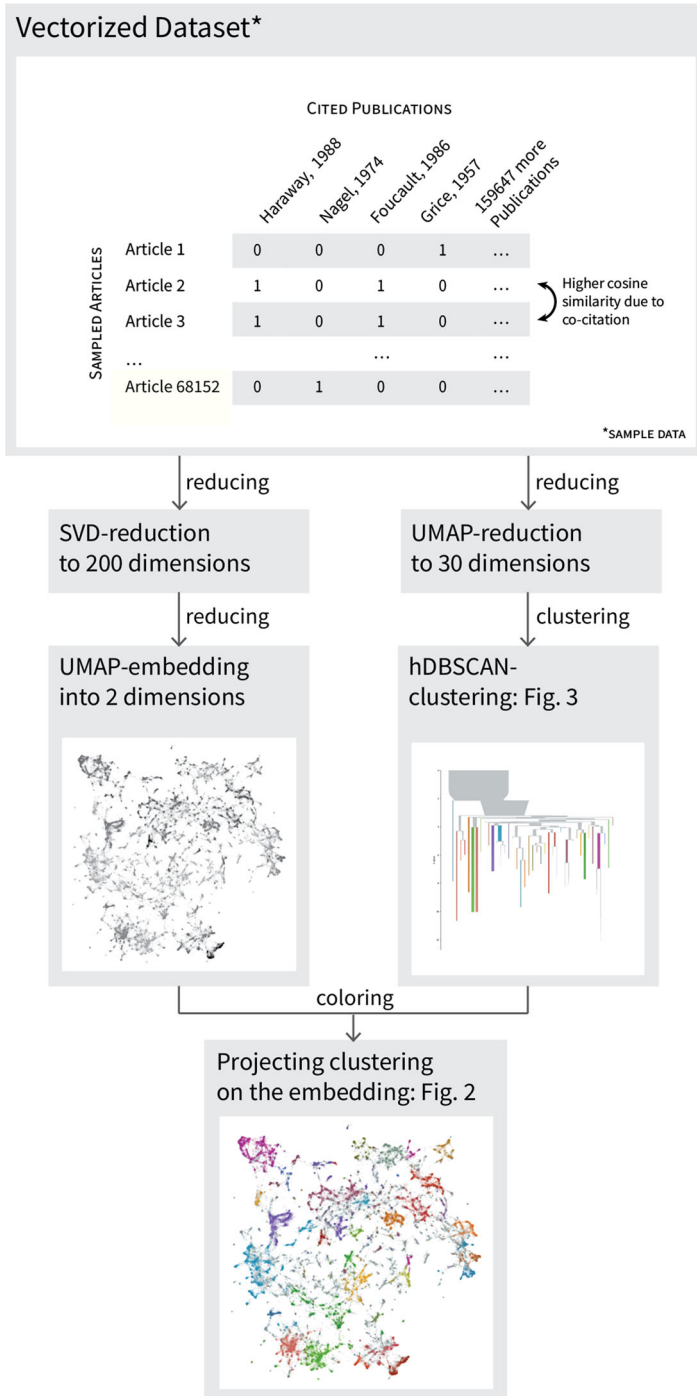


Fig. 1 Illustration of the processing of the citation-vectors with SVD, umap and hDBSCAN and the generation of Figs. 2 and 3. Graphic made with Adobe Illustrator

components (in our case 400 of them), we remove all the variances in the data that do not align themselves with any larger patterns. This makes the two dimensional picture we create afterwards with UMAP much clearer, as a lot of confounding information, is removed. We could skip this step altogether. Doing so would result in a more “feathery” or “cloudy” map, which would leave the general regions nonetheless recognizable. For our clustering we will not conduct SVD to avoid losing data, and apply UMAP directly to the full dataset. UMAP (McInnes et al. 2018), is a recently proposed⁵ non-linear dimensionality-reduction technique, which means that it can grasp much more complicated relations than simple linear correlations. To initiate it, we first need a similarity-measure between our texts. In our case, we choose cosine similarity.⁶ Umap now constructs a weighted k-neighbour graph, which means that for each article, it looks up the $k = 50$ nearest other articles according to our similarity measure, and converts this measure to a connection between the two of a certain strength. Two papers which are very similar in that they both cite e. g. several works by Donna Haraway, would be connected in our graph, and their connection would be very strong. As we do this for all the articles, this graph will give a good approximation of the structure that we are interested in. The next step is simply to give a representation of that graph in a low-dimensional space, which can either be directly perceived by us as a two dimensional mapping, provided in (Fig. 2) or can be comfortably handled by the clustering-algorithm. For this, UMAP uses a force-directed layout: All articles are assigned coordinates, and then moved around like balls connected with springs, which push and pull them until an optimal state is reached, in which strongly linked articles are in close proximity.

Clustering was conducted using the hDBSCAN-algorithm (hierarchical Density-Based Spatial Clustering of Applications with Noise, McInnes et al. 2017) on an embedding to thirty dimensions produced by UMAP from the unreduced data, as hDBSCAN is quite sensitive to datasets of higher dimensionality. hDBSCAN builds a minimal spanning tree based on mutual reachability distance between the datapoints.⁷ Based on that tree, a hierarchy of cluster-memberships is constructed. A condensed form of that tree forms the basis of the visualization in Fig. 3. hDBSCAN takes the parameter of minimum cluster size, which in this case was set to 300. Using this parameter, the algorithm removes links from the tree by increasing the minimum weight of links $\lambda = \frac{1}{distance}$, and checks at every juncture in the tree, whether the parts that remain are larger than the minimum cluster size. If they are, they are kept for consideration as a cluster, if they are not, they are declared noise. That datapoints break off the clusters all the way, until only the most central points of the clusters remain, is the reason for the pointy structures in Fig. 3. The parts of the tree which are the most resilient to changes of lambda are considered clusters by hDBSCAN. The groups of papers which are kept together before breaking up into the clusters we finally choose—the grey vertical bars in 3—can be considered as potential metaclusters. For each cluster the most frequent nouns, noun phrases, the most frequently cited journals,

⁵ UMAP has until now mainly been used to handle large RNA sequencing datasets in bioinformatics (for details see e. g. Becht et al. 2018).

⁶ See Huang (2008) for an account of different similarity measures for text-classification.

⁷ A very good explanation of this is provided in the documentation (McInnes et al. 2019).

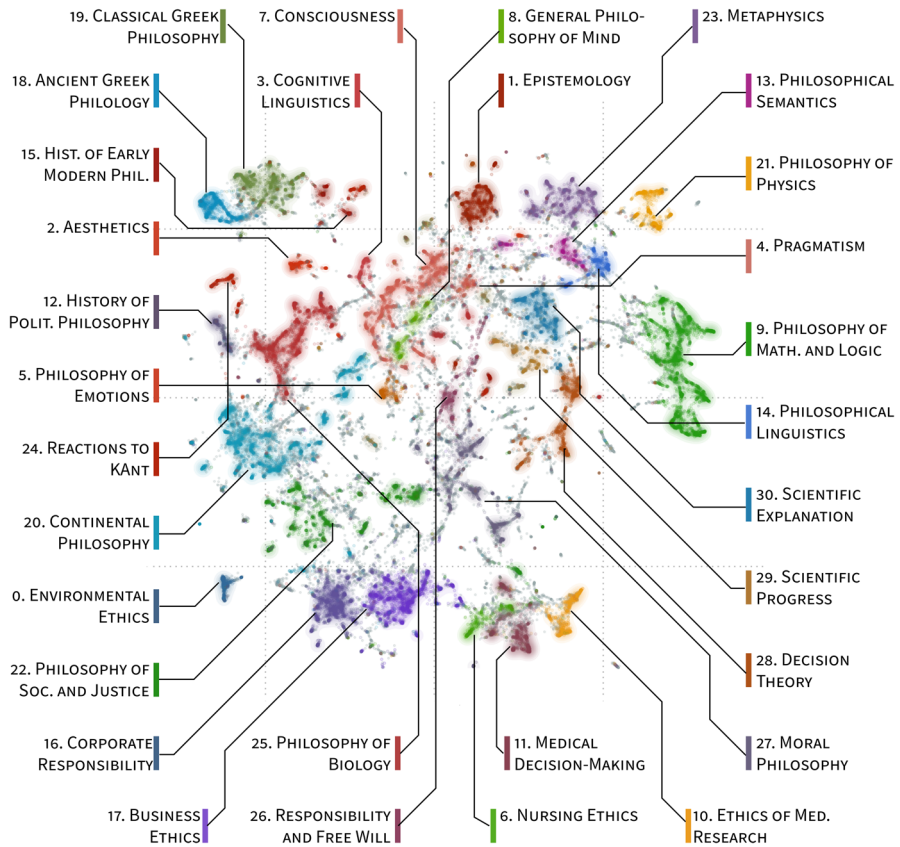


Fig. 2 Each point in the graphic represents one of the 68,152 articles. They are arranged according to a umap embedding, a process that tries to position articles with similar citation-patterns close to each other. Points are colored according to their cluster, while points which were not assigned a cluster were left grey. The labels have been derived from the most common terms in the abstracts and titles and the publications with the highest citation counts in the clusters. More information about each cluster can be accessed in the clustering report: *Online Resource 1*. The graphic was plotted using the ggplot2 package in R, and post-processed in Adobe Illustrator

and most cited individual publications were extracted. A full report is available as online resource: *Online Resource 1.pdf*. Based on this data the clusters were identified and named. As in some cases the limited space does not allow for the exposition necessary to fully communicate the contents of the cluster, the reader is kindly asked to refer to the clustering-report for further information. It is of particular importance to specify that the absence of a cluster in the diagrams does neither mean that it does not exist, nor necessarily that it is missing from the data, only that it is not large or coherent enough to form a cluster at the chosen settings, which were aimed at drawing out only the largest possible structures in the data. If we wanted to find more, smaller and more coherent clusters, to establish for example the relations between subfields, this would be easily possible with a smaller value for the minimum cluster size, but would make comprehensive visualization in the context of a paper impossible.

3 Results

3.1 General observations

The results are presented in two ways, the map in Fig. 2 and the clustering tree represented in Fig. 3.⁸ The first thing that stands out when viewing Fig. 2 is that there are usually some connecting links between adjacent clusters. As articles were clustered based on the similarity of their bibliographies, we can think of them as groups of scholarly co-engagement. It therefore makes sense to say that philosophy is quite cohesive, as there are very few fully unconnected areas. It should be noted that articles disagreeing on a specific subject-matter are likely to end up in the same cluster, as they will engage with the same literature. The presented method can not be used to differentiate the stances that are actually taken on that subject, unless there is so little interaction between the disagreeing parties that there occurs no bibliographic coupling at all. In this context it is important to mention that there is no normative judgement passed. Clusters can be called heterogenous or homogenous while remaining fully agnostic on the question whether these are good or bad things.

The cluster that breaks off the earliest from the rest in the clustering tree are the environmental ethics (0), presumably because of their very interdisciplinary nature. The next significant branches from the clustering-tree are the clusters of applied philosophy in business (16, 17) and medical (6, 11, 10) contexts, forming a meta-cluster, consisting of multiple clusters of applied philosophy.

Above this large group, there seems to run a string of practical philosophy, from its Continental proponents (more on that later) on the left of the graphic, over the spread out cluster of philosophy of society and justice (22) to the cluster of moral philosophy. This cluster is connected via the cluster of moral agency and free will (26) to the metacluster of philosophy of mind (7, 8, 5) at the upper center of the graphic. This again seems to be part of a large cluster of theoretic philosophy that runs from philosophy of mathematics (9), which the algorithm identified on the right, philosophy of language (13, 14) to philosophy of mind. We also find a large metacluster of philosophy of science (29, 30), which shows a similar composition to that identified by Kreuzman (2001), and which, as his findings already suggested, does not directly connect to epistemology (1). And on the upper left we find a group of historical clusters (18, 19, 15, 12, 24).

As the UMAP algorithm that produced the layout for this map has to make many tradeoffs in the approximation of the high-dimensional citation-structure, judgements about the distances of the clusters can be put on more solid grounds by considering the clustering tree presented in Fig. 3. This tree also allows us to compare the sizes of the clusters by looking at their relative widths at their base. As we could already suspect from the UMAP-embedding, several clusters form together little trees of their own, which suggest closer relations between them. We can for example easily make out the groups of practical philosophy and history of philosophy, and a gradient from more theoretical to more practical subjects, which is not very surprising. To illustrate an area where there might be less agreement between common disciplinary self-knowledge

⁸ The corresponding code that produces the graphics is provided in *Online Resource 5.ipynb*.

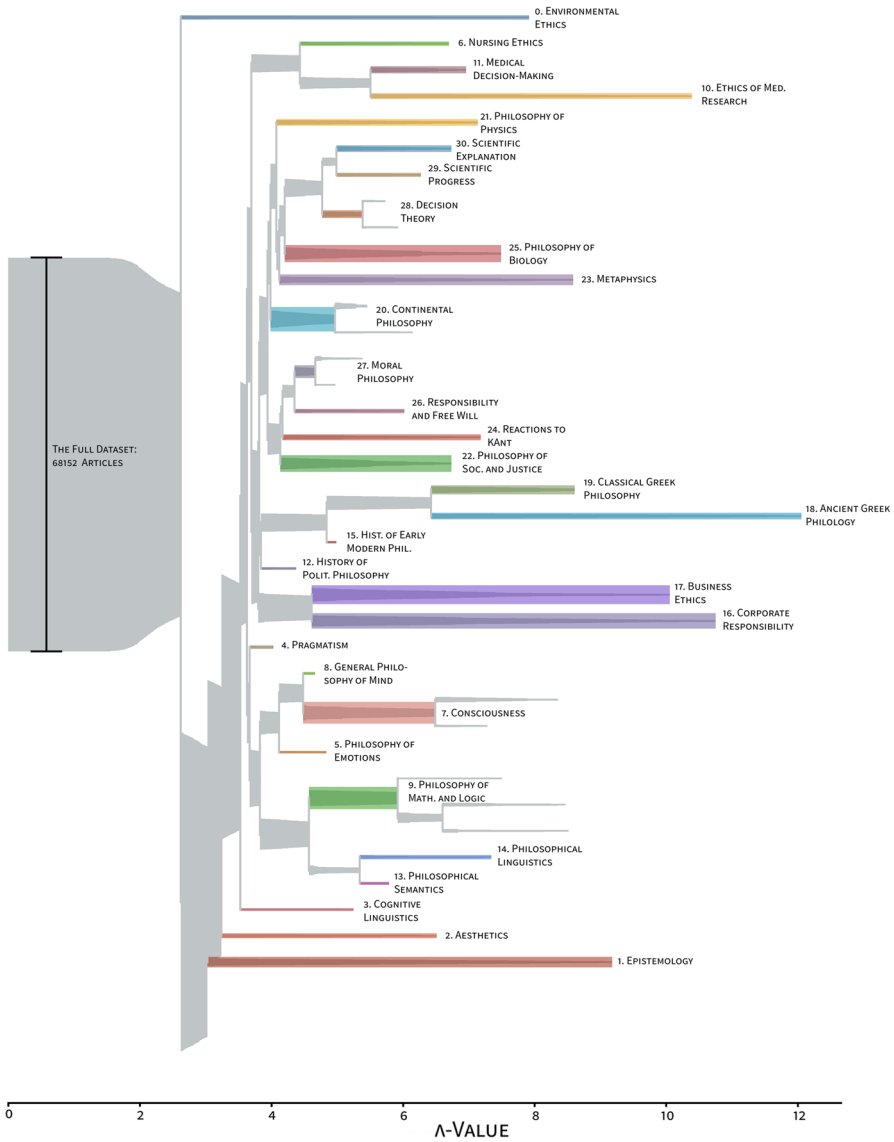


Fig. 3 The condensed clustering tree produced by hDBSCAN. λ represents the weight of the edges. As lambda increases, edges with a weight less than λ are dropped meaning that datapoints either break off the existing clusters, or form new clusters, as clusters that persisted at lower values for λ break apart. Plotted in python with matplotlib and post-processed in Adobe Illustrator

and the picture suggested by this analysis, the question of analytical and Continental philosophy is now considered.

3.2 The analytical/Continental divide

We do not find a cluster for analytic philosophy, which is coherent with claims about its heterogenous structure (Preston 2004, p. 460; May 2002, p. 425; Glock 2008, p. 217) but might be disappointing for philosophers who invoke the importance of such a field. This does of course not mean that there never *was* such a cohesive field, only that it doesn't form a cohesive structure in a sample centered around the the early two 2000s.

If we on the other hand identify the cluster which is defined by the high citation counts of Merleau-Ponty, Heidegger, Garfinkel, Bourdieu, Arendt, Butler, Husserl, Anderson, Hardt, Derrida and Foucault with Continental philosophy,⁹ as for example the description of Continental philosophy given by Levy (2003, p. 287) or West (1996, p. 1) would suggest, we see that there is indeed such a thing as Continental philosophy. This might run contrary to some accounts (Glendinning 2006, p. 13; West 1996, p. 1; Critchley 2001, p. 32) that maintain (with more or less strong qualifications) that the label "Continental philosophy" is problematic or completely misguided, as the thing it is supposed to correspond to is far to heterogeneous to be fit under one label. But from our analysis so far it should be clear that Continental philosophy is homogeneous enough to form a cluster at the same level of analysis at which philosophy of mind already splits up into sub-clusters, which suggests that there is at least quite a bit of cohesiveness to its scholarly co-engagement.

This claim needs to be qualified though: As the clustering tree shows, there is some structure in the Continental philosophy-cluster, which might be drawn out with other settings of the algorithm. But on the whole the picture that emerges is clear: When considering the philosophical output of the last fifty years as a whole, Continental philosophy emerges as a separatable entity in the literature, which nonetheless has multiple links to the rest of philosophy, specifically to the cluster that has been termed Philosophy of Society and Justice (22), and to the large clusters of philosophy of biology (25), which on closer inspection seems to include a lot of general history and philosophy of science. Seen as a whole, Continental philosophy seems to be well integrated into practical philosophy. If one wants to insist on the metaphor of a divide when talking about recent philosophy, one also needs to adopt that terminology for multiple other disciplinary divisions, like between epistemology and theory of science. It seems more sensible to strongly qualify claims about a divide, at least when talking about the last thirty years within anglophone philosophy.

4 Limitations and further work

There are various limitations that should be taken into account when interpreting these results and when looking at the pictures. While there is some reason to be optimistic about PhilPapers being reasonably well suited for our purpose, at least as far as anglophone philosophy is concerned, one can be a bit sceptic about how well it matches up with the WoS-database. The results returned by the WoS query were roughly half the

⁹ While of course not all of the mentioned might themselves be identified as philosophers.

amount that was indexed by the PhilPapers archive. This should not discourage us too much: The Web of Science database is put together with considerations of academic impact in mind. As one of our selection criteria was the number of citations that an article had gotten, we can expect that our sample of reasonably highly cited articles from WoS does not strongly deviate from one that would be drawn from PhilPapers itself.¹⁰

But we can of course be wary of the decision to look only at articles that were cited more than three times itself. On the one hand, one might claim that this approach does indeed give a better picture of philosophy than a random sample would give, as it does incorporate the value judgements (via citation) of philosophers. On the other hand, for the very same reasons it incorporates more or less inadvertent exclusionary practices on various dimensions, for example language (Schwitzgebel et al. 2018) or gender.¹¹ These effects are hopefully minimized by using roughly one twentieth of the total population of articles as our sample, as elite-effects can be expected to get smaller with scale, but they should certainly be kept in mind when making judgements based on the presented visualizations.

It should also be considered that the combination of SVD, UMAP and hDBSCAN for the production of scientometrics mappings is to my best knowledge a new technique. For this reason, it calls for formal evaluation to make its results comparable to similar methods, before we can be fully comfortable with the presented findings.

5 Summary

Disputes about the structure of a discipline are a frequent, albeit informal part of academic life. Large scale analysis of the citation-structure of a discipline can help to ground such disputes by providing data-driven metaphors for disciplinary structure, which are presumably more reliable than the intuitions of individual scholars. For philosophy this approach discourages a metaphoric of universal divisions that run through the whole discipline, but encourages one that stresses a strong interconnect-edness of many smaller and larger subfields. In particular the salient metaphor of a divide between analytic and Continental philosophy is not borne out by the structure of the literature, although Continental philosophy, in contrast to analytical philosophy, does form a coherent group of scholarly co-engagement.

Acknowledgements Open access funding provided by University of Vienna.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

¹⁰ This is a claim that at the time of writing can not be quantified, as WoS does not allow access to queries of more than 100 000 articles. Some indication of how well our sample and PhilPapers match up is given in *OnlineResource1*.

¹¹ See e.g. the comments by Healy (2013) on the fact that 3.6 percent of entries among the top 500 most cited publications in his data were written by women, while 6.3 percent were by a single philosopher, David Lewis. There is some evidence that these effects get worse with a stronger focus on elite journals (Schwitzgebel and Jennings 2017).

References

- Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.4314>.
- Bourget, D. (2010). Paperless philosophy as a philosophical method. *Social Epistemology*, 24(4), 363–375. <https://doi.org/10.1080/02691728.2010.499180>.
- Bourget, D. (2019). *Journal archive—PhilPapers*. Retrieved February 26, 2019, from <https://philpapers.org/journals>.
- Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies*, 170(3), 465–500. <https://doi.org/10.1007/s11098-013-0259-7>.
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404. <https://doi.org/10.1002/asi.21419>.
- Boyack, K. W., & Klavans, R. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, 68(4), 984–998. <https://doi.org/10.1002/asi.23734>.
- 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>.
- Buckner, C., Niepert, M., & Allen, C. (2011). From encyclopedia to ontology: Toward dynamic representation of the discipline of philosophy. *Synthese*, 182(2), 205–233. <https://doi.org/10.1007/s11229-009-9659-9>.
- Critchley, S. (2001). *Continental philosophy: A very short introduction*. Oxford: Oxford University Press.
- Glendinning, S. (2006). *The idea of Continental philosophy: A philosophical chronicle*. Edinburgh: Edinburgh University Press.
- Glock, H. J. (2008). *What is analytic philosophy?*. Cambridge: Cambridge University Press.
- Healy, K. (2013). *A co-citation network for philosophy*. Retrieved February 28, 2019, from <http://kieranhealy.org/blog/archives/2013/06/18/a-co-citation-network-for-philosophy/>.
- Higgins, A., & Smith, B. (2013). A citation based view of the ontology community in philosophy. In H. Davis (Ed.), *Proceedings of the 5th annual ACM web science conference*, ACM Digital Library, Association for Computing Machinery.
- Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth New Zealand computer science research student conference (NZCSRSC2008)* (Vol. 4, pp. 9–56), Christchurch, New Zealand.
- Katzav, J. (2018). Analytic philosophy, 1925–69: Emergence, management and nature. *British Journal for the History of Philosophy*, 26(6), 1197–1221. <https://doi.org/10.1080/09608788.2018.1450219>.
- Kreuzman, H. (2001). A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science. *Scientometrics*, 50(3), 525–539.
- Levy, N. (2003). Analytic and Continental philosophy: Explaining the differences. *Metaphilosophy*, 34(3), 284–304.
- May, T. (2002). On the very idea of Continental (or for that matter Anglo-American) philosophy. *Metaphilosophy*, 33(4), 401–425. <https://doi.org/10.1111/1467-9973.00237>.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbSCAN: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205. <https://doi.org/10.21105/joss.00205>.
- McInnes, L., Healy, J., & Astels, S. (2019). *How HDBSCAN Works—hdbSCAN 0.8.1 documentation*. Retrieved February 28, 2019, from https://hdbSCAN.readthedocs.io/en/latest/how_hdbSCAN_works.html.
- McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform manifold approximation and projection for dimension reduction*. [arXiv:1802.03426](https://arxiv.org/abs/1802.03426). Retrieved February 27, 2019.
- McLevey, J., & McLroy-Young, R. (2017). Introducing metaknowledge: Software for computational research in information science, network analysis, and science of science. *Journal of Informetrics*, 11(1), 176–197. <https://doi.org/10.1016/j.joi.2016.12.005>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. Accessed: 2019-02-26.
- Preston, A. (2004). Prolegomena to any future history of analytic philosophy. *Metaphilosophy*, 35(4), 445–465.

- Schwitzgebel, E. (2012). *The splintered mind: The Ghettoization of Nietzsche*. Retrieved May 17, 2018, from <http://schwitzsplinters.blogspot.co.at/2012/08/the-ghettoization-of-nietzsche.html>.
- Schwitzgebel, E., Huang, L. T. L., Higgins, A., & Gonzalez-Cabrera, I. (2018). The insularity of anglophone philosophy: Quantitative analyses. *Philosophical Papers*, 47(1), 21–48. <https://doi.org/10.1080/05568641.2018.1429741>.
- Schwitzgebel, E., & Jennings, C. D. (2017). Women in philosophy: Quantitative analyses of specialization, prevalence, visibility, and generational change. *Public Affairs Quarterly*, 31, 83–105.
- Shafiei, M., Wang, S., Zhang, R., Milios, E., Tang, B., Tougas, J., et al. (2007). Document representation and dimension reduction for text clustering. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 770–779). IEEE.
- Underhill, D. G., McDowell, L. K., Marchette, D. J., & Solka, J. L. (2007). Enhancing text analysis via dimensionality reduction. In *IEEE international conference on information reuse and integration, 2007 (IRI 2007)* (pp. 348–353). IEEE.
- Waltman, L., Boyack, K. W., Colavizza, G., & van Eck, N. J. (2019). *A principled methodology for comparing relatedness measures for clustering publications*. [arXiv:1901.06815](https://arxiv.org/abs/1901.06815). Retrieved February 27, 2019.
- 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>.
- West, D. (1996). *An introduction to Continental philosophy*. Cambridge: Polity Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.