Modeling disciplinary structure with uniform manifold approximation and projection

Noichl, Maximilian

noichlmax@hotmail.co.uk
Universität Wien, Österreich

Philosophy seems have been slower in the adoption of digital methods for investigative purposes than other humanities. But the expanded view which digital methods allow does make sense when broad questions of intellectual structure and change, particularly in philosophy of science (Pence, Ramsay, 2018) or history of philosophy (Betti et al. 2019) are concerned.Assessments of disciplinary structures, which are commonly found in philosophical texts tend to focus on the personal and argumentative relations between small numbers of prominent actors, from which views about the disciplines as a whole are deduced.

To supplement such detailed accounts, this project proposes the use of Uniform Manifold *Approximation and Projection* (McInnes et al. 2018) for the mapping and clustering of disciplines. Benchmarks and comparisons to similar methods are produced, and the method is applied to medium-large to large samples of papers from the disciplines of philosophy and history. In particular, it tries to answer questions about the (often stereotypical) relations between specific subfields, the relation between gender and academic subfield, and the prominence of different subfields in public debate.

Preliminary results are already available for philosophy. As the establishment of disciplinary boundaries of philosophy is a notoriously hard problem, the adopted sampling strategy was as expansive as possible. PhilPapers, an expansive index of recent philosophy, has compiled a list of 1349 journals of philosophy, which links, at the time of writing to 1782816 indexed articles. The list was downloaded and compiled into a Web of Science query, to get the citation data for individual papers. 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. Only records were sampled that were cited at least four times, which resulted in a total sample of n=75942 articles, with 1194451 citations events to 159647 unique sources.

The citation-data of the articles was treated like the words in a standard computational text classification problem: For the map-layout, citation-vectors were reduced with singular value decomposition (SVD) to remove noise. This data was in turn transformed with uniform manifold approximation

and projection (UMAP) into a two dimensional map, which will not only be presented as a poster, but is made available in interactive online form, (https://homepage.univie.ac.at/noichlm94/full/zoom_final/index.html) of which the graphic below ought to give an idea. Each dot of the mapping represents a paper, which is positioned relative to all other papers according to similarities in the sources it cites. The clustering was produced using hDBSCAN5 on a 30-dimensional UMAP-embedding of the data. Clusters can be interpreted as groups of scholarly co-engagement. The clusters were identified using the most frequent words from the abstracts of the papers, the most cited sources, and the most common venues of publication.

For philosophy, the preliminary results suggest that (a) the divide into analytical and continental philosophy is generally overstated. While continental philosophy, in contrast with multiple accounts (West, 1996; Glendinning, 2006), does form a distinct cluster at an appropriate level of detail, and therefore is quite coherent in terms of scholarly co-engagement, analytical philosophy fails to do so, validating recent scholarship (Preston, 2004; Glock, 2008) (b) Using a two-sided binomial test on a small-cluster-solution, 135 of 170 clusters show significant (p < 0.0003) deviations from a 1:1 gender-ratio. Of those 135, 2 clusters were dominated by female, 133 by male authors. (c) There is a clear relation between certain academic fields and the amount of public attention they generate. In particular medical ethics are frequently discussed by the public.
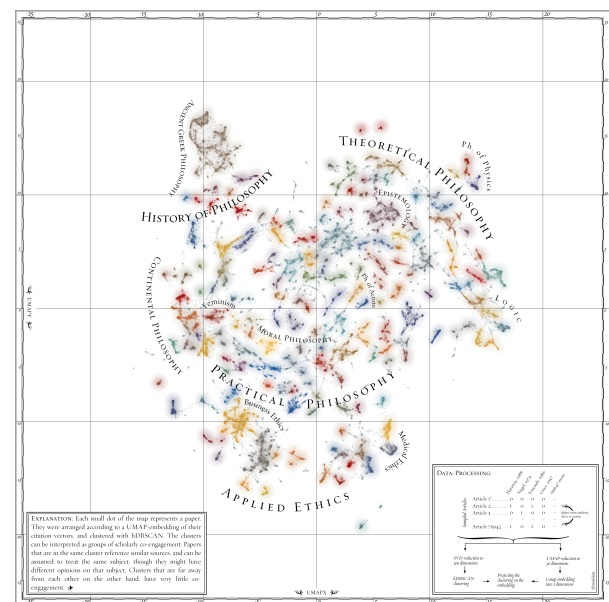

Figure 1: Map-layout.

Bibliography

Betti, A. / Van Den Berg, H. / Oortwijn, Y. / Treijtel, C. (2019): History of Philosophy in Ones and Zeros. In

Curtis M. & Fischer E. (eds.), Methodological Advances in Experimental Philosophy. Bloomsbury: London.

Healy, K. (2013): *A Co-Citation Network for Philosophy* (accessed Feb 28, 2019).

Hobbs, V. (2014): Accounting for the Great Divide: Features of Clarity in Analytic Philosophy Journal Articles. *English for Academic Purposes*, 15, 27–36.

Glendinning, S. (2006): *The Idea of Continental Philosophy: A Philosophical Chronicle;* Edinburgh University Press: Edinburgh.

Glock, H.-J. (2008): *What Is Analytic Philosophy?;* Cambridge, MA.

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.

McInnes, L. / Healy, J. / Melville, J. (2018): UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*:1802.03426.

McInnes, L. / Healy, J. / Astels, S. (2017): Hdbscan: Hierarchical Density Based Clustering. *The Journal of Open Source Software*, 2 (11), 205.

Pence, C. / Ramsey, G. (2018): How to Do Digital Philosophy of Science. *Philosophy of Science* 85 (5): 930–41.

Preston, A. (2004): Prolegomena to Any Future History of Analytic Philosophy. *Metaphilosophy*, 35 (4), 445–465.

Weingart, S. B. (2015): Finding the History and Philosophy of Science. *Erkenntnis*, 80 (1), 201–213.

West, D. (1996): *An Introduction to Continental Philosophy;* Polity Press: Cambridge, MA.