Epistemic Diversity as Distribution of Paper Dissimilarities

Jochen Gläser[1], Michael Heinz[2] and Frank Havemann[2]

[1] *Jochen.Glaser@ztg.tu-berlin.de*
Center for Technology and Society, TU Berlin, Hardenbergstr. 16-18, Berlin, 10623 (Germany)

[2] *{Michael.Heinz, Frank.Havemann}@ibi.hu-berlin.de*
Berlin School of Library and Information Science, Humboldt-University of Berlin, Dorotheenstraße 26, 10099 Berlin (Germany)

Abstract

We continue our quest for measures of epistemic diversity that fit the inherent properties of thematic structures in science. Starting from theoretical considerations, we argue that currently available measures of diversity are not applicable to the epistemic diversity of published scientific knowledge because topics are fluid and overlap. Consequently, we abandon attempts to assign papers to topics and instead explore opportunities to measure diversity based on paper dissimilarities. Considerations of the exploitation of information and signal-to-noise ratios in networks of papers let us dismiss an earlier attempt to base a dissimilarity measure on the resistance distance between papers in the network of papers and their cited sources. In this paper, we explore a dissimilarity measure based on papers' 'views' on the whole network, with the 'view' of a paper consisting of all other papers in the network ranked according to the length of their shortest paths to the paper. We present test results on the diversity of topics, journals and country outputs for information science (2008) as well as on the diversity of country outputs in astronomy and astrophysics (2010).

Conference Topics

Methods and techniques; Indicators

Introduction

The epistemic diversity of research – the diversity of empirical objects, methods, problems, or approaches to solving them – has become a matter of concern for science policy. Attempts by science policy to increase the selectivity of research funding and the growth in strength and homogeneity of incentives for universities have led to concerns about an undue reduction of the diversity of research. Several specific warnings refer to the UK's research assessment exercise (Gläser et al., 2002, Molas-Gallart & Salter, 2002, Rafols et al., 2012). A similar concern has been raised in Germany, where profile-building activities at all universities may make the small subjects disappear (HRK, 2007). Laudel & Weyer (2014) observed in the Netherlands that universities' uniform responses to political signals contributed to the disappearance of one field and the stagnation of another.

Discussions about dangers to the epistemic diversity of research have in common that they lack both theoretical backing and empirical evidence. Epistemic diversity is an ill-understood topic in science studies. It is rarely clear what the concept is intended to refer to, how epistemic diversity might affect research, and how it can be operationalized. Theoretical reasoning drawing on analogies to biodiversity assumes diversity is good for science (e.g. Rafols et al., 2012). However, arguments lack empirical grounding, and no specific arguments about necessary and sufficient levels of diversity or about dangers of too much diversity can be made. The empirical studies of interdisciplinarity (e.g. Bordons et al., 2004; Rafols & Meyer, 2007; Rafols et al., 2012) were forced to use rather coarse indicators such as the journal classification of the web of science, and could not theoretically justify the measures they applied.

The aim of our paper is to present a systematic approach to the measurement of diversity that derives possible bibliometric measures of diversity from properties of the system whose diversity is to be measured, namely scientific knowledge.

We start from a theoretical definition of 'topics' in science and demonstrate that the properties of topics do not match the built-in assumptions of current indicators. While this does not necessarily invalidate the indicators, the assumptions underlying the measurement of diversity in science must be made explicit, and their applicability be argued. We suggest two additional strategies that may alleviate the problems resulting from the mismatch between properties of topics and prerequisites of indicators. The first strategy abandons the explicit identification of topics and measures the diversity of paper networks rather than scientific knowledge. We propose a measure of paper similarity that takes some of the properties of scientific knowledge into account, and demonstrate our approach by applying the measure to two data sets. The second strategy, which is outlined in this paper but not applied, uses the same similarity measure for determining the disparity of topics, thereby enabling the application of existing diversity measures.

Theoretical background

In the most general sense, 'diversity' is the property of a system, namely its heterogeneity, which is caused by the disparity of its elements. Among the many aspects of a science system to which the concept diversity can be applied, we are interested in the diversity of published scientific knowledge. Other aspects of a field's diversity such as the diversity of informal knowledge, instrumentation, empirical objects, or scientific training of researchers, will not be considered here. The *epistemic diversity of a research field* is thus defined here as the diversity of published knowledge claims about scientific problems, solutions, empirical objects, approaches and methods, which are communicated by the field's researchers in publications.

The definition of epistemic diversity as a property of published knowledge suggests using bibliometric methods for its measurement. These methods must support the reconstruction of knowledge structures from publications in a way that is both valid (i.e. returns knowledge structures researchers work with) and supports the measurement of diversity. Fulfilling both requirements is made difficult by inherent properties of knowledge structures in science. In the following, we first discuss the built-in assumptions of current measures of diversity. We then argue that properties of scientific knowledge and of its representation in publications do not meet these assumptions, and discuss opportunities to reconstruct knowledge structures from publications and to measure the epistemic diversity of research.

Built-in assumptions of current approaches to the measurement of diversity

Diversity has been an important topic of biological and environmental research for some time. These fields are mainly concerned with the impact of diversity on the stability and development of biotopes and species. Two approaches to the measurement of biodiversity can be distinguished:

a) The diversity of biotopes[1] composed of several species is measured with a three-level hierarchical approach. Biotopes are considered as consisting of species, which in turn consist of individuals. Three factors contribute to the diversity of such a system, namely

- variety (the number of species in the biotope),
- disparity (the extent to which the species differ from each other), and
- evenness (the distribution of individuals across the different species).

Depending on the research question, these factors can be assessed separately (e.g. if only the number of species is measured) or be combined in synthetic measures such as Shannon's Entropy (combining variety and evenness) or the Rao-Index (combining all three measures).

---

[1] A biotope is a physical environment (habitat) with a distinctive assemblage of conspicuous species (Olenin & Ducrotoy, 2006: 22).

This approach to diversity is applied in fields outside the biosciences as well (see Rafols et al., 2012, Stirling, 2007). It requires that

- the system whose diversity is to be measured can be analytically decomposed in three levels (system, categories, and elements),
- the contribution of differences between individuals of the same species to the biotope's diversity can be neglected,
- the categories can be constructed as disjunct by assigning each element to exactly one category or by fractional assignments of elements to categories, and that
- all categories share a property that can be used to calculate disparity.

b) The diversity of species composed of individuals is measured on the basis of a two-level approach. In this approach, variety and evenness become meaningless because there is no intermediate level of categories to which elements can belong. The only remaining basis for measuring the diversity of the system is the disparity of individuals. While this approach is used less frequently, it can be considered to be more fundamental because it conceptualizes diversity as the degree to which the elements of a system (here: a species) differ from each other. This approach is applicable as long as a system can be delineated and elements share a property that can be used to calculate disparity.

Both approaches share a premise concerning the disparity of categories and elements. Categories and elements are conceptualized as stable, and their pairwise disparities as independent, i.e. not affected by other categories respectively elements. New elements entering the system (i.e. individuals of a species being born or migrating to a biotope) do not affect the disparity between existing elements or between the categories, and new categories (i.e. species migrating to a biotope) do not affect the disparity between the categories or between the elements that are already present. The same applies to the disappearance of elements or categories.

Properties of topics in scientific knowledge

If the approaches to the measurement of diversity are to be applied to scientific knowledge, the system, categories and elements must be determined. For the three-level approach, the system would be the knowledge of a field, topics in this field would serve as categories, and knowledge claims (the claim for some empirical, theoretical or methodological statement to be true) would constitute the elements of the system. For the diversity measures discussed above to be applicable, these knowledge structures would need to fulfil the built-in assumptions of the measures. We therefore begin by briefly discussing the properties of scientific knowledge in its structures.

Scientific knowledge is produced by scientific communities whose members

- observe the community's shared body of knowledge,
- interpret this knowledge in the light of their own research experience,
- identify gaps in that knowledge and design research processes for producing the knowledge that closes the observed gap, and
- offer their interpretation and the new knowledge to their community.

The interpretation of the community's knowledge and claims about new knowledge are fully or partially shared by some members of the community. We define a topic as *a focus on theoretical, methodological or empirical knowledge that is shared by a number of researchers and thereby provides these researchers with a joint frame of reference for the formulation of problems, the selection of methods or objects, the organization of empirical data, or the interpretation of data* (on the social ordering of research by knowledge see Gläser, 2006). This definition resonates with Whitley's (1974) description of research areas but abandons the assumption that topics form a hierarchy. The only demand the definition makes is that some scientific knowledge is perceived similarly by researchers and influences their decisions.

Due to this nature as shared and collective perspectives, topics have structural and dynamic properties that affect the opportunities for measurement. *Structural properties* include the following:

1) All topics are *emergent meso- or macro-structures*, i.e. they are collective-level products of autonomous interpretations and uses of knowledge by individual researchers.

2) From this follows that topics are *local* in the sense that they are primarily topics to the researchers whose decisions are influenced and who contribute to them, and only secondarily topics to those colleagues who are outside observers.

3) Given the multiple objects of knowledge that can serve as common reference for researchers, it is inevitable that topics *overlap*. Overlaps are ubiquitous because any research is likely to address several topics at once, e.g. by including theories about an object, methodologies for investigating it, and empirical information about an object. They also occur when a knowledge claim belongs to several topics at once (e.g. formulae used in bibliometrics belonging to mathematics but also expressing bibliometric relationships).

4) Knowledge has a *fractal structure* (e.g. van Raan, 2000), and topics can have any size between small (emerging topics that in the beginning may concern just two or three researchers) and very large thematic structures such as bibliometrics. The 'size' of a topic can be defined in various ways – as scope (range of phenomena covered), level of abstraction (which is again linked to the range of phenomena covered), or number of research processes or researchers influenced by it. In all these dimensions there is a continuum from very small to very large topics.

5) The degree to which knowledge influences researchers' actions, and the strength of links between new findings and existing knowledge that are constructed by researchers, also vary between 'very weak' and 'very strong'. As a result, the '*distinctiveness'* of topics varies. Some topics are unambiguously seen as being different from other knowledge by most researchers of a field and are thus well separated from surrounding knowledge, while others are much less pronounced.

These structural properties of topics let them form an inconsistent poly-hierarchy for which not even meaningful levels can be determined. This also implies that no field or collection of papers has exactly one definite thematic structure. Different perspectives can be applied to fields and collections of papers and will return different topical structures. Topics may overlap in their boundaries or pervasively. They vary considerably in their size and 'distinctness', i.e. the extent to which they actually constitute a shared concern of researchers.

Dynamic properties of topics are shaped by their role in the knowledge production process. As coinciding perspectives of researchers, topics are perpetually changing. Researchers constantly revise their perspectives on the existing knowledge and thus the relationships of their perspectives to those of their colleagues. They also utilize and contribute to more than one topic (e.g. theoretical, methodological and empirical ones). Hence, their production of new knowledge may instigate at least one and in many cases all of the following changes:

* Enrichment: Since new knowledge is added to the system, the community's knowledge on a topic is likely to grow.

* Restructuring: The new knowledge is linked to existing knowledge and thereby links existing knowledge, i.e. the density of connections in the system of knowledge increases.

* Reduction: The new knowledge may devalue existing knowledge by proving it to be wrong or may reduce it by subordinating it to a generalisation.

Through these processes, the size of topics, their distinctness and relations between them are constantly changed. New topics may emerge at any time, and existing topics may disappear or radically change.

Representation of knowledge in publications and reconstructions of topics

Since bibliometric methods reconstruct knowledge structures from publications, the representation of knowledge in publications provides the opportunities and constraints for a bibliometric measurement of diversity, which we now discuss in more detail. In the sociology of science, knowledge claims are treated as the basic unit through which new knowledge is communicated (e.g. Cozzens, 1985, Pinch, 1985). Knowledge claims are claims that some new knowledge produced by the author is true; a publication usually contains several such claims.

For the new knowledge claims to be added to the community's body of knowledge, they must be used by other community members in their subsequent knowledge production. This requires the new knowledge to be available to all potential users, which is achieved by publication. With each publication, researchers construct

- an account of the state of the current knowledge on a topic,
- the claim that there is a specific gap in that knowledge,
- the claim to have developed an approach whose application can close that gap,
- the new knowledge produced with this approach, which is claimed to close the gap, and
- in many cases conclusions concerning implications of the new knowledge including the necessity of further specific research (Gläser, 2006: 125-126, Swales, 1986: 45).

These claims embed the new knowledge that is offered to the community in the existing knowledge. However, they do so selectively and *ad hoc*. The claims in a publication are organised in a way that maximises the chances of the new knowledge's further use by emphasizing originality, relevance, validity and reliability of the new knowledge. Links to the existing knowledge are crafted to further this impression.

The new knowledge claims shape subsequent knowledge production processes if they inform the formulation of problems, choice of methods or interpretation of results by readers of the publication. If they do so, the researchers using them are likely to indicate the link of the new knowledge they offer to these knowledge claims, thereby treating them as part of the community's knowledge. This 'elementary process' of adding knowledge causes the dynamic properties described in the previous section. If a new knowledge claim is added, the community's knowledge becomes enriched, and its structure changes because the claim creates new links between, reinforces or remove existing links. New knowledge claims may also invalidate existing claims or subsume them to more general statements if they are used by other community members in this way.

Consequences for the measurement of diversity

The properties of knowledge claims and topics affect the opportunities to reconstruct topics from publications with bibliometric methods, i.e. by using properties of publications such as authors, journals, references, or terms. To begin with, no method for the bibliometric reconstruction of individual knowledge claims has been proposed so far. Knowledge claims are represented in series of sentences and clauses that are distributed across a publication. Reconstructing them would be a task for linguistics but is still impossible for that field, too.

Bibliometric methods are better suited for the reconstruction of topics because the latter are larger and span many publications. However, from the properties of topics described earlier follows that none of the bibliometrically usable properties of a paper can be assumed to be thematically homogeneous in the sense of representing only one topic. Since research processes are influenced by and address more than one topic, topics overlap in research processes, publications (and thus references), terms, journals, and authors. Furthermore, researchers apply their individual perspectives on the scientific knowledge when constructing and linking topics, which is why links to topics may occur unpredictably in a variety of scientific fields. Consequently, any finite sub-set of papers is unlikely to include all publications addressing a specific topic, which means that any hierarchy of topics is also only partially covered by the paper set.

Owing to the mismatch between properties of publications that can be used for the reconstruction of topics and the representation of topics in publications, bibliometric methods inevitably reduce the complexity of the underlying knowledge structures. This is not a problem in itself because all models reduce complexity. The question is not how the reduction of complexity can be avoided but whether a specific reduction of complexity is appropriate to the purpose. Answers to this question should be part of a bibliometric methodology that links specific purposes of topic reconstruction to specific strategies that are applied. The absence of such a methodology is one of the major obstacles for bibliometrics.

When we apply these methodological considerations to the measurement of epistemic diversity, we can distinguish three strategies for solving the problems posed by properties of scientific topics. The first strategy, which has been applied in all attempts to measure epistemic diversity so far, constructs distinct topics to which papers are assigned. The three-level approach is then used for the measurement of diversity.

A second possible strategy would be to construct overlapping topics to which papers belong partially. In order to apply three level-diversity measures, the topics would have to be made disjunct by fractionalising the papers. The disparity of topics would need to be measured based on the difference in paper membership. While this strategy still has some problems in the case of pervasive overlaps of topics, it would create a more precise representation of topics and still enable the application of three-level diversity measures.

The third strategy, which we apply in the remainder of the paper, circumvents the problem of topic reconstruction by applying the two-level approach. Since knowledge claims cannot be reconstructed from publications, the strategy measures paper diversity as a proxy for knowledge diversity. This strategy requires a similarity measure for published papers, which should reflect the properties of thematic structures in science discussed above.

Methods and Data

Network-based measures of paper similarity

Diversity measures for the two-level approach aggregate the pairwise similarities of all elements. Among the many ways in which the similarity of two papers in a network can be determined, we need to find those that strike a balance between utilizing as much information as possible and avoiding the inclusion of irrelevant information that contaminates the measure.

Bibliographic coupling is well-established, and is commonly considered as one of the best bibliometric measures of paper similarity (Ahlgren & Jarneving, 2008: 274-275). The strength of bibliographic coupling between two papers can be used directly as a measure of their similarity. However, bibliographic coupling is not a useful measure for the similarity of papers that are not coupled. All these papers must be considered equally dissimilar, which they are certainly not. Thus, bibliographic coupling is unsatisfactory as a measure of paper similarity in networks.

An alternative to using bibliographic coupling is the utilization of all connections in a network, e.g. by measuring similarity as resistance distance in networks of papers and their cited sources or in bibliographic coupling networks. In this approach, indirect links between papers are taken into account, i.e. information about the whole network is utilized for the calculation of all pairwise paper similarities (see Gläser et al., 2013 for an example). However, this approach inevitably uses information about detours through a network – i.e.

about connections that exist and can technically be made but are not meaningful in terms of paper similarities. In other words, the measure is distorted by paths that do not reflect thematic similarity. Furthermore, our own experiments showed the measure to favour papers with a high degree. Finally, using all paths in a paper network for the measurement of its diversity makes the measure particularly sensitive to changes in the network structure. If measures of paper similarity are based on the resistance distance, each paper that is added to the network changes the resistance distance and thus the similarities of all papers in the network. This is an extremely unrealistic assumption about the impact of new publications on the epistemic diversity of a field.

Between the use of only information about direct coupling and the use of information about all possible connections between papers lie measures such as length of the shortest path between two nodes. This measure makes little sense in networks of papers and their cited sources because each reference two papers have in common creates a path of the length two between them. For networks in which links reflect the relative strength of bibliographic coupling, the length of shortest paths captures more information.

By determining the length of the shortest path between two papers in a network, other connections are taken into account indirectly by dismissing them as longer paths. Still, the environment of a paper is largely neglected by such a measure. However, the length of shortest path can be used to construct an indirect measure of paper similarity that takes the environment of papers into account. We can construct the 'view' of a paper on its environment by ranking all other papers in the network according to their distance to that paper. The 'view' describes how dissimilar other papers in the network are in terms of their shortest paths. The similarity between two papers can be defined as the similarity of the two papers' 'views' on the network, which is measured by calculating the rank correlation of the two lists.

Thus, we measure the similarity of two papers by:

- determining the shortest paths between all pairs of papers in a bibliographically coupled network (weighted with the arccosine of Salton's Cosine),
- creating a 'view' of each paper by ranking all other papers according to increasing lengths of their shortest paths,
- calculating the similarity of two papers as the rank correlation (Spearman) between the two lists, and
- transforming the rank correlation in a similarity measure.

This measure, which can be interpreted as the similarity of the 'views' of the two papers on their scientific environment, avoids the influence of degrees. It is similar to the use of "preferences" in an "affinity" system by Balcan et al. (2012) in their construction of overlapping endogenous communities.

Data

To test our measure, we used two data sets. The first data set is the main component of publications (articles, letters and proceedings papers) in six information science journals, which consists of 492 papers (see Havemann et al., 2012 for a description of this data set). The second data set is the main component of 14,770 publications (articles, letters, and proceedings papers) published 2010 in 53 astronomy and astrophysics journals (see Havemann et al., 2015 for a description of this data set). For each data set, we constructed and analysed the bibliographic coupling network.

Methods

For each data set, we calculated pairwise paper similarities as transformed Spearman's rank correlation of the papers' 'views' on the network. The 'view' of a paper p_i on the network is the vector of shortest paths between p_i and the papers p_1 to p_n of the network. Thus, the dissimilarity of two papers – their distance – is calculated as

$$dist\ \left(view(p_i i), view(p_j j)\right) = 1 - \frac{r_{sp}\left(view(p_i), view(p_j)\right) + 1}{2}$$

Where r_{sp} is the Spearman's rank correlation coefficient of the two views.

We tested this similarity measure on our information science data set by using it for a Ward clustering and comparing the best matching Ward clusters to three topics we had previously identified by inspecting titles and keywords of the articles.

We then calculated the distributions of paper similarities for country subsets and journal subsets of papers in both data sets, and used the median of the distributions as single-number value of the subset's diversity.

Our diversity measure also enables the construction of 'collective views', i.e. of 'views' of paper sets on each other. We exploited this opportunity in a third step and constructed similarities between countries and journals in information science.

Results

Information science

Our Ward clustering with the similarity measure led to results that compare well to previous experiments with other algorithms (Table 1).

Table 1. Salton's Cosine of precision and recall of pre-defined information science topics by five algorithms.[2]

Table	*MONC*	*HLC*	*FHC*	*RDDC*	*SPBC*
h-index	0.71	0.93	0.59	0.92	0.95
Bibliometrics	0.79	0.82	0.83	0.87	0.86
Webometrics	0.58	0.60	0.46	0.65	0.53

The three best performing algorithms – HLC, RDDC and SPBC – perform best for the h-index, good for bibliometrics including the h-index, and worst for Webometrics. These differences may be linked to the topics' internal diversity (Figure 1). Internal diversity is lowest for the h-index (all papers are very similar) and highest for webometrics (a high proportion of webometrics papers is not very similar). The differences in internal diversity may explain the differential success of algorithms in recapturing the topics.

---

[2] MONC= Merging overlapping natural communities, HLC=Hierarchical link clustering, FHC=Fuzzification of hard clusters (see Havemann et al., 2012). RDDC= Ward clustering with a similarity measure using the rank correlation of 'views' based on the resistance distance in direct citation networks (Gläser et al., 2013). SPPC= Ward clustering with a similarity measure using the rank correlation of 'views' based on the length of shortest paths in bibliographic coupling networks (algorithm presented in this paper). Among the three topics, bibliometrics also includes the h-index papers.

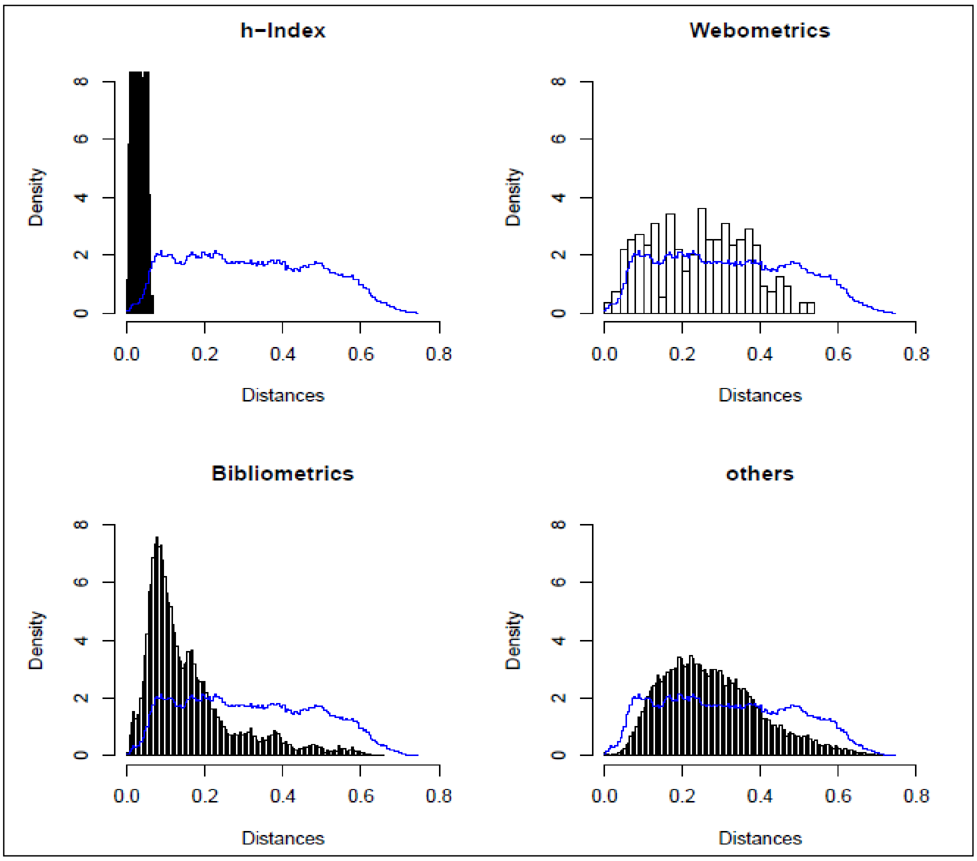

Figure 1. Internal diversity of three topics in the information science network (the blue lines represent the distribution for the whole network, the areas always equal one).

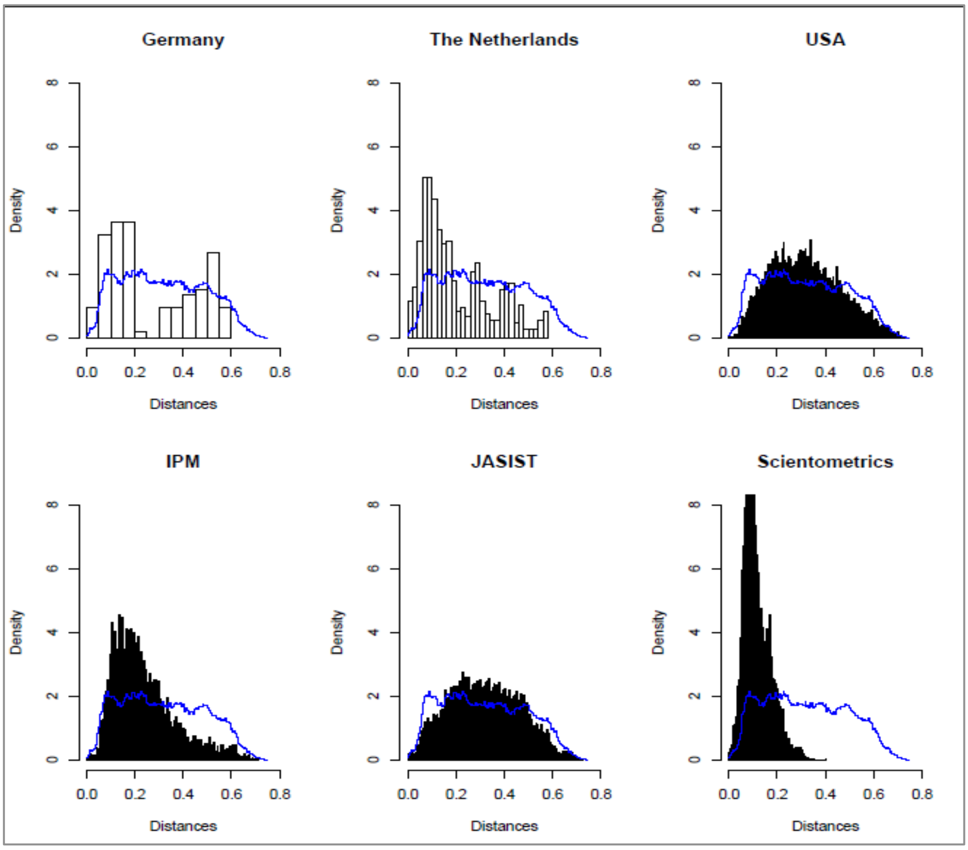

Figure 2. Diversity of information science publications from three countries and three journals.

Figure 2 shows the diversity of information science publications in three journals and of three countries. According to these distributions of distances,

a) Dutch information science publications are less diverse than the few German publications and the publications from the USA; and

b) *Scientometrics* was the least diverse (most focused) journal, followed by *JASIST* and *Information Processing and Management*.

Astronomy and astrophysics

The astronomy and astrophysics publication network is less diverse than the information science network. Taking the median as a single-number measure of diversity, the information science network (median = 0.32) is much more diverse than the astronomy and astrophysics network (median = 0.27). Owing to space limitations, we can provide only one comparison. Figure 3 compares the distribution of paper similarities for Chilean and US-American publications. Astronomy and astrophysics publications from Chile appear to be much less diverse (much more concentrated on one or few topics) than those from the USA.

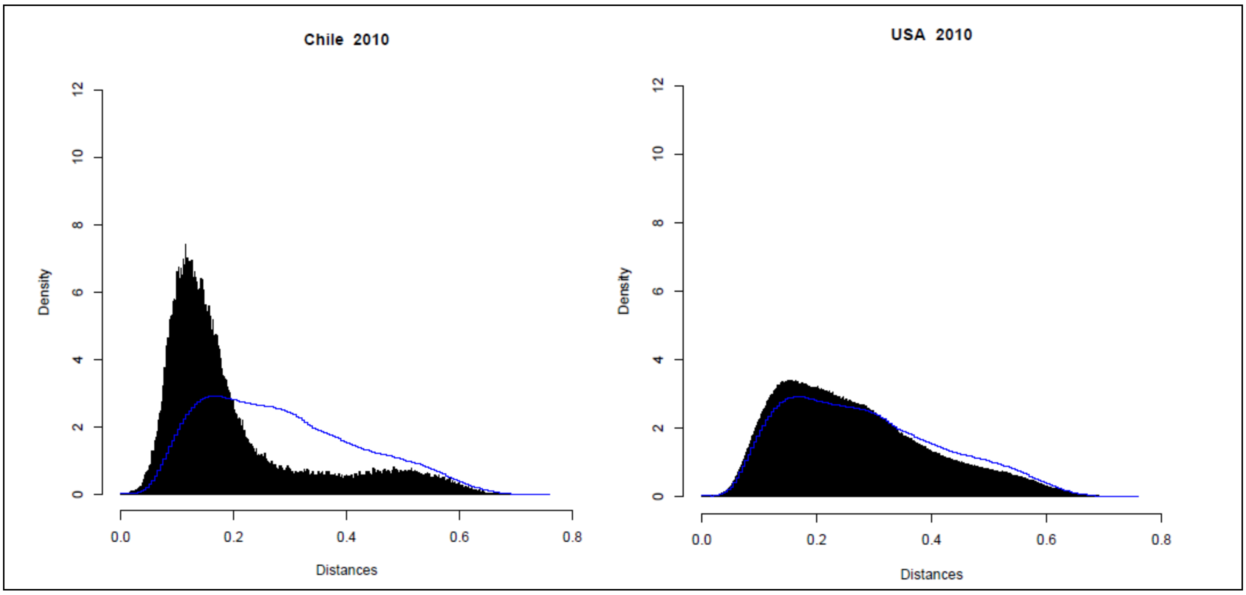

Figure 3. Diversity of astronomy and astrophysics publications from Chile and the USA (the blue lines represent the distribution for the whole network).

Discussion

A small but noxious problem for the application of our diversity measure is the occurrence of direct citations between publications from the same year. Direct citations can be considered a strong indicator of thematic similarity. However, it is not known how strong an indicator a direct citation is, and how it should be treated in comparison with bibliographic coupling of two publications. Our current solution is to add the citing and cited publication to each other's reference lists, i.e. integrating direct citation into bibliographic coupling. This solution is, however, as arbitrary as any other solution would be.

A more consequential limitation stems from our use of networks of papers as models of published knowledge. Adding a node with at least two links to a network indirectly changes connections between all nodes. This is not true for added knowledge, which can induce changes in similarities that remain local in that they affect only the knowledge to which it links directly. Although the length of the shortest path between two papers is not as sensitive to changes in networks as the measure we tried before (resistance distance), it remains to be seen whether time series of diversity constructed with our distance measure can be interpreted. Since the literature in most fields keeps growing, time series of diversity have to cope with ever-growing paper networks.

Finally, a third limitation is inherent to our measure. Measuring the diversity of any set of papers with the approach suggested in this paper requires the set of papers to be embedded in a connected subgraph. If a research organisation has publications in many unrelated fields (as most universities do, providing an aggregate measure of the diversity of this organisations published output would be impossible. However, such an aggregate measure is likely to be meaningless in any case.

Conclusion

While further tests are of course necessary, the diversity measure proposed in this article appears to enable comparisons of paper sets from topics, journals, specialised organisations, or countries. The measure appears to use enough information to provide meaningful results without being sensitive to the noise created by network connections that have no bearing on the similarity of two papers. It is also compatible with sociological findings that ground the publication process in an author's personal experience and perspective. The 'view' of a paper on the network can easily be interpreted as the scientific perspective of its author.

Our discussion of diversity measures and their applicability to the epistemic diversity of published knowledge suggests two lines of further work. First, the problem of time series must be solved, i.e. the diversity of a field must be measured for networks of different sizes. This requires assessing the sensitivity of our diversity measure for changes in networks that are unrelated to epistemic diversity.

Second, a solution must be found for the measurement of diversity with a three-level approach. This is both theoretically and practically important because changes in the diversity of research are caused by the selective growth and shrinking of topics. Understanding the role of epistemic diversity for research requires causally attributing changes in the epistemic diversity to such processes of growth and decline, which in turn requires linking publications to topics. The obvious solution is making topics disjoint by fractionally assigning papers to overlapping topics. However, this does not solve all problems posed by thematic structures in science. Consider the following simple example: A paper on the h-index is simultaneously a paper in bibliometrics because the topic h-index is fully included in bibliometrics. How would one assign such a paper to the two topics?

Developing three-level measures for the diversity of overlapping topics might mean abandoning all established measures, and might prove a very challenging task.

References

Ahlgren, P. & Jarneving, B. (2008). Bibliographic coupling, common abstract stems and clustering: A comparison of two document–document similarity approaches in the context of science mapping, *Scientometrics, 76*, 273-290.

Balcan, M.-F., C. Borgs, M. Braverman, J. Chayes & S.-H. Teng. (2012). Finding endogenously formed communities. arXiv:1201.4899v2.

Bordons, M., F. Morillo & I. Gómez (2004). Analysis of cross-disciplinary research through bibliometric tools. In: H. F. Moed , W. Glänzel & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research*, Dordrecht: Kluwer, 437-456.

Cozzens, S. E. (1985). Comparing the Sciences: Citation Context Analysis of Papers from Neuropharmacology and the Sociology of Science. *Social Studies of Science 15*(1), 127-153.

Gläser, J. (2006). *Wissenschaftliche Produktionsgemeinschaften: Die soziale Ordnung der Forschung*. Frankfurt am Main: Campus.

Gläser, J. G. Laudel, S. Hinze & L. Butler (2002). Impact of Evaluation-based Funding on the Production of Scientific Knowledge: What to Worry About, and how to Find Out. Report to the BMBF. Retrieved January 20, 2015, from http://www.laudel.info/wp-content/uploads/2013/pdf/research%20papers/02expertiseglaelauhinbut.pdf.

Gläser, J., F. Havemann, M. Heinz & A. Struck. (2013). Measuring the diversity of research using paper similarities. In: S. Hinze & A.Lottmann (Eds.), *Translational twists and turns: Science as a socio-economic endeavor. Proceedings of the 18th International Conference on Science and Technology Indicators, Berlin, Germany, September 4 — 6, 2013*, Berlin, 130-139.

Havemann, F., J. Gläser, M. Heinz & A. Struck (2012). Identifying overlapping and hierarchical thematic structures in networks of scholarly papers: A comparison of three approaches. *PLoS ONE, 7*(3), e33255.

Havemann, F., Gläser, J. & Heinz, M. (2015). A link-based memetic algorithm for reconstructing overlapping topics from networks of papers and their cited sources. *15th International Conference on Scientometrics and Informetrics*, Istanbul, 29 June -3 July 2015.

Hochschulrektorenkonferenz (HRK) (2007). Die Zukunft der kleinen Fächer: Potenziale – Herausforderungen – Perspektiven, Bonn: Hochschulrektorenkonferenz. http://ipts.jrc.ec.europa.eu/home/report/english/articles/vol66/ITP1E666.html.

Laudel, G. & E. Weyer. (2014). Where have all the Scientists Gone? Building Research Profiles at Dutch Universities and its Consequences for Research. In: Richard Whitley and Jochen Gläser (Eds.), *Organizational Transformation And Scientific Change: The Impact Of Institutional Restructuring On Universities And Intellectual Innovation*, Bingley, UK: Emerald Group Publishing Limited, 111-140.

Molas-Gallart, J. & A. Salter (2002). Diversity and Excellence: Considerations on Research Policy. IPTS Report.

Olenin, S. & J.-P. Ducrotoy (2006). The concept of biotope in marine ecology and coastal management. *Marine Pollution Bulletin 53*, 20–29.

Pinch, T. (1985). Towards an Analysis of Scientific Observation: The Externality and Evidential Significance of Observational Reports in Physics. *Social Studies of Science 15,* 3-36.

Rafols, I., L. Leydesdorff, A. O'Hare, P. Nightingale & A. Stirling (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy, 41*, 1262-1282.

Rafols, I. & M. Meyer (2007). How cross-disciplinary is bionanotechnology? Explorations in the specialty of molecular motors. *Scientometrics, 70*(3), 633-650.

Stirling, A. (2007). A general framework for analyzing diversity in science, technology and society. *Journal of The Royal Society Interface, 4*, 707–719.

Swales, J. (1986). Citation analysis and discourse analysis. *Applied Linguistics, 7*, 39-56.

Van Raan, A. F. J. (2000). On growth, ageing, and fractal differentiation of science. *Scientometrics,* 47, 347-362.

Whitley, R. (1974). Cognitive and social institutionalization of scientific specialties and research areas. In: R. Whitley (ed), *Social Processes of Scientific Development* (pp.69–95). London.