Map of Science with Topic Modeling: Comparison of Unsupervised Learning and Human-Assigned Subject Classification

Arho Suominen
Innovation, Policy & Economy, VTT Technical Research Centre of Finland, P.O.Box 1000, Espoo 02044, Finland. E-mail: arho.suominen@vtt.fi

Hannes Toivanen
Innovation, Policy & Economy, VTT Technical Research Centre of Finland, P.O.Box 1000, Espoo 02044, Finland and School of Business, Lappeenranta University of Technology, P.O.Box 20, Lappeenranta 53851, Finland. E-mail: hannes.toivanen@vtt.fi

The delineation of coordinates is fundamental for the cartography of science, and accurate and credible classification of scientific knowledge presents a persistent challenge in this regard. We present a map of Finnish science based on unsupervised-learning classification, and discuss the advantages and disadvantages of this approach vis-à-vis those generated by human reasoning. We conclude that from theoretical and practical perspectives there exist several challenges for human reasoning-based classification frameworks of scientific knowledge, as they typically try to fit new-to-the-world knowledge into historical models of scientific knowledge, and cannot easily be deployed for new large-scale data sets. Automated classification schemes, in contrast, generate classification models only from the available text corpus, thereby identifying credibly novel bodies of knowledge. They also lend themselves to versatile large-scale data analysis, and enable a range of Big Data possibilities. However, we also argue that it is neither possible nor fruitful to declare one or another method a superior approach in terms of realism to classify scientific knowledge, and we believe that the merits of each approach are dependent on the practical objectives of analysis.

Introduction

A central challenge for the cartography of scientific knowledge is the creation of valid and accurate coordinates. This paper discusses the choice of the origin of coordinates in order to make a map of scientific knowledge, and, in particular, demonstrates the advantages of unsupervised learning-assigned coordinates over those created by human reasoning.

Human-assigned metadata, such as subject category classification of articles or journals, has been the dominant source of coordinates in science maps (even when cartographers have relied on cocitation information) (Börner, 2010). However, classification of scientific knowledge with such metadata is subject to several well-known weaknesses. Preexisting categories of science provide a finite definition of new knowledge, fitting knowledge that is by definition infinite and new to the world into preexisting categories and coordinates, for example, Small (2004). They are best at monitoring the behavior of known and defined bodies of knowledge, but lend themselves poorly—if at all—to correctly identifying the emergence of truly new epistemic bodies of knowledge. In short, human-assigned subject categories are akin to using a rearview mirror to predict where a fast-moving car is heading.

The literature on structuring science focuses on classification and mapping, which should not be considered synonymous (Klavans & Boyack, 2009). Classification of science—the process of separating science into different partitions—is a precondition of the existing mode of scientific dialog. The need to define research fields and to assign journals and publications to them stems from the need to create an information retrieval system that would help scholars find relevant information. As Glänzel and Schubert (2003) argue, correct classification of publications for scientific fields is a necessity for credible scientometrics.

At the core of this article is the problem that scientometrics is applied as a system of measurement (of scientific

Received April 16, 2015; revised May 19, 2015; accepted June 8, 2015

© 2015 The Authors. Journal of the Association for Information Science and Technology published by Wiley Periodicals, Inc. on behalf of ASIS&T • Published online 23 September 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23596

knowledge) while its standard measures—subject classification of scientific publications—remain problematic. Unlike agreed and internationally standardized measures, such as the metric system, several challenges remain for the classification of scientific knowledge into subject areas, the fundamental one being that there is no clear, readily available, and agreeable solution: knowledge that is new to the world cannot, and as we argue, should not be fitted into historical classification schemes. Another matter altogether is that science is supposed to create a mirror of the natural world: Human classification of scientific knowledge is used to make scientists' observations and claims about the natural world comprehensible and communicable to other humans. It remains debatable how adequately such language and classifications mirror the actual natural world.

A more practical challenge for classification of scientific knowledge with human-assigned metadata is its character as a social convention and routine, which increasingly renders it a historically derived classification system. Starting with the division and formation of new faculties from theological ones at medieval universities, most scientific classification systems used today are historically derived, and their evolution has been characterized by revision and adaptation as excessively large conceptual challenges emerge.

A case in point is the Web of Science (ISI [Institute for Scientific Information; now Thomson Reuters]-WoS) journal- and article-level subject classification system, which is trusted and perhaps the most widely used in the world for mapping scientific knowledge. It is in fact a library classification scheme, devised for efficient retrieval of information rather than to establish valid and credible coordinates for mapping scientific knowledge. Indeed, this and other information retrieval-based classification systems lack the capability to produce consensus measures for scientometric studies (Glänzel & Schubert, 2003).

Recently, computing advances have made text-mining techniques available that offer new approaches to defining coordinates for science maps. Data mining, and specifically text mining, opens new avenues for unsupervised or semisupervised classification methods, as it classifies scientific text based on content, foregoing human-given labels. Indeed, text-mining methods are promising tools in classifying fields of science (Glenisson, Glänzel, & Persson, 2005) and are currently undergoing rapid development. In practice, text mining seeks to identify words or phrases that explain possible underlying structures and relationships in the data uncovered through distribution analysis, association rules, or different clustering approaches (Feldman & Sanger, 2006).

One promising unsupervised classification method is topic modeling, which Blei, Ng, and Jordan (2003) suggested to be a useful tool for extracting information from textual data and later shown by Wei and Croft (2006) to outperform more traditional methods. To date, the literature on topic models has focused on the computer science aspect of the methods; their applicability to the scientometric field has only been established recently (e.g., Yau, Porter, Newman, & Suominen, 2013). One significant remaining challenge in applying unsupervised learning more broadly is automatically labeling the classifications that emerge in such a way that the results are easily interpreted (Mei, Shen, & Zhai, 2007).

These new methods, such as topic modeling, enable the definition of coordinates for science maps by generating classification categories of scientific knowledge via unsupervised learning methods. Aided by computing algorithms, scientific texts are essentially classified into subject categories that best comprise the variety of epistemic bodies addressed or advanced within the corpus. The subject categories are retrieved based on what is discussed in scientific publications, and no effort is made to force those publications into preexisting categories. The central novelty of unsupervised-learning methods in classifying scientific knowledge is that they virtually eliminate the need to fit new-to-the-world knowledge into known-to-the-world definitions.

Motivated by the possibilities of unsupervised classification and the challenges of existing classification methods, we analyze science publications with topic modeling, showing an example of unsupervised classification. We analyze whether visualizing the semantic space of a scientific publication corpus creates a meaningful classification, which can provide real value for research management. We also look at how existing high-level tree of knowledge-type classifications can be linked to semantic classifications to create user-friendly labeling for the topical space. Our objective is to uncover how unsupervised classification methods can be used to classify scientific documents, and if by overlaying tree of knowledge-type classifications (ISI-WoS) of science onto unsupervised learning results, we are able to uncover meaningful connections between the two.

To move beyond the existing classification systems, we propose modeling the relationship of science documents via the unsupervised classification of text. Taking advantage of the semantic relationship of documents, we map semantic classes with existing Organisation for Economic Co-operation and Development (OECD) classifications (OECD, 2007), which are an aggregate created from ISI-WoS article-level classifications, to create a map of Finnish science from 1995 to 2011. In addition, we overlay topics with indicators of relative size and growth in the national research landscape. Finally, we discuss the consistency of the results in light of the research trends in Finland and globally.

The paper is organized as follows: The following section explores the background of science classification and mapping, as well as the application of unsupervised learning to science classification. The section, Data and Preprocessing presents our methods, followed by a section describing the results. The paper concludes with a Discussion section.

Background

Mapping Scientific Publications

Science maps define and visualize elements in science to enable an orientation within the knowledge landscape, usually by deriving coordinates by classifying disciplines, fields, and subjects into themes or topics. By definition, a map positions entities in relation to other elements on the map based on a measured distance, usually the physical distance of the entities. Commonly used distance measures for science maps are cocitation of articles, coclassification of articles, cocitation of journals, cocitation of authors, and use of common words (coword analysis) (e.g., Bassecoulard & Zitt, 1999; Börner, Chen, & Boyack, 2003; Boyack, Klavans, & Börner, 2005; Braam, Moed, & Van Raan, 1991; Griffith, Small, Stonehill, & Dey, 1974; Leydesdorff & Rafols, 2009; Moya-Anegón et al., 2004; Small, 1973). All of the aforementioned methods approach science publications by dimension reduction. Scholars try to reduce the high-dimensional space of semantic text into a space with fewer dimensions, extracting only some of the features of a publication. The selected features, such as whether two papers have been cocited, are then used as a basis for creating a measure of distance between two elements.

As Glänzel et al. have argued: "The delimitation of scientific subfields is one of the central issues in both information science and bibliometric research. The classification of scientific literature into appropriate subject fields is, moreover, one of the basic preconditions of valid bibliometric analysis" (Glänzel, Schubert, & Czerwon, 1999, p. 427). The established practice for structuring scientific publications has been to rely on meta-information embedded in scientific publictions, usually describing the context and nature of a publication, as well as documenting its authorship (people, institutions, countries) and other information. Starting from "statistical bibliography" or even older studies using, for example, citation indexes, students of scientometrics have used metadata, such as publication year, citations, human-given subject categories, and keywords, to analyze and map scientific publications. For a historical review, refer to Hood and Wilson (2001).

From early on, the actual text embedded in documents was also of interest, but the lack of computational power made large-scale analysis of semantic text a challenge and the use of metadata offered an easy proxy measure to reduce the dimensionality of science. Frequently used methods include human-assigned subject categories, analyzing citation chains or genealogies, and coword analysis of keywords.

Journal subject categories and different article-level classifications might be the best-known methods to reduce the dimensionality of science. Starting from the journal maps used by Bassecoulard and Zitt (1999) and Boyack et al. (2005), using journal disciplinary categories to reduce dimensionality has been an interesting avenue of research (e.g., Leydesdorff, Carley, & Rafols, 2013; Leydesdorff & Rafols, 2009; Moya-Anegón et al., 2004; Moya-Anegón et al., 2007; Zhang, Liu, Janssens, Liang, & Glänzel, 2010). However, the use of such subject categories has also been debated intensively (Toivanen & Suominen, 2014).

As Glänzel et al. (1999) argue, much of the macro-level research on science has relied on the classification of journals into specific subject categories, but the ability of subject categories to describe the content of documents is limited. Similarly, Leydesdorff and Rafols (2009) have called into question the usability of ISI subject categories. The major problem here is that by giving subject categories to journals or articles and then using that as a proxy for the content, we have created a crude measure of science (Pudovkin & Garfield, 2002).

In addition to journal categories, science has been mapped with cocitations. Small (1973), Griffith et al. (1974), Small (1993), and Small (1999) clustered science onto macro-level maps with cocitations, assuming the shared intellectual focus, made explicit through common citations, sufficiently enables dimension reduction though statistical methods, for example, looking at citation chains and collapsing clusters of publications connected by citations, and then clustered with statistical methods. The accuracy of cocitations can be enhanced with the use of metadata and in particular the use of word-profiles. Braam et al. (1991, p. 252) argue that "[p]ublication groups identified by cocitation clusters from different years that have high word-profile similarity, can be considered approximately as different phases of the same specialty" To an extent, this cocitation analysis approach, in cases enhanced with metadata or lexical analysis, is, among subject categories, one of the dominant approaches to reducing the dimensionality of a science database.

Words in the publication have also been used for dimensionality reduction. Most prominently, coword approaches—studies using the co-occurrence of words—have been proposed as an alternative method to cocitations. Among others, Rip and Courtial (1984), Leydesdroff (1989), Peters and van Raan (1993b), Peters and van Raan (1993a), and Ding, Chowdhury, and Foo (2001) have used the co-occurrence of words in keywords, titles, or abstracts to create maps of a given science area. Coword approaches work with a matrix representation of co-occurrence, which is statistically modeled. For a practical analysis, the number of terms remains limited, as modeling a large co-occurrence matrix is challenging even with current statistical analysis programs, thus excluding significant human judgment in selecting terms to be included in the analysis and forcing the publication into a category based on a relatively low dimensional representation of the actual content.

Developments in data or text mining and information retrieval have made taking advantage of semantic text a practical approach. Glenisson et al. (2005) argued that text mining methods show promise as a valuable tool in mapping fields of science. In text mining, we seek to identify words or phrases that could explain possible underlying content and structures in the data based on co-occurrence data analyzed by distribution analysis, association rules, or different clustering approaches (Feldman & Sanger, 2006). Novel text-mining methods create value by being able to create practical categories directly from semantic text, rather than using preordained categories, keywords, or citations. Among the plethora of methods for clustering semantic text, Hofmann (1999) suggested that topic modeling would be a useful tool for creating text-based clusters from large textual collections. Wei and Croft (2006) later confirmed that topic models outperform more traditional, cluster-based approaches.

In this paper, we use topic modeling to reduce the dimensionality of Finnish science to clusters of latent topical areas. Our objective is to uncover transitions and diversity in Finnish science. By using topic modeling, we work around the challenges of human-given labeling and enable an unsupervised method to draw out latent topics based on the semantic text, excluding the meta-information embedded in each publication.

Unsupervised Learning and Topic Modeling

Unsupervised learning is a subfield of machine learning that produces an outcome based on an input with no feedback from its environment. While in supervised or reinforced learning processes the machine is directed to learn to produce the correct outcome, unsupervised learning relies on the formal framework that enables the algorithm to find patterns in the input that extend beyond noise.

The majority of unsupervised learning methods rely on a probabilistic model of the input data. The machine estimates a model that represents the probability distribution for an input based either on previous inputs or independently. Unsupervised models can be used, for example, to detect outliers or for classification tasks, as in topic modeling.

Latent Dirichlet allocation (LDA) is a topic model that draws out latent patterns in semantic text. Topic models refer to a number of different algorithms, where LDA is one of the best-known algorithms. In the field of information retrieval, researchers first proposed a probabilistic model, "probabilistic latent semantic indexing" (PLSI) (Hofmann, 1999), which is capable of drawing latent patterns in semantic text. pLSI models each word in a document as a sample from a mixture model, where the mixture components are multinomial random variables that can be viewed as representations of topics. Using a similar approach, but improving on the limitations of pLSI, Blei et al. (2003) suggested a method named latent dirichlet allocation (LDA). LDA is a three-layer Bayesian model that is now widely used in different applications, such as text mining, bioinformatics, and image processing.

In 2007, Blei and Lafferty showed the usability of topic models in modeling the structure of science. Blei and Lafferty (2007, p. 18) noted that topic models ". . . can extract surprisingly interpretable and useful structure without any explicit 'understanding' of the language by computer." The basic idea behind the model is that each document in a corpus is a random mixture over latent topics, and each latent topic is characterized by a distribution over words. In the LDA model, each document is a mixture of a number of topics based on the words attributable to each of the topics. LDA allows us to uncover these latent probability distributions based on the semantic text used in the document, thus classifying the documents based on the latent patterns within them. For a detailed explanation of the algorithm, refer to, for example, Blei and Lafferty (2009).

Recent studies have evaluated the usability of unsupervised learning in classifying science. For example, Talley et al. (2011) created a map of National Institutes of Health grants using topic modeling and a two-dimensional map created via graph-based clustering. Nichols (2014) analyzed the interdisciplinarity of National Science Foundation awards using topic modeling. Klavans and Boyack (2014) used website data to structure the thematic landscape of nonprofit organizations. Boyack et al. (2011) analyzed 2.15 million MEDLINE publications comparing nine different clustering approaches. Liu et al. (2010) used a hybrid approach, integrating semantic and citation analysis to clustering, creating a visualization and ranking journals using the PageRank algorithm. Yan, Ding, Milojević, and Sugimoto (2012) analyzed science publications, drawing out dynamic changes in science communities. Yau et al. (2013) analyzed the precision and recall of topic modeling using expert opinion created clusters.

Data and Methods

We follow a research design where we first preprocess existing WoS raw data, then use LDA to abstract text to draw latent patterns from the data. Finally, we incorporate the LDA results into the already existing metadata for analysis and visualization. This research design is described in detail in the following subsections and illustrated in Figure 1.

Data and Preprocessing

We use a bibliometric data set for the years 1995–2011 obtained from the ISI-WoS where at least one of the publication authors has Finnish affiliation. The usual caveats associated with using ISI-WoS data apply; most important in the Finnish case is the fact that the increasing indexing of conference proceedings and abstracts since late 1990 is responsible for much of the publication growth in the late 1990s.

The data were obtained via two methods. First, the data until 2010 were delivered as tagged XML data with the full article-level information as recorded in the ISI-WoS. Data delivery was done by Thomson Reuters in August 2012. These data were updated via web access to the WoS, by which means data for 2011 and 2012 were added to the first data set. The data were subsequently preprocessed with VantagePoint software. The time series of the data is described in Table 1. Final data were limited to articles, conference proceedings, abstracts, and reviews, and only include publications with an abstract. This totaled 144,081 records between 1995 and 2011.

The data set was extracted from the database to a comma-separated file including the unique identifier and abstract text. The unique identifier was extracted and used to enable subsequent data mining results to be merged with original

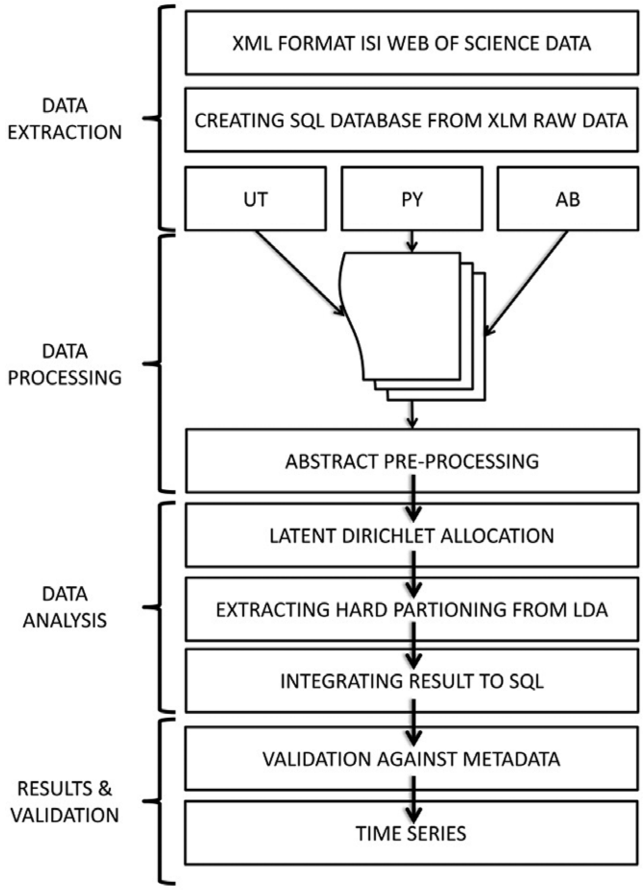
DATA EXTRACTION
XML FORMAT ISI WEB OF SCIENCE DATA
CREATING SQL DATABASE FROM XLM RAW DATA
UT
PY
AB
DATA PROCESSING
ABSTRACT PRE-PROCESSING
DATA ANALYSIS
LATENT DIRICHLET ALLOCATION
EXTRACTING HARD PARTIONING FROM LDA
INTEGRATING RESULT TO SQL
RESULTS & VALIDATION
VALIDATION AGAINST METADATA
TIME SERIES

FIG. 1. Graphical representation of the research design. Abbreviations: UT unique identifier of a record, PY publication year, and AB abstract.

TABLE 1. Time series of data used for the study.

Year	Records in raw data	Records after preprocessing
1995	3,330	3,318
1996	4,009	3,999
1997	4,951	4,933
1998	7,451	7,421
1999	7,576	7,543
2000	8,098	8,069
2010	8,249	8,231
2002	8,395	8,371
2003	8,788	8,769
2004	9,205	9,188
2005	9,327	9,312
2006	9,947	9,930
2007	10,452	10,430
2008	10,859	10,837
2009	11,312	11,292
2010	10,859	10,841
2011	11,273	10,172
TOTAL	144,081	142,656

Source: ISI-WoS; Authors.

TABLE 2. Setting for the LDA algorithm.

Variable	Value
The maximum number of iterations	20
The convergence criteria for variational inference	1e-6
The maximum number of iterations of variational EM	100
The convergence criteria for variational EM	1e-4
Alpha	Estimated
Topic initialization	Random seed

publication metadata. The abstract text was used as raw data for the analysis and served as a control variable.

Prior to analysis, the abstract texts were preprocessed with a Python script. The script first checks the data validity by, for example, checking that all the records truly have an abstract text. The data were also manipulated with the Python script to remove stopwords, tagged acknowledgment texts, punctuations, and terms containing numbers or consisting solely of numbers. Terms that occurred only once in the whole data set were also removed at this stage. After all the previously mentioned terms were removed, the text was tokenized and each token was transformed into a corresponding number to further reduce the complexity of the data. After all the cleaning processes, the data set comprised 142,656 records with 95,664 unique tokens.

Data Analysis

The practical LDA analysis was performed by implementing variational expectation-maximization (EM) for LDA by Blei et al. (2003). Table 2 presents the settings for the algorithm. We experimented with other settings, but after a qualitative estimation of the results we kept the values in Table 2.

Before applying LDA to the problem, we note the basic assumptions behind LDA and topic modeling. The assumptions include the following:

1. Documents are exchangeable in a corpus.
2. Words are exchangeable in a document ("bag-of-words" assumption, meaning no syntactic or proximity relationship information used).
3. A topic is modeled as a multinomial distribution of words from some basic vocabulary.
4. Words in a document arise from a number of latent topics.

The first two assumptions state that LDA ignores document order in a corpus and word order in a document. This assumption is made although models have been suggested that go beyond the "bag-of-words" assumption (Wallach, 2006). The third assumption defines the meaning of a topic, which is a distribution throughout a dictionary. For example, high probability words in a sports topic may be baseball, football, athlete, etc. Consequently, we usually use the top few words to represent a given topic. The last one assumes that each word has one or several corresponding latent topics with which it is associated, and given the particular topic, the word is drawing from that topic's distribution.

LDA has been shown to produce a good estimation of the latent pattern in a given corpus; however, there are two important limitations. First, it is difficult to measure the topic correlations between each of the topics. Second, LDA requires a fixed number of topics, which are usually unknown to the researcher. In this study we made efforts to control the influence of these limitations.

The correlation between topics, or research fields in this case, is important, but in this study we used LDA narrowly as a hard partitioning tool. By hard partitioning, we refer to the fact that even though LDA produces a probability distribution that gives a document a probability value to be classified in each of the topics, we only attributed the highest probability topic of a given publication, thus forcing each document to belong to only one topic. We used LDA as a method of creating mutually exclusive classes rather than creating a clusters of documents that are not mutually exclusive.

LDA also needs a fixed number of topics, requiring the researchers to have some idea of the possible bounds of latent features in the text. Some efforts were made to use metrics to estimate which number of topics would best fit the corpus used as data. For example, a perplexity value has been used to evaluate the performance of an LDA model with different K values (Blei et al., 2003), thus having an idea of what would be the best possible number of topics describing a given corpus. However, the algorithmic model fit can be of limited value when humans interpret the thematic area of each of the topics produced. In fact, "[t]raditional metrics are, indeed, negatively correlated with the measures of topic quality" (Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009, p. 8) and instead, the number of topics should be fitted with an evaluation of real-world performance. To estimate the real-world performance of different K values, we used a trial-and-error approach, where we tested several K values and evaluated each of the results based on the coherence of terminology in a specific topic. After a trial-and-error phase and the subsequent evaluation of quality, the number of topics used for the study was fixed to 60.

Probability distributions for words and documents were analyzed using R statistics. R was used to assign each document a topic based on the highest probability of the document belonging to a topic. The word probability distribution for each topic was visualized with a wordcloud through the R package wordcloud.

Using ISI-WoS article-level subject categories, we retrieved the OECD major classification (OECD, 2007) for each document. We then created an asymmetrical co-occurrence matrix of topics for OECD major classifications based on how topical hard classifications and OECD major classifications co-occur. The co-occurrence matrix was deconstructed into network data and imported into Gephi for visualization and further analysis. In the network data, the nodes were topics and OECD major classifications, while the edges were the linkages among them. The edge weight was defined as the number of co-occurrences and node weight as the number of documents classified into each topic and OECD major classification class. Gephi was used to visualize the results, using ForceAtlas 2 algorithm in LinLog mode. Finally, the network was clustered using the modularity algorithm by Blondel, Guillaume, Lambiotte, and Lefebvre (2008). This clustering process creates aggregate clusters of OECD classes and unsupervised learning topics.

To further analyze each of the topics, we calculated the overall growth percentage of topics and yearly Finnish publications overall. For each year we then compared whether a topic grew faster than science in Finland overall. This was aggregated to a growth index (GI) value, where for each year a topic was assigned a value of −1 for growth slower than science overall, 0 for equal growth, and 1 for growth faster or as fast as science overall. The yearly values were summed through the time series into an index value. Each topic can thus have a growth index ranging from −16, if the topic grew slower than science overall in each year of the analysis, to 16, if the topic grew faster than or as fast as science overall in each year of the analysis.

Results

For the 60 topics, we use the probability distribution of documents to topics to find the largest probability for each document belonging to a topic, and create a hard partitioning of documents. This process assigns each document to only one topic. We then use the hard partitioned documents with existing metadata on the publication year to create a time series of thematic topics in Finnish science. In order to secure a truly automated classification approach, we have not created human-assigned names for the 60 topics; rather, they are identified by number. However, the theme or topical area of each topic is described with a wordcloud drawn from all the abstracts assigned to a topic, provided in Appendix S1.

Map of Science

Table 3 depicts the classification of Finnish science into 60 topics. Using 60 topics, we are able to draw out relevant disciplinary areas for Finland that are sufficiently detailed to point toward meaningful areas of science.

Merging the hard classification topic assignment of each document with the OECD major classification of each publication, we create a network visualization to illustrate the thematic classes of the topics and validate the results. In practice, we created a network visualization of topics to OECD classifications seen in Figure 2, which depicts how topics created via unsupervised learning are related to historically created ISI-WoS article-level subject and OECD major classifications. As expected, many of our classifications adhere to the OECD classifications, but there are important differences and exceptions, as will be discussed here.

DOI: 10.1002/asi

TABLE 3. Descriptive statistics of the thematic topics created by the LDA algorithm. Abbreviations used are C for Count, G for Growth and GI for Growth Index.

Cluster	Count	C (avg.)	C (SD)	G (avg.)	G (SD)	GI
Topic 1	3,203	188.4	78.5	14%	25%	4
Topic 2	3,931	231.2	66.6	8%	12%	–4
Topic 3	1,071	63.0	19.5	16%	36%	2
Topic 4	1,751	103.0	23.9	8%	20%	–2
Topic 5	524	30.8	8.5	9%	39%	–4
Topic 6	1,843	108.4	17.9	4%	16%	–4
Topic 7	1,533	90.2	29.0	9%	22%	2
Topic 8	1,432	84.2	18.3	–2%	14%	–8
Topic 9	5,657	332.8	186.5	29%	48%	8
Topic 10	1,537	90.4	36.4	19%	43%	6
Topic 11	1,659	97.6	14.0	1%	19%	–6
Topic 12	2,372	139.5	40.0	10%	27%	–2
Topic 13	1,409	82.9	22.4	9%	14%	4
Topic 14	3,991	234.8	102.0	22%	47%	2
Topic 15	2,214	130.2	31.0	8%	20%	–2
Topic 16	3,065	180.3	70.0	17%	44%	2
Topic 17	1,443	84.9	15.8	2%	28%	–2
Topic 18	1,856	109.2	24.9	–2%	20%	–12
Topic 19	1,887	111.0	50.6	18%	30%	8
Topic 20	1,232	72.5	12.0	1%	17%	–6
Topic 21	2,265	133.2	23.4	4%	13%	–2
Topic 22	1,370	80.6	16.5	6%	24%	–2
Topic 23	4,958	291.6	81.8	11%	22%	2
Topic 24	1,474	86.7	46.2	15%	30%	4
Topic 25	1,396	82.1	13.5	5%	24%	0
Topic 26	2,801	164.8	63.2	13%	19%	8
Topic 27	2,343	137.8	25.8	1%	15%	–6
Topic 28	3,530	207.6	115.1	60%	122%	4
Topic 29	1,704	100.2	27.0	8%	18%	2
Topic 30	2,376	139.8	20.3	1%	12%	–6
Topic 31	2,010	118.2	39.1	13%	32%	0
Topic 32	3,021	177.7	99.9	19%	28%	10
Topic 33	2,011	118.3	25.2	3%	16%	–6
Topic 34	1,229	72.3	20.1	6%	27%	–8
Topic 35	1,863	109.6	28.8	8%	18%	2
Topic 36	3,316	195.1	65.9	11%	20%	2
Topic 37	2,836	166.8	57.4	21%	45%	0
Topic 38	1,281	75.4	16.2	3%	21%	–2
Topic 39	1,425	83.8	24.2	8%	17%	0
Topic 40	1,855	109.1	22.0	6%	13%	–4
Topic 41	2,257	132.8	56.6	15%	33%	4
Topic 42	778	45.8	8.5	4%	20%	0
Topic 43	3,320	195.3	43.4	6%	15%	0
Topic 44	1,441	84.8	22.5	3%	23%	–4
Topic 45	2,963	174.3	38.6	6%	17%	–2
Topic 46	3,198	188.1	25.0	4%	14%	–2
Topic 47	2,475	145.6	54.6	15%	29%	0
Topic 48	2,946	173.3	55.1	9%	14%	2
Topic 49	5,202	306.0	207.2	27%	45%	14
Topic 50	4,310	253.5	114.4	20%	47%	0
Topic 51	2,669	157.0	22.6	4%	10%	4
Topic 52	2,656	156.2	60.4	15%	38%	2
Topic 53	2,454	144.4	34.3	8%	21%	–2
Topic 54	1,342	78.9	17.9	–3%	19%	–4
Topic 55	2,303	135.5	48.4	17%	39%	2
Topic 56	2,933	172.5	66.1	8%	14%	0
Topic 57	3,467	203.9	80.3	13%	24%	4
Topic 58	2,758	162.2	49.3	9%	17%	–2
Topic 59	1,961	115.4	20.3	4%	25%	–4
Topic 60	2,549	149.9	25.5	1%	15%	–4

Source: ISI-WoS; Authors.

For the network data, we employ a modularity algorithm by Blondel et al. (2008) to aggregate the topics. The algorithm produces five communities: medical research, biological research, chemical and physical sciences, Earth science, and a community containing information and communication technology-related science and social sciences. The size distribution of different communities can be seen in Figure 3.

The division visible in Figure 3 shows how LDA classifies the corpus in relation to the ISI-WoS and OECD classifications, and a complete list of OECD classifications and topic classifications into communities is given in Table 4. Figure 3 shows how many OECD main classifications or LDA-generated topics cover each of the main communities. It shows large differences in terms of granularity of classification schemes: for established and large (by volume) areas of science, such as natural sciences and medical research, topic modeling produces a more refined classification than the OECD classification. Medical research is covered by 10 OECD fields but has more than 26 LDA topics. Similar results are visible in communities covering chemistry and physical sciences, biological research, and Earth science. In the case of information and communications technology (ICT) and social sciences, LDA produces more generalized results and the OECD classes produce more specialized results.

Analyzing the results, Figure 2 highlights the differences of the OECD and unsupervised learning-based mappings. OECD mapping aggregates research into large classes, whereas topic modeling creates a number of small thematic areas. These topics position themselves at the intersection of several OECD classifications. A case in point would be Topic 2: clustered with biological research, the topic is mapped between basic medical research and clinical medicine; the topic's wordcloud positions the research at cell-level research, with terms referring to basic medical research and clinical work. Similarly interesting is Topic 22, focusing on bone implants, bioactive materials, and clinical studies, and then mapped close to chemistry and biological sciences, but clustered with medical research.

In Figure 2, educational sciences are clustered with medical research, which could be interpreted as a misclassification. However, Finnish educational sciences have a strong component focusing on medical research or cross-cutting elements. A case in point would be a research article from 1999 published in the *International Journal of Science Education* entitled "Spontaneous concept maps aiding the understanding of scientific concepts" (Slotte & Lonka, 1999) where the medical field was a point of interest for educational research. This cross-disciplinary approach is better visualized in Figure 2, where educational sciences are clustered with medical research but actually located near social sciences. Similarly, the network visualizes how the information and communication technology topics are related to the social sciences cluster. Even though the cluster results merged social sciences and humanities with ICT, the network visualization presented a clear division between the positions of engineering-related and social science-related

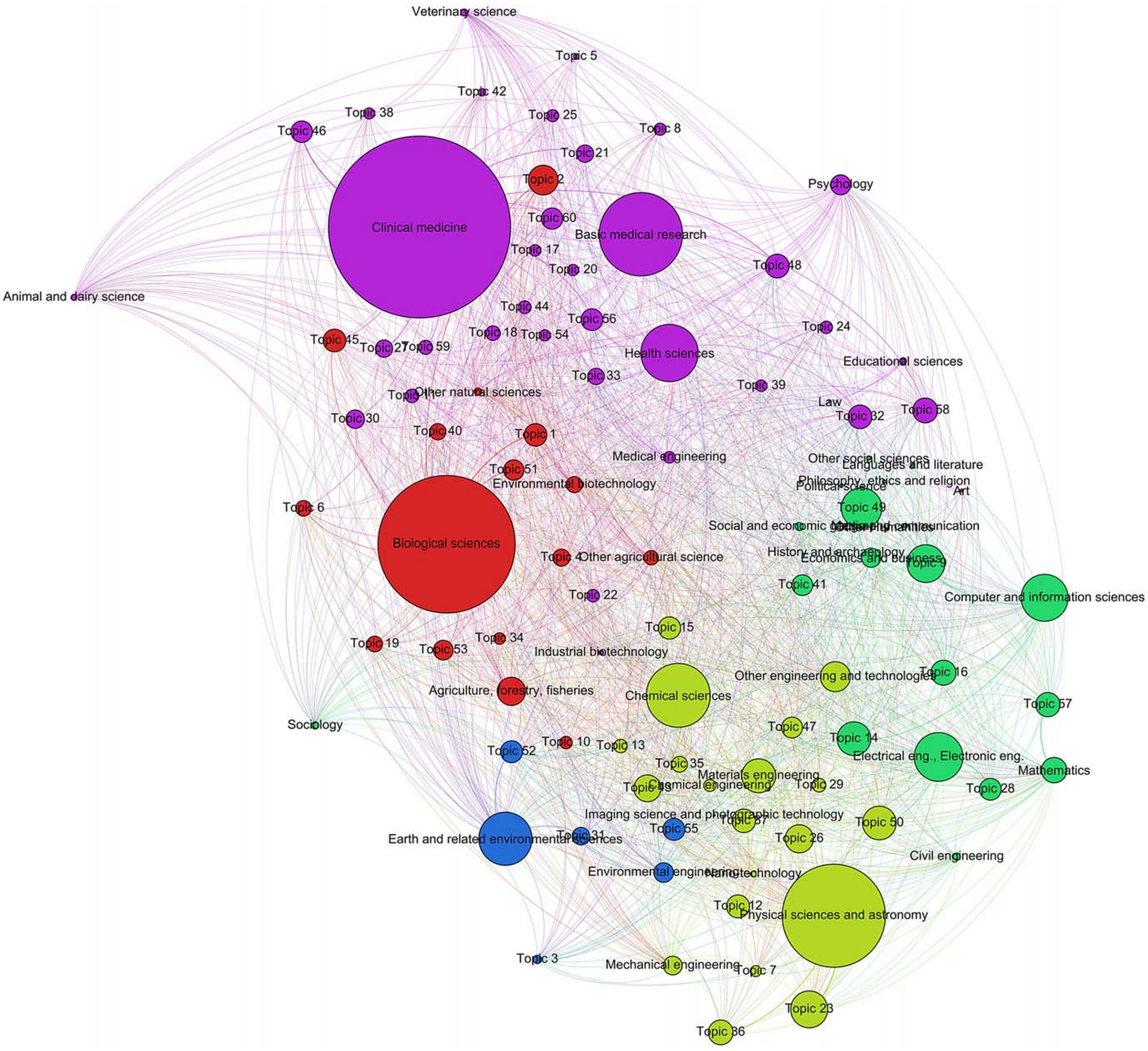

FIG. 2. Network visualization of topics and OECD major classifications assigned to publications. Node weight is based on the count of publications assigned to a given node. Edge weight is based on the co-occurrence of links between a topic and OECD major classification. Node and edge colors are based on a modularity algorithm in which each community is assigned a specific color. Source: ISI-WoS; Authors. Full sized picture in Appendix S2. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

topics. Similar examples that illustrate the thematic position of a topic in the network are industrial biotechnology and medical engineering, both classified in the medical research community but located near engineering. Thus, topic modeling and network visualization can depict interdisciplinary connections, and distinguish between "core" topics of a community and their interfaces with other communities.

When evaluating the position of nodes in Figure 2 we should consider the limitations produced by the ForceAtlas 2 algorithm used to create the visualization. ForceAtlas 2 is a force-directed layout where nodes repulse each other and edges attract the nodes they are connected to (Jacomy, Heymann, Venturini, & Bastian, 2011). The position of nodes is defined ultimately by the energy model of the algorithm. The objective of the ForceAtlas 2 algorithm is to produce a readable spatial arrangement, limiting true overlaps of nodes to support reading. This impacts the position of nodes and questions to what extent we can rely on node positions showing true intersections and which is produced by the algorithm in favor of readability.

Through a qualitative analysis of Figure 2, we argue that topic modeling captures problem-driven themes while the OECD classification structures science through the tree of knowledge disciplinary approach. Our findings from

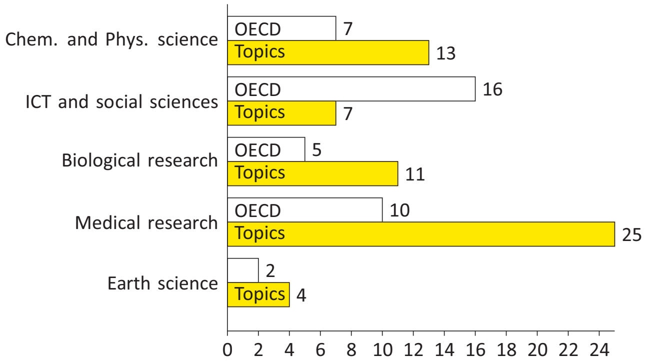

FIG. 3. Bar chart showing the division of OECD classes and topics in different communities. White bars give the number of OECD classes in a community. Yellow bars give the number of thematic topics in a community. Source: ISI-WoS; Authors. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 2 create the narrative to explain Figure 3. The difference in the number of classes rather indicates the difference in approach.

Dynamic Changes in the System

Classification of scientific knowledge is an obligatory step for a detailed analysis of the evolution of research systems, as it is fundamental to understanding which areas of the scientific field are key components of growth, decline, or stagnation. Indeed, any plausible analysis of a research system is required to relate to its epistemic properties, triggering the issue of classification of knowledge (Toivanen & Suominen, 2014).

Whereas the traditional approach is to plot year-to-year changes through ISI-WoS subject categories, we present annual change through the communities based on LDA. The overall publication volume in the data set grows by ~8% annually. This growth percentage is partly due to the large increase in publication volumes from 1995 to 1998. Overall, publication volumes grew by 21% from 1995 to 1996, by 23% from 1996 to 1997, and by 50% from 1997 to 1998. To some extent, this large growth in publishing is due to increased indexing of conference proceedings and abstracts in the ISI-WoS. If we remove the values of these outliers from the data set, actual annual growth is, on average, 3%. This is closer to the reported values of science publication expansion (Veugelers, 2010). Table 3 shows the individual average growth and standard deviation values for each of the topics. Not excluding the high-growth years, resulting in some of the topics having a large standard deviation in growth, most of the topics show growth on average.

We further control the time series of the topics by adjusting the growth of topics against overall growth in scientific publishing. We look at how topics grow yearly against the overall growth in scientific publishing and count the number of years that a topic grows faster than science overall. In Table 3, 24 of the topics grow faster than science overall, but 27 of the topics are stagnant or in decline compared with the expansion of science production in Finland. Nine topics have a growth index of 0. To highlight extremes, Topic 18 has the smallest index value, −12, indicating that it grew slower than research overall in nearly every year of the study. Similarly, Topic 8 shows decline with a growth index of −8. On the other extreme, Topic 49 shows an index value of 14, with almost year-to-year high growth, and Topic 32 has a growth index of 10. The median growth index value for the topics is 0.

The topics declining most are related to the medical field. These include Topic 8, which is defined by terms such as "rats," "effects," "treatment," and "mg," as well as Topic 18, which is defined by "cell(s)," "expression," "tissue," and "epithelial." The topics with the most growth are on the intersection of social sciences and ICT (Topic 49), defined by "social," "policy," and "management," and education and medical research (Topic 32), defined by "education," "health," and "nursing."

We used the modularity algorithm communities to represent a time series of development as seen in Figure 4.

Figure 4 describes the dynamics of the communities. The overall growth trend in each line corresponds with the overall increase in research publication volume. Each line also shows the transition in 1997–1998, where the research output of Finland nearly doubled, and if we exclude the Earth science community, increased appropriations raised the volume of publications for several years after 1998. Although not so clearly visible in lines other than the medical research community, a decrease in output was registered in the early 2000s, which may be explained in part by the plateauing of government increases in research and development (R&D) funding at that time. However, this decrease was short-lived, as publication volumes increased steadily in the 2000s. The notable case in the graph is ICT and the social sciences community, which shows a stronger growth dynamic than the other communities, most likely due to the "Nokia phenomenon" in the Finnish economy and innovation system.

Indeed, the ICT and social sciences community emerges as the true growth component in Finnish research. Looking at the entire time series, this community is almost equal in size to the Earth science community in the mid-1990s, but by 2009 the ICT and social sciences community was about 10 times bigger than Earth science. Much of this growth can be attributed to the changes in Finnish science and technology policy, specifically the strong focus on technology and especially ICT, and the subsequent emphasis on knowledge-based business—in short, the rise of Nokia. Only one topic in the community is clearly a social science theme and the other refers to ICT. The social science theme has a high growth index of 14 and grows on average by 27% yearly. If we aggregate the ICT topics, their growth index is 10, with a yearly average growth of 20%. It seems that both ICT and social sciences grow similarly.

Its decline from 2009–2010 can be partly attributed to a data problem, as the disciplines in the community have a

TABLE 4. Descriptive statistics of the thematic topics created by the LDA algorithm.

Classification	OECD classifications	Topics
Chem. and Phys. Science	Chemical engineering, Chemical sciences, Materials engineering, Mechanical engineering, Nanotechnology, Other engineering and technologies, Physical sciences and astronomy	Topic 12, Topic 13, Topic 15, Topic 23, Topic 26, Topic 29, Topic 35, Topic 36, Topic 37, Topic 43, Topic 47, Topic 50, Topic 7
ICT and social sciences	Civil engineering, Computer and information sciences, Economics and business, Electrical eng., Electronic eng., History and archeology, Imaging science and photographic technology, Languages and literature, Law, Mathematics, Media and communication, Other Humanities, Other social sciences, Philosophy, ethics and religion, Political science, Social and economic geography, Sociology	Topic 14, Topic 16, Topic 28, Topic 41, Topic 49, Topic 57, Topic 9
Biological research	Agriculture forestry fisheries, Biological sciences, Environmental biotechnology, Other agricultural science, Other natural sciences	Topic 1, Topic 10, Topic 19, Topic 2, Topic 34, Topic 4, Topic 40, Topic 45, Topic 51, Topic 53, Topic 6
Medical research	Animal and dairy science, Art, Basic medical research, Clinical medicine, Educational sciences, Health sciences, Industrial biotechnology, Medical engineering, Psychology, Veterinary science,	Topic 11, Topic 17, Topic 18, Topic 20, Topic 21, Topic 22, Topic 24, Topic 25, Topic 27, Topic 30, Topic 32, Topic 33, Topic 38, Topic 39, Topic 42, Topic 44, Topic 46, Topic 48, Topic 5, Topic 54, Topic 56, Topic 58, Topic 59, Topic 60, Topic 8
Earth science	Earth and related environmental sciences, Environmental engineering	Topic 3, Topic 31, Topic 52, Topic 55

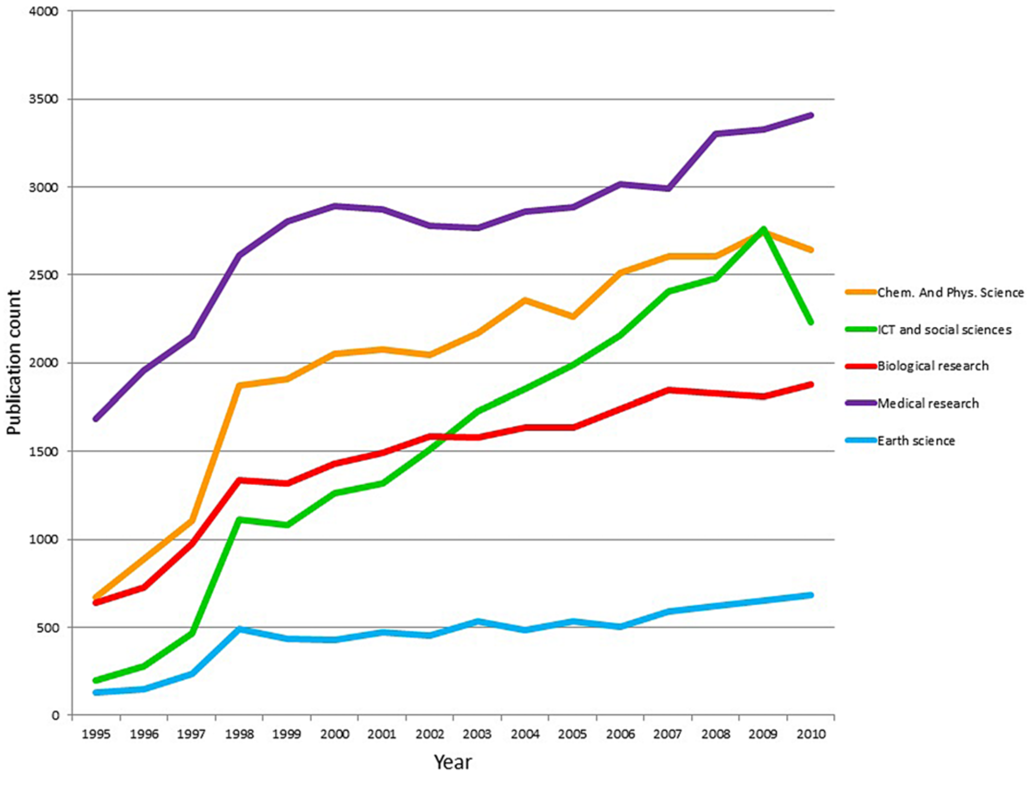

FIG. 4. Time series graphics for communities created by the modularity algorithm. In the figure, the time series of topics aggregated in the same communities were summed to create a time series of the publication count of each community. Source: ISI-WoS; Authors. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

propensity to publish through conference proceedings, which are added to the database with a time lag. The other part of the explanation for the decline is the significant decrease in corporate R&D in ICT, which reduced the amount of private funding for ICT.

In Finland, the Ministry of Education and Culture and the Academy of Science analyze the state and impact of Finnish science through publication volumes and citations (Lehvo & Nuutinen, 2006; Muhonen, Leino, & Puuska, 2012). As we compare our results, it becomes clear that while the bibliometric evaluations of national stakeholders are fine-tuned to produce metrics on organizational efficiency, they lend themselves poorly to showing epistemic changes. The unsupervised learning results presented here classify epistemic changes, truly uncovering latent patterns in science.

Discussion

The central objective of this paper has been to demonstrate and validate an unsupervised learning-based map of science, and to discuss its potential advantages and disadvantages vis-à-vis maps based on human-assigned classifications of science publications. From the theoretical perspective of sociology and philosophy of science, it is relatively easy to compare the advantages and disadvantages of human-assigned and unsupervised learning-based classifications. By definition, science is about new-to-the-world knowledge, expanding the existing body of knowledge. Human reasoning-based classifications of science, such as journal- or article-level subject classifications or the university faculty organization, are inherently historical by nature, whereas machine learning-based unsupervised learning methods generate classifications of science solely based on the available corpus. Thus, put simply, human reasoning-based classification schemes are about fitting new knowledge into a historical classification framework, whereas unsupervised learning-based classifications—as exemplified in this paper—are about generating a new classification that best describes the dimensions of the used corpus.

In practice, however, it is much more difficult to argue the superiority of one method over the other. We would be inclined to suggest that it may even be impossible to judge, and we question the fruitfulness of such a debate. Both human and machine-generated classification schemes of science are navigational tools that reduce the complexity of the inherently complex body of scientific knowledge instantiated in publications. Any perception of the superiority of one method is dependent on the intended use of the classification. Historical classification schemes are useful for information retrieval, because they make it easier to define what one is searching for. Ahistorical classification schemes more credibly identify novel areas of research that depart from existing knowledge, and are probably more useful in identifying emerging research topics. Machine learning classification schemes are also dependent on the input data, and are thus not entirely credible in establishing stable classification frameworks. Indeed, our paper raises the question of whether all maps of science are temporally and thematically bound, and suggest that it may be impossible to create a universal map of science. This is to be confirmed by future studies extending a single country and taking on the full copy of databases such as the WoS or Scopus.

In our opinion, the greatest potential value of machine learning-based science classifications is practicality. They make it much easier to generate high-accuracy special-purpose maps, for example, to zoom into special areas of research or to detect emerging fields, and so forth, and they allow for a broader set of statistical analysis than historical classification schemes. Machine learning classification also enables rapid generation of maps from large-scale data, thus enabling the possibility of creating a range of alternative mapping versions from the same data with changed processing and classification parameters. As such, they stand to benefit not only scholars as they try to analyze and navigate their own research environment more accurately, but also research management, science and technology policy, and corporate strategy development.

A central challenge for the cartography of scientific knowledge is the creation of valid and accurate coordinates, and in this regard, human-assigned and machine learning-generated classifications have important differences. Machine learning methods, such as the unsupervised learning methods used in this study, enable the classification of science without existing disciplinary or journal classification boundaries. In contrast, human-assigned classifications rely on historical classification schemes, lending themselves best to narrowly focused research outlets and confined research problems. However, the transformation of scientific practices presents fundamental challenges for such classification principles. Mutations of research problems (e.g., grand challenges), teams and publications, the emergence of multidisciplinary research, and new types of "meta" and "mega" science publications such as PLoS One, which works outside of traditional classification boundaries, pose a challenge for traditional metrics—but less so for the machine learning-based methods used in this study.

Because human-assigned classification of articles or journals is the dominant source for coordinates in science maps (Börner, 2010), such maps should be used to validate and analyze maps created with automated classification. This has been a central effort of this paper, and we have demonstrated how the two classification methods produce highly overlapping results in part, while also differing importantly. Automated classification detects epistemic connections between disciplines that are not otherwise automatically visible, the case in point being our discussion of the interconnectedness of Finnish ICT research and social sciences.

A central observation derived from Figures 2 and 3 is that the topics are mapped at the intersections of the tree of knowledge-type classes. Latent topics emerge based solely on data provided, clearly creating themes at different abstraction levels. Unsupervised learning produces problem or research question-driven themes, uncovering epistemic structure rather than organizational efficiency. The results provided serve the purpose of identifying thematic components of science in stability, growth, or decline. The analysis shows clear areas of decline and growth, but also the stable baseline of science.

Our study has limitations. Due to limitations in presenting the results in a format suitable for journal publication, we restricted our results to hard partitioning. This limits how our study can visualize interdisciplinarity. If we were to take advantage of a document that is partitioned into multiple topics, we could identify important links between topics, transitions of basic research into more applied studies, or the adoption of computer science tools as an enabling technology in the sciences. Thus, using the soft partitioning of LDA would allow for an even richer view of the science map, but would require visualization tools to enable users to take advantage of the results. We question whether multidimensional results such as these are useful in print format.

The capability to draw more elaborate multidimensional models of science is interesting, but to keep the structures at a practical level for human interpretation, we obtained 60 topics with a hard partitioning classification. For thematically broad data sets, such as the one used in this paper, we can increase the number of topics significantly, creating a more refined representation. Using measures such as perplexity to evaluate the number of topics would result in a higher number of topics, as the machine is able to find differences in details humans cannot identify. Chang et al. (2009) showed that metrics do not capture human-interpreted topic coherence, and the choice of the number of topics remains a qualitative task. However, when the topic modeling results are further analyzed and, for example, clustered by hierarchical clustering, the number of topics should be drawn from the capabilities and requirements of the selected analysis tools.

The LDA approach is also limited by the bag-of-words approach to the corpus. This approach loses some of the contextuality of the text. A number of approaches can be used to circumvent the bag-of-words limitation, such as preprocessing the text using *n*-grams or using event extraction algorithms to find underlying contextualities in the text. A more aggressive approach to preprocessing, most significantly using *n*-grams and phrases, could also improve matters. However, recent results by Yau et al. (2013) question the usefulness of aggressive preprocessing, and as the wordclouds in Appendix S1 seem practical, we decided to perform minimal preprocessing and to accept the appearance of some terms that are not useful in the wordclouds.

Acknowledgment

Tekes—the Finnish Funding Agency for Technology and Innovation-supported this work with the research grant (2004/31/2011) awarded for the project “Co-evolution of knowledge creation systems and innovation pipelines”. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of scientific journals. Scientometrics, 44(3), 323–345.

Blei, D.M., & Lafferty, J.D. (2007). A correlated topic model of science. The Annals of Applied Statistics, 1(1), 17–35.

Blei, D.M., & Lafferty, J.D. (2009). Topic Models. In A.N. Srivastava & M. Sahami (Eds.), Text mining: Classification, clustering, and applications (10th ed., pp. 71–94). Boca Raton FL USA: Taylor and Francis.

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research: JMLR, 3(2003), 993–1022.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

Boyack, K.W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. Scientometrics, 64(3), 351–374.

Boyack, K.W., Newman, D., Duhon, R.J., Klavans, R., Patek, M., Biberstine, J.R., . . . Börner, K. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. PLoS One, 6(3), e18029.

Börner, K. (2010). Atlas of science: Visualizing what we know. Cambridge, MA: MIT Press.

Börner, K., Chen, C., & Boyack, K.W. (2003). Visualizing knowledge domains. Annual Review of Information Science and Technology, 37(1), 179–255.

Braam, R.R., Moed, H.F., & Van Raan, A.F. (1991). Mapping of science by combined cocitation and word analysis: II: Dynamical aspects. Journal of the American Society for Information Science, 42(4), 252–266.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., & Blei, D.M. (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), Advances in neural information processing systems (pp. 288–296). Curran Associates, Inc..

Ding, Y., Chowdhury, G.G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using coword analysis. Information Processing & Management, 37(6), 817–842.

Feldman, R., & Sanger, J. (2006). The text mining handbook: Advanced approaches in analyzing unstructured data. New York: Cambridge University Press.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics, 56(3), 357–367.

Glänzel, W., Schubert, A., & Czerwon, H.J. (1999). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. Scientometrics, 44(3), 427–439.

Glenisson, P., Glänzel, W., & Persson, O. (2005). Combining full-text analysis and bibliometric indicators. a pilot study. Scientometrics, 63(1), 163–180.

Griffith, B.C., Small, H.G., Stonehill, J.A., & Dey, S. (1974). The structure of scientific literatures. II: Toward a macro-and microstructure for science. Social Studies of Science, 4(4), 339–365.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In F. Gey, M., Hearst, & R., Tong (Eds.), Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 50–57). Berkley CA USA: ACM.

Hood, W.W., & Wilson, C.S. (2001). The literature of bibliometrics, scientometrics, and informetrics. Scientometrics, 52(2), 291–314.

Jacomy, M., Heymann, S., Venturini, T., & Bastian, M. (2011). ForceAtlas2, a continuous graph layout algorithm for handy network visualization. Medialab Center of Research, 560.

Klavans, R., & Boyack, K.W. (2009). Toward a consensus map of science. Journal of the American Society for Information Science and Technology, 60(3), 455–476.

Klavans, R., & Boyack, K.W. (2014). Mapping altruism. Journal of Informetrics, 8(2), 431–447.

Lehvo, A., & Nuutinen, A. (2006). Finnish science in international comparison: A bibliometric analysis. Publications of the Academy of Finland, 15(6).

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. Journal of the American Society for Information Science and Technology, 60(2), 348–362.

Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. Scientometrics, 94(2), 589–593.

Leydesdroff, L. (1989). Words and cowords as indicators of intellectual organization. Research Policy, 18(4), 209–223.

DOI: 10.1002/asi

Liu, X., Yu, S., Janssens, F., Glänzel, W., Moreau, Y., & De Moor, B. (2010). Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database. Journal of the American Society for Information Science and Technology, 61(6), 1105–1119.

Mei, Q., Shen, X., & Zhai, C. (2007). Automatic labeling of multinomial topic models. In P. Berkhin, R. Caruana, & X.D. Wu (Eds.), Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 490–499). San Jose, CA, USA: ACM.

Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Munoz-Fernández, F.J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. Scientometrics, 61(1), 129–145.

Moya-Anegón, S.G., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F.J. & Herrero-Solana, V. (2007). Visualizing the marrow of science. Journal of the American Society for Information Science and Technology, 58(14), 2167–2179.

Muhonen, R., Leino, Y., & Puuska, H.-M. (2012). Suomen kansainvälinen yhteisjulkaiseminen. Opetus- ja kulttuuriministeriön julkaisuja 2012:4.

Nichols, L.G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. Scientometrics, 100(3), 741–754.

OECD. (2007). *Revised field of science and technology (fos) classifications in the frascati manual* (DSTI/EAS/STP/NESTI(2006)19/FINAL ed.). Paris.

Peters, H., & van Raan, A.F. (1993a). Co-word-based science maps of chemical engineering. Part II: Representations by combined clustering and multidimensional scaling. Research Policy, 22(1), 47–71.

Peters, H., & van Raan, A.F. (1993b). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. Research Policy, 22(1), 23–45.

Pudovkin, A.I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. Journal of the American Society for Information Science and Technology, 53(13), 1113–1119.

Rip, A., & Courtial, J.P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. Scientometrics, 6(6), 381–400.

Slotte, V., & Lonka, K. (1999). Spontaneous concept maps aiding the understanding of scientific concepts. International Journal of Science Education, 21(5), 515–531.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science, 24(4), 265–269.

Small, H. (1993). Macro-level changes in the structure of cocitation clusters: 1983–1989. Scientometrics, 26(1), 5–20.

Small, H. (1999). Visualizing science by citation mapping. Journal of the American Society for Information Science, 50(9), 799–813.

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. Scientometrics, 60(1), 71–79.

Talley, E.M., Newman, D., Mimno, D., Herr, B.W., II, Wallach, H.M., Burns, G.A., . . . McCallum, A. (2011). Database of NIH grants using machine-learned categories and graphical clustering. Nature Methods, 8(6), 443–444.

Toivanen, H., & Suominen, A. (2014). Epistemic integration of the European research area: The shifting geography of the knowledge base of Finnish research, 1995–2010. Science and Public Policy, scu066. http://spp.oxfordjournals.org/content/early/2014/11/23/scipol.scu066.short

Veugelers, R. (2010). Towards a multipolar science world: Trends and impact. Scientometrics, 82(2), 439–456.

Wallach, H. (2006). Topic modeling: Beyond bag-of-words. In W. Cohen & A. Moore (Eds.), Proceedings of the 23rd International Conference on Machine Learning (p. 977–984). Pittsburgh, PA: ACM.

Wei, X., & Croft, W.B. (2006). LDA-based document models for ad-hoc retrieval. In E.N. Efthimiadis, , S. Dumais, D. Hawking, & K. Järvelin (Eds.), Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (p. 178–185). Seattle, WA, USA: ACM.

Yan, E., Ding, Y., Milojević, S., & Sugimoto, C.R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. Journal of Informetrics, 6(1), 140–153.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. (2013). Clustering scientific documents with topic modeling. Scientometrics, 100(3), 767–786.

Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on isi category classification. Journal of Informetrics, 4(2), 185–193.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Wordcloud representation of the latent topics presented.

Appendix S2. Full sized version of FIG 2.