Noname manuscript No.
(will be inserted by the editor)

Identification and analysis of the highly cited knowledge base of sustainability science

R.K. Buter · A.F.J. Van Raan

Received: date / Accepted: date

Abstract We investigated the interdisciplinary 'pillars' of scientific knowledge on which the emerging field of sustainability science is founded, using a bibliometric approach and data from the Web of Science database. To find this scientific basis, we first located publications that represent a relevant part of sustainability science and then extracted the set of best cited publications, which we called the HCKB. To find the research orientation in this set we inspected the occurrence of fields and contrasted this with the occurrence of fields in other publication sets relevant to sustainability science. We also created a network of co-cited HCKB publications using the seed set citations, extracted communities or clusters in this network and visualised the result. Additionally, we inspected the most cited publications in these HCKB clusters. We found that themes related to the three pillars of SD (environment, economy, and sociology) are all present in the HCKB, although social science (not including economics) is less visible. Finally, we found increasing diversity of fields and clusters in the citations of the seed set, indicating that the field of sustainability science is not yet moving into a more transdisciplinary state.

Keywords Sustainability science · Knowledge base · Bibliometric analysis, · Network visualisation · clustering, trends · transdisciplinarity

1 Introduction

Sustainable development is a complex, multifaceted topic that has political, environmental, economical, and sociological aspects. Although all these aspects are necessary for any successful approach, it is also clear clear that scientific research (both fundamental and applied) has an important role to play as well (Dodds 1997; Kates et al 2001; Clark and Dickson 2003; Clark 2007). Recently, this resulted in an interdisciplinary science field, 'sustainability science', which organises and supports scientific research into specific, focused topics, while also encourages and promotes integration of these individual results in a larger context (Komiyama and Takeuchi 2006; Keitsch 2010).

The inherent interdisciplinary character of sustainability science and the many publications on SD topics that have been published over the years, make it interesting to investigate on what 'pillars' of scientific knowledge this research is based. In bibliometric terms, we understand this as the set of publications which are cited by publications that we consider to represent sustainability science. We call this cited set the 'knowledge base' (Braam et al 1991) of sustainability science. So, in this publication we present a bibliometric analysis of the knowledge base (KB) of scientific publications from peer-reviewed journals that signal a strong relation with sustainable development or sustainability, and that we can thus consider to be a representative sample of sustainability science. We want to answer a number of questions with our investigation of this KB. First, we want to get a general impression of the research topics in the KB, which we refer to as the *research focus* of this set. Second, we are interested in the clusters of related publications that exist in the KB, as well as the different research foci in these clusters. Finally, we want to know if the interdisciplinary character of sustainability science (as seen through the KB) is changing over time. We think this investigation is interesting for at least three reasons. First, it will show what scientific results have been used to support sustainability science, so we can get an impression of the original research context in which these publications have been written (as opposed to the research focus in which they are used). Second, it pro-

R.K. Buter ✉ · A.F.J. Van Raan
Centre for Science and Technology Studies, Leiden University
PO Box 905, 2300 AX Leiden, The Netherlands
E-mail: buter@cwts.leidenuniv.nl

vides a basis for the monitoring of changes in the use of the KB, which may hint at changes in the field of sustainability science. Finally, our efforts could be used as the basis for a 'compass' (Lee 1993) to guide SD initiatives in a policy context: for example to find experts on particular topics, or organisations within a network of excellence.

In this paper we often use term 'research focus' to indicate the combination of topics that differentiates one set of publications from other sets. The research focus of a set of publications is part of a larger context, which we call the 'common ground' (Stalnaker 2002) of a field: the *presumed* background knowledge required to understand a publication or (by extension) to participate in a particular area or field. For example, a publication can refer to concepts that are not explicitly explained in the publications, but are still required to be understood in order to grasp the message of the publication. So, common ground is an intangible thing. However, tangible aspects of it are present in the network of topics present in a set of publications. We can even find it in expectations regarding the format of titles (Buter and Van Raan 2011). As we note in Buter (2012, p.29), common ground can also be revealed by 'exemplars' like the highly cited publications, as well as words, important authors and organisations. In this paper we will use a combination of citations, words, and field names in order to get an impression of what the research focus of a set of related publications is.

To investigate the knowledge base, we first have to operationalise SD in terms of scientific publications. The multifaceted nature makes this difficult and there will be many difficult corner cases. For example, take research into new building materials. Such materials could be lighter, yet isolate better, and use fewer toxic materials in their production chains. Clearly, such materials would be a good fit for sustainable urban development projects. However, whether research into such materials should be labelled as research relevant to SD in general is a matter of discussion. The basic research into the development of these materials was probably *not* done with the (specific) intent to use it in SD projects. We do not think we can resolve this. So, we opt for a more pragmatic approach instead, partially based on Clark and Dickson (2003) who point out that many interactions between science and SD are problem driven (*ibid*, p.8059). Hence, we will use as a representative sample of sustainability science those publications explicitly labelled so by their authors. Since a title of a publication is the most important instrument for stating the intent of a publication and is constructed to attract the attention of readers (Buter and Van Raan 2011; Soler 2007; Noyons 1999), we consider a title as the most important carrier for intended purpose, stronger even than an abstract. Thus, we will use as the representative set of sustainability science those publications with a word starting with 'sustainab' in their title.

After we selected these publications, we can collect the cited publications. For a number of reasons however, we do not want to use *all* cited) publications. One reason is that the role of citations varies: for example, some are included for social or political reasons, while others are there to indicate a relation based on content. Also, self-citations[1] may be included to self-promote research and not necessarily to indicate a direct connection to the research present in the citing publication (De Solla Price 1981). For a comprehensive overview of the different opinions on the role of citations we refer to Moed (2005, p.194). In addition, since one publication usually cites more than one other papers, a KB of is usually (much) larger than the set of citing publications. Moreover, many publications in a KB are cited only once or twice. Thus, we focus on the subset of highly cited publications. As already noted by Small (1978), such highly cited publication can be considered 'exemplars', 'pointers' to, or 'hubs' of knowledge. Even more, Peters et al (1995) showed that not only is there a link between the (highly) cited publication and the citing publications, but *also* that the citing publications are content-wise related to each other if they share references to such highly cited publications. As a result, we think that such a highly cited set of publications is a good representation of the research focus of sustainability science.

Many authors have expanded on the diverse challenges within SD. Some have taken a systemic approach from an educational perspective, like Elliot (2006). Others have taken a specific part of SD, like environment and economics, and expanded on that (Pearce et al 1990). With respect to the interactions between SD and science, many have taken an abstract, qualitative approach. Komiyama and Takeuchi (2006) for example, used their extensive experience to differentiate between three levels ('global', 'social', and 'human') and used this to classify problems in sustainability, not only at these three levels, but also where these levels interact. Clark and Dickson (2003) used their knowledge of the field to emphasise the pragmatic character of sustainability science, the interaction between scholars and practitioners, and the associated challenges and complexities. Kates et al (2001) stressed the interaction between nature and society and proposed a set of research questions which can be addressed by sustainability science. Kajikawa (2008) held an interesting middle ground, and combined a qualitative and quantitative analysis to arrive at a research framework for sustainability science. Our paper also features such a quantitatively driven approach: we apply a well-tested combination of citation analysis and topical analysis, using clustering and mapping. Our approach was pioneered by Braam et al (1991) and used in practice by for example Noyons et al (1999). Kajikawa et al (2007) used a similar approach to analyse publications

[1] A self-citation is a citation to a publication with at least one author that is also the author of the citing publication.

that have words starting with 'sustainab' in their title *or* abstract. Still, our approach differs on important points, as will be explained in the next section. In addition, our focus is on the analysis of the highly cited knowledge base and not on the representative sample of sustainability science publications itself. Still, we will briefly compare and contrast their results in our conclusions.

2 Materials and methods

The data for our analyses were extracted from the Web of Science (WOS)[2], a bibliographic database that covers a large part of the global output published in peer-reviewed science journals. The database contains important elements of publications, such as title, abstract, affiliations, keywords, and journal. Additionally, it contains the citations of publications, many of which to publications that are also in the WOS. Our institute (CWTS, Leiden University) has used the WOS to create a system that is fully dedicated to bibliometric analyses and indicator calculation. We have added many corrections to its parent (the WOS), such as affiliation and author names, and we also make use of a substantially improved publication identification algorithm. As a result, the CWTS database is better suited for advanced bibliometric analyses.

We already outlined our approach for finding publications that we consider as representative for sustainability science as well as the core of its KB, but we will repeat it here in more detail. A basic idea is that the title a very important instrument to signal the intended purpose of a publication. Hence, we collected publications from WOS published between 1999 and 2008 that have words in their title which start with 'sustainab' and have thus have a high chance that they are dealing with topics on sustainability or sustainable development. We call this collection of publications the *seed set*. Note that this set is not the complete set of sustainability science, but like Kajikawa et al (2007), we assume that it is a good representation of sustainability publications. Still, 'sustainable' and 'sustainability' can have different connotations in different contexts and we cannot be certain that this approach only finds those related to topics on SD. However, if a particular connotation has enough cited publications, than this should become visible in the knowledge base as a disparate set of publications in our analyses. After having created the seed set, we then collected publications *cited* by those in the seed set. This is the knowledge base (KB) of the seed set which we call the 'seed-KB', containing all cited publications of the seed set (i.e. the citing publications). For reasons we noted above (different grounds to cite, sparsity of a KB), we focused on a core of highly cited publications using two citation thresholds: one for the number of times a publication is cited *locally* by the seed set; and another to express how many times the publication is cited in *general* by the WOS. To arrive at the proper general citation threshold, we have to take into account that sustainability research cites papers from many fields and that this creates a bibliometric problem: since there are large differences in the lengths of the reference lists in different fields, there are also large differences in the expected number of citations for publications in fields. As a result, we cannot use an *absolute* minimum number of citations to express 'highly cited' in general. Instead, we must use a *field-normalised* citation indicator. Our institute has much experience with such indicators and therefore we chose the field-normalised citation measure CPP/FCSm[3], which was developed and used by CWTS in most of its bibliometric studies (Moed et al 1995; Van Raan 1996). We calculated CPP/FCSm values for the period 1999-2008. A value of 1.0 for the CPP/FCSm means that a publication is cited precisely at the average of the field, and thus, publications that have a value larger than 1.0 are cited above average. However, what value can be considered as the threshold between well cited and highly cited, is a more difficult question. In the past, we have experimented with the distribution function of the CPP/FCSm values in a set (Buter et al 2010a), but it remains difficult to arrive at a completely objective approach. Therefore, we used our experience in the application of bibliometric indicators in research performance assessment studies to settle on the following: at least five citations from the seed set and a CPP/FCSm larger than 1.5[4]. Applying these two thresholds to the seed-KB resulted in a subset of highly cited publications of the seed-KB, which we call the HCKB.

Our HCKB-approach differs from the approach taken by Kajikawa et al (2007) who analysed the maximally connected component within their seed set, i.e. the largest set of publications linked by citation. The advantage of our approach is that we can extract unlinked clusters of cited publications, that are still highly relevant for research covered by the seed set. However, we can expect the HCKB to be

[2] The WOS database has been licensed by CWTS from its producer, Thomson Scientific.

[3] CPP is the average number of citations per publication, without self-citations, in any set of publications (e.g., the publications of a research group, or the publications in a specific set). FCS (field citation score) is the average number of citations per publication, again without self-citations, for a whole field, worldwide. It acts as an international reference level and in the impact calculations it is used as a normalisation for CPP, yielding the field-specific impact indicator CPP/FCSm. Almost always a research group or any other larger set of publications covers more than one field and therefore we use the average FCSm. For the calculation of FCSm the same publication and citation counting parameters (e.g. time window, article types) are used as in the case of CPP. For the calculation of this indicator, the WOS journal subject categories (JSCs) are used as proxies for fields.

[4] Since a CPP/FCSm value of 1.0 means that the impact is precisely the field-average, a value of 1.5 indicates a 50% higher number of citations than the average. Based on our experience with performing bibliometric analyses, we know that a value of 1.5 can be used to identify highly cited publications.

strongly connected, because of the high minimal degree of the papers[5].

We found it interesting to include in some of our analyses the set of all publications from the period 1999–2008 that cite publications in the HCKB. This set has overlap with the seed set, but is a not a super set of the seed set because some publications in the seed set do not cite any publication in the HCKB (we will return to this issue, below). However, the set of all publications citing the HCKB can be considered as part of the larger research environment of the seed set, as it shares the interest in the HCKB. In our analyses below, we sometimes refer to this set as the 'all-citing set'. Another set we used in some of our analyses, is the knowledge base of the set publications from the period 1999–2008 published in *journals* that have a word in their title starting with 'sustainab'. Currently, the WOS covers twelve such journals[6] with a total of $3,556$ publications up to 2008. The journal that appeared first in the WOS is the *Journal of Sustainable Agriculture*, appearing in 1991 and having 970 publications up to 2008 in the WOS. The other journals appeared later, mostly from 1999 and onward. Due to the focus of these journals, we can assume that their publications are related to sustainability science. However, as opposed to our seed set, crucial contributions to sustainability science that appeared outside these journals (most notably from prestigious, multidisciplinary outlets such as *Science*, *Nature*, and *PNAS*[7]) are not covered in this set. To get at the core of the KB of this set, we required at least two citing publications from the journals; and since this already reduced the number of KB publications considerably, we did not require an additional minimum value for the CPP/FCSm to represent the global impact. We call the resulting set of cited publications the 'journals' knowledge base' (JKB).

We created in Figure 1 a schematic overview of the different publication sets used in this paper. The size of the ellipses do not precisely represent the size of the sets, as we merely wanted to express how the sets overlap. The outer set is the whole database (WOS) and all other sets are subsets of the WOS. As we see in the figure, the seed-KB partially overlaps with the seed set itself, because there are some highly cited publications that have a word starting with 'sustainab' in their title. Additionally, the seed set overlaps with the set of publications in the sustainability journals, as well as the JKB. The figure also shows that the HCKB is a complete subset of the seed-KB and partially overlaps with the JKB.

We investigated the structure of the HCKB by visualising and partitioning the networks of papers in the HCKB. For this we used the open-source networking tool 'Gephi' (Bastian et al 2009). This tool allows the investigation and manipulation of large graphs and contains functionality for spatialising, filtering, navigating, manipulating, partitioning and clustering. The Gephi tool uses the 'Louvain' algorithm by Blondel et al (2008) to partition networks. This algorithm is designed to quickly find partitions in large, weighted networks and has been tested on algorithms with over one hundred million nodes. The partitioning is done in two phases. At the start, all nodes are in their own community. The first phase then iteratively assigns nodes to the community of their next neighbour (according to some random ordering of the nodes) if the gain of modularity function is positive, otherwise it leaves the node in its original community. The second phase uses the communities found in the first phase and builds a network from these communities. Then, the first phase can be reapplied to this new network. The algorithm stops if no changes are made in the community structure. The final result is said to have a "maximum of modularity" (*ibid*, p.4). To layout the network we used a combination of the layout algorithm by Hu (2005), which is good at quickly structuring related communities, and the *Force Atlas* algorithm by Bastian et al (2009), which is good at spreading out these communities over a larger area, without distorting the general community layout created by the first algorithm. For more information on these algorithms, we refer to their papers. Note that we do not strive for accuracy in our efforts to partition and lay out the network, but are interested in the communities and their relative distance to get an impression of HCKB.

To create the clusters and the visualisation, we used the co-citation network of the HCKB *on* the seed set, i.e. the network of HCKB publications created by the reference lists of the publications in the seed set. Two publications are cocited if there is one publication that cites them both, while two publications are bibliographically coupled if they share at least one citation to a paper. A co-citation network represents the organisation of a KB by its use of the citing papers, while a bibliographically coupled network represents the organisation of a set of publications by way of how they use their (shared) KB. Co-citation and bibliographically coupled networks can also be used to get an impression of the coherence of a set, by looking at how strongly the network is connected: if there are many unconnected components in the network, then we consider the network loosely coupled.

As noted above, in our calculation of the CPP/FCSm, we use the Journal Subject Categories (JSCs) in the WOS as proxies for science fields. JSCs are certainly not perfect with

[5] At least five citations from within the seed set corresponds to a degree of at least five.

[6] At the time of our analysis, these were (in alphabetical order): *Agronomy for Sustainable Development*; *Environmental Progress & Sustainable Energy*; *International Journal of Agricultural Sustainability*; *International Journal of Sustainable Development and World Ecology*; *International Journal of Sustainable Transportation*; *Journal of Renewable and Sustainable Energy*; *Journal of Sustainable Agriculture*; *Journal of Sustainable Tourism*; *Proceedings of the Institution of Civil Engineers-Engineering Sustainability*; *Renewable & Sustainable Energy Reviews*; *Sustainability Science*; and *Sustainable Development*.

[7] Proceedings of the National Academy of Sciences of the United States of America.

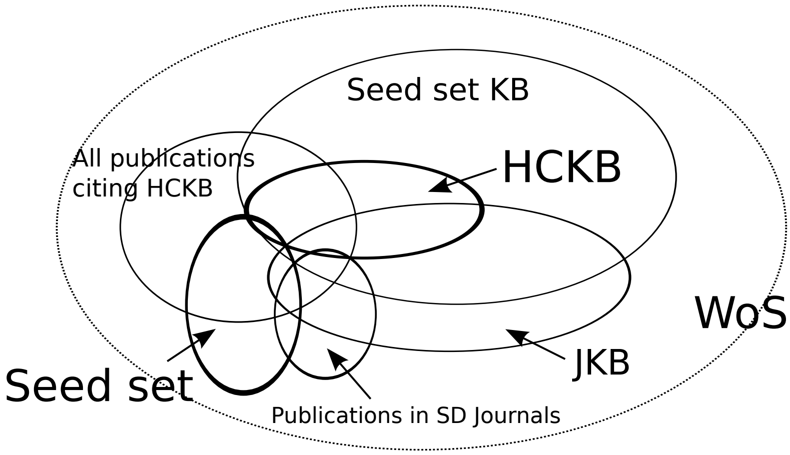

Fig. 1 A schematic overview of the different sets of publications used in the article: the WOS, containing the seed set, the KB of this set ('Seed set KB'), its subset the HCKB, the all-citing set ('All publications citing HCKB'), the set of publications from the 'sustainable' journals ('Publications in SD Journals'), and the KB from this set (JKB).

respect to the delineation of related research, and they provide a rather high-level classification to represent the whole of science[8]. Nevertheless, they are a readily available categorisation of related publications, their definition is standardised in a transparent way, they are stable over time, and their use is consistent with our use of the CPP/FCSm.

One way to get an impression of the research orientation in the HCKB, is by comparing it with the other sets. For this, we use both fields and clusters, as they capture the research context of the publications in different ways: fields are more general and high-level, while clusters are more local, because (in our case) clusters are based on the use of publications in reference lists. To compare the research focus in these subdivisions of publications (fields and clusters), we plotted the share of publications as a graph (see for example Figure 3, below). This approach was also used in Noyons and van Raan (1998) to compare the shift in cognitive orientation in the research area of 'Optimisation'. A spike in this graph implies that there is an orientation on a specific field or cluster in one set that is not present in the base set (the HCKB, in our case), and a valley implies that the HCKB focuses on research not represented in the other set.

To assess the content of the publications in the clusters, we used frequently occurring noun phrases (NPs) from titles and abstracts. Noun phrases consist of a noun and its accompanying modifiers, like 'red ball' or 'complex algorithm'. However, not all frequent phrases are informative: a general phrase like 'paper' occurs frequently in many publications but does in general not convey much information. Added to that, even specific phrases are not always useful: they could be part of the jargon in some research area and used in many publications of that area and thus be as non-informative as 'paper' is in general. So, to identify research topics in a cluster, we should prefer frequent, informative phrases. To find such phrases, we used the following approach. First, we applied a standard list of 'non-informative' phrases[9] that we have developed over the years, including phrases such as 'paper', 'result', and 'improved method'. Then for the remaining NPs, we calculated the relative occurrence: $p_i(\text{NP}) = n_i(\text{NP}) / \sum_i n_i(\text{NP})$, where $n_i(\text{NP})$ is the number of publications for some NP in cluster i. We used this to calculate a Shannon entropy value: $-\sum p_i(\text{NP}) \ln(p_i(\text{NP}))$. If an entropy value for some NP is low then the NP is in for some cluster and is a probably an informative phrase describing research specific to that cluster (Quinlan 1993; Buter and Noyons 2005; Buter et al 2010a).

We use interdisciplinarity and transdisciplinarity to refer to different states in the development of a scientific field (Gibbons et al 1994; Van den Besselaar and Heimeriks 2001; Buter 2012). This development starts with a multidisciplinary approach in which different contributions from fields address a single topic, but without exhibiting integration beyond that. Then, an interdisciplinary approach differentiate from the former by the development of concepts and methods particular to the topics being addressed, as well as an increasing overlap between the different contributions (for example, manifested by citing the same basic literature). Still, interdisciplinary research maintains strong links to the originating disciplines. Finally, a transdisciplinary state is reached when the approach has become integrated, up to a point that the research is more or less self-sustained and newly developed, shared concepts and knowledge has become the basis for further research. To measure changes in the interdisciplinary character of sustainability science, we can look at the distribution of references over fields and clusters over time. We also used the Shannon entropy to capture this distribution in a single metric, similar to the approach in Buter and Noyons (2005); Rafols and Meyer (2009); Buter et al (2010b), and also bears relation to that of Larivière and Gin-

[8] At the time we conducted our analyses, there were 246 JSCs.

[9] This is a list composed by us on the basis of the results of many CWTS studies with NPs from scientific publications; it can be requested for research purposes from the first author of this publication.

Year	P	Growth
1999	674	100
2000	789	117
2001	751	111
2002	787	117
2003	946	140
2004	965	143
2005	1,192	177
2006	1,245	185
2007	1,410	209
2008	1,835	272
Total	10,594	-

Table 1 The number of publications P per year containing a word starting with 'sustainab' in their title and the indexed growth (year 1999 = 100).

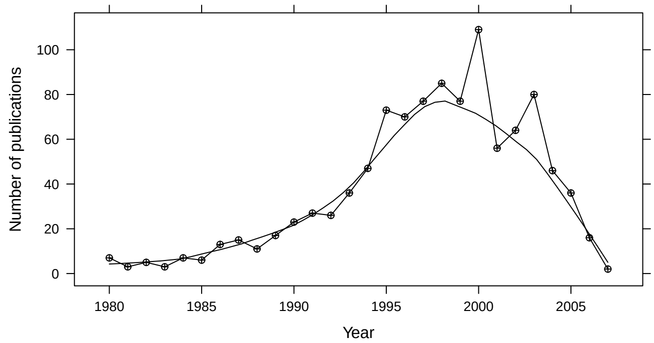

Fig. 2 The number of HCKB publications per year and a smoothed version of the distribution.

gras (2010). We consider a research area to become transdisciplinary if its researchers start to merge concepts and methods from different disciplines into new concepts and methods. If a research area is successful in becoming transdisciplinary, then it should become less and less dependent on other fields, i.e. extract less and less knowledge from other fields and thus show an increasingly lower diversity value. In other words, the research in a transdisciplinary area becomes increasingly self-supporting and specialised. Thus, this would correspond to lower diversity values for the reference lists of the citing publications.

3 Results and Discussion

3.1 General

We start with some general results. In Table 1 we show the number of (WOS covered) publications (P) in our seed set, broken down by year, as well as the indexed growth (year 1999 = 100). The table shows that the number of publications that have words starting with 'sustainab' in their title, is steadily growing: from 674 publications in 1999 to nearly triple that amount (1,835 publications, 272%) in 2008. In total, the seed set contains 10,594 publications.

The seed set publications cite 66,294 different publications. As noted before, many publications are cited only once or twice by seed set: 55,013 are cited only once, 11,281 publications are cited at least twice, and only 1,379 are cited by five or more[10]. After calculating the field-normalised impact of the publications, we found that 1,038 publications are cited by more than five seed publications and also have a CPP/FCSm value above 1.5. This set is the highly cited knowledge base (HCKB). We did not restrict the publications in the HCKB to the period 1999–2008, because important, highly cited papers may be published before 1999. Figure 2 displays the number of HCKB publications broken down by year and shows that most of the 1,038 cited publications were published between 1994 and 2004. This is not surprising because authors prefer to cite recent work and the seed set contains publications from the period 1999–2008. We consider the peak in 2000 followed by a sharp decline 2001 a database artefact and do not attach any significant meaning to it. So, the smoothed version of the distribution probably shows a more accurate distribution.

One way to measure the coherence of the HCKB, is by looking at the number of seed publications that *share* a reference to a publication in the HCKB, i.e. how many seed publications are bibliographically coupled if we use the HCKB to create the couplings. As it turns out, 29% (3,063 out of 10,594) of the publications in the seed set share one or more references to HCKB publications, comparable to the percentage of papers in the maximum connected component of Kajikawa et al (2007). Still, this number appears to be small, but if we take into account that only 42% (4,484 out of 10,594) of the seed set publications are bibliographically coupled *in general*, i.e. by using the *whole* WOS instead of the HCKB to create the couplings, we see that the seed set is already a loosely coupled network of publications. Then, if we take the subset of bibliographically coupled seed publications, the HCKB couples over two thirds (68%, 3,063 out of 4,484), a much higher share than the 29% of the total number of seed set publications. To us, this reinforces that the HCKB is a valid representation of the core of the scientific basis of sustainability research.

3.2 Fields

Table 2 shows the prevalent fields (in terms of the JSCs of the WOS) in the HCKB, and the corresponding number of pa-

[10] Note that we only indicate the number of citations from the seed set; the number of citations from all WOS covered publications may be larger.

pers for those fields in other sets[11]: the seed set, the set of all publications citing the HCKB ('all-citing'), and the JKB. At the top of Table 2, we find fields related to economical and environmental research. Further down, we also find agriculture, chemistry, and (ranked thirteenth) sociology. Also, we see that the JKB is much more focused on *Energy & Fuels*, *Agronomy*, *Soil Science* and *Plant Sciences* than the HCKB. Additionally, the focus on environment and ecology appears to be larger in the all-citing set than in the HCKB.

To further investigate these differences, we created Figure 3. The figure shows the share of fields in the different sets in the same order as in Table 2, with labels attached to some of the most important outliers (spikes as well as valleys). The large spikes confirm the focus in the JKB on the fields *Energy & Fuels*, *Agronomy*, *Soil Science*, and *Plant Sciences*. Most likely, these spikes represent the specialisation of the sustainability journals at the basis of this set, such as the *Journal of Sustainable Agriculture* and *Renewable & Sustainable Energy Reviews*. Less apparent in Table 2, but now better visible is the valley labelled *Economics*, meaning that this research is more specific to the HCKB than to other sets (even the seed set). *Ecology* on the other hand has a higher relative coverage (about 15%) in the set of all publications citing the HCKB (the all-citing set), and shows an interesting spread of coverage in the other sets, notably the seed set with a coverage of about 7%. The fourth field, *Environmental Studies* is covered better than *Ecology* in the seed set, but worse in the all-citing set. *Management* is another field that is covered more in the all-citing set, but much less in the seed set, while *Engineering, Environmental* and *Water & Freshwater Biology* is covered better by the seed set. The *Multidisciplinary Sciences*, which covers journals like *Science* and *Nature*, have (as could be expected) a higher coverage in the HCKB than in the seed set and the all-citing set (Nature and Science publications are often highly cited). However, the JKB also shows a rather low coverage of publications in *Multidisciplinary Sciences*. We think this can mean two things. First, the high coverage of the fields we mentioned above could overwhelm the ratio of publications from *Multidisciplinary Sciences*. Alternatively, it may mean that such papers have a less prominent role in the JKB, perhaps because those publications are more focused and more 'normal' science in the sense of Kuhn[12].

[11] In Figure 1, we also show the 'seed set KB' and 'publications in SD journals', but we find both these sets less relevant for our investigation: the first because it is too wide, and the second set because we want to focus on the KB of sustainability science and have used the publications in SD journals only to obtain the JKB.

[12] In Kuhn (1970) a distinction is made between phases of 'paradigm shifts' and 'normal science'. The former is a phase in which large changes in understanding of particular topics results in a quick build-up of publications, while the latter is the phase between such changes, in which most scientific publications are produced.

Some fields are not in the top most common fields in the HCKB, but are covered well by other sets. For instance, the fields *Thermodynamics* (#9 with 149 publications) and *Mechanics* (#11 with 92 publications) are covered well by the JKB. This is understandable, because one of the twelve journals at the basis of the JKB is *Renewable & Sustainable Energy Reviews*. Also, we find *Psychology* ranked 14th in the all-citing set, with 2,719 publications; however, it has no prominent coverage in the other sets. From this we may hypothesise that psychologists are interested in sustainability issues (such as topics on human behaviour) so they cite the HCKB, but that psychology is not strongly represented in the scientific basis of sustainability research, otherwise it would have had a stronger position in the HCKB. Thus, the difference in field coverage for the various sets is understandable from the content and orientation of the publications on which the sets are based. We think that this confirms that the HCKB represents the broad disciplinary range of sustainability research rather well.

3.3 Clusters

We move on to the analysis of the clusters in the HCKB. Using the Louvain algorithm described above, we find that the HCKB can be divided into 12 communities or clusters. We list these clusters in Table 3, together with the number of publications that make up the cluster, the number of citations from the seed set and the all-citing set, as well as the overlap with publications in the JKB to indicate where the HCKB and the JKB share research themes. The last column contains the most informative NPs. We extracted these NPs from the titles and abstracts of the clustered papers using the methodology described above for the selection of informative NPs, using an entropy value of less than 0.8 and at least four occurrences[13]. Clusters 6 and 10 did not yield any informative NPs and for these clusters we manually selected the NPs shown in Table 3. The descriptive label in the third column is our interpretation of these NPs and the most dominant journals and fields in the cluster. The table illustrates how the research foci in the clusters differ. Based on these clusters and their descriptions, we note that prevalent ones in the HCKB are a combination of economical and ecological themes, while the core social themes appear limited to the cluster we labelled *Rural and urban studies* (4). This prevalence is in line with what we found when inspecting the most fields in Table 2. The largest clusters are *Ecosystems and diversity* (4) and *Ecological economics and policy* (7), which both represent about 20% of the HCKB publications. The largest overlap with the JKB publications is (not surprisingly) with *Agricultural aspects* (9), but there is also

[13] These numbers were a result of a limited number of experiments with different values for these parameters.

#	Field	n_{HCKB}	n_{seed}	$n_{all\text{-}citing}$	n_{JKB}
1	Environmental Sciences	228	2,003	12,523	435
2	Economics	200	757	9,808	160
3	Ecology	194	890	15,336	300
4	Environmental Studies	183	1,400	5,579	243
5	Agriculture Multidisciplinary	86	513	2,000	202
6	Multidisciplinary Sciences	86	254	1,721	69
7	Planning & Development	82	621	2,918	83
8	Management	79	271	8,781	34
9	Business	75	174	6,629	15
10	Geography	40	528	2,564	32
11	Energy & Fuels	40	453	1,620	634
12	Engineering Environmental	34	722	1,937	56
13	Sociology	33	203	2,716	62
14	Forestry	30	397	2,757	15
15	Biodiversity Conservation	25	77	3,253	18
16	Urban Studies	22	357	1,405	23
17	Agronomy	22	252	2,204	332
18	Fisheries	18	120	1,876	0
19	Engineering Chemical	17	327	1,937	85
20	Marine & Freshwater Biology	17	164	4,231	3
21	Social Sciences Mathematical Methods	16	17	1,027	2
22	Soil Science	16	124	3,436	236
23	Water Resources	15	569	1,856	52
24	Operations Research & Management Science	15	91	1,686	21
25	Psychology Multidisciplinary	14	46	2,263	7
26	Social Sciences Interdisciplinary	13	100	1,135	7
27	Statistics & Probability	12	9	604	1
28	Chemistry Multidisciplinary	12	229	2,740	8
29	Public Environmental & Occupational Health	12	226	2,364	2
30	Plant Sciences	12	141	4,626	127
31	Construction & Building Technology	11	184	184	15

Table 2 The twenty most frequently occurring science fields in the HCKB, together with the corresponding number of publications in the seed set, the set of all publications citing the HCKB, and the JKB.

significant overlap with the two largest clusters. We will get back to differences between sets below, when we introduce Figure 5. We can also see differences in age of the cited and citing publications: the average age in the *Econometrics* cluster for example is 1991, well below the average of 1997; and while most of the citing publications are from 2004, we see that those for the *Green chemistry and engineering* are slightly younger. We think this means that the former cluster consists mostly of classics[14], while the latter is more representative of current, 'normal' science.

In Figure 4 we show a visualisation of the co-citation network of the HCKB and the clusters in a two-dimensional map, based on the co-citation of HCKB publications in the seed set. The size of the nodes is relative to the CPP/FCSm value of their corresponding publications. We indicated each cluster with a different colour and labelled these clusters using the cluster number from Table 3. Additionally, an edge that connects a node in one cluster with a node in another cluster, has a colour that is a mix of the two colours of the nodes. This for example results in the brown area between the yellow cluster *Australian agricultural research* (11) on the top left and the red cluster *Ecological economics and policy* (7) in the middle. The hub (connecting node) between these two clusters is King et al (2000), which is a review on sustainability indicator development and use in Australia in relation to agricultural production systems, and is thus clearly a link between the Australian research context of the yellow cluster and the larger environment of sustainable agriculture that is represented by the red cluster. The blue cluster, *Green chemistry and engineering* (8) has an elongated shape, occupying an area from top to bottom: at the top, the focus is on 'green energy', moving through 'green engineering' in the middle, until we reach 'green chemistry' at the bottom. The green cluster *Business and management* (5) on the right, contains the largest nodes, indicative for the largest CPP/FCSm values. The largest, Baron and Kenny (1986), is a true classic on the distinction between moderators and mediators in psychology, and it is not surprising that this publication also has many connections with the purple cluster *Rural and urban studies* (3) at the top of the visualisation, which represents a large part of the social science pub-

[14] These are relatively old publications that are used to indicate the general research background of a publication, instead of being relevant to the topics discussed in the citing publications.

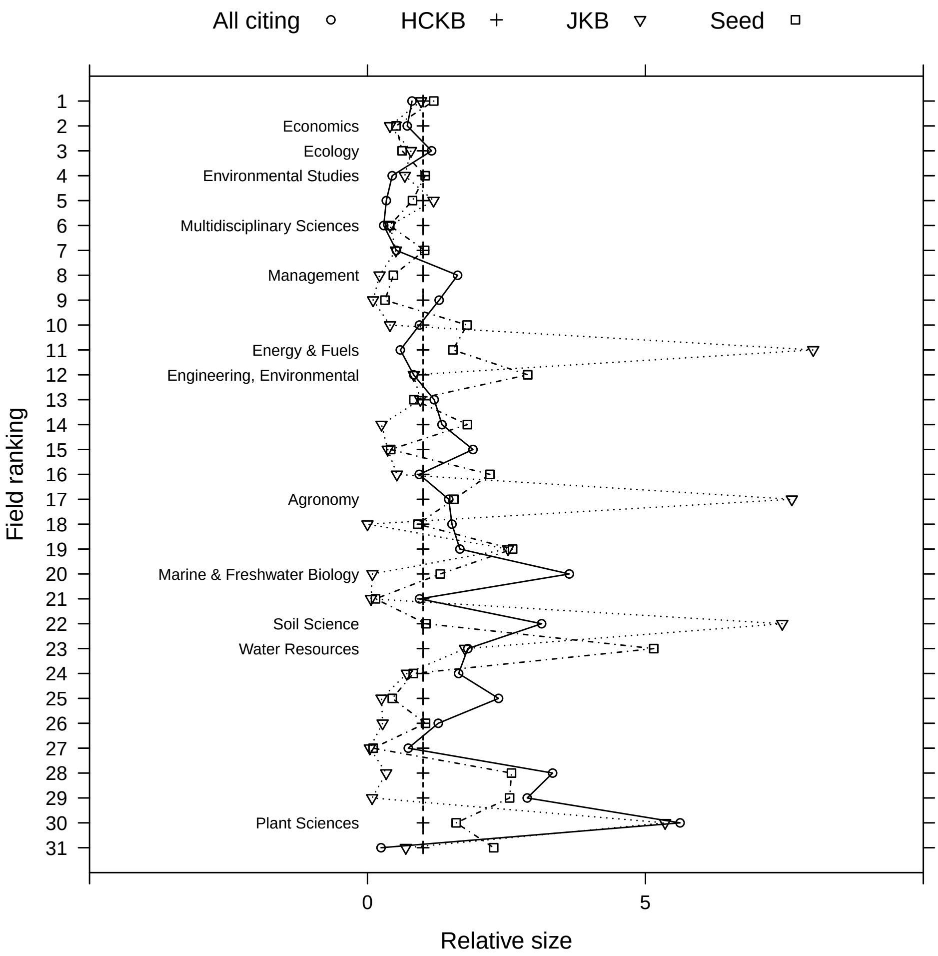

Fig. 3 Relative publication counts from Table 2, same rank numbers(#) as in Table 2. Some important outliers have been labelled with the field to which they belong.

lications in the HCKB. The cluster with the largest number of Nature and Science publications, however is *Ecosystems and diversity* (4) and not cluster 5.

Both Table 3 and Figure 4 show a number of (to us) unexpected clusters: *Polymer science* (6), *Membranes and purification* (10) and *Health programs* (12). All these clusters only have a small number of publications (3, 4 and 17). In the visualisation of Figure 4, clusters 10 and 12 are clearly separated from the core of the field, but cluster 6 is hidden among the nodes of the red cluster *Business and management* (7). Still, the themes of these three clusters, represented by the NPs in Table 3, appear to make sense in the context of sustainability science. In fact, Kajikawa et al (2007) feature a *Health* cluster that has a similar research focus as *Health programs*. Figure 4 shows that the cluster *Econometrics* (1) is also separated from the core of the network of papers. It is attached mostly to cluster *Business and Management* through publications on public debt and government spending. Probably, this is due to a particular connotation of the word 'sustainabile', in the context of public spending. In addition, the cluster represents a body of mostly theoretical and rather old (as can bee seen in Table 3) work, clearly not written with the intent of use in a sustainability science context.

These outliers and a cluster such as *Australian agricultural research* which represents the interest of a research community in a particular geographical context, affirms that the clusters are a result of the *use* of the HCKB publications by the seed set. In other words, the context which we have visualised in Figure 4 is the context of use, which is mostly different from the research context in which the HCKB publications have been published. This also becomes apparent when when we look at Figure 5, which shows a 'spectral' analysis similar to Figure 3, but instead of fields we now compare the (relative) number of publications in the HCKB *clusters* with the number of citing publications from the seed set and the all-citing set, as well as the number of JKB publications in the cluster. The figure illustrates for example that the interest in papers from the *Business and management* (5) cluster is largest from the all-citing set of publications

#	Label	P	%	C_{seed}	C_{all}	O_{JKB}	Y_{HCKB}	Y_{seed}	NPS
1	Econometrics	30	2.9	66	5,092	0	1991.2	2004.8	unit root, deficit
2	Forests and conservation	93	9.0	316	10,449	18	1996.6	2004.3	forest, ecosystem, landscape, tree, fire
3	Rural and urban studies	130	12.6	543	9,021	32	1997.1	2004.8	behavior, consumption, city, politics
4	Ecosystems and diversity	212	20.6	967	37,104	52	1997.9	2004.7	ecosystem, species, fishery, resilience
5	Business and management	108	10.5	396	29,507	17	1994.9	2004.4	firm, organization, resource based view, competative advantage
6	Polymer science	3	0.3	12	47	0	1997.0	2004.7	resin, linseed oil, hydroxyethyl
7	Ecological Economics and policy	198	19.2	846	7,906	65	1997.4	2004.7	energy, economy, natural capital, ecological footprint
8	Green chemistry and engineering	52	5.0	212	8,759	15	2000.3	2005.7	energy, industry, biomass, ionic liquid
9	Agricultural aspects	139	13.5	576	6,459	73	1997.9	2004.8	sustainable agriculture, farm, sustainability indicator
10	Membranes and purification	4	0.4	10	402	0	2000.0	2006.8	membrane bioreactor, domestic wastewater
11	Australian agricultural research	45	4.4	92	1,415	5	2000.4	2003.5	pasture, water, sustainable grazing system, persistence
12	Health programs	17	1.7	53	612	0	1998.8	2005.7	organization, community, program, intervention
Total or average		1,037	100	4,089	116,773	277	1997.5	2004.9	

Table 3 Clusters of the HCKB. The columns indicate, from left to right, the cluster number *#*; the name of the cluster (*Label*); the absolute (*P*) and relative (*%*) number of publications; the number of citations to the HCKB cluster from the seed set (C_{seed}) and the all-citing set (C_{all}); the overlap (O_{JKB}) between the HCKB cluster and publications in the JKB; the average year of publication in the cluster (Y_{HCKB}) and the average year of the citing seed publication (Y_{seed}); and finally the most informative NPs for that cluster.

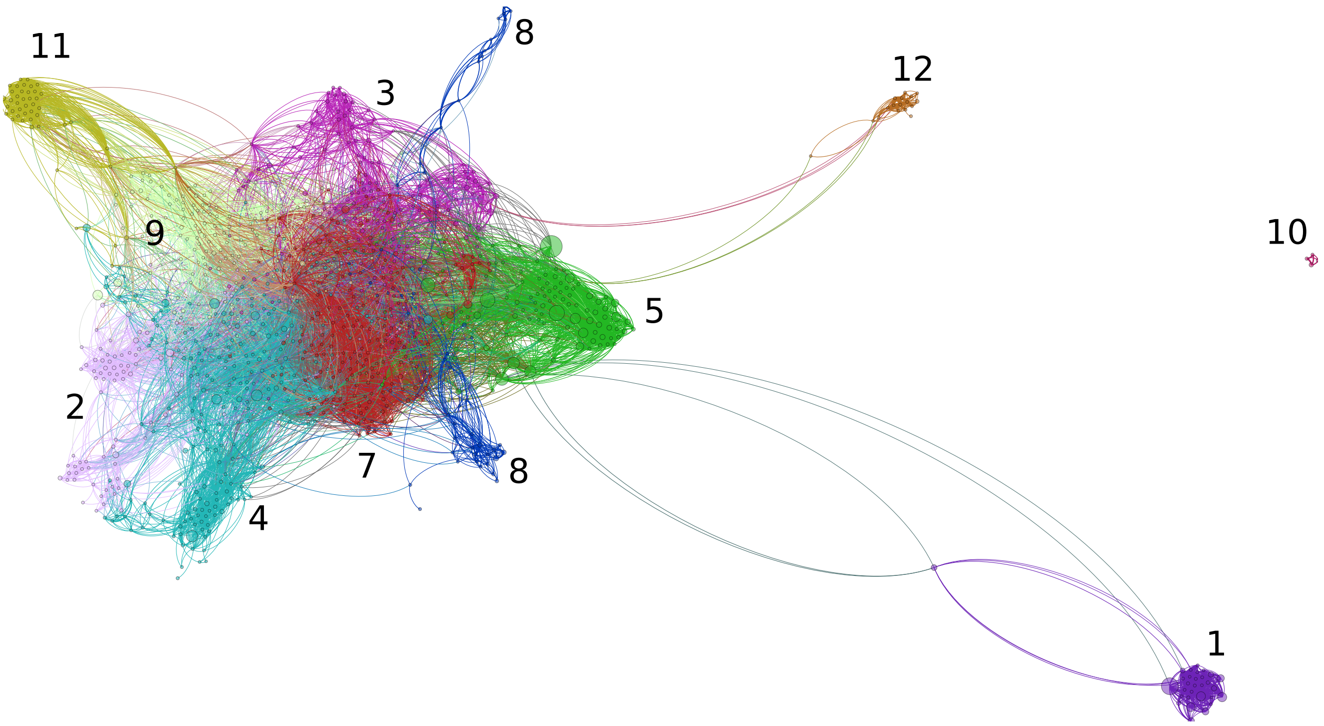

Fig. 4 A visualisation of the co-citation network of the HCKB, using the seed set to create the co-citation links. The colours of the nodes indicate the different communities or clusters in the HCKB, while the colours of the edges are a combination of the colours of the connected nodes. The clusters have also been labelled with their corresponding identifier from Table 3. The size of the nodes is relative to the CPP/FCSm value of their corresponding publications.

and that this set also has a relatively high interest in the *Econometrics* publications, which we found to be separated from the core of the network, above. The unexpected clusters *Polymer science* (6) and *Membranes and purification* do not have any overlap with the JKB publications, but are used mostly by respectively the all-citing set and the seed set. Also, although the focus of the *Australian agriculture research* cluster is clearly similar to parts of the JKB, the corresponding publications do not appear as much in the JKB. The most used journal in this cluster is for example the *Australian journal of Experimental Agriculture*, which has no direct connection with sustainability research in its title.

Since the HCKB is made up of highly cited publications that can be thought of as exemplars of larger bodies of knowledge, it is interesting to look at what exemplars (publications) in clusters are used most by the seed set. This will

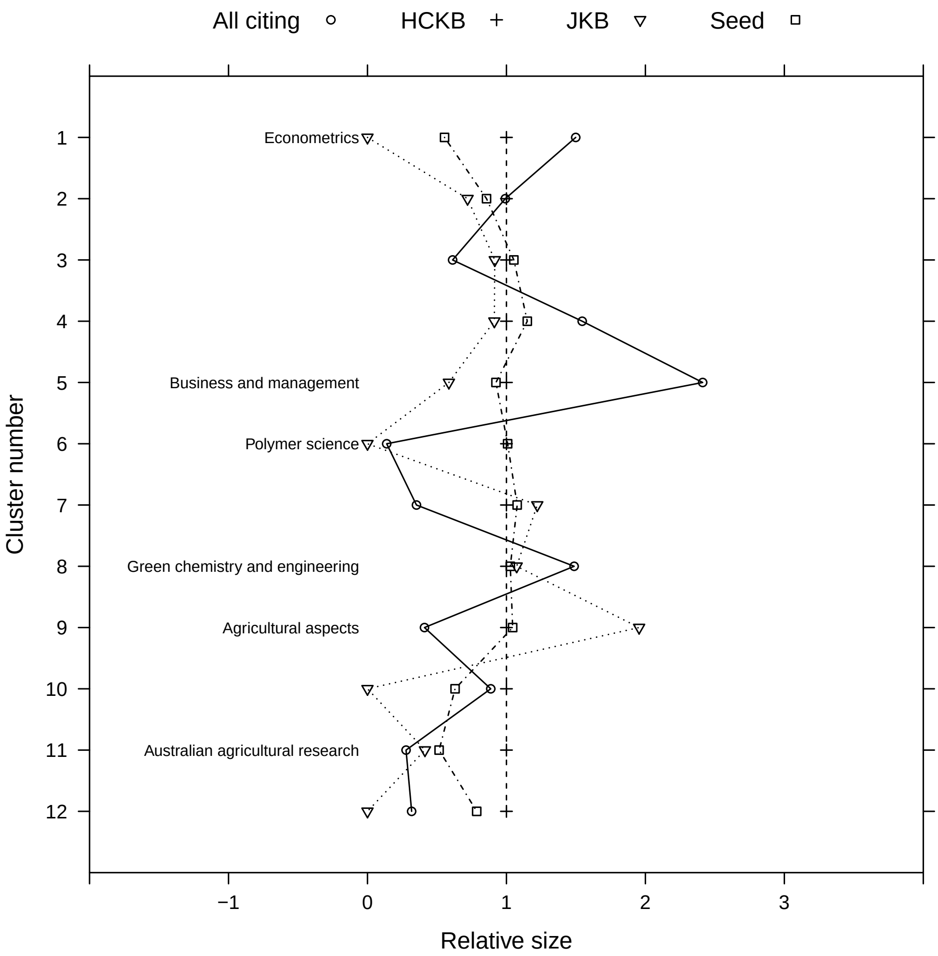

Fig. 5 Share of the publications in the clusters from the HCKB, and the number of citing publications from the seed set, the set citing all HCKB publications, and the JKB. We note that the horizontal axis also includes a tick at -1, although size cannot be less than 0; however, we used the space on the left of the figure to highlight a number of interesting results.

give us an impression of what the pillars of sustainability research look like. We do this by showing two results: in Table 4 the overall best cited publications (in the seed set); and in Table 5 the best cited publications published after 1999. This separation is useful, because in the second table will not contain true (old) classics. To prevent the tables becoming too large, we only list the single most cited publication of a cluster. In Table 4 we again see that the titles of the cited publications are in line with the research focus. But we also note that their focus appears to be fundamental in character, by which we mean they deal with basic concepts and definitions instead of case studies. In addition, we immediately spot a true classic in the *Econometrics* cluster, published in 1986. Not all publications are classics or fundamental, though: cluster 11 for example features a case study published in 2003 that reported on experiments “to develop more profitable and sustainable pasture systems” Michalk et al (2003, p.861). Turning to Table 5, we see that there is a general shift from concepts and definitions, toward reviews and publications on tools and frameworks. There are additional case studies as well, for clusters 5 and 10. What we learn from both tables is that the knowledge base indeed consists of pillars and that the generality of these papers mostly correlates with the age of the publications.

3.4 Diversity

One of the things we learn from the results above, is that the HCKB represent the interdisciplinarity character of sustainability science. Moreover, Figure 4 shows a strongly connected network. So, another question is how this interdisciplinarity is visible in the seed set and how that changes over time. To do this, we calculated the diversity of fields in seed set reference lists, both in general (WOS) and for HCKB citations. In addition, we calculated the diversity of clusters in these lists. And for comparison, we also calculated the diversity of fields in which publications of the seed set are

#	Cluster	Title	Journal	Year	C	I	N
1	Econometrics	On the limitations of government borrowing - A framework for empirical testing	American Economic Review	1986	113	8.7	34
2	Forests and conservation	What is ecosystem management	Conservation Biology	1994	371	25.0	19
3	Rural and urban studies	Sustainable Development - A critical-review	World Development	1991	164	6.9	51
4	Ecosystems and diversity	The value of the world's ecosystem services and natural capital	Nature	1997	1061	18.9	85
5	Business and management	Green and competitive - Ending the stalemate	Harvard Business Review	1995	316	21.6	35
6	Polymer science	Studies on a newly developed linseed oil-based alumina-filled polyesteramide anticorrosive coating	J. of Applied Polymer Science	1999	23	2.6	11
7	Ecological economics and policy	On the intergenerational allocation of natural-resources	Scandinavian J. of Economics	1986	147	11.4	37
8	Green chemistry and engineering	Indicators of sustainable development for industry: a general framework	Process Safety And Environmental Protection	2000	19	2.4	19
9	Agricultural aspects	Is agricultural sustainability a useful concept?	Agricultural Systems	1996	45	7.6	36
10	Membranes and purification	Critical flux concept for microfiltration fouling	J. of Membrane Science	1995	214	25.5	9
11	Australian agricultural research	Sustainable grazing systems for the Central Tablelands, New South Wales	Australian J. of Experimental Agriculture	2003	18	5.6	14
12	Health programs	Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy	Health Education Research	1998	65	7.1	26

Table 4 The best cited publication in a cluster, by number of citations from the seed set (N). Also shown is the title of the publication, the year of publication, the number of citations in the WOS (C) (between 1980-2008) and the impact (CPP/FCSm, indicated by I).

#	Cluster	Title	Journal	Year	C	I	N
1	Econometrics	US Deficit Sustainability: A New Approach Based On Multiple Endogenous Breaks	J. of Applied Econometrics	2000	11	1.6	7
2	Forests and conservation	Forest Certification - An Instrument To Promote Sustainable Forest Management?	J. of Environmental Management	2003	27	4.2	13
3	Rural and urban studies	Are We Planning For Sustainable Development? An Evaluation Of 30 Comprehensive Plans	J. of The American Planning Association	2000	27	4.8	21
4	Ecosystems and diversity	Tracking The Ecological Overshoot Of The Human Economy	PNAS	2002	85	2.2	31
5	Business and management	Assessing The Sustainability Performances Of Industries	J. of Cleaner Production	2005	5	1.7	14
7	Ecological economics and policy	Strategic Sustainable Development - Selection, Design And Synergies Of Applied Tools	J. of Cleaner Production	2002	40	4.8	37
8	Green chemistry and engineering	Indicators Of Sustainable Development For Industry: A General Framework	Process Safety And Environmental Protection	2000	19	2.4	19
9	Agricultural aspects	Assessment Of The Contribution Of Sustainability Indicators To Sustainable Development: A Novel Approach Using Fuzzy Set Theory	Agriculture Ecosystems & Environment	2001	21	2.6	18
10	Membranes and purification	Fouling Transients In Nominally Sub-Critical Flux Operation Of A Membrane Bioreactor	J. of Membrane Science	2002	62	11.5	6
11	Australian agricultural research	Sustainable Grazing Systems For The Central Tablelands, New South Wales	Australian J. of Experimental Agriculture	2003	18	5.6	14
12	Health programs	Is Sustainability Possible? A Review And Commentary On Empirical Studies Of Program Sustainability	American J. of Evaluation	2005	10	2.7	12

Table 5 The best cited publication in a cluster, by number of citations from the seed set (N) and published in 2000 or later. Also shown is the title of the publication, the year of publication, the number of citations in the WOS (C) (between 1980 and 2008) and the impact (CPP/FCSm, indicated by I).

published. Figure 6 represents the results. In this figure we scaled to values between 0 and 1 (the maximum) in order to be able to compare the curves better. We see that the diversity of the fields in which the seed sets are published is mostly constant and perhaps slowly declining. However, the curve for the diversity of fields in citations to the HCKB (the top plot) is growing and sharper than in the general WOS context. We could explain this by assuming that the topic of sustainability is spreading over more clusters, but then the top curve would have to grow as well. Instead, we think that this shows that the field is still integrating new knowledge from different fields and not yet converging towards knowledge from particular fields. With respect to the diversity of clusters in the reference lists, we see a more erratic growth: the first period shows a decline, while the second period shows a sharp growth followed by a sharp decline again. This is for us difficult to understand and deserves more attention in the future.

4 Conclusions and future research

We presented a bibliometric analysis of a set of highly cited publications from the knowledge base of sustainability science using data from the Web of Science database (WOS). We started by identifying publications that specifically indicate their intent to present research related to sustainability or sustainable development, by selecting publications that have a word that starts with 'sustainab' in their title. From

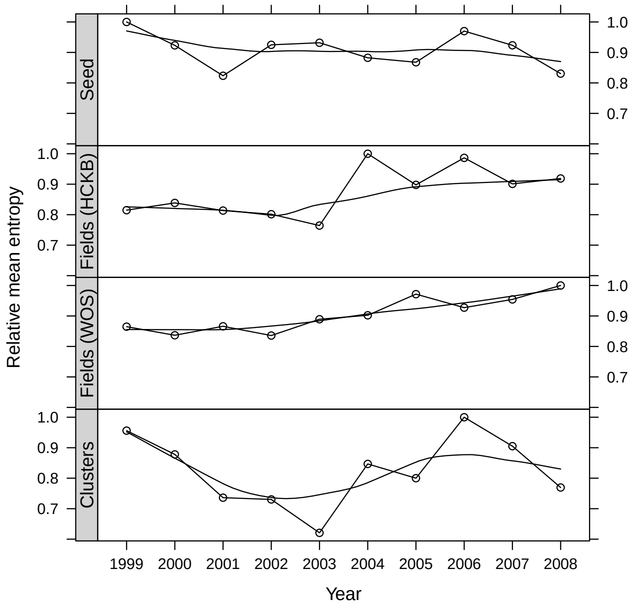

Fig. 6 Diversity of fields in which the seed set publications appear ('Seed'), the diversity of fields in citations from the seed set to the HCKB publications ('Fields (HCKB)') and to to the WOS in general ('Fields (WOS)'), as well as the diversity of clusters in citations from the seed set ('Clusters'). All are measured as the (mean) entropy of the occurrences of fields and clusters and scaled between 0 and 1 for better comparison of the slopes of the curves.

the resulting 'seed set' of publications we then identified the knowledge base, which are the publications cited by the seed set. Finally, we selected the strongly cited publications from this knowledge base, resulting in a strongly connected set of papers which we called the highly cited knowledge base (HCKB).

First we investigated the research orientation of the HCKB by comparing the coverage of fields in the HCKB with the seed set, the set of all publications citing the HCKB, and the knowledge base of the publications published in the specialised sustainability journals that we called the JKB. We found that the focus in the HCKB is mostly on economics and ecology, but not as much on social sciences. In addition, we found that in particular the orientation of the JKB was different from the HCKB. The JKB shows an orientation on energy and agriculture, but contains less research on for example economics and management. Moreover, the coverage of important multidisciplinary journals such as Nature and Science (in this study represented by the subject category *Multidisciplinary Sciences*) is relatively low in the JKB. We conclude that the research published in these specialised journals does not represent the full, interdisciplinary spectrum of sustainability research and that the JKB may be more representative of a knowledge base of 'normal science' than the HCKB.

Next, we presented a clustering of the HCKB, using a co-citation network of the HCKB publications in the seed set, representing the *use* of the HCKB in the seed set. We found twelve clusters for which we identified the research focus in these clusters by investigating the informative NPs in these clusters in addition to the most prominent journals and fields. The visualisation of the network and the clusters within the network showed that the HCKB is mostly a strongly connected network, with some small exceptions. We also indicated some interesting hubs between clusters. Again, we found a focus on economics and ecology and less on social sciences. In addition, we noted that much attention of the important journals (like Nature and Science) has gone to publications on ecosystems and diversity issues. We also used clusters to show on what exemplars these clusters are 'built', by listing the most cited publication in a HCKB cluster, measured by the number of citations from the seed set. Again, this confirmed the identified research focus of the clusters and also showed that younger HCKB publications focus more on case studies than the older, better cited 'classics'. This analysis again confirmed that sustainability science is published in a wider context than the set of journals with 'sustainab' in their title, as for example featured by a cluster on sustainability in the context of Australian agriculture.

Finally, we investigated the change in diversity of the seed set and the references in the seed set. We learned that the diversity of the fields in which seed set publications appear is stable and perhaps slowly declining, but that the diversity of the fields to which these seed publications refer is increasing. The diversity of the references to publications in particular clusters was more erratic and we suggested that this may be an interesting topic for further study. From this we conclude that the interdisciplinary approach to solve the complex issues in sustainability science is still in development, and that researchers are still integrating knowledge from a increasingly richer palette of fields. As a result, we do not see a trend towards a state of transdisciplinary research, but if the field of sustainability science becomes more established, we expect the diversity of references to decrease, as citations would go to a more confined (transdisciplinary) environment.

Although orientation and methodology of our paper differs on important points from Kajikawa et al (2007), there are a number similarities which warrant a quick comparison of results. Kajikawa et al. find more sustainability publications because they also look for 'sustainab' words in the abstract of publications, whereas we only consider the title, which we regard as an even stronger indicator for the intent of authors. Using this extended set of sustainability publications Kajikawa et al. focus on 15 clusters that have a distribution of publications over clusters that is similar to ours. However, our clusters contain combinations of topics, whereas the research focus in their clusters is more specific, resulting from the use of HCKB publications instead of the sustainability publications themselves. For example, biodiversity and fishery are in our result combined in one cluster (*Ecosystems and diversity*), whereas Kajikawa et al. have two (*Fisheries* and *Forestry (biodiversity)*. In addition, Kajikawa et al. do not feature an *Econometrics* cluster, whereas we are for example missing a cluster with a focus on wildlife. In a further study by Kajikawa (2008), the clusters from Kajikawa et al (2007) are combined into ten general topics with an integrated research focus, and these are in fact more similar to our clusters. However, where Kajikawa (2008) arrives at the result by examining the literature, we arrive at this result using a quantitative approach based on the co-citation network of the HCKB.

Future research may expand the results by using a broader seed set, for example by including publications that are bibliographically coupled with our current seed set, and then compare results with our current ones to see what the effect is on the research focus. Next, because we found that the context of the use of (cited) publications and the original research context of the citing publications differs, we are also interested in a more direct comparison of clusters of co-cited publications from the HCKB and clusters of bibliographically coupled seed publications. In addition, it would be interesting to focus even more on the time dimension, for example by analysing changes in research topics and research networks within sets of papers, which would allow us to identify trends and hypes. Finally, we have left actors (authors, research groups, institutes) out of the scope of this paper, yet they are of course the true 'enabling' elements of sustainability science and therefore would merit attention in a future study as well.

Acknowledgements We kindly acknowledge the help of Bert de Wit, MSc. and Prof. Herman Eijsackers for their help on earlier versions of this paper. In addition, we are grateful for the valuable comments by the anonymous referees of this paper, which helped to improve the focus and quality of our manuscript..

References

Baron R, Kenny D (1986) The moderatormediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology 51(6):1173

Bastian M, Heymann S, Jacom M (2009) Gephi: An open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Weblogs and Social Media

Blondel VD, Guillaume J, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10):P10,008

Braam RR, Moed HF, Van Raan AFJ (1991) Mapping of science by combined co-citation and word analysis. I. Structural aspects. Journal of the American Society for Information Science 42(4):233–251

Buter R (2012) Scientific structures in context - identification and use of structures, context, and new developments in science. PhD thesis, Leiden University, Leiden

Buter RK, Noyons ECM (2005) Using reference structures to evaluate co-word structures: First explorations. In: Proceedings of the 10th international conference of the international society for scientometrics and informetrics, Stockholm, pp 260–264

Buter RK, Van Raan AFJ (2011) Non-alphanumeric characters in titles of scientific publications: an analysis of their occurrence and correlation with citation impact. Journal of Informetrics 5(4):608–617

Buter RK, Noyons ECM, Van Raan AFJ (2010a) Identification of converging research areas using publication and citation data. Research Evaluation 19(1):19–27

Buter RK, Noyons ECM, Van Raan AFJ (2010b) Searching for converging research using field to field citations. Scientometrics 86(2):325–338

Clark WC (2007) Sustainability science: A room of its own. Proceedings of the National Academy of Sciences 104(6):1737–1738

Clark WC, Dickson NM (2003) Sustainability science: The emerging research program. Proceedings of the National Academy of Sciences of the United States of America 100(14):8059–8061

De Solla Price D (1981) The analysis of scientometric matrices for policy implications. Scientometrics 3(1):47–53

Dodds S (1997) Towards a 'science of sustainability': Improving the way ecological economics understands human well-being. Ecological Economics 23(2):95–111

Elliot JA (2006) An introduction to Sustainable Development. Routledge perspectives on development, Routledge, Abingdon

Gibbons M, Limogenes C, Nowotny H, Schwartzman S, Scott P, Trow M (1994) The new production of knowledge. Sage Publications

Hu Y (2005) Efficient, high-quality force-directed graph drawing. Mathematica Journal 10(1):3771

Kajikawa Y (2008) Research core and framework of sustainability science. Sustainability Science 3(2):215–239

Kajikawa Y, Ohno J, Takeda Y, Matsushima K, Komiyama H (2007) Creating an academic landscape of sustainability science: an analysis of the citation network. Sustainability Science 2(2):221–231

Kates RW, Clark WC, Corell R, Hall JM, Jaeger CC, Lowe I, McCarthy JJ, Schellnhuber HJ, Bolin B, Dickson NM, Faucheux S, Gallopin GC, Grübler A, Huntley B, Jäger J, Jodha NS, Kasperson RE, Mabogunje A, Matson P, Mooney H, Moore B, O'Riordan T, Svedin U (2001) Sustainability science. Science 292(5517):641–642

Keitsch MM (2010) Sustainability and science – challenges for theory and practice. Sustainable Development 18(5):241–244

King C, Gunton J, Freebairn D, Coutts J, Webb I (2000) The sustainability indicator industry: where to from here? a focus group study to explore the potential of farmer participation in the development of indicators. Australian Journal of Experimental Agriculture 40(4):631642

Komiyama H, Takeuchi K (2006) Sustainability science: building a new discipline. Sustainability Science 1(1):1–6

Kuhn TS (1970) The structure of scientific revolutions. University of Chicago Press, Chicago

Larivière V, Gingras Y (2010) On the relationship between interdisciplinarity and scientific impact. Journal of the American Society for Information Science and Technology 61(1):126–131

Lee KN (1993) Compass and Gyroscope: Integrating Science And Politics For The Environment. Island Press, Washington

Michalk D, Dowling P, Kemp D, King W, Packer I, Holst P, Jones R, Priest S, Millar G, Brisbane S, Stanley D (2003) Sustainable grazing systems for the central tablelands, new south wales. Australian journal of experimental agriculture 43(7-8):861–874

Moed H, Bruin RD, Van Leeuwen T (1995) New bibliometric tools for the assessment of national research performance: Database de-

scription, overview of indicators and first applications. Scientometrics 33(3):381–422
Moed HF (2005) Citation Analysis in Research Evaluation, Information Science and Knowledge Management, vol 9. Springer, Dordrecht
Noyons ECM (1999) Bibliometric mapping as a science policy and research management tool. PhD thesis, Leiden University, Leiden
Noyons ECM, van Raan AFJ (1998) Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. Journal of the American Society for Information Science 49(1):68–81
Noyons ECM, Moed HF, Luwel M (1999) Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. Journal of the American Society for Information Science 50(2):115–131
Pearce D, Barbier E, Markandya A (1990) Sustainable Development - Economics and Environment in the Third World. Edward Elgar, Aldershot
Peters HPF, Braam RR, Van Raan AFJ (1995) Cognitive resemblance and citations relations in chemical engineering publications. Journal of the American Society for Information Science 46(1):9–21
Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann
Rafols I, Meyer M (2009) Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. Scientometrics 82(2):263–287
Small HG (1978) Cited documents as concept symbols. Social Studies of Science 8(3):327–340
Soler V (2007) Writing titles in science: An exploratory study. English for Specific Purposes 26(1):90–102
Stalnaker R (2002) Common ground. Linguistics and Philosophy 25(5):701–721
Van den Besselaar P, Heimeriks G (2001) Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In: Paper for the 8th ISSI Conference for Scientometrics and Informetrics, Sydney
Van Raan AFJ (1996) Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. Scientometrics 36(3):397–420