

Creating an academic landscape of sustainability science: an analysis of the citation network

Yuya Kajikawa · Junko Ohno · Yoshiyuki Takeda ·
Katsumori Matsushima · Hiroshi Komiyama

Received: 28 February 2007 / Accepted: 7 May 2007 / Published online: 7 July 2007
© Integrated Research System for Sustainability Science and Springer 2007

Abstract Sustainability is an important concept for society, economics, and the environment, with thousands of research papers published on the subject annually. As sustainability science becomes a distinctive research field, it is important to **define sustainability clearly and grasp the entire structure, current status, and future directions** of sustainability science. This paper provides an academic landscape of sustainability science by analyzing the citation network of papers published in academic journals. A topological clustering method is used to detect the sub-domains of sustainability science. Results show the existence of 15 main research clusters: Agriculture, Fisheries, Ecological Economics, Forestry (agroforestry), Forestry (tropical rain forest), Business, Tourism, Water, Forestry (biodiversity), Urban Planning, Rural Sociology, Energy, Health, Soil, and Wildlife. Agriculture, Fisheries, Ecological Economics, and Forestry (agroforestry) clusters are predominant among these. The Energy cluster is currently developing, as indicated by the age of papers in the cluster, although it has a relatively small number of papers. These results are compared with those obtained by natural language processing. Education, Biotechnology, Medical, Livestock, Climate Change, Welfare, and Livelihood clusters are uniquely extracted by natural language pro-

cessing, because they are common topics across clusters in the citation network.

Keywords Sustainability science · Research on research · Citation network · Overview map · Network analysis

Introduction

Sustainability is an important concept for society, economics, and the environment (Lélé 1991; Goodland 1995; Christensen et al. 1996). Although the essence of the concept of sustainability has a long history dating back to JS Mill and TR Malthus (Goodland 1995), it has not been a significant issue in its present context until recently. In their book *The Limits to Growth*, Meadows et al. (1972) warned that our future development is limited and constrained by the growing world population and the depletion of natural resources. For the further development of society we must seek growth in a sustainable manner, as envisioned by the World Commission on Environment and Development (WCED) (1987), which proposed the concept of sustainable development in *Our Common Future* (also known as the Brundtland Report). These two publications invoked public interest in sustainability and sustainable development, posing challenges such as the management of contractive problems, for example growth versus limits, intergenerational versus intragenerational equity, and individual versus collective interests (Dovers 1993).

Sustainability science is becoming a distinct scientific field (Kates et al. 2001; Mihelcic et al. 2003; Clark and Dickson 2003; Reitan 2005; Komiyama and Takeuchi 2006). Currently, more than 3,000 papers are published in the field annually (Fig. 1). The number of annual publications is increasing linearly, and, therefore, the accumu-

Y. Kajikawa (✉) · J. Ohno · Y. Takeda · K. Matsushima
Institute of Engineering Innovation, School of Engineering,
The University of Tokyo, 2-11-16 Yayoi,
Bunkyo-ku, Tokyo 113-8656, Japan
e-mail: kaji@biz-model.t.u-tokyo.ac.jp

H. Komiyama
Integrated Research System for Sustainability Science (IR3S),
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-8654, Japan

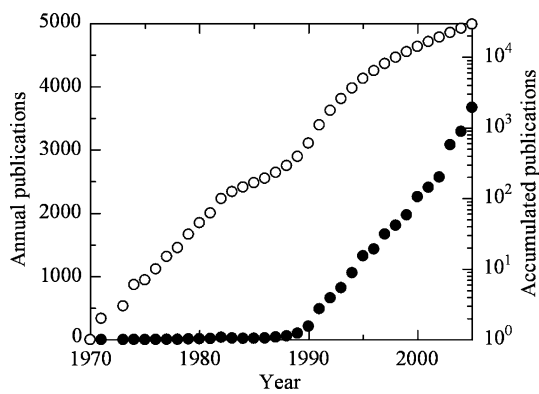


Fig. 1 Number of papers including “sustainable” or “sustainability” in the title or abstract. *Black circles* and *white circles* are the number of annual publications and the accumulated number of publications, respectively

lated number of publications is increasing exponentially. Since the late 1990s a variety of academic journals have been launched to meet both academic and social demand (Table 1). The multidisciplinary nature of sustainability

Table 1 A list of academic journals including “sustainable” or “sustainability” in their titles. The retrieval was performed using the Online Public Access Catalog (OPAC)

| Journal title | Year |
|--|------|
| Journal of Sustainable Agriculture | 1990 |
| Journal of Sustainable Forestry | 1993 |
| Journal of Sustainable Tourism | 1993 |
| Sustainable Development | 1993 |
| The International Journal of Sustainable Development and World Ecology | 1994 |
| Renewable & Sustainable Energy Reviews | 1997 |
| The Journal of Sustainable Product Design | 1997 |
| International Journal of Sustainable Development | 1998 |
| Journal of Sustainable Development in Africa | 1999 |
| Environment, Development and Sustainability | 1999 |
| International Journal of Agricultural Resources, Governance and Ecology | 2000 |
| International Journal of Sustainability in Higher Education | 2000 |
| International Journal of Environment and Sustainable Development | 2002 |
| International Journal of Technology Management & Sustainable Development | 2002 |
| International Journal of Agricultural Sustainability | 2003 |
| International Journal of Sustainable Energy | 2003 |
| World Review of Science, Technology and Sustainable Development | 2004 |
| Agronomy for Sustainable Development | 2005 |
| Sustainability: Science, Practice, & Policy | 2005 |
| Sustainable Humanosphere | 2005 |
| Sustainability Science | 2006 |

science is often emphasized (Komiya and Takeuchi 2006), and it is sometimes claimed that research involving novel schemes and techniques must be employed, extended, or invented (Kates et al. 2001). The scientific and technological basis of the concept remains unclear, however (Komiya and Takeuchi 2006).

There has been a long debate on the definition of sustainability (Brown et al. 1987; Barbier 1987; Simon 1989; Shearman 1990; Lélé 1991; Redclift 1992; Goodland 1995; Callicott and Mumford 1997). The Brundtland Report defined sustainable development as development that “meets the needs of the present generation without compromising the ability of future generations to meet their own needs.” Sustainability is lexically defined as “the ability to maintain something undiminished over some time period” (Lélé and Norgaard 1996). While sustainable development is associated with the human exploitation of nature, “sustainability” does not include such a connotation. In fact, the meaning of sustainability depends on the context, in which it is applied (Brown et al. 1987; Shearman 1990). We must keep in mind that sustainability is not a goal; it is a constraint on the achievement of other goals (Marcuse 1998). Marcuse (1998) gives the following example: a problem such as the world’s poor is not that their conditions cannot be sustained but that they should not be sustained. In short, sustainability is a prerequisite to attain a goal, which means different things to different people. Therefore, “sustainability” is polyphonic and polysemic, and the content may differ from context to context.

The vague definition of sustainability is not necessarily an obstacle at the nascent stage of research and development. To some extent, the value of the phrase lies in its broadness and its ability to stimulate vigorous and open discussion. It also allows people with conflicting positions in the environment-development debate to search for common ground, on which to compromise (Lélé 1991). In some situations, avoiding rigorous definition may have a fruitful outcome. WCED (1987) defined sustainable development in a manner that, although somewhat vague and inoperative, attracted wide attention and endorsement (Dovers 1993).

The vagueness in definition also conveys shortcomings in grasping the overall structure of sustainability science, however. It is, for example, difficult to answer the question “What is sustainability science, and what disciplines does it include?” Such a discourse is common for other young academic domains as seen in environmental studies (Soulé and Press 1998). **Efforts to offer a comprehensive understanding and definition of a research domain have conventionally been made by domain experts. But grasping the current status of sustainability science has become an urgent task because of the growing body of publications as shown in Fig. 1.**

To meet this challenge a computer-based approach can be used to complement the expert-based approach because it is compatible with the scale of information (Börner et al. 2003; Boyack et al. 2005). A citation-based approach, which is computer-based, operates on the assumption that citing and cited papers have similar research topics. By analyzing this citation network, we can comprehend the structure of a research domain constituting a larger volume of papers than we can read. In previous works, a citation-based approach has been applied to water resource management (Thelwall et al. 2006) and ecological economics (Costanza et al. 2004; Ma and Stern 2006). The objective of this paper is to provide an academic landscape of sustainability science by using citation network analysis as a computational support tool.

Data and method

Data

Assuming sustainability science in its historical context and current state to be reflected in academic publications, we collected a set of academic publications including “sustainability” or “sustainable” in their titles, abstracts, and keywords. We collected citation data for those publications from the Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI) compiled by the Institute for Scientific Information (ISI), because SCI and SSCI are two of the best sources of citation data. We used Web of Science, which is a Web-based user interface for ISI’s citation databases, and searched the papers using `sustainab*` as a query, where `*` represents a wildcard. The corpus thus acquired therefore contains papers that include both “sustainability” and “sustainable.” **A total of 29,391 such papers were retrieved.** We realized, however, that some of these papers might not be relevant to sustainability science because they were retrieved via the simple query described above. Therefore we focused on the maximum connected component, **which currently consists of 9,973 papers.** In other words, we regarded papers not citing other papers in the component as digressional from the mainstream of sustainability science and eliminated them. We checked whether those eliminated papers also formed a large network, but found that the second-largest connected component has only 35 nodes. We therefore considered it reasonable to focus on the maximum connected component to reveal the structure of sustainability science.

Method

Our analyzing procedure is illustrated schematically in Fig. 2. The retrieved data includes both connected com-

ponents and isolated nodes, as shown in Fig. 2a. The links in Fig. 2a are directional, i.e. citing and cited papers are distinguished. The data are then converted into a non-weighted, non-directed network, and the maximum connected component of the network is extracted as in Fig. 2b. The resulting maximum connected component has 9,973 nodes as described above. Finally, the network is divided into clusters using the topological clustering method (Newman 2004; Newman and Girvan 2004), as seen in Fig. 2c. The clustering algorithm is based on modularity Q , which is defined as follows (Newman 2004; Newman and Girvan 2004):

$$Q = \sum_{s=1}^{N_m} \left[\frac{l_s}{l} - \left(\frac{d_s}{2l} \right)^2 \right] \quad (1)$$

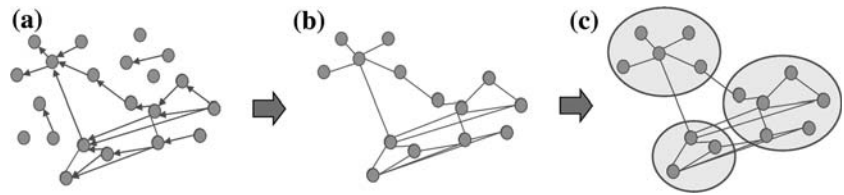
where N_m is the number of clusters, l_s is the number of links between nodes in cluster s , and d_s is the sum of the degrees of the nodes in cluster s . In other words, Q is the fraction of links that fall within clusters, minus the expected value of the same quantity if the links fall at random without regard for the clustered structure. Because a high value of Q represents a good division, we stopped clustering when ΔQ became minus. A good partition of a network into clusters means there are many intra-cluster links and as few as possible inter-cluster links. The clustered network is visualized by using a large graph layout (LGL) (Adai et al. 2004). LGL is based on a spring layout algorithm where links play the role of spring connecting nodes. As a result of this layout a group of papers citing each other is located in closer positions. In our visualization we hide inter-cluster links and only show the intra-cluster links for each cluster with the same color to clarify the position of each cluster.

After clustering the network, we analyzed the characteristics of each cluster by titles and abstracts of papers that are frequently cited by the other papers in the cluster, and also journals, in which the papers in the cluster were published. Papers in the maximum connected component were published in 1,255 journals, which reflects the diversity of the research domain of sustainability science. The distribution of the journals is not uniform, however; each cluster has a characteristic trend. We define the journal weight factor (JWF) of journal i in cluster s , JWF_{si} , as:

$$JWF_{si} = \frac{n_{si} n_{si}}{n_i n_s}, \quad (2)$$

where n_i , n_s , and n_{si} are the number of papers of journal i in the maximum connected component, the number of papers in cluster s , and the number of papers of journal i in cluster s , respectively. F_{si} becomes higher when we have more

Fig. 2 Schematic diagram of citation network analysis: **a** retrieved data; **b** the maximum connected component; **c** the maximum connected component after clustering



papers of the journal in the entire network and also in the cluster. Similarly, we can define the country weight factor (CWF) of country j in cluster s , CWF_{sj} , as:

$$CWF_{sj} = \frac{n_{sj} n_{sj}}{n_j n_s}, \quad (3)$$

where n_j and n_{sj} are the number of papers of country j and the number in cluster s , respectively. The age of a cluster was determined as 2006 minus the average publication year. Key topics of a cluster were identified from titles and abstracts of the top ten most cited papers in the cluster.

In addition to citation network analysis we used natural language processing (NLP) to analyze the structure of sustainability science. We employed NLP as a supplemental method for citation network analysis. As a citation network might have a citation bias, it was used to illuminate only one facet of sustainability science; NLP was expected to illustrate another facet. In NLP we first identified key terms that often appear in the abstracts of the 29,391 papers. We then measured the similarity between the extracted terms. Using the calculated similarity, the terms were merged into clusters. We analyzed those clusters on the assumption that they would reflect some aspect of the current status of sustainability science.

For term recognition we used the NC-value method to extract the key terms that frequently appear in the abstracts (Mima and Ananiadou 2000). The NC-value is a score for measuring the relevance of terms; it measures the relative importance of sequential words in the corpus by assuming that terms which include many words and frequently occur with other key terms have high plausibility as key terms in the corpus. The NC-value for the candidate string a , $NC\text{-value}(a)$, is given by:

$$NC\text{-value}(a) = 0.8 \times C\text{-value}(a) + 0.2 \times \text{Context}\text{-value}(a). \quad (4)$$

In Eq. (4), $C\text{-value}(a)$ is given when a is nested as

$$C\text{-value}(a) = \max\{1, \log_2 |a| f(a)\}; \quad (5)$$

otherwise,

$$C\text{-value}(a) = \max(1, \log_2 |a|) \left\{ f(a) - \sum_{b \in T_a} f(b) / P(T_a) \right\}, \quad (6)$$

where $|a|$, $f(a)$, T_a , and $P(T_a)$ are the length of a , its frequency of occurrence in the corpus, the set of extracted candidate terms that contain a , and the number of those candidate terms, respectively. In short, the $C\text{-value}$ has a high value when a term with long strings frequently appears in the corpus. Here, we assume that key terms have such characteristics. $\text{Context-value}(a)$ measures the frequency of the co-occurrence of a with another context word, b . Context words are nouns, adjectives, and verbs which frequently appear with key terms. We assume that the co-occurrence of a term with a context word increases the plausibility of the term as a key term in the domain. $\text{Context-value}(a)$ is given by:

$$\text{Context-value}(a) = \sum_{b \in Ca} f_a(b) \text{weight}(b), \quad (7)$$

where Ca is the set of distinct context words, $f_a(b)$ is the frequency of b as a context word of a , and $\text{weight}(b)$ is defined as $t(b)/n$; $t(b)$ is the number of terms the word b appears with and n is the total number of terms considered. We linguistically filtered sequential words constituted by nouns and combinations of noun and adjective, and extracted them. We then calculated the NC-value of those terms. We extracted key terms with a high NC-value in decreasing order.

After term recognition we counted the occurrence of those terms in each abstract. We then expressed the result by using a vector space model (VSM) (Salton et al. 1975). VSM encodes a collection of documents by a term-document matrix whose $[i, j]$ th element indicates the association between the i th term and the j th document. In our case, a term is a sequential word extracted by the NC-value method and a document is an abstract. **We calculated the similarity between two terms by the cosine of the angle between their vectors.** Briefly, we regarded the similarity of the terms to be high when they appeared in the same abstracts. Finally, those terms were clustered by the group average method using these cosine measures. After

obtaining the clusters, we manually annotated the names of the clusters.

Results and discussion

The citation network of sustainability science can be divided into 93 clusters, where the number of nodes in each cluster varies from three (the smallest clusters) to 1,584 (the biggest cluster, #1). Papers in each cluster are strongly coupled by intra-cluster citations. Cluster size, i.e. the number of nodes in each cluster, gradually decreases until the 15th cluster, and after the 30th cluster the number becomes negligible. In the following discussion, therefore, we focus on the top 15 clusters, which cover more than 80% of the papers in the network. Figure 3 visualizes the structures of the citation networks of the top 15 clusters. In this figure we assign the same color to intra-cluster links for each cluster. When the structure of a cluster in Fig. 3 is compact and round, it means that papers in the cluster have a strong tendency to cite other papers in the same cluster. Conversely, when a cluster is stretched and spiky, the cluster is closely related to other clusters located in that direction. When two clusters are near to each other, it means the papers in these two clusters cite each other. Table 2 summarizes the contents of each cluster.

Cluster #1 is the Agriculture cluster, in which sustainable agriculture is discussed. The Agriculture cluster has 1,584 papers in it and is the biggest among the 93 clusters. It is also the oldest among the top 15 clusters. Research topics include soil erosion, soil fertility, soil resilience, nutrients, food productivity, plant biodiversity, and so forth. Cluster #2 is the Fisheries cluster, in which the sustainability of world fisheries is discussed. The United States dominates this cluster, with a large CWF. Cluster #3 is Ecological Economics, in which economic indicators of sustainability are proposed and measured. The above three

clusters occupy central positions in the network because of their large volume (Fig. 3). The stretched and spiky shape of cluster #3 in Fig. 3 means that this cluster is closely connected to other clusters in the network. Cluster #4 is Forestry (agroforestry). Fertility, such as nitrogen and phosphorus content, is the main concern. Managing the competition between trees and crops for light, water, and nutrients is the key success factor for agroforestry systems. India and Brazil have high CWFs in this cluster, which reflects the importance of this research in those countries. As seen in Fig. 4, approximately half of the papers belong to the top four clusters. It is worth noting that the concept of sustainability originated in the context of sustainable yields for agriculture and renewable resources such as forests or fisheries and has subsequently been adopted as a broad slogan by the environmental movement (Lélé 1991). This historical background is a factor in the current central position of those clusters.

Cluster #5 is Forestry (tropical rain forest). Most papers in this cluster are written by authors in the US and discuss management and economic aspects of timber and non-timber forest products from tropical forests. Cluster #6 is the Business cluster, which is somewhat noisy because most papers discuss the sustainable competitive advantages of a firm. The topological position of the cluster in the citation network reflects this. Some papers definitely share the same context as the other categories, however, e.g. by linking environmental performance and economic performance. Cluster #7 is the Tourism cluster; the subject of sustainable tourism is controversial and the management of oceans and coasts in particular is deliberated. Cluster #8 is the Water cluster, in which wastewater treatment, water resource management, and the water cycle are key topics. It is noteworthy that China focuses on water research. Cluster #9 is the Forestry (biodiversity) cluster; Canada has the highest CWF and is predominant in this cluster. An important goal for research in the cluster is the conservation of biological diversity in forests.

Cluster #10 is the Urban Planning cluster, in which sustainable city and landscape planning are key topics. Social and political aspects of sustainability, for example planning and regulation, are also discussed. Cluster #11 is the Rural Sociology cluster, in which sustainability is closely associated with social issues. Key topics are agreement between the countries of the North and those of the South, rural development, local knowledge, and local food systems. Cluster #12 is the Energy cluster, which is the youngest among the top 15 clusters. In the Energy cluster no country has a value of CWF markedly higher than for other countries, which means that the sustainability of energy is a common and global problem, at least for the developed countries where scientific research is active. Cluster #13 is the Health cluster, in which the sustainability

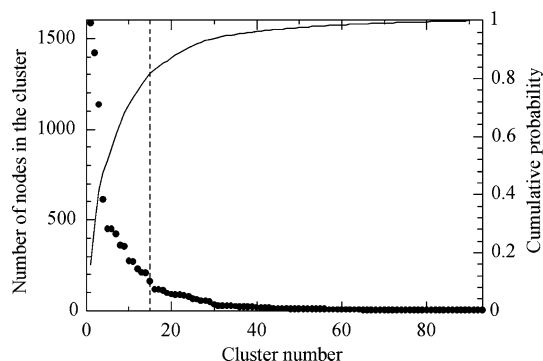


Fig. 3 Cluster size. Black dots are the number of nodes in each cluster. The line is the cumulative probability of the number of nodes. The dashed line is at a cluster number equal to 15

Table 2 Characteristics of the top 15 clusters in the citation network

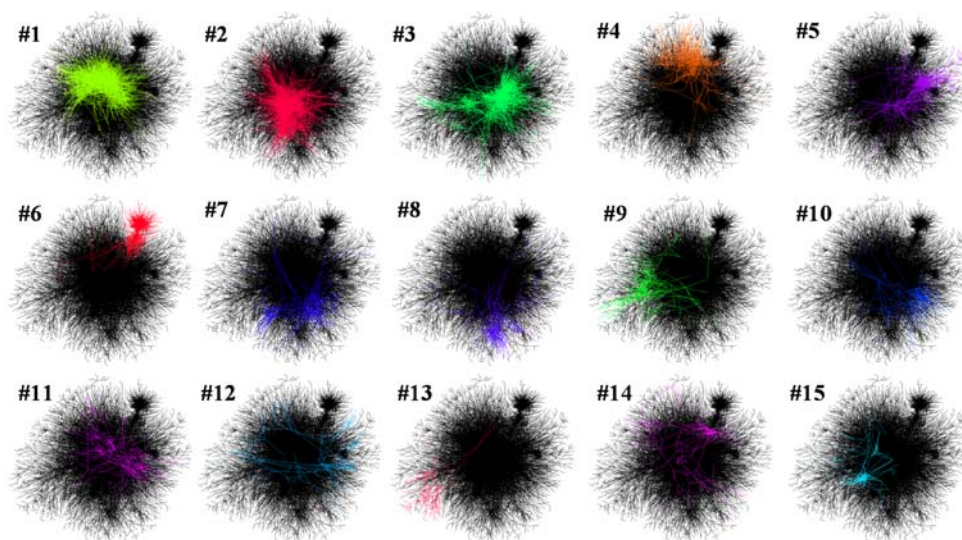
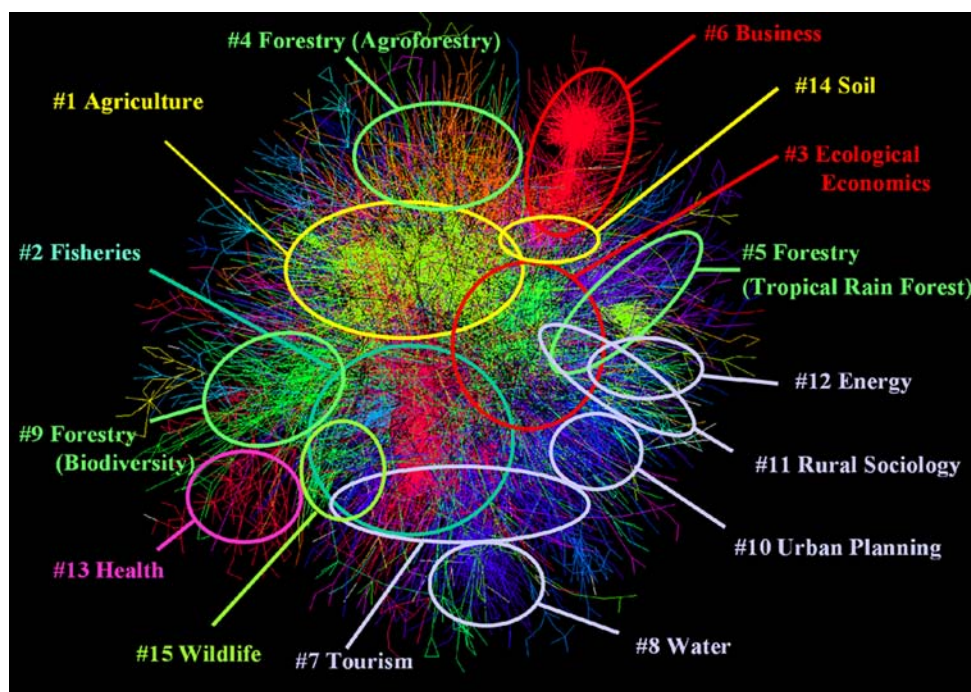
| No. | Cluster name | #Node | Age | Main journal | JWF | Main country | CWF | Key topic |
|-----|---------------------------------|-------|-----|---|------|--------------|------|-----------------------------------|
| #1 | Agriculture | 1584 | 7.1 | Agriculture, Ecosystems & Environment | 1.17 | USA | 7.50 | Soil |
| | | | | Journal of Sustainable Agriculture | 0.29 | Netherlands | 4.17 | Crop |
| | | | | Agricultural Systems | 0.28 | Australia | 1.64 | Biodiversity |
| #2 | Fisheries | 1419 | 5.5 | Ecological Applications | 4.34 | USA | 16.6 | Fish catch |
| | | | | Conservation Biology | 1.38 | Sweden | 2.81 | Marine |
| | | | | Marine Policy | 1.37 | Canada | 2.61 | Ecosystem |
| #3 | Ecological Economics | 1135 | 5.5 | Ecological Economics | 7.97 | USA | 3.90 | Natural capital accounting |
| | | | | Land Economics | 1.29 | England | 2.44 | Sustainability index |
| | | | | Resources Policy | 1.24 | Netherlands | 1.74 | Ecological footprint |
| #4 | Forestry (agroforestry) | 614 | 6.3 | Agroforestry Systems | 2.85 | India | 2.20 | Nutrient |
| | | | | Field Crops Research | 1.39 | Brazil | 1.42 | Soil |
| | | | | Nutrient Cycling in Agroecosystems | 1.08 | Germany | 1.41 | Nitrogen-fixation |
| #5 | Forestry (tropical rain forest) | 450 | 6.5 | Economic Botany | 4.07 | USA | 5.47 | Tropical forest |
| | | | | Forest Ecology and Management | 1.88 | England | 0.81 | Timber and non-timber forest |
| | | | | Conservation Biology | 0.72 | Spain | 0.61 | Harvest |
| #6 | Business | 450 | 5.5 | Strategic Management Journal | 9.56 | South Africa | 3.44 | Sustainable competitive advantage |
| | | | | Journal of Business Ethics | 3.59 | Brazil | 2.12 | Environmental performance |
| | | | | Academy of Management Review | 3.56 | USA | 2.10 | Natural resource |
| #7 | Tourism | 423 | 6.5 | Tourism Management | 9.88 | England | 1.02 | Eco-tourism |
| | | | | Ocean & Coastal Management | 9.69 | USA | 0.98 | Coastal management |
| | | | | Annals of Tourism Research | 6.21 | Scotland | 0.91 | Tropical country |
| #8 | Water | 361 | 5.5 | Water Science and Technology | 11.1 | China | 1.40 | Water resource |
| | | | | Water International | 6.03 | Switzerland | 1.24 | Waste water |
| | | | | Hydrological Sciences Journal | 3.83 | Germany | 1.05 | Water cycle |
| #9 | Forestry (biodiversity) | 353 | 5.4 | Forestry Chronicle | 20.3 | Canada | 13.1 | Forest management |
| | | | | Journal of Forestry | 4.72 | USA | 1.64 | Biodiversity |
| | | | | Canadian Journal of Forest Research | 2.95 | France | 0.51 | Ecosystem management |
| #10 | Urban Planning | 277 | 5.9 | Landscape Urban Planning | 2.63 | England | 3.38 | Sustainable city |
| | | | | Journal of Planning Education and Research | 2.45 | USA | 0.55 | Landscape planning |
| | | | | Regional Studies | 2.4 | Scotland | 0.39 | Regulation |
| #11 | Rural Sociology | 271 | 6.6 | Sociologia Ruralis | 6.50 | USA | 1.01 | Developing country |
| | | | | Rural Sociology | 5.31 | New Zealand | 0.99 | Rural development |
| | | | | American Journal of Alternative Agriculture | 1.3 | England | 0.89 | Local knowledge |
| #12 | Energy | 229 | 4.9 | Energy Policy | 9.17 | England | 0.32 | Hydrogen |
| | | | | Energy Sources | 6.42 | Netherlands | 0.31 | Biomass |
| | | | | International Journal of Hydrogen Energy | 4.45 | USA | 0.28 | Photovoltaic |
| #13 | Health | 211 | 5.8 | Health Policy and Planning | 13.1 | USA | 2.05 | Health program |
| | | | | Social Science & Medicine | 6.01 | Canada | 0.99 | Intervention |
| | | | | Tropical Medicine & International Health | 5.68 | Australia | 0.92 | Community |
| #14 | Soil | 208 | 5.5 | Australian Journal of Soil Research | 4.62 | Australia | 3.83 | Fertile soil |
| | | | | Indian Journal of Agronomy | 2.84 | USA | 0.87 | Organic matter management |
| | | | | Grass and Forage Science | 2.37 | Brazil | 0.73 | Cultivation |

Table 2 continued

| No. | Cluster name | #Node | Age | Main journal | JWF | Main country | CWF | Key topic |
|-----|--------------|-------|-----|-------------------------------|------|--------------|------|----------------|
| #15 | Wildlife | 161 | 5.9 | Geography in Higher Education | 2.88 | England | 1.10 | Wildlife |
| | | | | Oryx | 2.53 | USA | 0.66 | Hunting |
| | | | | Biodiversity and Conservation | 2.48 | Sweden | 0.07 | Forest mammals |

of health projects is discussed. The penetration of intervention into a population and community participation in health-care programs is essential for sustaining health.

Cluster #14 is the Soil cluster. Compared with the Agriculture cluster the Soil cluster is more technology-focused. In journals with a high JWF, however, detection of this

Fig. 4 Visualization of the citation networks of the top 15 clusters**Fig. 5** Visualization of the structure of sustainability science

cluster may be because of an emphasis on regional agricultural systems or a citation bias by which researchers in each country cite journals of their own countries. Cluster #15 is the Wildlife cluster, in which the impact of commercial hunting on forest mammals is investigated. Subsistence hunting by inhabitants and the sustainability of wildlife, especially mammals threatened by game hunting, are investigated.

In Fig. 5, we show the relative positions of these clusters to summarize the above results. We can use this image as an academic overview map of sustainability science. It is worth pointing out some implications of the map. As shown in Fig. 5, some clusters discussing related topics are located in relatively close positions. For example, the Business cluster (#6) is just above the Ecological Economics cluster (#3). The Soil cluster (#14) is in the proximity of the Agriculture cluster (#1). These proximities accord with the relatedness of topics in these clusters. The Forestry clusters (#4, #5, #9) are far from each other, however. This may reflect the diversity of topics in forest research. Agroforestry (#4) is close to Agriculture (#1), Tropical Rain Forest (#5) is near Rural Sociology (#11), and Biodiversity (#9) is near Wildlife (#15). Another view is also possible, however. These clusters (#4, #5, #9) treat similar topics, i.e. forestry and forest management. The citation gap among forestry clusters suggests the existence of a research gap, and the possibility of future collaboration among these clusters. This view might also be valid for Agriculture (#1) and Soil (#14). Papers in the Soil cluster are region-specific as shown in the main journals of Table 2; this may be because of different fields of specialization and research communities from those in the Agriculture cluster.

In the citation-based approach it is assumed that citing and cited papers have similar research topics. Citation behavior is motivated in different ways, however (MacRoberts and MacRoberts 1989), and the result therefore

reflect the cognitive structure of scholars in each research domain (Kajikawa et al. 2006). In other words, the citation map can be depicted as a result that must take these different motivations—for example citing papers having similar research topics, unrelated but prominent papers, and self-citations—into consideration. We therefore used NLP as a supplemental method for citation network analysis. We shall now look at the results obtained by NLP.

First, we checked the relevance of the NLP results. Generally, a large fraction of noisy terms is recognized by NLP, and this fraction increases as the number of extracted terms increases. We therefore checked the relevance of extracted terms by comparing them with the keywords designated by the authors. We defined the precision of the result as the fraction of terms according with keywords among all terms extracted by the NC-value method. The results are shown in Fig. 6. The precision was highest at approximately 1,000 terms and decreased as the number of extracted terms increased. Therefore in the following analysis we focused on 1,000 terms extracted by the NC-value method and analyzed the similarity among them. We used the group average method as a clustering method after pruning terms with a similarity threshold of 0.09 to reduce noise. As a result we obtained a dendrogram of 679 terms, as shown in Fig. 7. Some parts of the dendrogram are clearly divided into clusters. Because there is no common criterion for setting the threshold for statistical clustering, we manually set ad hoc criteria to recognize clusters and obtained 19 clusters, as shown in Fig. 7. Two clusters (cluster N_1 and N_2) consist of noisy terms and one cluster consists of generic terms (cluster M). There are clusters that can be divided at high similarity but cannot be divided at low similarity (A_1 – A_3 , E_1 – E_2 , and I_1 – I_3). Examples of the terms included in each cluster are shown in Table 3.

Comparing the results obtained by NLP (Table 3) with those by citation network analysis (Table 2), we can see similar clusters. Natural resource-related clusters such as Agriculture, Fisheries, Forestry, Water, and Biodiversity are extracted by both citation network analysis and NLP. These clusters are the central research domains of sustainability science. Clusters relating to Economics (Ecological Economics and Business) are also seen in both results. But some discrepancies exist. For example, the Tourism cluster in the citation network seems to be merged into the Ecological Economics cluster (cluster D) in NLP. This is because in the Tourism cluster the focus of discussion is often on its economic aspects. In NLP we have only one Forestry cluster (cluster F) whereas in the citation network there are three Forestry-related clusters (#4, #5, #9). This suggests the existence of a common terminology for these forestry research domains.

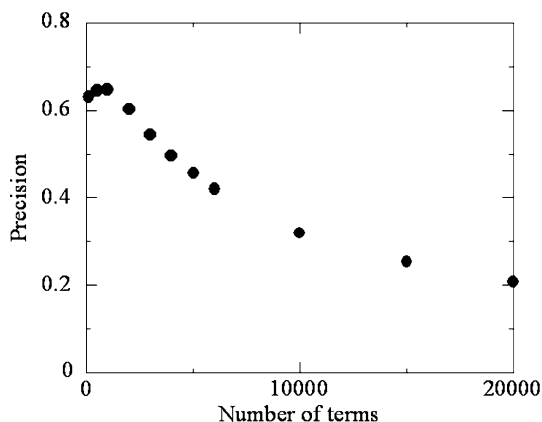
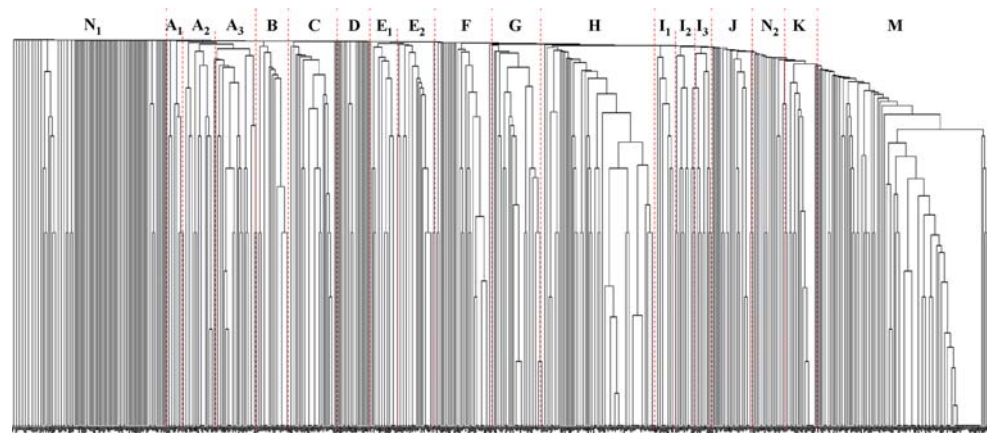


Fig. 6 Number of extracted terms and their relevance

Fig. 7 Dendrogram of key terms**Table 3** Clusters extracted by natural language processing

| # | Cluster name | Example of extracted terms |
|----|-------------------------|---|
| A1 | Education | Education, training, learning, skill, school, university, innovation |
| A2 | Biotechnology | Biotechnology, cell, protein, gene, cultivar, breeding, pesticide |
| A3 | Medical | Hospital, patient, care, disease, vaccine, infection, pathogen, insect, insecticide |
| B | Livestock | Livestock, rangeland, grassland, pasture, forage, cattle, sheep |
| C | Water | Water, river, groundwater, aquifer, wastewater, effluent, drainage |
| D | Ecological Economics | Ecology, economics, regulation, legislation, profitability, tourism |
| E1 | Climate change | Climate change, biosphere, planet, pollution, CO ₂ , temperature, emission |
| E2 | Energy | Energy, fuel, electricity, oil, hydrogen, biomass, vehicle, recycling |
| F | Forestry | Forestry, tree, timber, planting, fire, vegetation, logging, plantation, fire |
| G | Fisheries | Fishery, fish, fishing, ocean, sea, aquaculture, catch, harvest |
| H | Agriculture | Crop, rice, corn, plant, soil, fertility, nutrient, cultivation, erosion, topsoil |
| I1 | Business | Business, company, firm, customer, competitiveness, capability |
| I2 | Welfare | Welfare, well-being, safety, health, food, nutrition, diet, consumer |
| I3 | Livelihood | Livelihood, income, household, poverty, family, employment, consumption |
| J | Economics | Capital, market, investment, price, benefit, cost, labor, incentive, value |
| K | Biodiversity | Biodiversity, wildlife, hunting, preservation, bird, landscape, ecosystem |
| M | General | Sustainability, society, nature, future, goal, assessment, solution, system |
| N1 | Noise 1 | Interview, questionnaire, simulation, scenario, survey, database, history |
| N2 | Noise 2 | Access, dynamics, program, question, debate, trend, structure, contribution |

In addition to common clusters, some new clusters can be detected by NLP. These are clusters A₁–A₃ (Education, Biotechnology, Medical), B (Livestock), E₁ (Climate Change), I₂ (Welfare), and I₃ (Livelihood). Clusters A₁–A₃ are closely related to each other as shown in the dendrogram (Fig. 7). These clusters are associated with the Health cluster (#13) in the citation network, which implies that education and biotechnology are mainly discussed in the context of the sustainability of health programs. Cluster E₂ (Climate Change) is close to cluster E₁ (Energy) in the dendrogram. Climate change cannot be detected as a distinct cluster in the citation network but appears in the dendrogram by NLP at this position.

Clusters I₂ (Welfare) and I₃ (Livelihood) are also detected as distinct clusters.

Why do these clusters emerge? One explanation is that these clusters have terms that appear in most of the clusters in the citation network but with few appearances in each cluster. As terms such as education and welfare appear in each citation cluster in small quantities, we cannot detect them as independent clusters by citation network analysis. Nevertheless, distinct clusters are shown by NLP because these terms appear in large quantities across the entire corpus. These clusters that were originally extracted by NLP are therefore considered to be common terms for clusters in the citation network. Common clusters seem to

be formed around topics representing what we should sustain: agriculture, fish, water, forests, energy, biodiversity. Some of the clusters that originally appear in the citation network are sub-categories such as soil and wildlife. Clusters originally detected by NLP are more common and more human-rooted, for example welfare, livelihood, and education.

Finally, let us address the limitations of our research. In our approach, we collected the corpus by making a query. The results obtained by citation network analysis indicated that agriculture and fisheries occupy the largest fractions of sustainability science. On the other hand, energy, which is an unquestionably important area of research in sustainability, represents a relatively small fraction of research and is the youngest among the top 15 clusters. But we must note that usage among researchers of the term “sustainability” has been changing. Sustainability was used as a technical term in the early days but nowadays seems to be used to express the importance of global sustainability. It is plausible that clusters with a longer history (e.g. agriculture) have used “sustainability” as a technical term while the younger Energy cluster uses “sustainability” with the latter meaning. Therefore, changes in the definition of sustainability (or the usage of this word) may be behind these results. Debate on the definition and targets of sustainability will continue as a part of sustainability science.

Conclusion

Although sustainability is an important concept for society, economics, and the environment, its definition is unclear. The number of journals and papers on sustainability continues, nevertheless, to increase. For example, there are several journals on sustainability specializing in sub-domains of sustainability, for example agriculture, forestry, tourism, energy, and education. Over 3,000 papers on sustainability are currently published annually. Sustainability science is expected to integrate these sub-domains and to offer forums for discussion addressing the polyphonic and polysemic nature of sustainability.

This paper analyzed the current status of sustainability science and used a computer-based approach to provide a fundamental framework for future research. In this paper we visualized the structure of sustainability science by analysis of citations in relevant publications, and used a topological clustering method to detect the sub-domains of sustainability science.

Our citation analysis extracted 15 main research domains: Agriculture, Fisheries, Ecological Economics, Forestry (agroforestry), Forestry (tropical rain forest), Business, Tourism, Water, Forestry (biodiversity), Urban Planning, Rural Sociology, Energy, Health, Soil, and

Wildlife. Agriculture, Fisheries, Ecological Economics, and Forestry (agroforestry) clusters are predominant among these. The Energy cluster is currently developing. These results were compared with those obtained by natural language processing. Education, Biotechnology, Medical, Livestock, Climate Change, Welfare, and Livelihood clusters were uniquely extracted by natural language processing, because they are common topics across other sub-domains of sustainability science.

We hope that the journal *Sustainability Science* publishes updates on achievements in each domain and facilitates interdisciplinary quests, multidisciplinary efforts to integrate these, and transdisciplinary actions to change the real world. We also hope that our landscape serves to guide those who contribute to sustainability science and helps them move society in sustainable directions, based on a clear grasp of their current position and new directions to explore.

Acknowledgments We are grateful to anonymous referees for their valuable comments, owing to which our manuscript was greatly improved. We thank Associate Professor Hideki Mima for his help in natural language processing. We also thank Dr Ai Hiramatsu, Professor Yoshihisa Murasawa, and Professor Shuichiro Asao at The University of Tokyo for encouraging our research. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 18700240, 2006.

References

- Aday AT, Date SV, Wieland S, Marcotte EM (2004) LGL: creating a map of protein function with an algorithm for visualizing very large biological networks. *J Mol Biol* 340:179–190
- Barbier EB (1987) The concept of sustainable economic development. *Environ Conserv* 14:101–110
- Börner K, Chen C, Boyack KW (2003) Visualizing knowledge domains. *Annu Rev Info Sci Technol* 37:179–255
- Boyack KW, Klavans R, Börner K (2005) Mapping the backbone of science. *Scientometrics* 64:351–374
- Brown BJ, Hanson ME, Liverman DM, Merideth RW (1987) Global sustainability: toward definition. *Environ Manage* 11:713–719
- Callicott JB, Mumford K (1997) Ecological sustainability as a conservation concept. *Conserv Biol* 11:32–40
- Christensen NL, Bartuska AM, Brown JH, Carpenter S, D’Antonio C, Francis R, Franklin JF, MacMahon JA, Noss RF, Parsons DJ, Peterson CH, Turner MG, Woodmansee RG (1996) The report of the Ecological Society of America committee on the scientific basis for ecosystem management. *Ecol Appl* 6:665–691
- Clark WC, Dickson NM (2003) Science and technology for sustainable development special feature: sustainability science: the emerging research program. *Proc Natl Acad Sci* 100(14):8059–8061
- Costanza R, Stern D, Fisher B, He L, Ma C (2004) Influential publications in ecological economics: a citation analysis. *Ecol Econ* 50:261–292
- Dovers SR (1993) Contradictions in sustainability. *Environ Conserv* 20:217–222
- Goodland R (1995) The concept of environmental sustainability. *Annu Rev Ecol Syst* 26:1–24

- Kajikawa Y, Abe K, Noda S (2006) Filling the gap between researchers studying different materials, different methods: a proposal of structured keyword. *J Info Sci* 32:511–524
- Kates RW, Clark WC, Corell R, Hall JM, Jaeger CC, Lowe I, McCarthy JJ, Schellnhuber HJ, Bolin B, Dickson NM, Faucheux S, Gallopin GC, Grubler A, Huntley B, Jäger J, Jodha NS, Kasperson RE, Mabogunje A, Matson P, Mooney H, Moore III B, O'Riordan T, Svedin U (2001) Environment and development: sustainability science. *Science* 292(5517):641–642
- Komiyama H, Takeuchi K (2006) Sustainability science: building a new discipline. *Sustain Sci* 1:1–6
- Lélé S (1991) Sustainable development: a critical review. *World Dev* 19:6–7–621
- Lélé S, Norgaard RB (1996) Sustainability and the scientist's burden. *Conserv Biol* 10:354–365
- Ma C, Stern DI (2006) Environmental and ecological economics: a citation analysis. *Ecol Econ* 58:491–506
- MacRoberts MH, MacRoberts BF (1989) Problems of citation analysis: a critical review. *J Am Soc Info Sci* 40:342–349
- Marcuse P (1998) Sustainability is not enough. *Environ Urban* 10:103–111
- Meadows D, Meadows D, Randers J, Behrens W (1972) *The limits to growth*. Universe Books, New York
- Mihelcic JR, Crittenden JC, Small MJ, Shonnard DR, Hokanson DR, Zhang Q, Chen H, Sorby SA, James VU, Sutherland JW, Schnoor JL (2003) Sustainability science and engineering: the emergence of a new metadiscipline. *Environ Sci Technol* 37:5314–5324
- Mima H, Ananiadou S (2000) An application and evaluation of the C/NC-value approach for the automatic term recognition of multiword units in Japanese. *Terminology* 6:175–194
- Newman MEJ (2004) Fast algorithm for detecting community structure in networks. *Phys Rev E* 69:066133
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113
- Redclift M (1992) The meaning of sustainable development. *Geoforum* 23:395–403
- Reitan PH (2005) Sustainability science—and what's needed beyond science. *Sustain Sci* 1:77–80
- Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. *Commun ACM* 18:613–620
- Shearman R (1990) The meaning and ethics of sustainability. *Environ Manage* 14:1–8
- Simon D (1989) Sustainable development: theoretical construct or attainable goal? *Environ Conserv* 16:41–48
- Soulé ME, Press D (1998) What is environmental studies? *Bioscience* 48:397–405
- Thelwall T, Vann K, Fairclough R (2006) Web issue analysis: an integrated water resource management case study. *J Am Soc Info Sci Technol* 57:1303–1314
- World Commission on Environment, Development (WCED) (1987) *Our common future*. Oxford University Press, Oxford