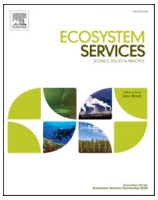
Ecosystem Services 61 (2023) 101525

Contents lists available at ScienceDirect

Ecosystem Services

journal homepage: www.elsevier.com/locate/ecoser



Full Length Article

Mapping linkages between ecosystem services and agricultural and landscape management using structural topic modeling of scientific literature

Gonzalo Villa-Cox [a,b,c,*], Santiago J. Bucaram-Villacís [d], Peter Goethals [e], Stijn Speelman [a]

[a] Department of Agricultural Economics, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, Ghent B-9000, Belgium
[b] ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Sociales y Humanísticas, Campus Gustavo Galindo Km. 30.5 Vía Perimteral, P.O. Box 09, 01-5863, Guayaquil, Ecuador
[c] ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Centro de Investigaciones Rurales, Campus Gustavo Galindo Km. 30.5 Vía Perimteral, P. O. Box 09-01-5863, Guayaquil, Ecuador
[d] Natural Resources & Disaster Risk Management Division, Inter-American Development Bank, USA
[e] Department of Animal Sciences and Aquatic Ecology, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, Ghent B-9000, Belgium

A B S T R A C T

This paper provides a systematic overview of existing literature on the interdependence between terrestrial Ecosystem Services (ES) and agricultural landscape management. To this end, we constructed a structured methodological approach that future researchers and other stakeholders may use to explore potential gaps for potential avenues for research. The framework starts with a constructed bibliometric dataset comprising 2478 paper abstracts (and associated metadata) published between 2010 and 2019 that contain references to specific ES classes and case study location references worldwide. Next, the dataset was used to construct a Structural Topic Model (STM), which features marginal effect estimation of metadata covariates on topic prevalence and topic prevalence correlation. The estimated topic graph structure is then used for community detection of interconnected topics via cluster identification, which allowed us to determine the current structure of the academic discourse on the interaction between terrestrial ES and agricultural landscape management. We expect this novel approach to enable identification of research priorities, prioritization of project portfolios for our field of interest, and reusable methodological guidelines that can be employed to target other fields of knowledge similarly.

1. Introduction

Paraphrasing the Millennium Ecosystem Assessment (MEA), Ecosystem Services (ES) are defined as "societal benefits obtained from ecosystems. They include provisioning, regulating, and providing cultural services that directly affect humans and the necessary services for maintaining direct services" (Reid et al., 2005). Other subsequent frameworks such as TEEB, UK-NEA, IPBES, and FEGS-CS further refined this definition by including elements such as "all the direct and indirect contributions of ecosystems to human well-being making human life both possible and worth living." All these concepts imply not only evident benefits like the provision of goods such as food and raw materials but also other services such as regulation (e.g., -pollination, pest regulation, climate regulation, natural hazard regulation, and water purification), cultural services (e.g., spiritual, recreational, aesthetic and educational), habitat (habitat for migratory species and viability of gene pools) and support services (e.g., hydrological cycle, nutrient cycling, soil formation). Early research efforts, such as Vitousek et al. (1997), show that the importance of these services for human welfare and their subsequent monetization is not a new concept. However, according to Fisher et al. (2009), they have attracted increased attention from the academic and policy community over the last two decades.

The focus of this paper will be narrowed to the body of literature that addresses the linkages between terrestrial ES and landscape and agriculture management. In the context of its application in the field of ES, Setten et al. (2012) characterize landscape management as "the development of strategies for the sustainable use of natural resources with a focus on ecological processes and governance at different spatial scales, by incorporating consideration of the value of ES in decision-making processes." Moreover, the authors define a set of challenges that characterize the limitations and difficulties that existing ES frameworks face to understand and tackle the issue of landscape management properly: 1) inability to fully understand and account for the complex, unique nature of landscapes; 2) the economic bias most ES frameworks impose

* Corresponding autor at: Department of Agricultural Economics, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, Ghent B-9000, Belgium.
E-mail addresses: gvilla@espol.edu.ec, gonzalo.villacox@ugent.be (G. Villa-Cox).

https://doi.org/10.1016/j.ecoser.2023.101525
Received 13 June 2022; Received in revised form 15 March 2023; Accepted 22 March 2023
Available online 8 April 2023

and their limitations in dealing with intangible and context-specific aspects of landscape dynamics; and, 3) lack of consideration for socio-cultural processes in shaping environmental attitudes and behavior. Moreover, Zhang et al. (2007) remark that effective management of ES in this context depends on both agricultural and landscape management. This argument is based on the effect that diversity, composition, and functioning of the surrounding landscape have on the flow of ES to and from agricultural systems. Therefore, it is necessary to consider the broader landscape context in addition to site-specific agricultural management practices to optimize these flows.

Conversely, Dale & Polasky (2007a) emphasize that agriculture is a significant land use, with an estimated 38% of global land destined for these purposes. In addition, agricultural management impacts landscapes by converting natural habitats to agricultural lands. The authors also remark on potential linkages with relevant ES. Namely, certain farming practices, such as tillage, can increase erosion and thus lead to land use change. Also, the use of agrochemicals in agriculture can impact landscapes and their management, as these chemicals can affect the productivity of the land and potentially lead to land abandonment or conversion to distinct purposes.

Following a similar narrative, Reid et al. (2005)claim that a holistic research approach to ES assessment is necessary; that is, to integrate ecological, economic, and institutional perspectives to identify and examine mid to long-term human impacts on ecosystems and the welfare effects of management policies. Nevertheless, there is still a lack of a consistent methodological approach that facilitates the determination of the credibility, comparability, and comprehensiveness of existing studies aimed at identifying both their political relevance and the gaps for future research (Ash et al., 2010; Seppelt et al., 2011). Progress in this regard has been made based on analytical approaches that utilize bibliometric datasets to map knowledge structures that provide multi-disciplinary insights for related fields of scientific research, such as ES, ecology, and climate change (Chen et al., 2020; Wang et al., 2014). For example, Wang et al. (2014) constructed a dataset based on 3004 papers from ISI-WOS from 1991 to 2012 to assess vulnerability to climate change. Similarly, Chen et al. (2020) selected a curated set of 1303 publications from ISI-WOS ranging from 1997 to 2016 to construct a holistic picture of the primary performance of ES literature, including influential journals, productive countries/territories and institutions, and popular methods used in the field. The authors found that forests, agricultural lands, and wetlands are among the landscapes that receive the most attention in ES and the most frequently employed methods in ES assessments.

In this paper, we start from a similar approach to construct a dataset from studies published from 2010 to 2019 using the ISI-WOS and Scopus search engines. Next, we filter the results using filters oriented to capture the linkages between ES and landscape agricultural management. The overall process resulted in a curated dataset of 2478 papers. It was possible to quantitively construct a set of additional metadata-based indicators that account for explicit ES and case studies referencing as done qualitatively by Seppelt et al. (2011) for a significantly smaller collection of papers. Further, our approach is directed at understanding the semantic structure of the literature embedded in our dataset. Cheng et al. (2018) follow a similar path by recurring to a Topic Model (TM) based on a bibliometric approach focused on the ISI-WOS platform and provided several justifications for this approach. They aim to introduce a TM into their discipline and develop an integrated framework for understanding the major research concerns and ecology, environment, and poverty clusters. Interestingly, they conclude that combining TM methods based on bibliometric data can assist in fully explaining research concerns and provide a basis for further investigations in their field. Further, they state that these joint methods are expected to be additionally applied in ecological and environmental economics and management fields.

We intend to build on these approaches and take further steps to apply these methods to our field of interest. This article focuses on developing an analytical guideline using Structural Topic Model, or STM, to achieve this aim (Roberts et al., 2013; Roberts et al., 2016). We proceed further by analyzing the estimation and post-estimation features of the STM to assist in identifying the current structure of the academic discourse on the interaction between ES with landscape and agricultural management, as well as to determine knowledge gaps in this body of literature. The overall objective of this manuscript is thus to provide a systematic overview of the uncovered structure and a methodological avenue that future researchers and other interested stakeholders may use to explore potential gaps in this structure, which can be exploited to extract potential avenues for research.

2. Materials and methods

2.1. Step 1: Buildup of the bibliometric dataset for the linkages between ES in agricultural and landscape management

Several contributions have aimed to understand the interconnection, relevance, and exhaustiveness of the existing literature that characterizes linkages between ES in agriculture and landscape management. As previously highlighted, one such example of a qualitative approach in this direction can be found in Dale & Polasky (2007b). Another similar example is provided by Seppelt et al. (2011), who conducted a quantitative review of 153 articles to assess the consistency in applications of the ES concept and the credibility and replicability of the results of those studies. Despite the relevance of these and similar studies, they result from subjective analytical processes, which tend to be based on a limited number of publications. Therefore, these authors created a series of indicators on the characteristics of discretionally selected samples of publications, which were taken to represent the existing literature. While no approach can genuinely be devoid of subjectivity, we aim to reduce this margin as much as possible by introducing a methodological framework. Thus, we start by incorporating a more comprehensive set of papers employing a bibliometric approach for building our dataset.

Our approach is based on constructing a bibliometric database using the ISI-WOS and Scopus online search engines, which aims at the previous decade of research (from January/2010 up to December/2019). It focuses solely on scientific article abstracts published as part of English papers during this time. As motivated by Gatti et al. (2015), the selected unit of analysis is the paper abstract, given a higher likelihood of containing a high density of words appropriate for inferring topics contained within a given article than using the full-text article. Moreover, choosing paper abstracts over full-text content is primarily driven by data availability and relative ease in extending our proposed methodological approach to other potential fields of knowledge.

First, a set of standard text mining filtering algorithms based on domain-specific knowledge specific was implemented to restrict the scope of the resulting bibliometric dataset adequately. Instead of targeting the overall ES discussion, the intended focus is narrowed to how papers addressing how ES is integrated into scientific studies about land use decisions, agricultural practices, and other relevant aspects of what have thematically denoted *landscape and agricultural management*. However, as noted by Syed & Spruit (2017), the decision to focus solely on paper abstracts might imply that increased attention is needed at all stages (abstract filtering, pre-processing, and model estimation & selection issues) given higher susceptibility of this data source to *noise terms* that might result in lower quality if not adequately addressed. With this in mind, we further systematically processed the initial query results using customized text filters applied to abstracts and titles for explicit references to ES classes derived from the CICES framework (Haines-Young & Potschin-Young, 2018). Then, we repeat the process and filter the resulting papers by searching for spatially explicit case study site references. This system allows further refinement of the resulting dataset by targeting only papers that address specific case studies and one or more ES. These processes result in a database comprising related metadata variables for 2478 paper abstracts. For further details on these

processes and their outcome, refer to Appendix 1.

2.2. Step 2: Buildup and estimation of the structural topic model

The STM approach was put forward by Roberts et al. (2013) and is characterized by using a variational Bayesian approach to estimate a correlated Topic model via a conditional probability process based on document metadata. Aside from improving topic accuracy and interpretability when compared to competing text mining approaches, Roberts et al. (2016) put forward the added advantage of accommodating marginal effect estimation of metadata covariates on topical prevalence, as well as the use of topic correlation structure to build a representation of estimated topics in the form of a network.

Based on the resulting dataset described in section 2.1, we built an STM by constructing a *corpus of documents* based on the set of processed abstracts in the dataset (the basic unit of analysis). There are a series of processes necessary to build this corpus from the raw input available in the bibliometric dataset before it can be of use for an STM:

1. Conduct the pre-processing steps described in Roberts et al. (2019) to help structure the corpus of documents into a set of root words and tokens: convert to lowercase, remove punctuation, remove numbers & English stop words, and stem using snowball algorithms[1].
2. Determine the words and tokens that should be removed from the corpus if they occur infrequently. Then, determine the resulting number of empty documents for removal. For example, Fig. 1 shows that 175 is a suitable threshold for removing words if they appear in less than 175 abstracts. This process removed over 121,772 tokens but no documents from the final corpus.

Syed & Spruit (2018) noted that applications of TM using asymmetric priors for *document–topic* distribution for abstract-based TM result in a significant increase in topic coherence scores and improved human topic ranking compared to symmetrical priors (which are commonplace in TM applications)[2]. Notably, a similar effect is accomplished within the STM approach by conditioning topic prevalence on metadata about the paper into the more traditional TM framework. Following Roberts et al. (2013), the Data Generating Process (DGP) comprises a set of $(w_1, w_2, \cdots, w_V)$ words associated with a set of $(T_1, T_2, \cdots, T_K)$ topics, which together form a set of $(D_1, D_2, \cdots, D_M)$ documents associated with a set of metadata covariate vectors $(X_1, X_2, \cdots, X_d, \cdots, X_M)$. Then, each document d is generated as follows:

For each document d, draw the document-specific $K \times 1$ topic prevalence vector θ_d, conditional on the $p \times 1$ metadata covariate vector X_d has the following distribution:

$$\theta_d | X_d\gamma, \Sigma \ \text{LogisticNormal}(X_d\gamma, \Sigma) \tag{1}$$

Where the prior distributions for matrix γ (of dimension $K \times p$) and matrix Σ (of dimension $p \times p$) are:

$$\gamma_k \ N\left(0, \begin{bmatrix} \sigma_1^2 & 0 & \cdots & \\ 0 & \sigma_2^2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \cdots & \sigma_p^2 & & \end{bmatrix}\right) \quad \text{for each } k \in \{1, \cdots, K\} \tag{2}$$

$$\text{where} \qquad \sigma^2_{t\in\{1,\cdots,p\}} \ \text{Gamma}(a_\gamma, b_\gamma)$$

$$\Sigma^{-1} \ \text{IW}(S, M)$$

Here, the hyperparameters σ_k^2, a_γ, b_γ and S are calibrated at the estimation stage to prevent over-fitting.

Document-specific topical content $V \times 1$ vectors $\beta_{d,k}$ (distribution over words) is then drawn from a base level μ and a topic-specific deviation τ_k, utilizing the following process:

$$\beta_{d,k} \propto \exp(\mu + \tau_k) \text{for each } k \in \{1, \cdots, K\}$$

With priors:

$$\begin{array}{l} \mu_v \ \text{Laplace}(\varphi_v, \pi_v) \\ \text{for each } v \in \{1, \cdots, V\} \\ \pi_v \ \text{Gamma}(a_\pi, b_\pi) \end{array} \tag{3}$$

$$\begin{array}{l} \tau_{k,v} \ \text{Laplace}(0, \pi_k) \\ \text{for each } v \in \{1, \cdots, V\} \\ \pi_k \ \text{Gamma}(a_k, b_k) \end{array}$$

Where the hyperparameters φ_v can be initialized to the empirical frequency of the word v in the vocabulary, and the others are set to avoid overfitting as well.

3. Then, for each word $v \in \{1, \cdots, V\}$ in each document $d \in \{1, \cdots, M\}$:
   a. Draw the word's topic assignment conditional on document distribution over topics:

$$z_{d,v} | \theta_d \ \text{Multinomial}(\theta_d) \tag{4}$$

   b. Conditional on the drawn topic, now draw a specific word for that topic:

$$w_{d,v} | z_{d,v}, \beta_{d,k=z_{d,v}} \ \text{Multinomial}\left(\beta_{d,k=z_{d,v}}\right) \tag{5}$$

The DGP described above can be estimated for any given STM with a fixed number of topics K using a semi-collapsed variational EM algorithm with deterministic spectral initialization (Roberts et al., 2016). Under this specification, metadata covariates for topical prevalence allow the observed metadata to affect the frequency with which a topic is discussed. For the case of continuous covariates, semi-parametric basis spline functions are used to configure the design matrix X_d and to allow for nonlinear partial effects. Based on these features, the model is specified by considering the following list of metadata covariates for the dataset introduced in subsection 2.1:

- *PY:* Year of publication.
- *ASJC.Knowledge.Classification:* Dummy variable for Knowledge Classification of source journal.
- *region.covar*: Dummy variable for geographical region referenced in the content of either the abstract or title of each paper in the dataset.
- *ecoserv.count*: Count of the number of specific ES references contained in the document following the CICES v5.1 nomenclature.
- *TC*: Total number of citations for each paper. This is a measure of overall paper prevalence in the body of literature.

Further, there is the issue of model selection. The question of how many topics must be estimated from a corpus of documents based on paper abstracts is complex in the TM literature, with no consensus yet. Given previously discussed considerations, a data-driven approach will be used to determine the number of topics K to improve the quality of the estimated K topic set. The best tradeoff between *semantic coherence* and *term exclusivity* is selected to achieve this goal (both estimated over a holdout sample of documents). semantic coherence is a criterion developed by Mimno et al. (2011) that is maximized when the most probables words in each topic frequently co-occur. These authors show that the metric correlates well with the human judgment of topic quality. However, Roberts et al. (2016) suggest that high semantic coherence

[1] This is a family of standard deterministic algorithms for computing root tokens. For more detail please refer to https://snowball.tartarus.org/texts/introduction.html.

[2] These authors also show that the choice between asymmetric/symmetric distributions does not imply comparable improvements on the topic–word distributions in TMs based on paper abstracts.

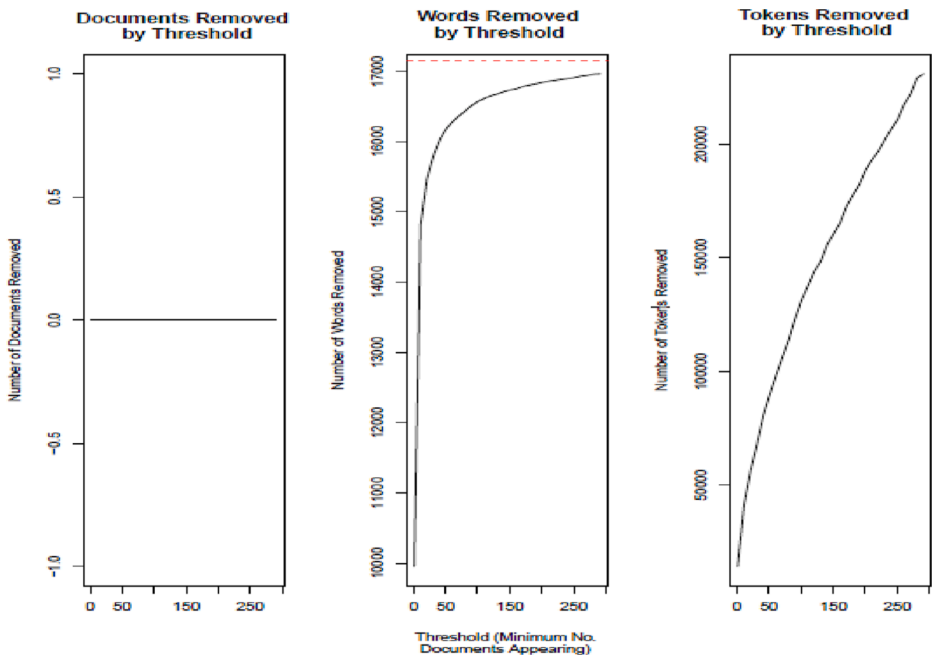


Fig. 1. Setup of token, term, and document removal threshold from the corpus of documents (paper abstracts).

can be obtained by selecting a small topic dominated by familiar words. Accordingly, they propose to measure topic quality through a combination of semantic coherence and exclusivity of words to topics. Thus, the adopted strategy consists in estimating a full STM model for different values of K using a random selection of close to 100 documents from the corpus as a holdout set for estimating each candidate value. Then, semantic coherence and term exclusivity are calculated on the remainder set of held-out documents to ensure the maximal predictive capacity of the selected model configuration. Once model selection based on tradeoffs between these two measures is carried out, the final STM will be estimated on the entire corpus of documents utilizing the selected K.

2.3. Step 3: Post-estimation in the structural topic model

After model estimation, several post-estimation processes are available. The following procedures provide metrics that can be used to guide and assist human interpretation, characterization, profiling, and validation of the set of K estimated topics:

1. The ability of STMs to condition topic prevalence on document metadata covariates is exploited to estimate their partial effects and analyze their marginal contribution to topical prevalence for both continuous and discrete covariates. It is worth noting that the resulting estimates will carry on global uncertainty propagation (which means that uncertainty of topic estimation will be included in the estimation of partial effects credible intervals). Each of these metrics will be used to assist human profiling and contextualizing the resulting set of K topics concerning all covariates used in the estimation process. It should be noted that this capability goes beyond standard TM approaches, which are still more prevalent in this literature.
2. The estimated model can predict $\widehat{\theta_d}$ at a document level. Two applications of this capability can be exploited:
   - o For each document d, the top 5 topics with the highest predicted posterior probability in $\widehat{\theta_d}$ are selected as the vector of predicted topics. These will then be contrasted with the references to ES classes in each document to build a *co-Occurrence matrix* with the count of cross-references between predicted issues and ES class references. A topic-wise inspection of the structure of this matrix is presented in Appendix 2 and provides additional criteria to support human profiling and interpretation of the topic.

   For each estimated topic, the top 10 documents (paper abstracts) having the highest Maximum A-Posteriori (MAP) estimate of the topic's $\widehat{\theta_d}$ value are identified[3]. The resulting set of paper abstracts can be considered "representative" for each of the resulting K topics. It should provide a vital context source for human validation and a deeper characterization of each topic. In other words, for each estimated issue, informed readers (experts) can investigate these ten abstracts and identify a *common theme* across each set of 10 papers.
3. Another useful post-estimation feature of STMs is that estimated topics are (or, more accurately, may be) correlated. In this sense, the presence of positive topic correlation entries in the estimated $\widehat{\Sigma}$ the matrix can be interpreted as a tendency for the co-occurrence of topics in the same documents (paper abstracts). Based on these, network representations of these relationships can be constructed to formally analyze this network in terms of which topics are addressed simultaneously and how such first-degree relationships can be compounded into indirect higher-degree relationships connecting issues. Depending on the resulting structure, these indirect connections can be characterized as meaningful *topic communities* that provide an additional layer of abstraction to construct the field of knowledge. This will be addressed in depth in the following subsection.

2.4. Step 4: Details of the structural analysis of the resulting topic network

The problem of estimating an undirected graph representation for the estimated $\widehat{\Sigma}$ matrix is addressed by recovering edges in a high-dimensional undirected graphical model: In these settings, observations are assumed to come from a multivariate normal distribution with

[3] In general, high MAP values for this estimator imply a higher probability of having a direct thematic correspondence to any given specific topic, irrespective of the values these estimates may have for other topics computed for the corpus (of paper abstracts).

a sparse precision matrix. The goal is to infer which elements of the precision matrix are non-zero corresponding to edges in a graph. T. Zhao et al. (2012) proposed transforming the topic proportions using penalized semiparametric Gaussian copulas to transform the data marginally. This weakens the gaussian assumption of the subsequent procedure. The graph representation is then computed using Meinshausen & Bühlmann (2006) procedure. Model selection for the scale parameter of the *L1 penalty* is performed using the *Rotation Information Criterion* (RIC), which estimates the optimal degree of regularization by random rotations. T. Zhao et al. (2012) note that this selection approach has strong empirical performance but can be sensitive to the under-selection of edges[4].

Once the resulting graph structure is estimated for the topical network, the following standard network metric statistics can be applied to provide insight into how these topics interact within the corpus of documents as follows:

- *edge.dens*: The edge density of a graph is the ratio between the number of edges and the total number of possible edges.
- *transitiv*: Transitivity measures the probability that the adjacent vertices of a given vertex are connected. Namely, it provides the possibility that node x is connected to node z, given that node x is connected to y and y is connected to z.
- *Centralization degree*: Compute a graph-level centrality score or graph $G := (V, E)$. The formula for this is:

$$C(G) = \frac{\sum_{v}^{|V|}\left[\max_{w}(c(w) - c(v))\right]}{|V|^2 - 3|V| + 2} \tag{6}$$

where $c(v)$ is the centrality of vertex v and $|V|$ is the total number of vertexes in graph G. Following Bonacich & Lu (2012), the measurement of $c(v)$ is implemented in two ways:

- o *hub.score*: The hub score of a vertex is defined from the principal eigenvector of $A \times A^{'}$, where A is the adjacency matrix representation of the estimated graph. It provides a measure of vertex centrality from the notion of the "popularity" of a vertex in terms of the number (and weight) of outgoing edges to other nodes. In addition, it allows the identification of *hub nodes* or nodes with a high hub score. These are nodes with relatively many connections to other hub nodes[5]. This measure's most centralized network structure is the graph with a single edge (and potentially many isolates), which becomes the reference structure for comparison.
- o *centre.betw*: Betweenness centrality is measured by the number of shortest paths (between any couple of nodes in the graphs) that pass through the target node[6]. In plain words, it provides a metric of the influence of a vertex over the flow of information between every pair of vertices under the assumption that information primarily flows over the shortest paths between them. Note that the most centralized network structure for betweenness is some version of the star graph, in-star, out-star, or undirected star.

Next, the estimated topic graph structure will be the basis for *community detection* of related topics. Newman & Girvan (2004) propose a method for this task by constructing clusters of topics based on the same betweenness notion described earlier but applied to edges between connected components (sets of nodes)[7]. The underlying idea relies on the notion that the betweenness score of edges connecting separate components is likely high, as all the shortest paths from one element to another must traverse through them. Thus, a dendrogram (a hierarchical map or tree) can result by iteratively removing the edges with the highest edge betweenness score. Once the final tree is obtained, the selected cutting level determines component membership, as with all other hierarchical clustering algorithms. The cutting criteria used here is the *modularity score* of the set of components (Clauset et al., 2004). Significantly, these methods will add metrics to the toolset that the STMs already incorporate to assist the interpretation and contextualization of each topic within the structure of academic discourse of the field.

Lastly, the uncovered structure has a final application that can assist future research efforts to position their intended contributions within the surveyed field of knowledge. Westgate et al. (2015) remark that research gaps can be identified in practice as pairs of themes that are unusually separate within the Corpus, both in terms of their thematic content and the articles in which they appear. Based on these insights, experts using the toolset developed in this paper can situate any prospective contributions in terms of how they fit within their thematic content and the ES they address, as well as the gaps they could address utilizing (at least) the following set of criteria:

1. Does the contribution address two specific topics for which there is no common edge but are both member topics of clearly identified communities?
2. Does it address two specific topics for which there is a common edge, but these are connecting topics between two adjacent topic communities?
3. Do the corresponding metadata covariates map towards an estimated partial effect different from what was identified in the current structure?
4. Are the ES addressed by the intended contribution not among the top co-occurrent for their corresponding communities and/or addressed topics?

3. Results

3.1. ES class reference structure

We begin with Fig. 2, which shows the frequency histogram for the 20 most frequent ES classes referenced in the database. It becomes clear that the singularly most prevalent ES class is *Cultivated plants, algae, and fungi,* which englobes references to any crops and fruits grown by humans for food (food crops), materials, or energy. Aside from this, other prevalent Provisioning ES classes are *Fibers and other materials from plants, algae, and animals for direct use, processing, or energy* and *Reared animals and their outputs (including aquaculture)*. Also, it is worthwhile to note that *Water Supply* is also a prevalent Provisioning ES class, which encompasses water availability for different types of consumption (drinking, irrigation, industrial use, etc.).

Moreover, several Maintenance/Regulation ES class references are also prevalent in the discussion. The caveat for these classes is that, as pointed out earlier, there is some degree of diffuseness among them. For instance, Fig. 2 shows that both *Pest and disease control, including invasive species* and *Biological control,* are prevalent. However, compared to the prevalence of the maintenance of biodiversity, *genetic diversity* both rank as less prevalent than the latter. Given the potential overlap of both classifications, this might be a drawback to adopting this method. However, we argue that it still provides a comprehensive overview of the most prevalent Maintenance/Regulation ES classes in this body of

[4] While admittedly a conservative metric, we chose it to maximize certainty that any uncovered links can be meaningful in assisting the identification process of the structure of the targeted body of research detailed at the end of this subsection.

[5] This is true for any undirected graph, for which this measure is equivalent to both authority score and eigen-centrality score measures at a node level. Thus, the notion of hub and authority are equivalent in this case.

[6] The betweenness centrality measure for each node is normalized by dividing the final count by $1/2[(N-1)(N-2)]$, where N is the total number of nodes in the graph.

[7] These are understood as edges connecting two nodes, each one belonging to a distinct component.

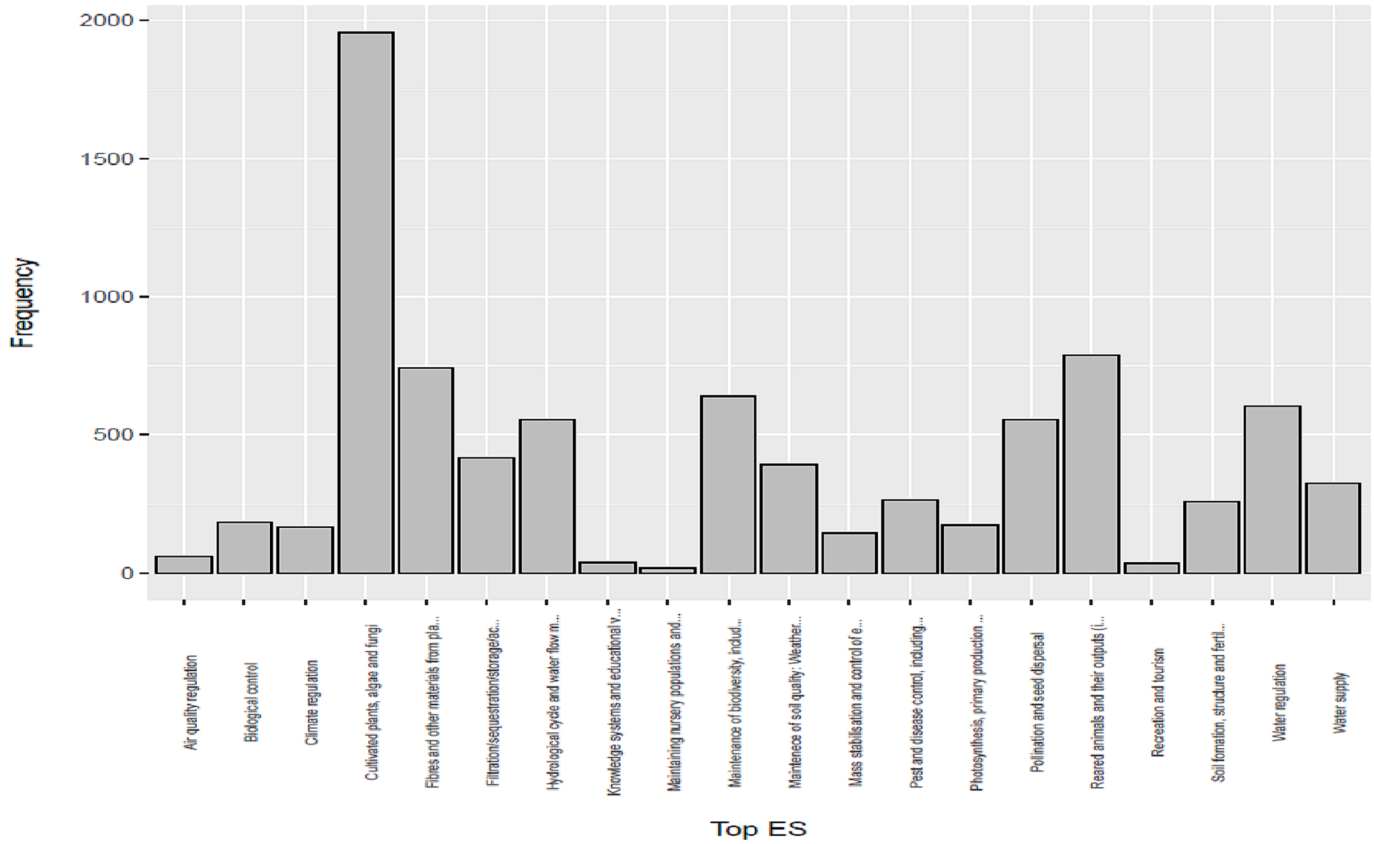


Fig. 2. Histogram of references for the top 20 ES classes referred to in the dataset.

literature. Also, other important ES classes are commonly associated with the academic discussion captured in this dataset: *Pollination and seed dispersal*, *Maintenance of soil quality: Weathering processes, decomposition, fixing and other processes*, and *Water Regulation*, among others. Further details on the structure of the estimated co-occurrence matrix for cross-references between estimated topics and references ES may be found in Appendix 2 (Table A.5).

3.2. Topic estimation results for the STM

Moving onto the results of the STM themselves, Fig. 3 shows a plot of the measures of Coherence and Exclusivity across different possibilities of the number of topics K estimated for a series of STMs on the corpus. The figure suggests that the best tradeoff between these measures occurs around $K = 15$, corresponding to the maximal arithmetic mean between both measurements[8]. Thus, this is taken to be the best K, and an STM is estimated on the total sample with the selected number of topics.

After the estimation step, a first glance of the resulting estimated topics is plotted in Fig. 4 (top 3 terms per topic), along with their overall prevalence in the corpus. Topics 9, 7, and 12 are the three most prevalent topics, while 5, 1, and 2 are the three least pervasive. A quick inspection of these shows that themes such as farm management and farmer issues, landscape, and climate change concerns seem prevalent in the literature. In contrast, others address wetland ES valuation, interactions amongst crops with pollination services, and forest ecosystems are less common.

Further, as mentioned in section 2.3, the STM provides a set of post-estimation scores that should assist expert interpretation, contextualization, and profiling of the resulting set of 15 topics. The first of such tools are *word clouds* compiled from the most frequent tokens associated with each estimated topic, then plotted along with the *top 2 thoughts* per topic (top abstracts, where each topic's posterior predicted probability is highest in the corpus). Further, as mentioned in section 2.3, the STM provides a set of post-estimation scores that should assist expert interpretation, contextualization, and profiling of the resulting set of 15 topics. The first of such tools are *word clouds* compiled from the most frequent tokens associated with each estimated topic, and then plotted along with the *top 2 thoughts* per topic (top documents, where each topic's MAP estimator for $\widehat{\theta_d}$ is highest in the corpus). As an example, Fig. 5 presents the case of Topic 9 (the most frequent topic in the corpus), which at first glance appears to unveil relation to policy and sustainability concerns about the adoption of sustainable agricultural practices, including participatory knowledge transfer, communal support structures for practice adoption, extension services, among others.

Naturally, these initial insights into each topic's discussion need to be broadened to enable deeper interpretation and contextualization for each estimated topic. To this end, instead of taking just the top 2 thoughts as done earlier, section 2.2 estimated the top 10 thoughts per topic. This led to analyzing 150 abstracts searching for common thematic content for each topic across their respective thoughts. The interested reader may find selected illustrations on these characterizations and tables that compile the complete list of extended descriptions and references for the corresponding thoughts in Appendix 2 (Tables A.1 through A.3).

Second, continuing Topic 9 as an illustration, Fig. 6 depicts all estimated partial effects for all available continuous covariates. Notably, there is a negative time trend that can be observed in the prevalence of

[8] This equates to maximizing the convex combination of semantic coherence and term exclusivity with equal preference for both measures. It's worthwhile to note that balancing these two criteria can be expected to induce some degree of repletion in the assignment of certain tokens to each estimated topic. However, as explained in detail in section 2.2, this is done by design to improve topic quality and correspondence with human interpretation.

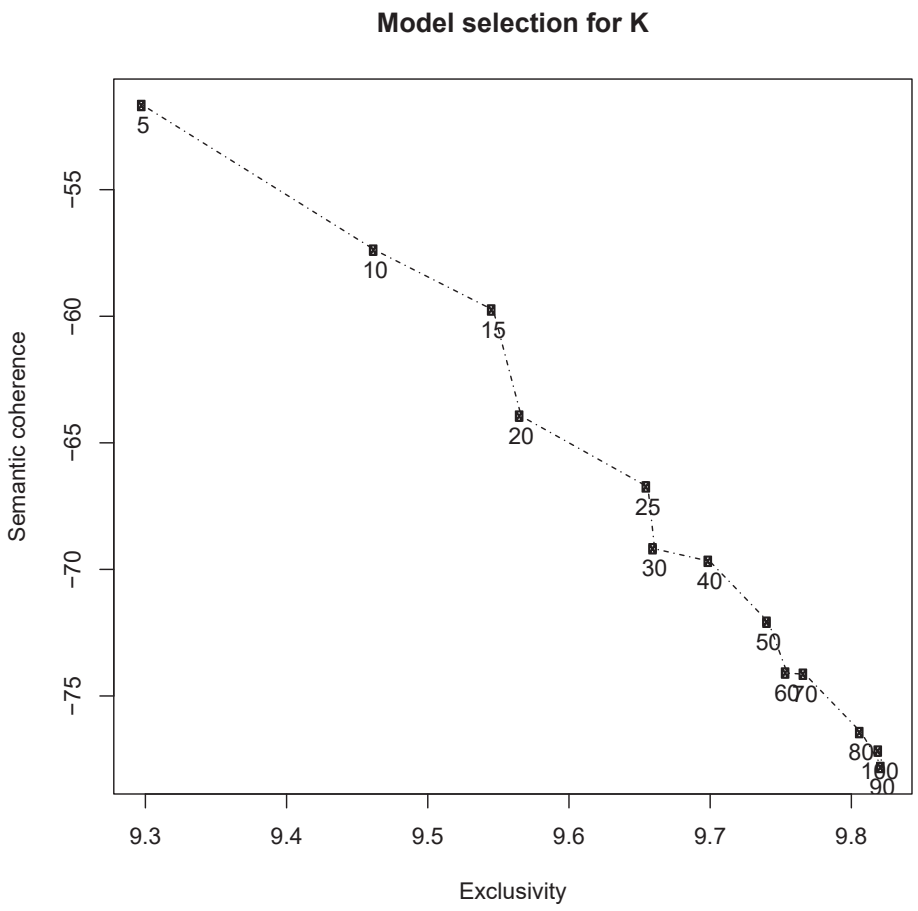


Fig. 3. Model selection for the number of topics of different STMs estimated on the corpus.

this topic, meaning that this topic has been decreasingly discussed in the literature from 2010 onwards (upper-left panel), equally prevalent in papers that address multiple ES and only comparatively more prevalent in relatively low cited papers.

Third, our list's other two discrete covariates provide interesting contrasts for all topics regarding what marginal effects illustrate. For example, the *ASJC knowledge classification* covariate on topic prevalence is shown in Fig. 7 for all estimated topics. Inspection of the figure shows a spread distribution concerning domain emphasis of topics in the body of literature between Environmental/Agricultural or Social/Decision sciences. For instance, Topic 9 cannot be statistically distinguished between these discipline domains, but it does show a slight skew toward the Environmental/Agricultural domain. There are, however, topics such as 5, 8, 13, and 14 whose prevalence cannot be distinguished between knowledge classifications. This is an interesting indication that these topics are currently being addressed in a multidisciplinary fashion in the existing body of literature. All other topics, however, show dispersion across the spectrum of the scope of scientific discussion (to a greater or lesser extent).

Further, Fig. 8 demonstrates the marginal effect of the regional reference covariate on topical prevalence. Continuing with the example discussed in this section, Topic 9 reflects a higher prevalence in papers focusing on case studies from Europe and North America. In general terms, however, nine topics do not report a significant marginal effect for this covariate for any regions shown in this figure. This suggests that these topics are equally prevalent across all areas in the database. At the same time, the remaining six tend to show higher prevalence in papers explicitly oriented towards one or more regions.

3.3. Estimation and structure of topic network

The next step is to make sense of the correlation structure of the estimated $\widehat{\Sigma}$ matrix produced by Meinshausen & Bühlmann (2006) (2006)'s algorithm (shown graphically in Fig. 5). Edge density is estimated at 0.192, which suggests a considerable degree of sparsity in the structure of this graph. Also, the transitivity coefficient for this network is 0.466, which conveys a fair probability that sensible clustering of topics can be constructed[9]. Another perspective of the estimated structure is provided by centrality measures computed for this network. The aggregate hub-centrality score is 0.162, and the betweenness-centrality score is 0.2484. Jointly, these scores suggest that the network does not resemble a star-like structure with one or more central nodes but rather further favors the notion of distributed communities (or clusters of) topics with some degree of connectivity.

Further, an adjacency matrix representation of the network depicted in Fig. 9 is used as the basis for the betweenness clustering algorithm described earlier. The resulting betweenness dendrogram is portrayed in Fig. 10 (left panel). The figure also represents the resulting topic communities (right pane) with an optimal modularity score computed at 0.451 and four topic clusters (A through D). Further insights into the resulting structure are provided in Table 1 via topic-specific centrality

[9] This implies there is close to 50% chance that any 2 nodes connected to a common node are connected between themselves.

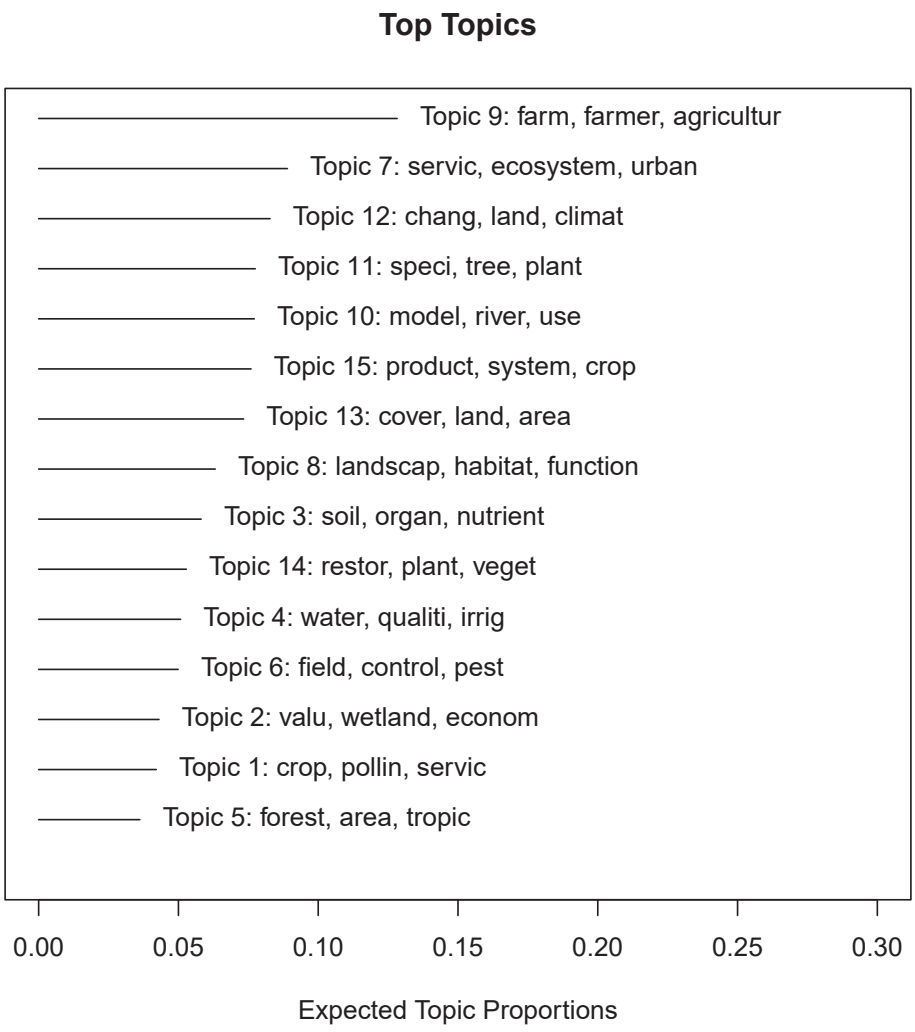


Fig. 4. Topic proportions resulting from the selected STM estimation.

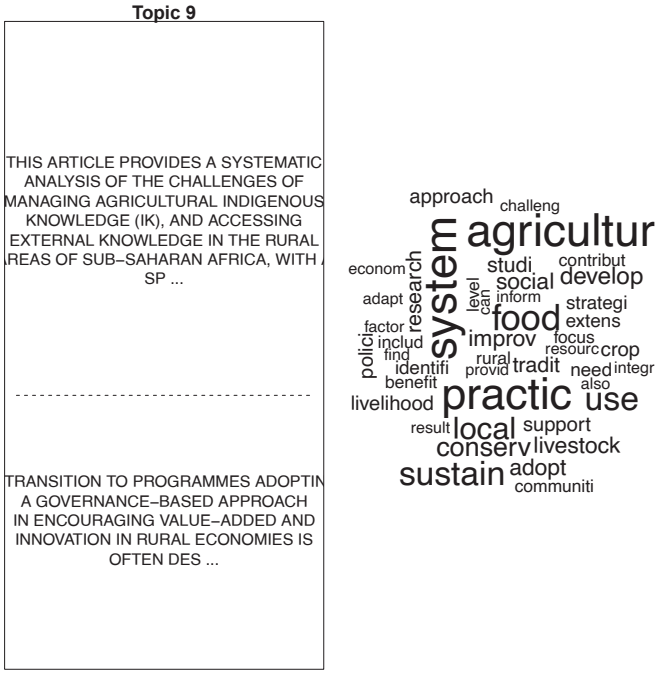


Fig. 5. Word cloud and top 2 thoughts for Topic 9.

measures arranged by community[10].

Notably, inspecting the resulting community/cluster structure is needed to operationalize the knowledge gap identification strategy in section 2.4. As mentioned before, these could be considered natural bridges between topic communities, since there already should be an existing body of literature that serves as a basis for papers that engage both topics simultaneously. Namely, there are three such connections (edges) between communities A and C (all connected through Topic 5), one connection between A and B, and one between communities B and C and communities C and D[11].

All figures discussed thus far are available as complete scorecards for all 15 estimated topics (Tables 1 through 3). These tables compile the top three tokens, the relative prevalence rank (labeled from 1 being the most prevalent topic, 15 being the least prevalent), as well summarized versions of the extended interpretation from Tables A.1 – A.3. Importantly, they also include community members with related connectivity metrics described earlier in this section.

[10] Particularly, hub and betweenness scores will be supplied to provide complementary perspectives of the connectedness degree for each topic and determine which ones may serve as pivots for other topics in the general body of the discussion. Table A4 in Appendix 2 provides a summary of all the measures described so far, organized by the estimated topic communities presented above.

[11] In Fig. 6, such edges connecting distinct communities are colored red, whilst "inner" edges within communities are colored black.

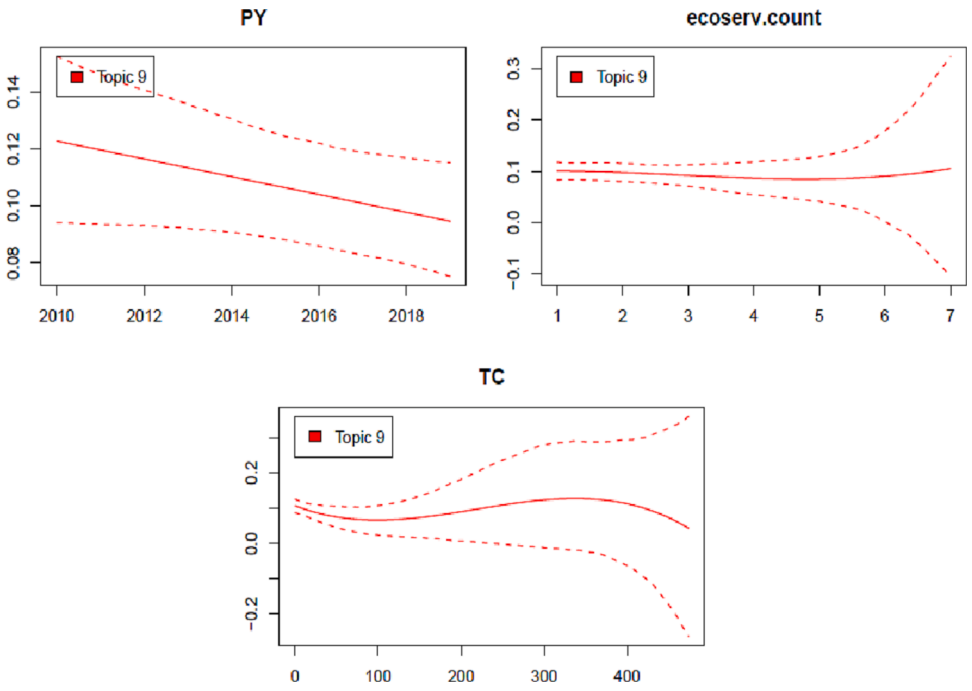


Fig. 6. Partial effects of continuous covariates for Topic 9.

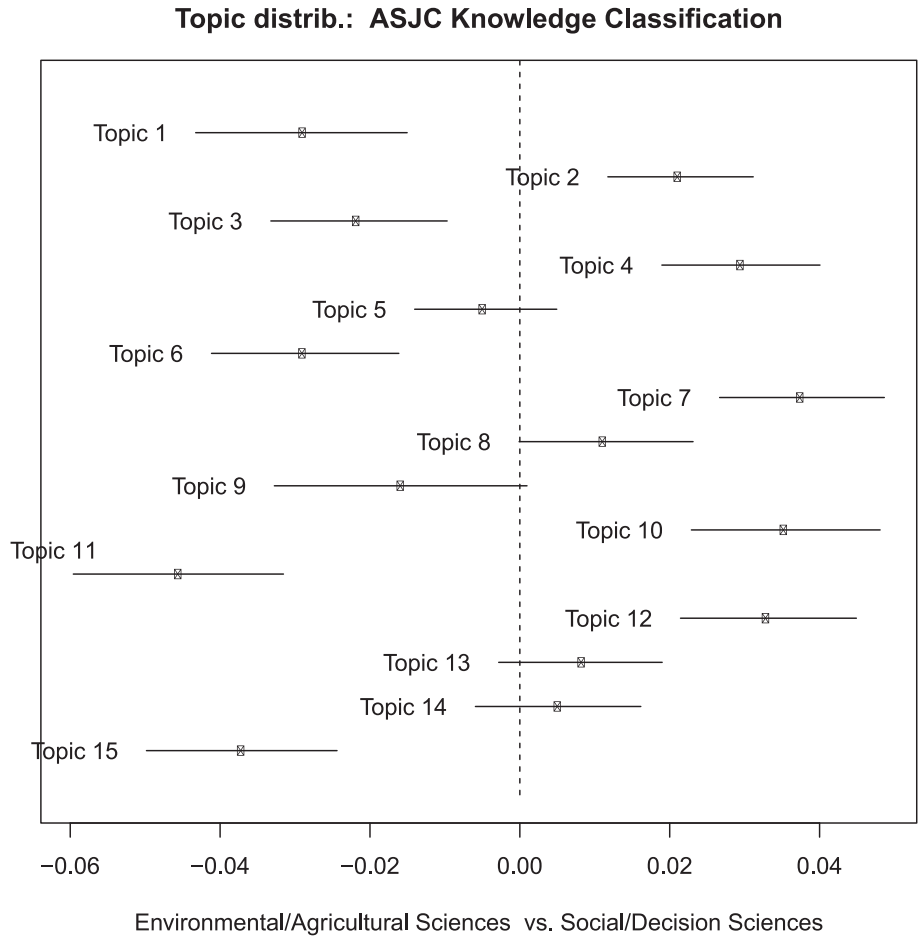


Fig. 7. Partial effect of ASJC knowledge classification for all estimated topics.

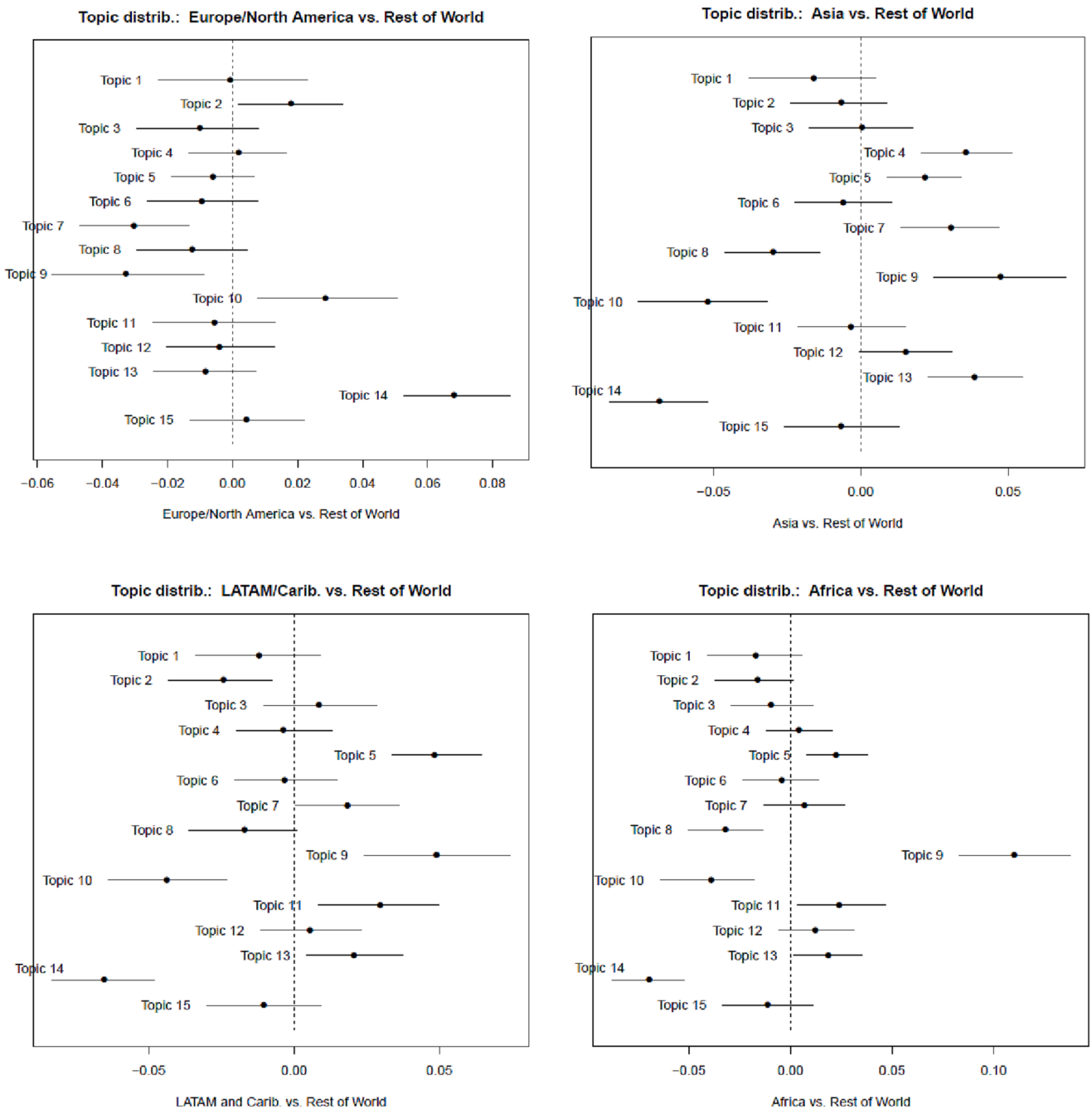


Fig. 8. Partial effect of regional references for all estimated topics, grouped per region of interest: Europe/North America (upper-left pane), LATAM/Carib. (upper-right) pane, Asia (lower-left pane), and Africa (lower-right pane).

4. Discussion

In this section, the insights that the discourse structure presented previously can help to unveil are contextualized. According to Meraj et al. (2022), the general field of ES continues to face significant challenges in understanding the impacts, tradeoffs/synergies, effective ways to assign monetary values to services, and integrating existing assessments into effective management strategies. We thus present a series of tables that compile the results described for all estimated communities and their respective member topics[12]. Then, we direct our attention toward the insights these can provide for our case study.

We start by addressing Community A (Table 4), which comprises mid-level and low-level prevalence topics 1, 6, 8, 11, and 14. Put together, avg. hub and betweenness scores suggest that this community

[12] Tables 1 through 4 present brief thematic summaries for all corresponding member topics, as well as the estimation and post-estimation results presented in section 3 and Appendix 2. These tables include the top 3 tokens, the top 4 co-occurring Regulation/Maintenance ES, the relative topic prevalence rank, relevant metadata partial effects plus network connectivity metrics.

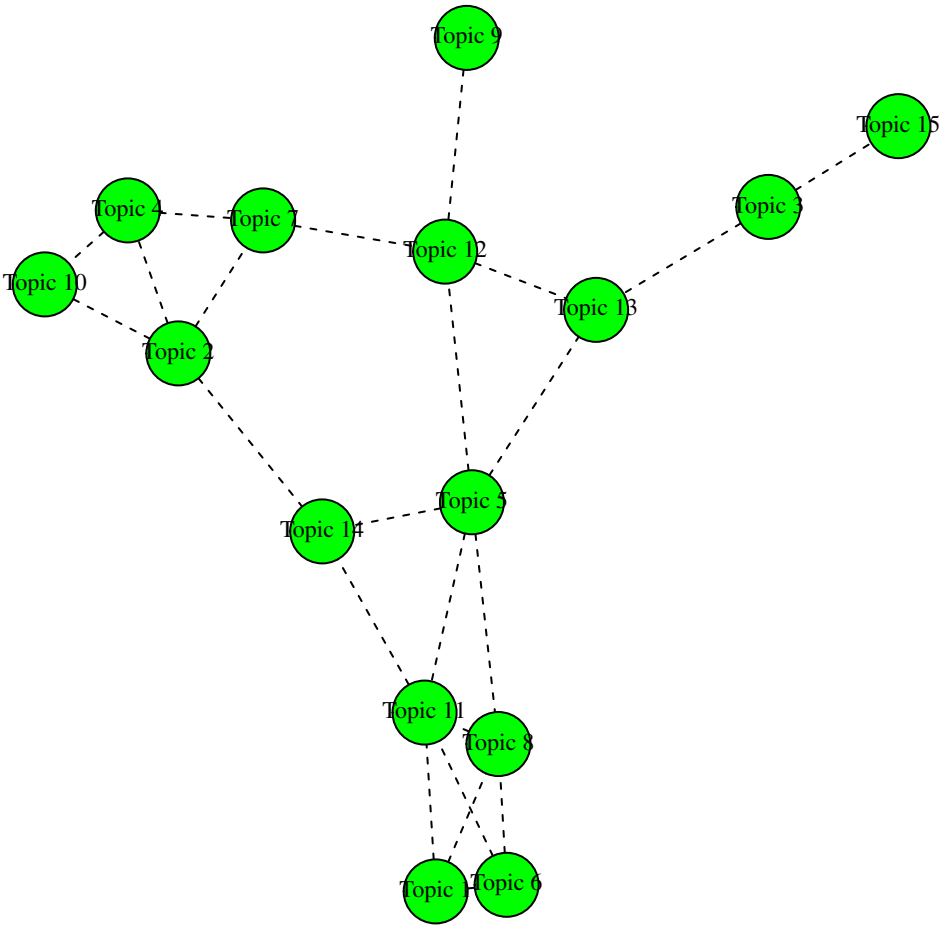


Fig. 9. Topic network structure based on the marginal topic proportion correlation matrix.

is relatively densely connected with the overarching study area, with several member topics acting as discussion hubs. Also, most research in this group is more frequent in the Environmental/Agricultural Sciences field. It exhibits no regional distinction on topical prevalence, except for the discussion on alterations of ecosystem functions that support biodiversity from increased intensive farming and related landscape changes (Topic 8). This topic shows orientation toward Social/Decision Sciences and seems more widespread in Asian, African, and LATAM & Carib contexts (though the latter is not statistically significant).

Topic 14 (restoration and maintenance management practices) follows a similar regional pattern. Still, it does not show an inclination in prevalence towards any field of knowledge and shows decreasing presence through the previous decade of research. Also, this topic has a relatively high betweenness score, shown in Fig. 6, indicating that it serves as a potential bridge between this community and communities C and B. However, hints towards potential gaps are also present within this community by deepening the link with other member topics. Doing so could blur the distinction between communities A, B, and C.

Further inspection of the connectivity structure of this community uncovers that Topic 11 (inventory of environmental functions/services provided by agroforestry and forest landscapes) has the highest values for both hub and betweenness scores. This suggests that it is among the most relevant central topics in the whole body of literature and has a significant role in overall topic connectivity in this field beyond its role in community A. Yet, despite this, there is a marked decreasing prevalence of this topic across the span analyzed in our dataset. However, Topic 8 is directly connected to it and is a relatively close second regarding connectivity and centrality. Curiously, Topic 6 (management practices related to organic agriculture and landscape complexity concerning pest and biological control) also acts as a hub but does not link this community, nor any other. Also, the remaining topic in this community, Topic 1 (pollination/seed dispersal and their relationship with crop productivity), shares this trait with Topic 6. Yet, this topic has shown an increasing prevalence in the literature towards the end of the past decade. Summarizing all these aspects, this community will thus be labeled: *Implications of agricultural landscape management on Regulation/Maintenance ES in terrestrial agroecosystems.*

The second presented community, termed *Socioeconomic and sustainability concerns in agricultural landscape management*, comprises topics 5, 9, 12 and 13 (labeled Community C) and is presented in Table 5. Most topics in this community are also among the most prevalent, having the best average prevalence and the clearest multidisciplinary inclination. These topics are broadly holistic, with a broader focus that often includes a discussion of the sociocultural/economic, policy, and sustainability concerns surrounding landscape and agricultural management in the context of overall ES discourse. In this regard, two scopes mainly dominate this cluster: 1) comparative assessments of impacts of landscape management and climate change to related ES provided by agroforestry systems (e.g., filtration/sequestration/storage/accumulation and erosion processes performed), and 2) biotic and abiotic processes that contribute to maintaining biodiversity.

Network connectivity for this community reflects the most “star-like” community since Topic 5 is a clear hub topic (with a distinctly high score), which is also central in the betweenness sense. The topic discusses the impacts of agroforestry systems on ES, which often result in tradeoffs and synergies between provisioning and maintenance/regulation of ES across different regional contexts. This topic is distinctly positioned to bridge gaps in the overarching discourse structure, as it

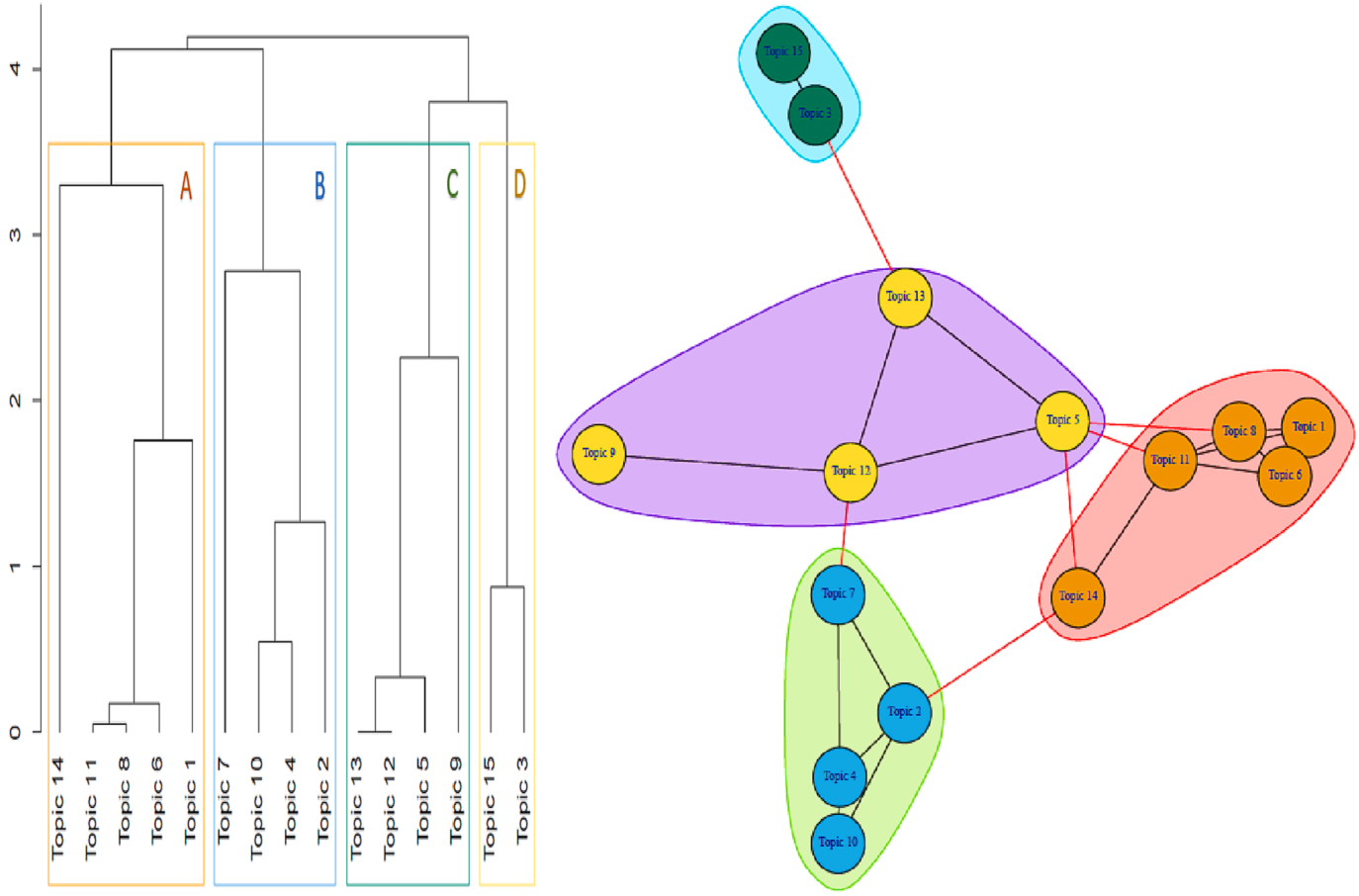


Fig. 10. Graphical representation (left pane) and betweenness dendrogram (right pane) of hierarchical clustering for estimated topics showing best community configuration.

Table 1
Detailed topic scorecard for interpretation, estimation results, community membership & connectivity metrics (Topics 1 through 5).

1	crop, pollin, servic	crop pollin	Impact of the spatio-temporal variability of pollination, seed dispersal and other related ES on crop production/productivity and maintenance of diverse plant communities.	14		Environmental / Agricultural Sciences	No distinction	A	0.6226	0.0000
2	valu, wetland, econom	wetland econom valu	Economic value of ES (e.g. pollination and those provided by wetlands and other terrestrial water bodies) and its assessment via of market-based and preference elicitation methods to inform policy planning related to landscape conversion.	13		Social / Decision Sciences	LATAM & Carib.	B	0.0173	0.1850
3	soil, organ, nutrient	soil	Relevance, measurement and physico-chemical/biological factors that influence soil health in soil quality assessment.	9		Environmental / Agricultural Sciences	No distinction	D	0.0099	0.1429
4	water, qualiti, irrig	water qualiti	Management options for water resources in coastal zones and terrestrial water bodies, including the use of hydraulic structures, the extraction of surface and groundwater for irrigation, as well as the assessment of the sustainability of agriculture through the use of tools such as total water footprint assessment.	11		Social / Decision Sciences	No distinction	B	0.0060	0.0275
5	forest, area, tropic	forest	Tradeoffs and synergies between related ES in both complex and homogenous agroforestry systems.	15		No distinction	No distinction	C	0.6038	0.3571

connects directly to other hub topics in Community A. That is, there is promise in expanding literature that further links this topic to supporting ecosystem functions and processes (Topic 11), the impact of landscape management practices (Topic 8), and restoration/maintenance practices (Topic 14). This aligns with Meraj et al. (2022), who remarks that despite the importance of these links for effective planning of conservation and restoration policies, such efforts are still not sufficiently available. Yet, our analysis also shows that if the distinction between communities A and C were to be blurred, explicit research paths towards holistic sustainability-centered assessments on

Table 2
Detailed topic scorecard for interpretation, estimation results, community membership & connectivity metrics (Topics 6 through 10).

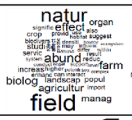
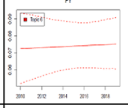
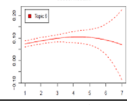
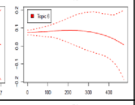
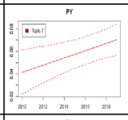
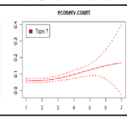
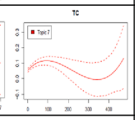
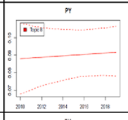
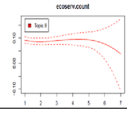
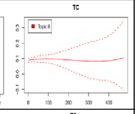
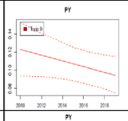
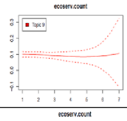
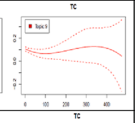
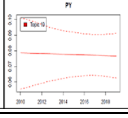
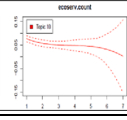
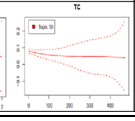
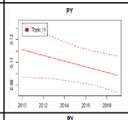
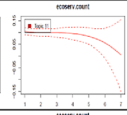
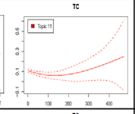
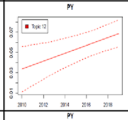
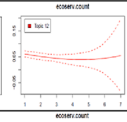
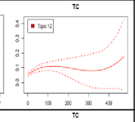
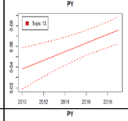
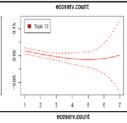
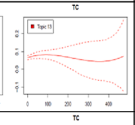
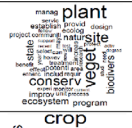
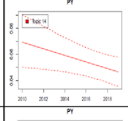
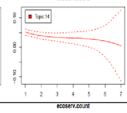
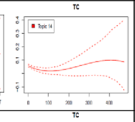
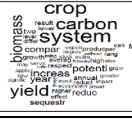
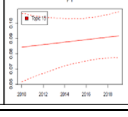
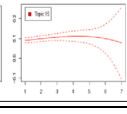
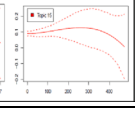
Topic #	*Topic estimation & interpretation*				*Partial effect summary*			ASIC Knowledge Classification Orientation	Regional emphasis	*Community & Network*		
	Top 3 tokens	Word cloud	Topic interpretation	Relative prevalence rank	PY	ecoserv.count	TC			Community membership	Topic Hub Score	Topic Betweenness Score
6	field, control, pest		Effects of organic farming practices and landscape complexity on natural pest control in crop production, including the potential negative and positive effects and the role of scale and spatial context.	12				Environmental / Agricultural Sciences	No distinction	A	0.7479	0.0000
7	servic, ecosystem, urban		Influence of landscape conversion (e.g. urban development, changes in cultural landscapes in protected areas and dam placement in river basins) on ES in non-agricultural landscapes, focusing on the importance of comprehensive spatial mapping/planning and the need for holistic and integrated management approaches.	2				Social / Decision Sciences	Europe / North America	B	0.0196	0.1465
8	landscap, habitat, function		Impact of agricultural intensification and related landscape changes on ecosystemic functions that support biodiversity through habitat loss and fragmentation, and the potential for sustainable management practices (e.g. fence construction, conservation buffer zones, and ecological corridors) to enhance protection and connectivity of biodiversity at appropriate spatial scales.	8				Social / Decision Sciences	Asia, Africa	A	0.8329	0.0733
9	farm, farmer, agricultur		Challenges to the adoption of sustainable agricultural management practices and potential solutions, including infrastructure and policy improvements, addressing intellectual property rights, and enhancing local/cultural knowledge sharing and linkages between research, extension services, and farmers through participatory approaches and supportive policies and institutions.	1				No distinction	Europe / North America	C	0.0030	0.0000
10	model, river, use		Development and use of analytical methods and frameworks for assessment of ES in agricultural and land use management, with a focus on the hydrological cycle and water regulation in terrestrial river basins.	5				Social / Decision Sciences	Asia, LATAM & Carib., Africa	B	0.0034	0.0000

Table 3
Detailed topic scorecard for interpretation, estimation results, community membership & connectivity metrics (Topics 11 through 15).

Topic #	*Topic estimation & interpretation*				*Partial effect summary*			ASIC Knowledge Classification Orientation	Regional emphasis	*Community & Network*		
	Top 3 tokens	Word cloud	Topic interpretation	Relative prevalence rank	PY	ecoserv.count	TC			Community membership	Topic Hub Score	Topic Betweenness Score
11	speci, tree, plant		Inventory of environmental functions and services provided by agroforestry and forest landscapes in various regions around the world.	4				Environmental / Agricultural Sciences	No distinction	A	1.0000	0.1886
12	chang, land, climat		Comparative impacts of land use and climate change on ES and the use of integrated modeling frameworks and policies that support multifunctional roles of agriculture and forests to promote sustainable development and adaptation.	3				Social / Decision Sciences	No distinction	C	0.0649	0.2985
13	cover, land, area		Soil erosion and weathering process as prevalent environmental problems threatening sustainable development and the various factors and processes that contribute to it, as well as potential mitigation strategies including vegetation heterogeneity, ecological restoration, and soil management practices.	7				No distinction	No distinction	C	0.0703	0.2637
14	restor, plant, veget		Management practices oriented towards restoration and maintenance of ES across heterogenous ecosystems (e.g. coastal dunes, wetlands, forests, agricultural lands).	10				No distinction	Asia, LATAM & Carib., Africa	A	0.1469	0.1960
15	product, system, crop		Management practices targeted at enhancing soil-related ES in agricultural and agroforestry systems, as well as potential benefits and factors affecting the response of crop yields to these practices.	6				Environmental / Agricultural Sciences	No distinction	D	0.0058	0.0000

implications of agricultural landscape management on Regulation/Maintenance ES in terrestrial agroecosystems would become more apparent.

Another central topic in this community is Topic 12, an increasingly prevalent topic with the second highest value for betweenness scores. It delves into the comparative impacts of land use and climate change on ES. It argues for using integrated models and policies to support sustainable development and adaptation in agricultural and agroforestry landscapes. Further, it is also notably positioned in the discussion within this community, as it is commonly observed in papers that immediately address neighboring topics.

One salient example of this is Topic 9, as both stand out among the most prevalent. However, despite being the single most frequent one, Topic 9 only exhibits a direct link to Topic 12, which is a second-order neighbor to all other topics in the community. This is further corroborated by the low (hub and betweenness) centrality measures for this topic. Put together. These facts remark its relative dis-connectedness from the rest of the discussion. In layman's terms, our results suggest that socioeconomic/institutional perspectives on sustainable management are often showcased in publications that tackle comparative assessment of land use and climate change in the context of this body of literature. Yet, these perspectives are relatively absent from other thematical content in this field of knowledge where Topic 12 is not explicitly boarded. Also, a clear gap within this community can be observed in the lack of direct connection between Topic 9 and Topic 13 (Fig. 6). This implies there is potential to link further discussions on

Table 4
Member topic interpretation and co-occurrence with top ES classes for Community A.

Community A:									
Implications of agricultural/landscape management on Regulation/Maintenance ES in terrestrial agroecosystems.									
Topic ID	Top 3 tokens	Summarized interpretation	Top co-occuring Regulation/ Maintenance ES	Time trend	Knowledge field orientation	Regional emphasis	Prevalence rank	Topic Hub Score	Topic Betweenness Score
1	crop, pollin, servic	Impact of the spatio-temporal variability of pollination, seed dispersal and other related ES on crop production/productivity and maintenance of diverse plant communities.	Pollination and seed dispersal Maintenance of biodiversity, including genetic diversity Pest and disease control, including invasive species Biological control	Increasing	Environmental/ Agricultural Sciences	No distinction	14	0.623	0.000
6	field, control, pest	Effects of organic farming practices and landscape complexity on natural pest control in crop production, including the potential negative and positive effects and the role of scale and spatial context.	Pollination and seed dispersal Pest and disease control, including invasive species Maintenance of biodiversity, including genetic diversity Biological control	Stable	Environmental/ Agricultural Sciences	No distinction	12	0.748	0.000
8	landscap, habitat, function	Impact of agricultural intensification and related landscape changes on ecosystemic functions that support biodiversity through habitat loss and fragmentation, and the potential for sustainable management practices (e.g. fence construction, conservation buffer zones, and ecological corridors) to enhance protection and connectivity of biodiversity at appropriate spatial scales.	Maintenance of biodiversity, including genetic diversity Pollination and seed dispersal Pest and disease control, including invasive species Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection	Stable	Social/Decision Sciences	Asia, Africa	8	0.833	0.073
11	speci, tree, plant	Inventory of environmental functions and services provided by agroforestry and forest landscapes in various regions around the world.	Maintenance of biodiversity, including genetic diversity Pollination and seed dispersal Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Pest and disease control, including invasive species	Decreasing	Environmental/ Agricultural Sciences	No distinction	4	1.000	0.189
14	restor, plant, veget	Management practices oriented towards restoration and maintenance of ES across heterogenous ecosystems (e.g. coastal dunes, wetlands, forests, agricultural lands).	Maintenance of biodiversity, including genetic diversity Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Water regulation Filtration/ sequestration/ storage/accumulation by micro-organisms, algae, plants	Decreasing	No distinction	Asia, LATAM & Carib., Africa	10	0.147	0.196

sociocultural/economic factors within sustainable adoption of management practices in papers that address erosion/weathering processes affecting Regulation/Maintenance ES related to the hydrological cycle.

These insights align with what both Setten et al. (2012) and Castro-Díaz et al. (2022) have remarked. The latter explicitly suggests the pressing need for context-explicit diverse transdisciplinary research and participatory approaches to study ES in Latin America. However, our results shed light beyond this by exposing that the past decade of research showed these traits could also apply to other contexts (such as Asian and African case studies). Moreover, a further gap comes from the

Table 5
Member topic interpretation and co-occurrence with top ES classes for Community C.

Community C: *Socioeconomic and sustainability concerns in agricultural/landscape management.*									
Topic ID	Top 3 tokens	Summarized interpretation	Top co-occuring Regulation/Maintenance ES	Time trend	Knowledge field orientation	Regional emphasis	Prevalence rank	Topic Hub Score	Topic Betweenness Score
5	forest, area, tropic	Tradeoffs and synergies between related ES in both complex and homogenous agroforestry systems.	Maintenance of biodiversity, including genetic diversity Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Pollination and seed dispersal	Stable	No distinction	No distinction	15	0.604	0.357
9	farm, farmer, agricultur	Challenges to the adoption of sustainable agricultural management practices and potential solutions, including infrastructure and policy improvements, addressing intellectual property rights, and enhancing local/cultural knowledge sharing and linkages between research, extension services, and farmers through participatory approaches and supportive policies and institutions.	Maintenance of biodiversity, including genetic diversity Water regulation Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Maintenece of soil quality: Weathering processes, decomposition, fixing and other processes	Decreasing	No distinction	Europe/ North America	1	0.010	0.000
12	chang, land, climat	Comparative impacts of land use and climate change on ES and the use of integrated modeling frameworks and policies that support multifunctional roles of agriculture and forests to promote sustainable development and adaptation.	Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Water regulation Maintenance of biodiversity, including genetic diversity Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants	Increasing	Social/ Decision Sciences	No distinction	3	0.065	0.299
13	cover, land, area	Soil erosion and weathering process as prevalent environmental problems threatening sustainable development and the various factors and processes that contribute to it, as well as potential mitigation strategies including vegetation heterogeneity, ecological restoration, and soil management practices.	Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Maintenance of biodiversity, including genetic diversity Water regulation Maintenece of soil quality: Weathering processes, decomposition, fixing and other processes	Increasing	No distinction	No distinction	7	0.070	0.264

link between Topic 13 and Topic 3 (from Community D). That is, combined with Topic 13′s increasing prevalence. There is potential to extend the integration of the latter discussion into the argument surrounding the effects of management practices on soil-specific ES not usually directly linked to climate change (e.g., maintenance of soil formation, structure, and fertility).

The third topic cluster, Community B, is composed of topics 2, 4, 7, and 10 (Table 6). Average rank prevalence here results from a mix of high/low ranked topics. Most temporal trends are either decreasing or maintaining prevalence in the evaluated period, except for Topic 7 (landscape conversion impacts in predominantly non-agricultural landscapes), which is increasingly more present in European & North American case studies. Another community member is Topic 4 (management options for water resources in coastal zones and terrestrial water bodies), which importantly shares an edge with all other topics in this community.

A common trait is that all member topics in this community are more prevalently conducted from a Social/Decision Science perspective. Moreover, estimated hub scores, in this case, suggest that there is a relatively low degree of connectivity in this community with the overall body of literature. However, Topic 2 (economic valuation of ES across distinct landscape configurations) and Topic 7 have the highest betweenness scores in the community, as they connect nodes with

Table 6
Member topic interpretation and co-occurrence with top ES classes for Community B.

Community B: *Socio-environmental analysis of ES management in terrestrial water bodies and landscapes.*									
Topic ID	Top 3 tokens	Summarized interpretation	Top co-occuring Regulation/ Maintenance ES	Time trend	Knowledge field orientation	Regional emphasis	Prevalence rank	Topic Hub Score	Topic Betweenness Score
2	valu, wetland, econom	Economic value of ES (e.g. pollination and those provided by wetlands and other terrestrial water bodies) and its assessment via of market-based and preference elicitation methods to inform policy planning related to landscape conversion.	Water regulation Hydrological cycle and water flow maintenance, including regulation of floods/ droughts and coastal protection Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants Maintenance of biodiversity, including genetic diversity	Decreasing	Social/ Decision Sciences	LATAM & Carib.	13	0.017	0.185
4	water, qualiti, irrig	Management options for water resources in coastal zones and terrestrial water bodies, including the use of hydraulic structures, the extraction of surface and groundwater for irrigation, as well as the assessment of the sustainability of agriculture through the use of tools such as total water footprint assessment.	Water regulation Hydrological cycle and water flow maintenance, including regulation of floods/ droughts and coastal protection Maintenece of soil quality: Weathering processes, decomposition, fixing and other processes Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants	Decreasing	Social/ Decision Sciences	No distinction	11	0.006	0.027
7	servic, ecosystem, urban	Influence of landscape conversion (e.g. urban development, changes in cultural landscapes in protected areas and dam placement in river basins) on ES in non-agricultural landscapes, focusing on the importance of comprehensive spatial mapping/planning and the need for holistic and integrated management approaches.	Maintenance of biodiversity, including genetic diversity Water regulation Hydrological cycle and water flow maintenance, including regulation of floods/ droughts and coastal protection Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants	Increasing	Social/ Decision Sciences	Europe/ North America	2	0.020	0.147
10	model, river, use	Development and use of analytical methods and frameworks for assessment of ES in agricultural and land use management, with a focus on the hydrological cycle and water regulation in terrestrial river basins.	Hydrological cycle and water flow maintenance, including regulation of floods/ droughts and coastal protection Water regulation Maintenance of biodiversity, including genetic diversity Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants	Stable	Social/ Decision Sciences	Asia, LATAM & Carib., Africa	5	0.003	0.000

communities A and C, respectively. Regarding regional prevalence, Topic 2 is skewed towards LATAM & Caribbean case studies, not unlike Topic 10 (development and use of analytical methods and frameworks for ES assessment). This skews towards Asian and African contexts.

These two topics' disposition within the identified community structure's overall body also reveals several interesting patterns. Perevochtchikova et al. (2021)showcase how economic valuations in Latin America are directed toward setting schemes for Payments for Ecosystem Services (PES). PES are primarily oriented towards conservation goals, hydrological modalities, and forestry systems and are dominated by traditional disciplinary approaches. We can see this with Topic 2 (the only topic explicitly focused on the valuation of ES). Its lack of hub centrality and connectedness with Topic 14 from community A suggests some prevalence in studies addressing valuation aspects of restoration and maintenance practices linked to agroforestry systems. Yet, such approaches are remarkably absent from other thematical content in other communities.

Topic 2′s direct edges reinforce the latter arguments with Topic 7

(assessments in predominantly non-agricultural landscapes) and Topic 4 (wetlands and other terrestrial hydrological systems) within this community. However, the direct edges between topics 2, 4, and 10 indicate that the integration of economic valuation aspects with comprehensive modeling/assessment frameworks (e.g., InVEST, OpenNESS) are readily jointly integrated into the discourse surrounding terrestrial hydrological systems. Even so, as earlier remarked, the lack of a direct link with Topic 9 (from Community C) indicates the pervasive need to include socio-cultural, economic, and institutional aspects at the center of these discourse clusters. Moreover, even more, the lack of this kind of holistic structure within the other topic communities is remarkable, which, as remarked by the authors mentioned above, can also be considered a salient research priority. Yet, our results highlight the specificities of the gap structure that might help guide such efforts. Based on these insights, topic community B can be comparably considered a niche body of literature: *Socio-environmental analysis of ES management in continental water bodies and terrestrial landscapes*.

Lastly, Table 7 summarizes the last topic cluster, Community D. The overall discussion body can be characterized by the interconnection of environmental health with the productivity and profitability of agro-forestry systems. In other words, this cluster includes papers mainly concerned with managing inputs into agricultural soils from an Agricultural/Environmental Sciences perspective. However, the problems analyzed in these studies not only need consideration of the impacts on the environmental integrity of the ecosystems but could also benefit from increasingly prevalent insights into the profitability and the welfare of agents involved in the management cycle. Therefore, there is an opportunity to produce more studies related to those problems from a Social/Decision Sciences perspective. The latter is a substantial gap that should be addressed to promote a holistic expansion of the knowledge frontier derived from these topics.

5. Conclusions

This paper presents a novel framework for analyzing the scientific knowledge concerning linkages between terrestrial ES in landscape and agricultural management, presented by estimated thematic communities. To this end, we have expanded the approaches put forward by authors such as Cheng et al. (2018) and Chen et al. (2020) through the novel systematic application of methods to generate this kind of understanding within our field of interest based on bibliometric approaches. In particular, using an STM is a step forward in providing a potentially more accurate alternative to the now standard (unconditional) TM that offers the possibility of conditioning topical prevalence on metadata covariates.

Another strength of our approach is the use of data-driven methods to determine the number of estimated topics and the use of data-driven clustering approaches on the derived topic network structure that results from using a correlated topical prevalence model structure to infer meaningful topic communities. This can be further exploited to identify gaps in the existing literature by inspecting key missing links within or between topic communities and exploring existing connections between communities. These characteristics cannot be understated in their potential to provide additional criteria to assist topical interpretation and contextualization of a field of interest. Also, while topical interpretation is an inherently subjective undertaking, we argue that using this kind of systematic approach can assist in reducing the degree of subjectivity that befalls in providing a comprehensive portrait of the state of the art of our field, which may complement existing traditional literature surveys. We thus advocate that the resulting structure can be useful to determine gaps for future research on the subject and other related fields of knowledge. We believe that subsequent applications of this methodology will allow us to systematically define the research priorities for our field of interest by quantitatively identifying research hubs and the

Table 7
Member topic interpretation and co-occurrence with top ES classes for Community D.

Community D:									
Soil management and crop production.									
Topic ID	Top 3 tokens	Summarized interpretation	Top co-occuring Regulation/ Maintenance ES	Time trend	Knowledge field orientation	Regional emphasis	Prevalence rank	Topic Hub Score	Topic Betweenness Score
3	soil, organ, nutrient	Relevance, measurement and physico-chemical/biological factors that influence soil health in soil quality assessment.	Maintenece of soil quality: Weathering processes, decomposition, fixing and other processes Soil fomation, structure and fertility Water regulation Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants	Stable	Environmental/ Agricultural Sciences	No distinction	9	0.010	0.143
15	product, system, crop	Management practices targeted at enhancing soil-related ES in agricultural and agroforestry systems, as well as potential benefits and factors affecting the response of crop yields to these practices.	Filtration/ sequestration/storage/ accumulation by micro-organisms, algae, plants Maintenece of soil quality: Weathering processes, decomposition, fixing and other processes Hydrological cycle and water flow maintenance, including regulation of floods/droughts and coastal protection Maintenance of biodiversity, including genetic diversity	Increasing	Environmental/ Agricultural Sciences	No distinction	6	0.006	0.000

corresponding gaps between them, research project portfolios, and optimization of the financial and human resources aimed at these activities.

Nonetheless, some pitfalls still need to be addressed to enhance applications of this type of toolset to map the structure of the scientific discourse in future applications. One such caveat was recognized by Cheng et al. (2018), who stated that the introduction of multilingual literature to the document corpus was hindered due to technical limitations in automatic translation processes. Furthermore, while our current approach relies on standardized, curated academic repositories such as ISI-WOS and Scopus databases as a basis of our bibliometric dataset, we concur with their insights into the limitations that this imposes to incorporate findings and recommendations from specific countries, territories, or regions (also remarked by Castro-Díaz et al. (2022) specifically for the field of ES). Lastly, Westgate et al. (2015) remark that although strategies for gap identification based on existing topical structures have the potential to unveil gaps in the current scientific discourse in a field of knowledge, the possibility that progress might also occur through spontaneous novel insights escapes this strategy. However, as they remark, this kind of insights are inherently less amenable to predictive methods and thus are less likely to be identified by approaches such as the one in this paper.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gonzalo Villa-Cox reports financial support was provided by SENESCYT, Ecuador.

Data availability

Data will be made available on request.

Acknowledgments

Gonzalo Villa-Cox was sponsored by Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT), Ecuador, during the development of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoser.2023.101525.

References

Ash, N., Blanco, H., Garcia, K., Brown, C., 2010. Ecosystems and human well-being: a manual for assessment practitioners. Island Press.

Bonacich, P., Lu, P., 2012. Introduction to mathematical sociology. Princeton University Press.

Castro-Díaz, R., Delgado, L.E., Langle-Flores, A., Perevochtchikova, M., Marín, V.H., 2022. A systematic review of social participation in ecosystem services studies in Latin America from a transdisciplinary perspective, 1996–2020. Science of The Total Environment 828, 154523.

Chen, W., Geng, Y., Zhong, S., Zhuang, M., Pan, H., 2020. A bibliometric analysis of ecosystem services evaluation from 1997 to 2016. Environmental Science and Pollution Research 27 (19), 23503–23513.

Cheng, X., Shuai, C., Liu, J., Wang, J., Liu, Y., Li, W., Shuai, J., 2018. Topic modelling of ecology, environment and poverty nexus: An integrated framework. Agriculture, Ecosystems & Environment 267, 1–14.

Clauset, A., Newman, M.E., Moore, C., 2004. Finding community structure in very large networks. Physical Review E 70 (6), 66111.

Dale, V.H., Polasky, S., 2007. Measures of the effects of agricultural practices on ecosystem services. Ecological Economics 64 (2), 286–296. https://doi.org/10.1016/j.ecolecon.2007.05.009.

Fisher, B., Turner, R.K., Morling, P., 2009. Defining and classifying ecosystem services for decision making. Ecological Economics 68 (3), 643–653.

Gatti, C.J., Brooks, J.D., Nurre, S.G., 2015. A historical analysis of the field of OR/MS using topic models. ArXiv Preprint. ArXiv:1510.05154.

Haines-Young, R., Potschin-Young, M., 2018. Revision of the common international classification for ecosystem services (CICES V5. 1): a policy brief. One Ecosystem 3, e27108.

Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the lasso. The Annals of Statistics 34 (3), 1436–1462.

Meraj, G., Singh, S.K., Kanga, S., Islam, M., et al., 2022. Modeling on comparison of ecosystem services concepts, tools, methods and their ecological-economic implications: A review. Modeling Earth Systems and Environment 8 (1), 15–34.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 69 (2 2), 26113. https://doi.org/10.1103/PhysRevE.69.026113.

Perevochtchikova, M., Castro-Díaz, R., Langle-Flores, A., Von Thaden Ugalde, J.J., 2021. A systematic review of scientific publications on the effects of payments for ecosystem services in Latin America, 2000–2020. Ecosystem Services 49, 101270.

Reid, W. V, Mooney, H. A., Cropper, A., Capistrano, D., Carpenter, S. R., Chopra, K., Dasgupta, P., Dietz, T., Duraiappah, A. K., Hassan, R., & others. (2005). *Ecosystems and human well-being-Synthesis: A report of the Millennium Ecosystem Assessment*. Island Press.

Roberts, M.E., Stewart, B.M., Tingley, D., Airoldi, E.M., et al., 2013. The structural topic model and applied social science. Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation 4, 1–20.

Roberts, M.E., Stewart, B.M., Airoldi, E.M., 2016. A Model of Text for Experimentation in the Social Sciences. Journal of the American Statistical Association 111 (515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684.

Roberts, M.E., Brandon, S.M., Dustin, T., 2019. stm: R Package for Structural Topic Models. Journal of Statistical Software VV(Ii), 42.

Seppelt, R., Dormann, C.F., Eppink, F.V., Lautenbach, S., Schmidt, S., 2011. A quantitative review of ecosystem service studies: Approaches, shortcomings and the road ahead. Journal of Applied Ecology 48 (3), 630–636. https://doi.org/10.1111/j.1365-2664.2010.01952.x.

Setten, G., Stenseke, M., Moen, J., 2012. Ecosystem services and landscape management: three challenges and one plea. International Journal of Biodiversity Science, Ecosystem Services & Management 8 (4), 305–312.

Syed, S., & Spruit, M. (2018). Selecting priors for latent Dirichlet allocation. *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, 194–202.

Syed, S., Spruit, M., 2017. Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. IEEE International Conference on Data Science and Advanced Analytics (DSAA) 2017, 165–174. https://doi.org/10.1109/DSAA.2017.61.

Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human Domination of Earth's Ecosystems. Science 277 (5325), 494–499. https://doi.org/10.1126/science.277.5325.494.

Wang, B., Pan, S.-Y., Ke, R.-Y., Wang, K., Wei, Y.-M., 2014. An overview of climate change vulnerability: a bibliometric analysis based on Web of Science database. Natural Hazards 74 (3), 1649–1666.

Westgate, M.J., Barton, P.S., Pierson, J.C., Lindenmayer, D.B., 2015. Text analysis tools for identification of emerging topics and research gaps in conservation science. Conservation Biology 29 (6), 1606–1614.

Zhang, W., Ricketts, T.H., Kremen, C., Carney, K., Swinton, S.M., 2007. Ecosystem services and dis-services to agriculture. Ecological Economics 64 (2), 253–260. https://doi.org/10.1016/j.ecolecon.2007.02.024.

Zhao, T., Liu, H., Roeder, K., Lafferty, J., Wasserman, L., 2012. The huge package for high-dimensional undirected graph estimation in {R}. Journal of Machine Learning Research 13, 1059–1062.