Technological Forecasting & Social Change 194 (2023) 122723

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Evolution and recombination of topics in Technological Forecasting and Social Change

Ivan Savin *

ESCP Business School, Madrid Campus, Spain
Institute of Environmental Science and Technology, Universitat Autònoma de Barcelona, Barcelona, Spain
Graduate School of Economics and Management, Ural Federal University, Yekaterinburg, Russian Federation

A R T I C L E I N F O

Keywords:
Computational linguistics
Knowledge recombination
Literature review
Machine learning
Topic modelling

A B S T R A C T

Technological Forecasting and Social Change (TFSC) is one of the main outlets in the literature on technological change. To assist its editors and future contributors in understanding the evolution of the journal, we review studies published between 1970 and 2022 identifying 25 main themes ranging from scenario foresight and forecasting methods that dominated the journal agenda in the first decades through innovation diffusion and patent analysis that gained popularity in 2006–2019 to social interaction and financial markets which experienced momentum in the last couple of years. We find that studies concentrated on more recent topics like firm performance, financial markets and environmental regulation have been cited more frequently and were contributed more often by scientists from China compared to the US. Inspired by the fact that studies recombining two or more topics are more impactful in terms of citations, we construct a graph of topics, both for the overall sample of 6240 studies reviewed and three periods of TFSC existence corresponding to different editors-in-chief. Our results illustrate knowledge complementarities explored in the journal so far and may indicate directions for further research.

1. Introduction

Technological Forecasting and Social Change (TFSC) is one of the most influential journals worldwide when it comes to the literature on technology forecasting and its discussion from business and management angles. Today it has an impact factor above ten and belongs to journals from top quartiles in areas of business, management, and urban planning in databases like Scopus and Web of Science (Kajikawa et al., 2022). Over its more than fifty years history and three generations of editors-in-chief the journal was naturally evolving with its main themes changing. According to its own aims and scope, it welcomes "*future studies … [that] interrelate social, environmental and technological factors*" with many articles devoted to financial and energy markets, firm performance, and environmental regulation. In this study we attempt to better understand how this evolution was taking place by analysing titles, abstracts, and keywords of more than 6000 articles using computational linguistics.

Given the large number of articles published in TFSC, reviewing them with conventional tools and human coders is very challenging (Callaghan et al., 2020). Fortunately, over the last years computational linguistics methods from the intersection of machine learning and natural language processing have been developed. Using these methods to review publications and analyse trends has become popular in the literature (Griffith and Steyvers, 2004; Mo et al., 2015; De Battisti et al., 2015; Lüdering and Winker, 2016; Ambrosino et al., 2018; Maier et al., 2018; Asmussen and Møller, 2019; Savin et al., 2022a; Savin and Teplyakov, 2022b). Recently these methods have been also applied to articles published in TFSC (Zhu and Cunningham, 2022; Kajikawa et al., 2022; Ashraf et al., 2022). Our contribution here is threefold. First, apart from describing the topics themselves, we study the temporal and spatial dimensions of those topics. Second, by building the graph of topics, we can reconstruct how these topics have been recombined to produce new insights over time.[1] The latter is important given the growing interdisciplinary nature of the journal and the fact that lately most discoveries were introduced by recombining existing knowledge (Youn et al., 2015; Savin, 2021), and that the most radical of them tend to recombine more

* C. de Arroyofresno, 1, Madrid 28035, Spain.
E-mail address: isavin@escp.eu.

[1] Note that Ashraf et al. (2022) using VOS viewer construct a graph of keywords for each topic based on their co-occurrence, but this does not allow to look on topics as a whole but keeps an isolated perspective on them.

https://doi.org/10.1016/j.techfore.2023.122723
Received 13 January 2023; Received in revised form 17 June 2023; Accepted 18 June 2023
Available online 28 June 2023

knowledge domains (Schoenmakers and Duysters, 2010). Finally, we study the relationship between the number of citations and the length of publication titles finding evidence for an inverted U-shaped relationship and indicating that the length of approximately 20 words has been associated with the highest count of citations per year.

More specifically, we derive 25 main themes covered in TFSC by applying structural topic modelling method to articles published between 1970 and 2022. We study how popularity of these topics was changing over time, from which countries authors have mostly contributed to the journal and which of the topics attracted more citations. Furthermore, we construct a graph of TFSC topics, both for the overall sample and three periods of TFSC existence corresponding to different editors-in-chief. This gives us a better view of the knowledge space covered in the TFSC journal, how this was changing over time and what research themes have been predominantly recombined depending on the period of its existence.

2. Data and methods

The sample of studies published in TFSC has been retrieved from Scopus database on December 22, 2022, and originally included 6731 studies (6307 of which are articles, while the rest are short surveys, letters, reviews, etc)· We excluded from further consideration publications with no abstract essential for our computational linguistic analysis reducing the sample to 6240 studies and removing most of editorial, letters and errata. As a result, we obtained the distribution of textual documents (we use titles, abstracts, and keywords) over time demonstrated on the left plot in Fig. 1. The number of publications was under 200 per year till around 2015 and increased dramatically afterwards reaching approx. 800 in 2021. The right plot of Fig. 1 shows the length of text per document demonstrating that in most cases we have between 200 and 300 words.

If we look on the coverage of countries where scientists who published in TFSC were affiliated,[2] we will find a list of 114 countries (top plot in Fig. 2) with the US, the UK and China being the top three states leading by a big margin and further followed by South Korea, the Netherlands, Italy, Germany, Spain and France. Interestingly, if we weight countries contribution by the number of citations per year the respective studies have received (bottom chart in Fig. 2), the US and China switch places (first and third) with the latter being the new leader. As we will show later, this has to do with the fact that authors affiliated in China contributed to TFSC on topics that attracted more citations.

To reveal hidden structure in our textual data, we use the topic modelling (TM) approach. In simple words, TM clusters words into topics based on how often any pair of words appears in the same texts (Blei, 2012; Savin et al., 2021). For example, if we see the words “pollution”, “pressure”, “quality”, “protection” in a topic labelled “environmental regulation”, it means that these words appear relatively more often in combination with each other and other words from this topic. Compared to simple count of keywords, TM has the advantage of considering words not in isolation, but accounting for their context, which can influence the meaning of the words. An advantage of structural topic modelling (STM) over classical TM is that it includes additional information about the publications, in our case the year of publication and the number of citations. Using additional data as covariates at the stage of estimating a topic model has proven to produce topics with higher predicting power and interpretability (He et al., 2009; Roberts et al., 2014; Speier et al., 2016). We apply STM using the associated R package by Roberts et al. (2019).

A necessary step before building a topic model is pre-processing of textual data. We used the standard steps described in recent literature (Aggarwal, 2018; Uglanova and Gius, 2020; Savin and Teplyakov, 2022a). In particular, the text documents were divided into separate elements (tokens); capital letters replaced; punctuation and stop words removed; and words converted to their dictionary form using Wordnet-based lemmatization engine (Fellbaum, 2005; Voutilainen, 2003); words that are too rare (i.e., that appear less or equal to 5 times[3] in all the documents) were subsequently removed; stable word sequences called n-grams have been additionally formed (e.g., “big_data”, “electric_vehicle”).[4] As a result, our final dataset contains 5577 unique words for building a topic model and 474,418 total word occurrences.

To determine the optimal number of topics, we run the model for 3 to 50 topics and record model performance on the following metrics (Savin et al., 2020): heldout log-likelihood (i.e. predictive power of the model), exclusivity (degree of overlap between popular words within each topic), and semantic coherence (the degree of co-occurrence of words from the same topic in text documents). Typically, increasing the number of topics tends to increase the model's predictive power and topic exclusivity, but reduces their semantic coherence. In Fig. A1 in the Appendix, we demonstrate that 25 topics allow us to achieve close to the maximum possible exclusivity while maintaining semantic coherence and predictive accuracy at reasonable levels.[5]

In order to construct a graph of topics, we define two topics as connected based on their cosine similarity, i.e. co-occurrence of those topics in the same TFSC publications, where presence of a topic in an article is proxied by its prevalence ranging from zero to one. Measuring this similarity between any pair of topics we obtain a symmetric matrix with ones on the main diagonal as cosine similarity of two identical vectors is one, and all other values bounded in [0,1). These values capture the potency of the relation between a pair of topics and can be interpreted as corresponding link weights in an undirected graph between these topics. In particular, a high weight implies that the two topics appear in many publications with a high prevalence, and they are strongly linked.

Since STM as a probabilistic model assigns small positive values to virtually every topic in a given text, the resulting graph represents a fully connected component. This result is in line with Shibata et al. (2011) earlier having shown that based on semantic information it can be difficult to distinguish unrelated groups of papers as they typically share many common terms. Therefore, following Savin et al. (2022b), we assess significance of each particular edge in the resulting graph and limit further analysis only to those that are statistically significant. Specifically, we follow Saracco et al. (2015) by building 1000 randomly generated counterparts to the empirical graph which exhibit on average

[2] Many articles published have authors affiliated in more than one country. To quantify share of a country, we identify the full list of unique countries mentioned in affiliations of each publication and then attribute equal shares to each of the countries.

[3] We chose the cut-off threshold in proportion to the size of our dataset. Specifically, Tvinnereim and Fløttum (2015) having around 2000 documents used earlier a cut-off value of 3, while Savin et al. (2022a) with approximately 200 thousand documents - a cut-off value of 50. Since there is no formula to estimate an optimal cut-off, the value of 5 looks a natural choice. Moreover, this cut-off value results in a data sample comparable to Zhu and Cunningham (2022) who focus on 5000 words considering a comparable dataset.

[4] Specifically, we concentrate on forming bi-grams by using gensim package in Python and the normalized pointwise mutual information (NMPI) score (see Bouma, 2009). This NMPI score ranges between −1 (means two words are never occurring together) through 0 (they occur together as often as expected, based on their independent probability) to +1 (if they occur only together). We form out of these words a bi-gram with a”_” symbol if NMPI score is above 0.65 that captures stable phrases like “renewable_energy” and “smart_city”.

[5] This number is considerably larger than what has been used by Zhu and Cunningham (Blei, 2012) or Kajkawa et al. (Hejazi et al., 2014). The larger number of topics should allow us to make more detailed analysis of the main themes in TFSC.An alternative approach to choose optimal number of topics that has been recently suggested in the literature is based on information criteria (Bystrov et al., 2023). Unfortunately, however, this method is not yet available in software packages on topic modelling in R or Python.

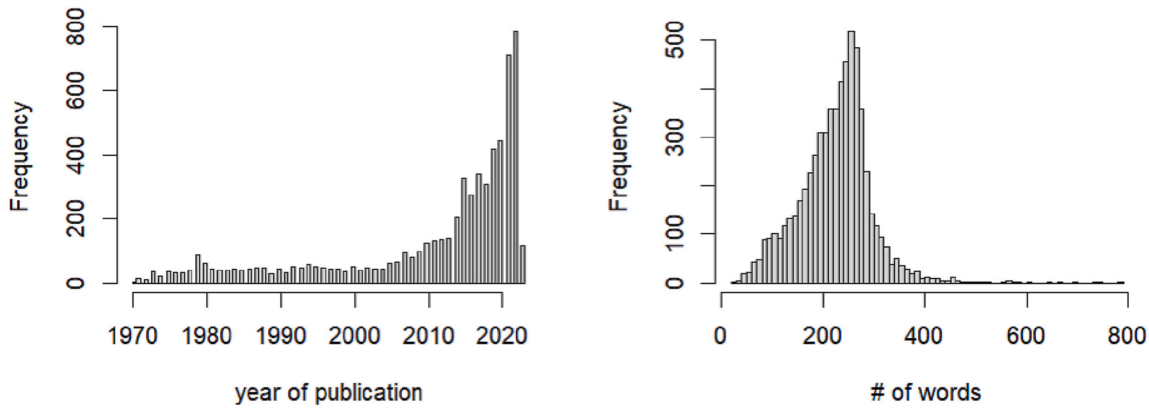

Fig. 1. Number of studies published in TFSC over time in our sample and the distribution of text length for our analysis.

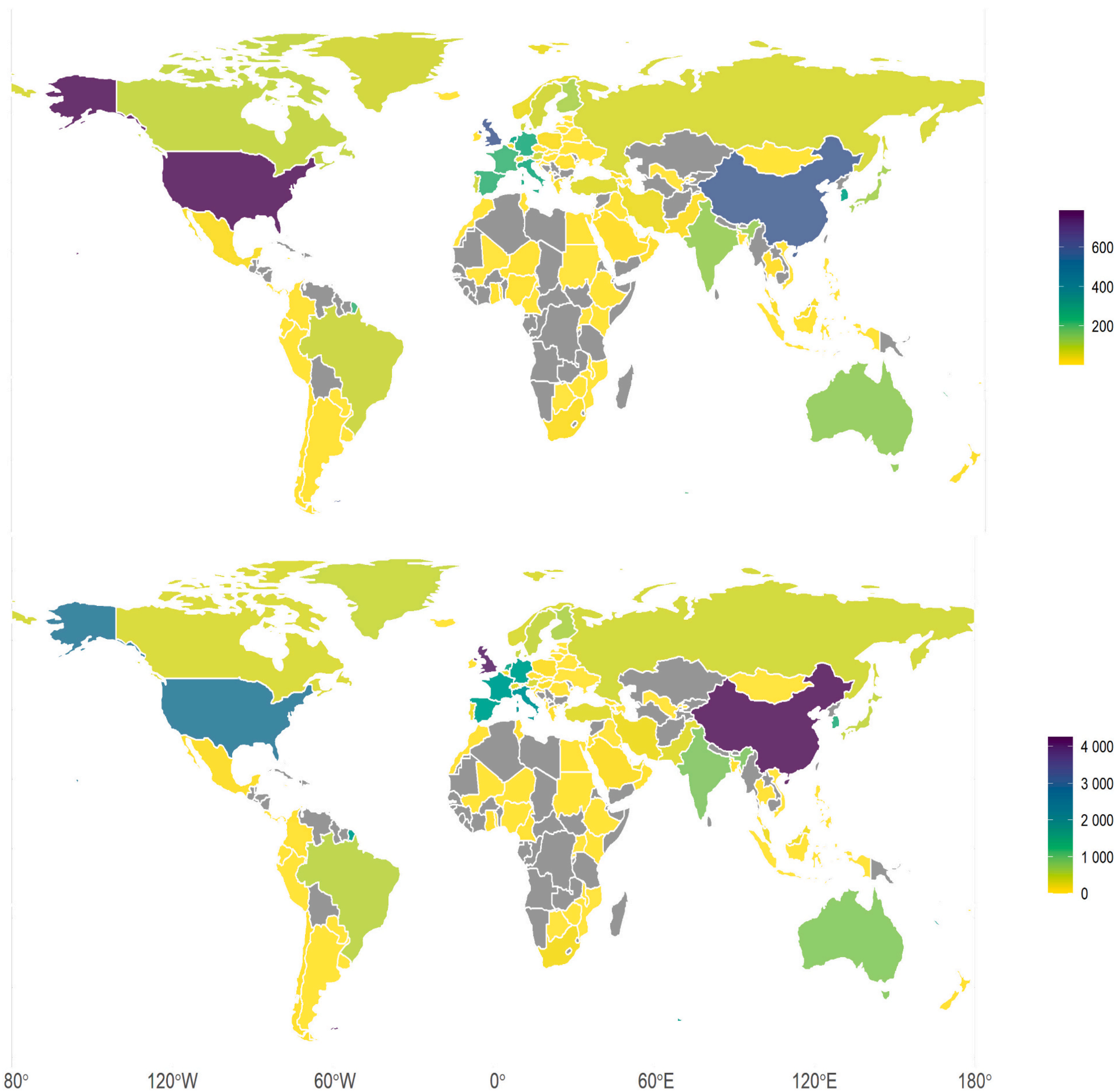

Fig. 2. Country coverage in TFSC.

Note: The top heatmap shows the global coverage where the studies with authors affiliated at more than one unique country are equally split between those countries. The bottom heatmap additionally takes the number of citations per year of that study into account.

the same degree distribution and the same weight distribution of each topic. Comparing the empirically observed weighted graph with these counterparts we preserve only those edges which weight falls in the 5 % (conventional significance level) of most outstanding edge weights. Therefore, the absence of an edge between any pair of topics in our henceforth analysis does not imply that they never appear together in any TFSC publication, but that the extent to which they co-occur does not meet the bar to be considered as significant. This allows us to better distinguish between strongly and weekly connected topics and form meaningful topic clusters.

Furthermore, since we earlier argued that radical innovations tend to recombine more knowledge domains (Schoenmakers and Duysters, 2010), we will test this result on TFSC studies by regressing the number of citations received on the number of topics recombined. For the latter we introduce a threshold of topic presence since, due to its probabilistic nature, STM assigns positive prevalences of each topic and a few percentages cannot yet reliably indicate that the study really addressed the topic. In the following we will use a threshold of 25 % to consider the topic to be present but experiments in the range 20 %–40 % produced qualitatively same result.[6] According to the chosen threshold, around 20 % of TFSC studies recombine two or more topics. The total number of citations received in our sample is in the range between 0 and 1920. Given that the articles differ considerably by the publication year, in line with Ashraf et al. (2022), we normalize the number of citations per year since publication (reducing the range to [0,320]).

3. Results

The resulting topics are presented in Table 1 and visualised in Fig. 3 as word clouds. Furthermore, Figs. 4 and 5 summarize how topic prevalences changed over time and which attracted larger number of citations per year. In Table 2 next to most frequent and exclusive words for each topic and an illustrative title of a paper with the highest prevalence of the topic we also provide concise topic labels we have formulated after studying titles, abstracts and keywords of top twenty documents with highest prevalence of the respective topics.

Topic 1 (Tx henceforth stands for topic x) on desirable futures has the largest prevalence in our sample of documents (7.5 %) and is most generic covering questions of future societies, ageing and creativity, among others. This topic was one of the dominant ones in TFSC in the first two decades of its existence (Fig. 5), but then its share has plummeted (from about 20 % in 1970s to less than 5 % today) and attracted relatively few citations per year (Fig. 6). A similar comment applies to T4 on scenario foresight and T6 on forecasting methods. These three topics can be called "traditional" topics to TFSC since they dominated in its agenda for a long part of the journal's history. To the contrary, T13 on industrial evolution, T17 on China (covering a wide range of issues related to productivity, emissions, agriculture, etc.), T18 on financial markets and T25 on environmental regulation have experienced increasing attention in TFSC more recently and tend to be cited on average considerably more often. Likewise, T9 on socio-technical transitions (describing transformative changes in many countries and industries), T16 on big data in healthcare (which encompasses many literature-based discoveries applied to the biomedical domain) and T22 on triple helix (focusing on the interaction between academia, private sector, and government) significantly increased their share in the journal in the last decade and have high citation rates. Other topics, such on technological roadmapping (T7), innovation diffusion (T10) and ICT (T21) have been relatively stable over time and received citations close to the journal average (5.25 per year per document in our whole

[6] Considering prevalences above 40 % is difficult as <5 % of papers have topic share above this bar (see Fig. A2 in the Appendix). In contrast, topic prevalences below 20 % are ubiquitous and often do not reflect that the paper is focusing on this topic.

Table 1
25 main topics in TFSC based on titles, abstracts, and keywords.

	Topic label	Most discriminating terms and illustrative titles	Topic proportion
T1	Desirable futures	creativity, creative, problem, ethical, people, old, peace, age, think, conflict "Undesirable versus desirable societies"	7.5 %
T2	Social interaction	intention, medium, online, perceive, brand, crowdfunding, acceptance, privacy, trust, attitude "Understanding player behavior in online games: The role of gender"	5.4 %
T3	Strategic management	decision, decision_make, strategic, fuzzy, decision_making, project, uncertainty, criterion, evaluation, risk "Some guides for the selection of a decision-making strategy"	5.4 %
T4	Scenario foresight	delphi, foresight, expert, scenario, opinion, participatory, future, exercise, quantitative, judgment "Preparing for the future: Development of an 'antifragile' methodology that complements scenario planning by omitting causation"	5.3 %
T5	STI policy	national, private, government, public, european, country, policy, international, instrument, asia "Science, technology and innovation policy in Russia and China: Mapping and comparisons in objectives, instruments and implementation"	4.9 %
T6	Forecasting methods	logistic, curve, forecast, prediction, time_series, substitution, forecasting, probability, accuracy, bass "Notes on forecasting a chaotic series using regression"	4.9 %
T7	Technology roadmapping	technology, roadmapping, nanotechnology, iot, adoption, roadmap, internet_thing, roadmaps, emerge, device "Roadmapping a disruptive technology: A case study The emerging microsystems and top-down nanosystems industry"	4.6 %
T8	Modelling methods	model, simulation, numerical, modelling, neural_network, agent, optimization, computer, algorithm, lotka_volterra "MISS-E: A method of modelling and simulation of dynamic systems"	4.5 %
T9	Socio-technical transition	ecosystem, actor, blockchain, socio_technical, transition, ecosystems, governance, niche, systemic, regime "Understanding institutional capacity for urban water transitions"	4.4 %
T10	Innovation diffusion	innovation, diffusion, product, radical, market, adopter, innovative, innovator, marketing, innovate "The legitimation strategies of early stage disruptive innovation	4.2 %
T11	Economic growth	cycle, growth, wave, economic, population, economy, world, long, evolutionary, historical "Kondratieff waves in global invention activity (1900–2008)"	4.1 %
T12	Knowledge acquisition	capability, collaboration, knowledge, collaborative, organizational, absorptive_capacity, alliance, open, partner, ambidexterity "Does open innovation always work? The role of complementary assets"	4.0 %
T13	Industrial evolution	industry, supply_chain, business, digitalization, digital, manufacture, manufacturing, chain, circular_economy, semiconductor	3.9 %

(continued on next page)

Table 1 (*continued*)

	Topic label	Most discriminating terms and illustrative titles	Topic proportion
		“Digitalization, agility, and customer value in tourism	
T14	Technological change	technological, material, progress, change, technical, production, scale, space, commercial, military “Technological change in the military aircraft industry”	3.7 %
T15	Firm performance	firm, smes, small_medium, corporate, performance, size, high_tech, ownership, family, responsibility “No new tricks for old dogs? Old directors and innovation performance”	3.6 %
T16	Big data in healthcare	medical, analytics, big_data, healthcare, bibliometric, health, discovery, health_care, hospital, intelligence “Literature-related discovery (LRD): Potential treatments for cataracts”	3.6 %
T17	Studies on China	china, productivity, efficiency, carbon_emission, spatial, province, trade, carbon, regional, region “Green total factor productivity of dairy cow in China: Key facts from scale and regional sector”	3.6 %
T18	Financial markets	finance, investment, financial, bitcoin, africa, stock, investor, labor, bank, return “The Effects of Central Bank Digital Currencies News on Financial Markets”	3.5 %
T19	Energy and climate	electricity, energy, renewable_energy, wind, solar, nuclear, climate, renewable, power, mitigation “Energy perspectives into the next millennium: From resources scarcity to decarbonization”	3.4 %
T20	Patent analysis	patent, patent_invention, intellectual_property, biotechnology, convergence, citation, cluster, classification, invention, right “Patent keyword network analysis for improving technology development efficiency”	3.2 %
T21	ICT	ict, telecommunication, communication, emergency, mobile, service, mobile_phone, internet, tourism, disaster “Competing risk model for mobile phone service”	3.1 %
T22	Triple helix	entrepreneurship, entrepreneurial, entrepreneur, smart_city, venture, student, education, university, teach, academic “A tension lens for understanding entrepreneurship-related activities in the university”	2.9 %
T23	Sustainable development	agricultural, agriculture, food, sustainability, farmer, sustainable, water, rural, waste, sdgs “Bundling agricultural technologies to adapt to climate change”	2.6 %
T24	Transport	vehicle, transport, electric_vehicle, car, transportation, mobility, hydrogen, automobile, road, autonomous “Travel patterns and the potential use of electric cars - Results from a direct survey in six European countries”	2.0 %
T25	Environmental regulation	green, environmental, eco, regulation, protection, quality, regulatory, pollution, friendly, degradation “Disentangling the causal structure behind environmental regulation”	1.8 %

Note: The terms shown are those that are the most frequent as well as exclusive to each topic. These words are estimated by the STM package using FREX (FRequency and EXclusivity) measure (Roberts et al., 2019). Illustrative titles are chosen from the ten documents with the highest topic prevalence. In the Supplementary Information we provide titles of all ten studies with highest prevalence for each topic.

sample).

If we look on which topics researcher affiliated in different countries tend to focus (Table 2), we will notice that the US, France and Germany have a larger share in “traditional” topics of TFSC like T1 (Desirable futures), T4 (Scenario foresight) and T6 (Forecasting methods), which are cited less frequently than TFSC studies on average. Researchers from China, in contrast, worked more on the topics devoted to China (T17), financial markets (T18) and firm performance (T15), which are cited considerably more often. This explains that even though the US-affiliated researchers lead in the total number of contributed studies, they lose the lead once citations are considered (Fig. 2). Similarly, researchers from Spain contributed more on well cited topics related to social interaction (T2), firm performance (T15) and knowledge acquisition (T12). Other countries in Table 2 have among their three most popular topics more and less (than on average) cited topics at the same time.

Having classified the publications into topics, we can now assess the extent to which those topics were recombined in TFSC studies, i.e. how often these topics appear in the same articles. In line with Savin et al. (2022b), we assume that the more the two topics are used jointly in the same studies, the more they complement each other. An example can be the study by Donbesuur et al. (2020) where the authors study how different strategies of knowledge acquisition (T12), namely technological and organizational innovation activity, can foster firm performance (T15). Results of OLS regressions explaining the number of citations per year are provided in Table 3. Since number of citations over time tends to increase (among others due to the greater number of scientific outlets available today compared to decades ago), we also add in a separate model as a control the year of publication. Once we take the year of publication into account, the number of topics becomes positive and significant indicating that studies recombining more knowledge bases attract more citations. This result holds also if instead of OLS we use a quantile estimator robust to non-normal distribution of the citations count.

Furthermore, since earlier in the literature it has been argued that articles with shorter titles tend to be cited more frequently (Letchford et al., 2015; Savin and van den Bergh, 2021), we add to the regression models in Table 3 the length of the study titles. This length ranges in our sample from 1 to 36 words. Since plotting the number of citations against the length of title hints at a possibility of an inverse U-shapes relationship, we further add a squared terms of title's length to capture that nonlinear pattern. As one can see, the coefficient for the length of title is positive and significant, but the squared term is indeed negative and significant confirming the presence of a non-linear relationship. Similar to Marson and Savin (2015) exploring a non-linear relationship, we estimate the title length corresponding to the maximum number of citations (by equating the first derivative of the quadratic function capturing the relationship between citations and title's length). For OLS estimate this value equals $-0.35/-0.02 \approx 18$, while for quantile regression $-0.13/-0.006 \approx 22$. Thus, we find that in the range between 1 and 18 words, longer titles in our sample of TFSC articles are associated with a higher performance in terms of citations, while in the range of 22–36 words, an additional word in the title is associated with less citations. This relationship is robust even if we control with dummy variables for the topics covered in the TFSC studies. Thus, even after controlling for the different citation trends among topics we still find the non-linear relationship between the citations and title's length.

At first, this result looks surprising as it somewhat contradicts prior evidence obtained based on many journals and hypothesizing that papers with shorter titles may be easier to understand, and hence attract more citations. However, if one looks closer on TFSC, one can realize that studies published here often are more technical in nature (e.g., suggesting a new methodology for technology forecasting or describing

Fig. 3. Word clouds of 25 topics in TFSC.

Note: The font size reflects the probability of the respective word given the topic, while darker colour indicates higher exclusivity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

socio-technical transition in a particular country or sector). Hence, eight out of ten articles with highest number of citations per year in this journal have a subtitle: "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations" by Wang et al. (2018), "Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries" by Belhadi et al. (2021), and "China's manufacturing locus in 2025: With a comparison of 'Made-in-China 2025' and 'Industry 4.0'" by Li (2018), to name a few. Our result demonstrates that shorter titles do not necessarily attract more citations. Instead, keeping the title around 20 words long and providing in it more details about the paper may be an optimal strategy to maximize own citation count.

Now let us return to the issue of topic recombination addressed earlier. As we demonstrated in Table 3, studies that combine more than one topic attract significantly more citations. Hence, capturing instances where topics have been systematically recombined is vital not only because this allows to look at an overall picture of the knowledge space covered in the TFSC journal, but also indicates areas where more impactful studies have been produced. In Fig. 7 we construct a graph on the overall sample of 6240 studies in TFSC, while in Fig. 8 we repeat this step on three different time windows: 1970–2004, 2006–2019 and 2021–2022. We chose these periods for two reasons. First, they roughly correspond to the periods of three different editors-in-chief (Hal Linstone, Fred Phillips, and joint editorship by Scott Cunningham and Mei-

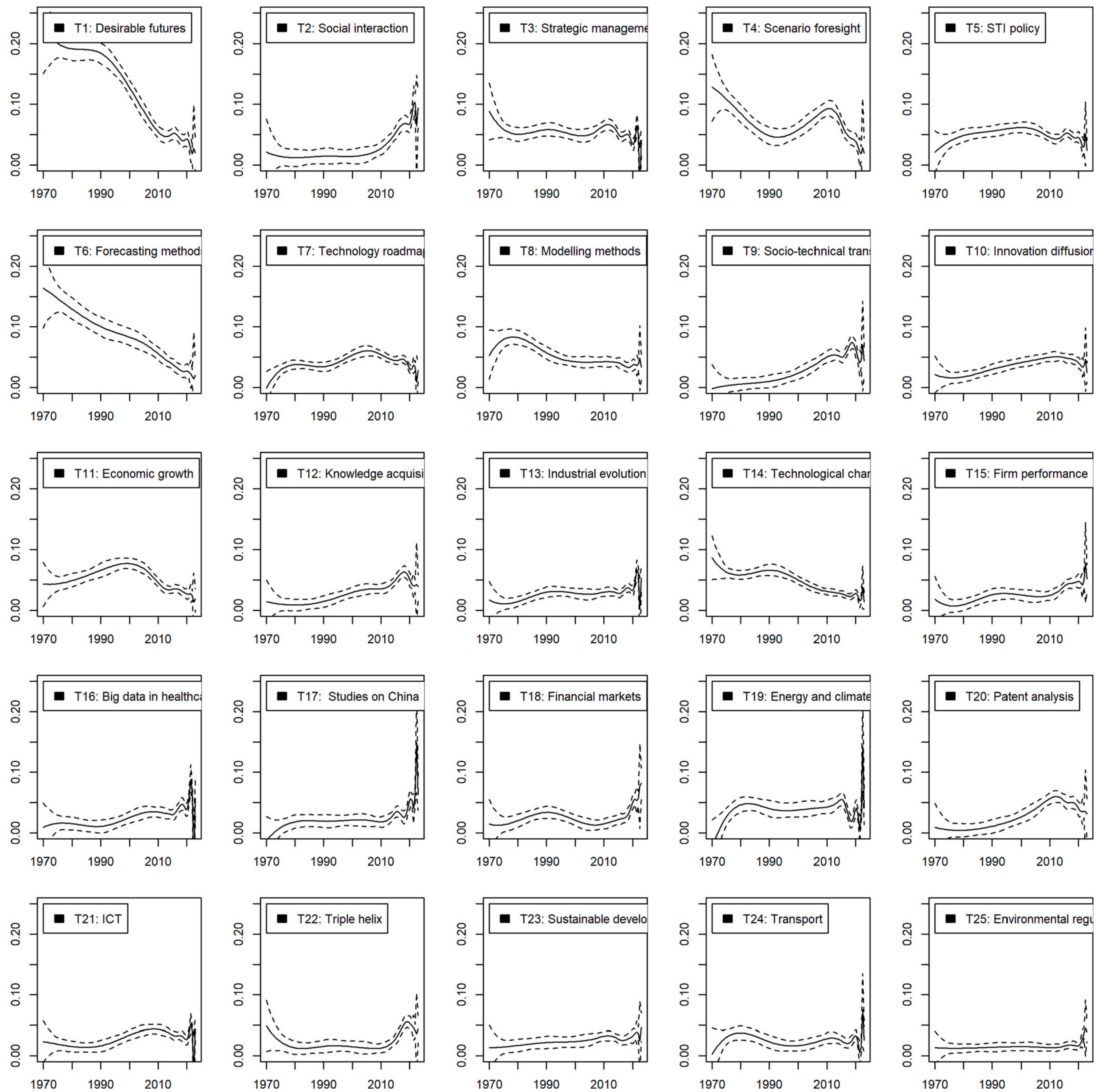

Fig. 4. Change in shares of topics in TFSC over time.
Note: To establish presence of a nonlinear relation between the period of publication and the topic prevalence, the former is converted into a set of dummies for each year. Coefficients generated by a regression where the outcome variable is the topic prevalence in each publication. An estimate with a 95 % confidence interval above zero indicates a significant prevalence of that topic in the respective year.

Chih Hu). Second, despite very different time length but due to the growing number of publications in TFSC over time, these periods cover rather comparable number of studies in our sample: 1398, 2730 and 1609, respectively.

From Fig. 7 we see that the 25 topics are represented with seven connected components (clusters). The largest component encompassing eight topics is about modelling various dynamic processes (therefore the most connected topic there is about modelling, T8): innovation diffusion (T10), social interactions (T2), socio-technical transitions (T9), knowledge acquisition (T12) and industrial evolution (T13). The two smallest topics in this component on transport (T24) and environmental regulation (T25) are related here since they are among the most dynamic areas of research, as exemplified by Pasaoglu et al. (2016) and Mahmood et al. (2022).

The second largest component in Fig. 7 with six topics is about foreseeing technological and economic trends. Its largest topics are on scenario foresight (T4) and STI policy (T5) being further related to triple helix (T22) and through patent analysis (T20) to economic growth (T11) and financial markets (T18). In this cluster the role of state and university-industry collaboration to stimulate technology and economy are mainly discussed. A good example is by Krammer (2017) who documents the role of STI policies to develop competitiveness in applied

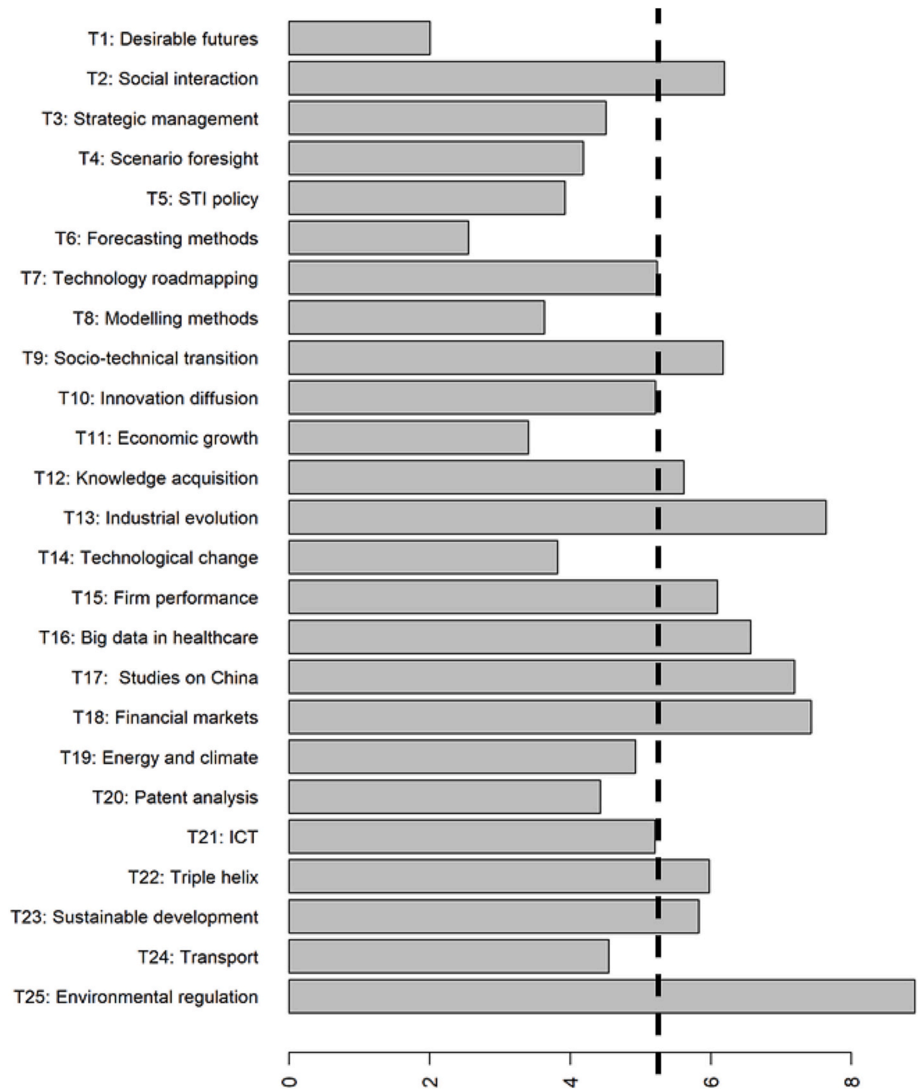

Fig. 5. Average number of citations per year per document for each topic.
Note: The number of citations per document belonging to a topic is calculated by multiplying the publication's number of citations per year with the topic prevalence and taking the average. The dashed line indicates the average number of citations per year an article in TFSC from our sample has received (5.25).

Table 2
Three most popular topics for researchers affiliated in the countries with highest contribution to TFSC.[a]

Most popular topics	China	United Kingdom	USA	Italy	France	Germany	Spain	Netherlands	South Korea
1	T17 (18.1 %)	T2 (8.1 %)	T1 (11.9 %)	T12 (6.6)	T2 (9.4 %)	T4 (12 %)	T2 (8.4 %)	T9 (12.1 %)	T20 (11.5 %)
2	T18 (6.4 %)	T6 (7.8 %)	T6 (8.0 %)	T13 (6.2 %)	T1 (6.4 %)	T9 (6.6 %)	T15 (7.6 %)	T1 (6.5 %)	T7 (7.5 %)
3	T15 (6.2 %)	T9 (7.7 %)	T3 (5.9 %)	T5 (5.6 %)	T12 (6.2 %)	T3 (5.9 %)	T12 (6.5 %)	T10 (6.2 %)	T2 (6.7 %)

Note: In brackets percentage of the respective topic out of all 25 topics to which scientists from respective countries have contributed to is provided.
[a] In Table A1 in the Appendix we provide a table with percentages of all 25 main topics in TFSC per nine countries with highest contribution to TFSC.

research (patents) and subsequently in final goods export.

The other five components in Fig. 7 are much smaller and encompass between one and three topics. Let us first look on those clusters containing three topics each. One consists of "traditional" topics in TFSC on desirable futures (T1), forecasting methods (T6) and technological change (T14). A recent example combining these topics is presented by Luo (2021). Another cluster of three topics is about technology roadmapping and its application to energy, climate and sustainable development (see Amer and Daim (2010) for an example).

One of the clusters with two topics is on strategic management (T3) and ICT (T21) and can be illustrated by Ominde et al. (2021) on optimising delivery of ICT projects. Another connected pair of topics is on firm performance (T15) and big data in healthcare (T16). It can be illustrated by Nayak et al. (2021) showing the value of emerging technologies like internet of things and cloud technologies for firms active in health insurance. Finally, the only isolated topic in Fig. 7 is T17 on China demonstrating that issues covered by TFSC in relation to China are very diverse and over the period of fifty years were not limited to one or few topics discussed above. As we show in Fig. 8, however, it does not imply that it has always been so.

Turning to Fig. 8, first thing one realizes is that the number of connected components as well as the number of significant edges in the graph of topics is not constant. In the first period its density was 21 % falling subsequently to 10 % (in 2006–2019) and 11 % (2021–2022),

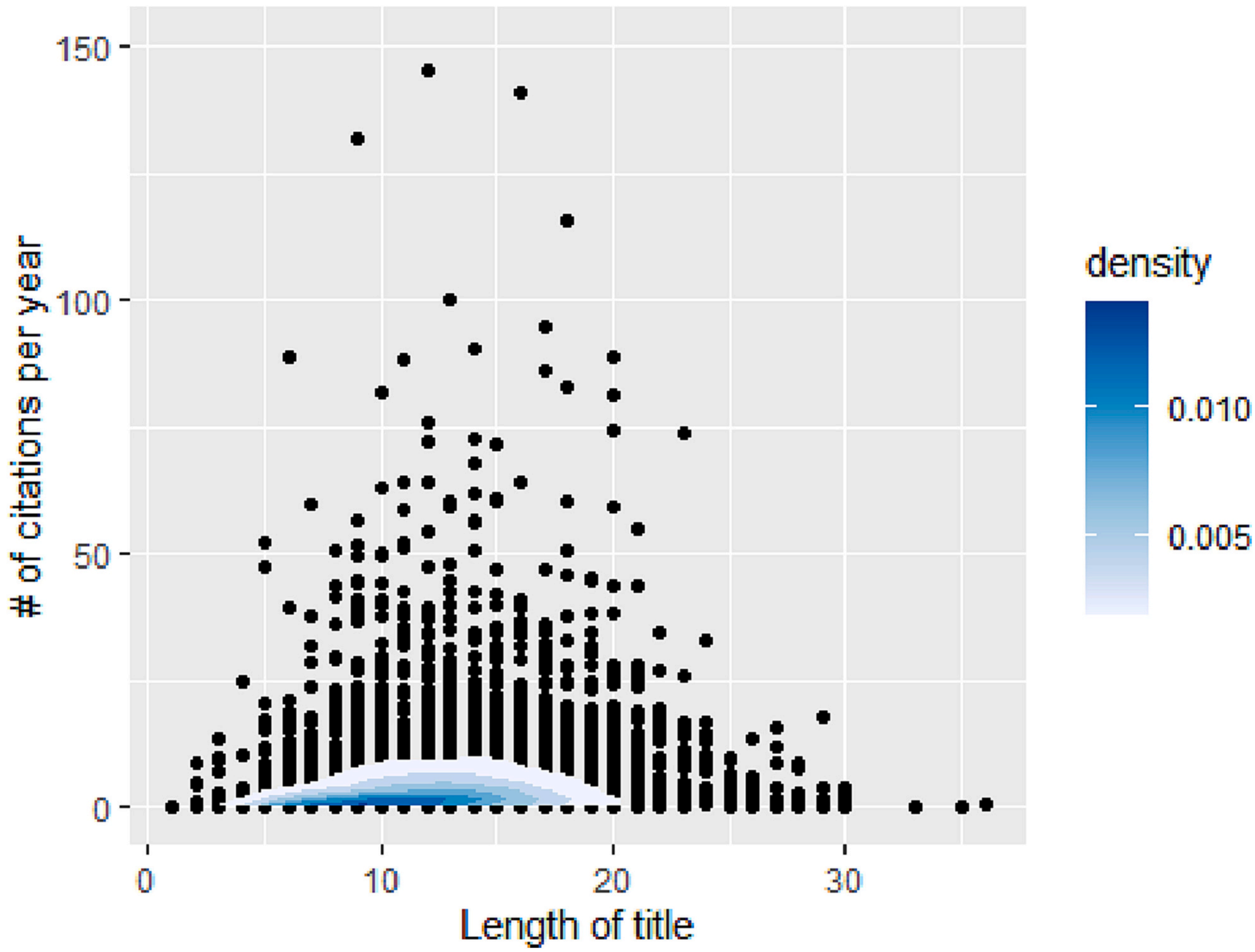

Fig. 6. Relationship between number of citations per year and length of the title.
Note: The 2D-density plot illustrates where most observations are concentrated.

Table 3
Regressing number of citations per year on the length of title and the number of topics covered.

	OLS					Quantile regression (50th percentile)				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Constant	4.96***	−363.6***	−347.5***	−339.8***	−229.5***	−2.41***	−182.3***	−170.7***	−171.1***	−134.3***
	(0.25)	(1.81)	(1.96)	(1.98)	(2.27)	(0.11)	(3.99)	(3.76)	(4.24)	(4.11)
Number of topics in a study	0.29	0.40**	0.41**	0.42**	2.33***	0.04	0.15***	0.14***	0.12***	2.13***
	(0.21)	(0.21)	(0.21)	(0.21)	(0.86)	(0.10)	(0.05)	(0.04)	(0.03)	(0.56)
Year of publication		0.18***	0.17***	0.17***	0.12***		0.09***	0.09***	0.09***	0.07***
		(0.01)	(0.01)	(0.01)	(0.01)		(0.002)	(0.002)	(0.002)	(0.002)
Length of the title (in words)			0.06**	0.35***	0.25**			0.05***	0.13***	0.07***
			(0.03)	(0.12)	(0.12)			(0.007)	(0.03)	(0.02)
Length of the title squared (in words)				−0.010**	−0.012**				−0.003**	−0.003**
				(0.004)	(0.004)				(0.001)	(0.001)
Topic dummies included	No	No	No	No	Yes	No	No	No	No	Yes
R squared	0.01	0.06	0.06	0.07	0.10	0.01	0.09	0.09	0.09	0.10

Note: Standard errors reported in parentheses. *p < 0.10; **p < 0.05; ***p < 0.01. Topic dummies refer to dummy variables equalling 1 whenever a given topic is found to have a prevalence beyond a 25 % threshold and zero otherwise.

while the number of components increased from 3 to 6 and 4, respectively. In the first period managed by Hal Linstone, the largest connected component encompassed 21 topics, and this was facilitated not only by linkages with "traditional" topics in TFSC like T1, T5 and T6 but also by - at that time emerging - topics T17 on China and T22 on triple helix (all these topics had eight edges or more in the graph). Studies devoted to China tended to be focused more in that period on topics related to financial markets (T18), economic growth (T11), patent analysis (T20) and transport (T24), among others. This graph also illustrates that many emerging topics in TFSC (such as on sustainable development or environmental regulation) have been originally developed in the journal not isolated but in close connection to one or several more traditional themes (like technology roadmapping or scenario foresight).

In the second period governed by Fred Philipps we observe traditional topics, T1 and T6 in particular, losing their share in favour of new topics like knowledge acquisition (T12), innovation diffusion (T10) and

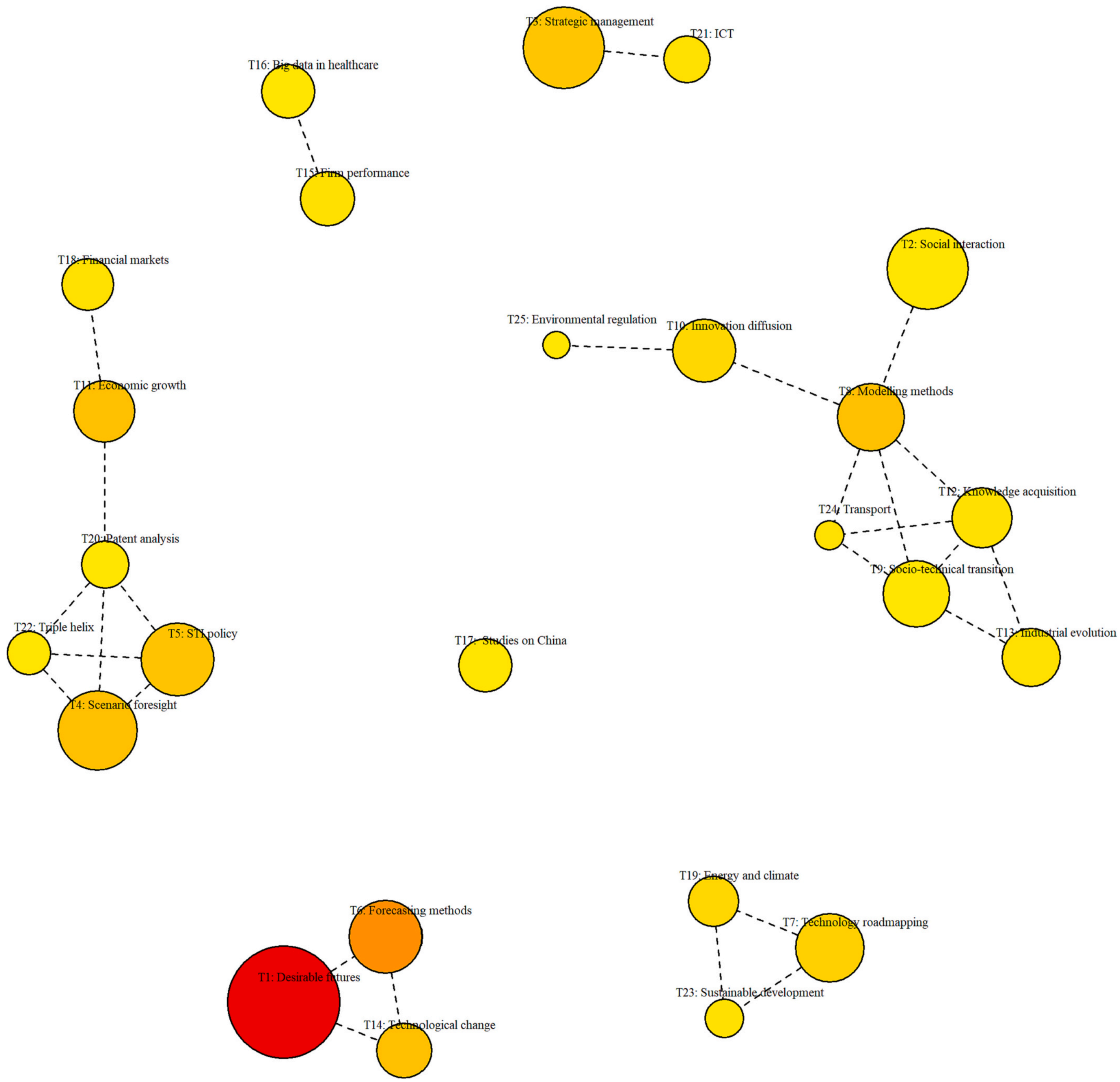

Fig. 7. Network of topics for the period 1970–2022.
Note: The size of the nodes represents the percent of the corpus of textual documents belonging to the topic, while the presence of an edge captures the fact that the two topics tend to significantly co-occur in TFSC publications. Darkness of the colour corresponds to the share of topics in the period 1970–2004 indicating more traditional themes in TFSC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

patent analysis (T20), which become also among the most connected and central topics in that period. The lower density of the graph indicates that many topics start to be developed more independently (for example, T15 on firm performance, T16 on big data in healthcare and T18 on financial markets). The earlier discussed topic on China (T17) is recombined systematically with fewer topics: T11 on economic growth, T12 on knowledge acquisition and T13 on industrial evolution. This reminds the result obtained by Savin et al. (2022b) for patents in service robotics: when this class of technologies has been nascent in the 1970s–1980s, it was developed in close connection to more established hardware and software technologies of industrial robots; but decades after when these service robotic technologies matured, they started to be developed independently and focus on specific application areas (like medicine, logistics, agriculture).

Considering TFSC publications in the last a couple of years under the joint management by Scott Cunningham and Mei-Chih Hu, we observe formation of a separate component devoted to modelling various dynamic processes (recall the largest component on Fig. 7) which is not anymore connected either to "traditional" topics of TFSC on forecasting or topics related to technological and economic trends (T4 on scenario foresight and T5 on STI policy). This component illustrates formation of a new focus in TFSC coverage. At the same time, to the largest topics in the knowledge graph now belong T2 on social interaction, T17 on China (here closely interconnected with the cluster on modelling transition processes) and T18 on financial markets. The latter relates technology (STI policy, patent analysis) and environmental (energy and climate,

Fig. 8. Networks of topics for the periods 1970–2004, 2006–2019 and 2021–2022.
Note: The size of the nodes represents the percent of the corpus of textual documents belonging to the topic, while the presence of an edge captures the fact that the two topics tend to significantly co-occur in TFSC publications. Darkness of the colour corresponds to the share of topics in the period 1970–2004 indicating ˙traditional˙ themes in TFSC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sustainable development) topics demonstrating the growing role to novel financial instruments (e.g., green finance, digital currencies).

To sum up, looking on the topics in a graph like presented in Fig. 8 one can consider topics not as isolated but in combination with other themes with which the given topics tended to be recombined over time: for example, the topic on sustainable development has been originally mostly discussed together with economic growth (see for example Simon, 1981 doing empirical analysis of the role of population growth on scarcity of resources), later - with energy and climate (exemplified by Hejazi et al., 2014 discussing global water projections using integrated assessment modelling), and most recently - with technological change (see Cascante et al., 2022 for a case study of aquaculture innovation). This helps to better understand the discourse of distinct topics in the TFSC journal.

4. Conclusion

To conclude, we identified 25 main themes published in TFSC over more than fifty years of its existence ranging from such “traditional” topics like scenario foresight and forecasting methods that dominated the journal agenda in the first decades through innovation diffusion and patent analysis that gained popularity in 2006–2019 to social interaction and financial markets which experienced momentum in the last couple of years. We discuss how these topics vary in terms of countries that have contributed most and in terms of citations these topics attracted. It turns out that studies concentrated on more recent topics like firm performance, financial markets and environmental regulation have been cited much more frequently and were contributed more often by scientists from China compared to the US. Interestingly, we find in TFSC an inverse U-shaped relationship between the number of citations and the length of title: studies with titles around 20 words long receive most citations. This contradicts earlier findings obtained in other journals demonstrating that shorter (and simpler) title is not always better. Furthermore, we construct a graph of TFSC topics, both for the overall sample of 6240 studies reviewed and the three periods of TFSC existence corresponding to different editors-in-chief. Having demonstrated that studies recombining two or more topics are more impactful in terms of citations, we illustrate which pairs of themes have been predominantly addressed depending on the period of its existence.

Our results can serve as guidance for future contributors of the TFSC journal and to the journal's editorial board helping to derive ideas on how to complement existing themes in a novel way to understand the evolution of the journal. Our paper further illustrates the tools from computational linguistics like STM facilitating fast and objective analysis of large number of studies, and how these can be further enriched with methods from graph theory and scientometrics.

Data availability

Data will be made available on request.

Acknowledgement

Ivan Savin acknowledges support from the ERC Advanced Grant from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (grant agreement no. 741087).

Appendix A

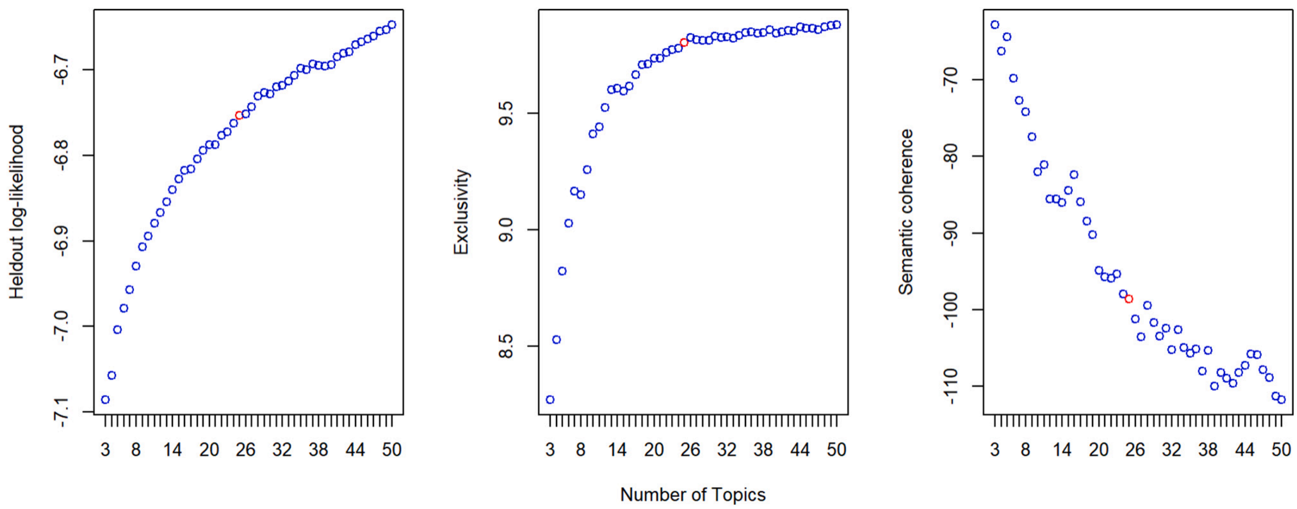

Fig. A1. Model performance depending on the number of topics.

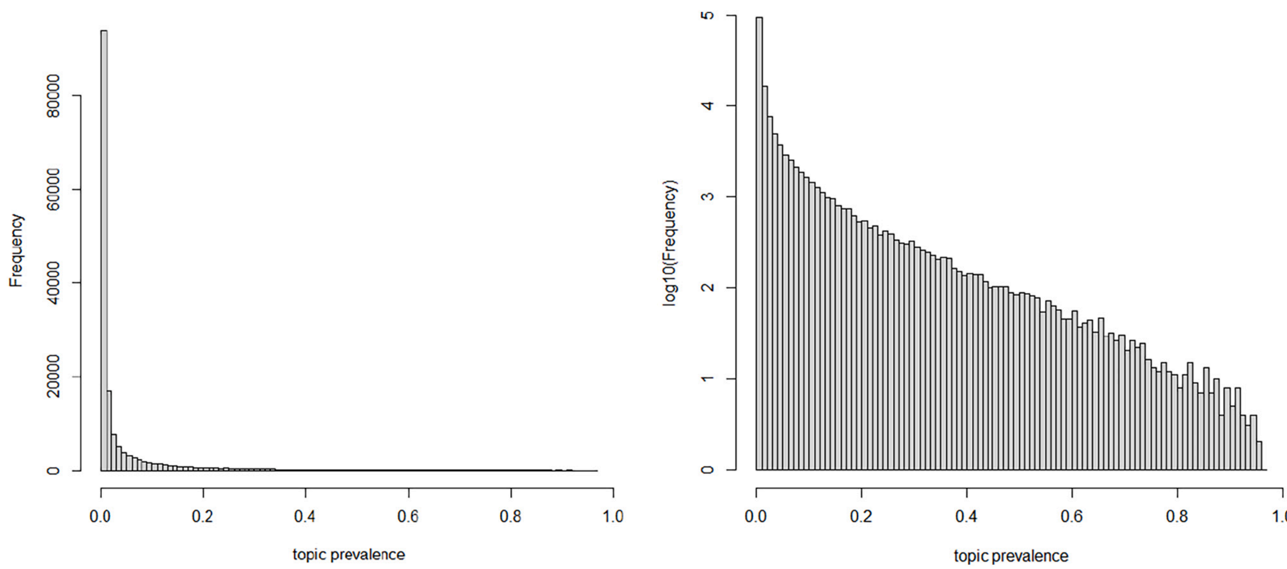

Fig. A2. Distribution of the resulting topic prevalences among papers (right plot is in log10 scale).

Table A1
Share of the main topics for researchers affiliated in the countries with highest contribution to TFSC.

	USA	UK	China	France	Italy	Germany	Spain	Netherlands	South Korea
T1	11.9	6	1.2	6.4	3.3	5.3	3.5	6.5	2.1
T2	3	8.1	5.5	9.4	5	5.7	8.4	3	6.7
T3	5.9	6.1	3.7	4.9	5.4	5.9	5.5	5.3	4.4
T4	5.4	7.8	1	4.4	4.3	12	5	6	2.7
T5	4.8	4.7	3.9	4.2	5.6	4.1	5.6	4.9	5.2
T6	8	2.1	2.7	1.9	3.5	2.6	2.2	3	5.7
T7	5.1	4.6	3.4	4.2	4	4.7	2.8	5.9	7.5
T8	4.7	2.9	4.1	3.4	4.7	3.5	3.5	3.9	5.4
T9	2.5	7.7	2.1	6.1	4.3	6.6	3.5	12.1	2.9
T10	3.6	4.1	3.9	4.4	5.2	4.9	5.4	6.2	4.9
T11	5.3	3	2.6	2.5	3.8	2.9	2.8	4.5	2.9
T12	2.7	4.7	4	6.3	6.6	3.6	6.5	4	3.6
T13	2.7	4.6	3.4	5.9	6.2	4.5	3.9	2.7	3.7
T14	5.5	2.8	3	2.8	3.5	2.9	2.6	3.6	2.5
T15	2.5	3.6	6.2	4.9	4.4	2.3	7.6	2	3.5
T16	4.3	4	3.2	3.9	4.6	3	3.5	3.6	4.2
T17	2.5	1.9	18.1	1.6	1.8	1.8	1.8	1.7	2.7
T18	2.3	3.4	6.4	4.4	3.6	3.1	4.4	1.7	2.9
T19	3.8	3.3	4.3	3	2.8	3.5	1.8	4.4	1.2
T20	2.8	1.7	4	2.3	3.8	4.8	2.9	2.1	11.5
T21	3.2	3	2.4	2.8	3	2.4	4.7	2.1	7.6
T22	2.1	3.7	1.5	3.8	4.4	2.1	5.9	3.2	1.6
T23	2.3	2.2	3.3	2.5	2.4	2.8	2.5	3.9	1.1
T24	1.7	2.1	1.7	1.3	2	3.2	1.6	2.5	2.2
T25	1.4	1.7	4.3	2.6	1.7	1.7	2.3	1.3	1

Note: Results are provided in percentages so that each column (country) sums up to 100. Colour corresponds to the percentage value (ranging from red for small values to green for large values).

References

Aggarwal, C.C., 2018. Text preparation and similarity computation. In: Machine Learning for Text. Springer, New York, pp. 17–30.

Ambrosino, A., Cedrini, M., Davis, J., Fioria, S., Guerzoni, M., Nuccio, M., 2018. What topic modeling could reveal about the evolution of economics. J. Econ. Methodol. 25 (4), 367–377.

Amer, M., Daim, T.U., 2010. Application of technology roadmaps for renewable energy sector. Technol. Forecast. Soc. Chang. 77 (8), 1355–1370.

Ashraf, R., Khan, M.A., Khuhro, R.A., Bhatti, Z.A., 2022. Knowledge creation dynamics of technological forecasting and social change special issues. Technol. Forecast. Soc. Chang. 180, 121663.

Asmussen, C.B., Møller, C., 2019. Smart literature review: a practical topic modelling approach to exploratory literature review. J. Big Data 6 (1), 1–18.

Belhadi, A., Kamble, S., Jabbour, C.J.C., Gunasekaran, A., Ndubisi, N.O., Venkatesh, M., 2021. Manufacturing and service supply chain resilience to the COVID-19 outbreak: lessons learned from the automobile and airline industries. Technol. Forecast. Soc. Chang. 163, 120447.

Blei, D.M., 2012. Probabilistic topic models. Commun. ACM 55 (4), 77–84.

Bouma, G., 2009. Normalized (pointwise) mutual information in collocation extraction. In: Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology, GCSL 09.

Bystrov V., Naboka V., Staszewska-Bystrova A. and P. Winker (2023). Choosing the number of topics in LDA models - a Monte Carlo comparison of selection criteria, arXiv 2212.14074.

Callaghan, M.W., Minx, J.C., Forster, P.M., 2020. A topography of climate change research. Nat. Clim. Chang. 10 (2), 118–123. https://doi.org/10.1038/s41558 019 0684.

Cascante, M.Q., García, N.A., Fold, N., 2022. The role of external forces in the adoption of aquaculture innovations: an ex-ante case study of fish farming in Colombia's southern Amazonian region. Technol. Forecast. Soc. Chang. 174, 121185.

De Battisti, F., Ferrara, A., Salini, S., 2015. A decade of research in statistics: a topic model approach. Scientometrics 103 (2), 413–433. https://doi.org/10.1007/s11192-015-1554-1.

Donbesuur, F., Ampong, G.O.A., Owusu-Yirenkyi, D., Chu, I., 2020. Technological innovation, organizational innovation and international performance of SMEs: the moderating role of domestic institutional environment. Technol. Forecast. Soc. Chang. 161, 120252.

Fellbaum, C.D., 2005. WordNet and wordnets. In: Brown, K. (Ed.), Encyclopedia of Language and Linguistics, Second edition. Elsevier, Oxford, pp. 665–670.

Griffith, T., Steyvers, M., 2004. Finding scientific topics. Proc. Natl. Acad. Sci. U. S. A. 101, 5228–5235.

He, Q., Chen, B., Pei, J., Qiu, B., Mitra, P., Giles, C.L., 2009. Detecting topic evolution in scientific literature: how can citations help? CIKM'09. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, pp. 957–966.

Hejazi, M., Edmonds, J., Clarke, L., Kyle, P., Davies, E., Chaturvedi, V., Wise, M., et al., 2014. Long-term global water projections using six socioeconomic scenarios in an integrated assessment modeling framework. Technol. Forecast. Soc. Chang. 81, 205–226.

Kajikawa, Y., Mejia, C., Wu, M., Zhang, Y., 2022. Academic landscape of Technological Forecasting and Social Change through citation network and topic analyses. Technol. Forecast. Soc. Chang. 182, 121877.

Krammer, S.M.S., 2017. Science, technology, and innovation for economic competitiveness: the role of smart specialization in less-developed countries. Technol. Forecast. Soc. Chang. 123, 95–107.

Letchford, A., Moa, H.S., Preis, T., 2015. The advantage of short paper titles. R. Soc. Open Sci. 2, 150266.

Li, L., 2018. China's manufacturing locus in 2025: with a comparison of "made-in-China 2025" and "industry 4.0". Technol. Forecast. Soc. Chang. 135, 66–74.

Lüdering, J., Winker, P., 2016. Forward or backward looking? The economic discourse and the observed reality. Jahrb. fur Natl. Stat. 236 (4), 483–515.

Luo, J., 2021. Forecasting COVID-19 pandemic: unknown unknowns and predictive monitoring. Technol. Forecast. Soc. Chang. 166, 120602.

Mahmood, N., Zhao, Y., Lou, Q., Geng, J., 2022. Role of environmental regulations and eco-innovation in energy structure transition for green growth: evidence from OECD. Technol. Forecast. Soc. Chang. 183, 121890.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., Adam, S., 2018. Applying LDA topic modeling in communication research: toward a valid and reliable methodology. Commun. Methods Meas. 12 (2–3), 93–118.

Marson, M., Savin, I., 2015. Ensuring sustainable access to drinking water in sub-Saharan Africa: conflict between financial and social objectives. World Dev. 76, 26–39.

Mo, Y., Kontonatsios, G., Ananiadou, S., 2015. Supporting systematic reviews using LDA based document representations. Syst. Rev. 4 (1).

Nayak, B., Bhattacharyya, S.S., Krishnamoorthy, B., 2021. Explicating the role of emerging technologies and firm capabilities towards attainment of competitive advantage in health insurance service firms. Technol. Forecast. Soc. Chang. 170, 120892.

Ominde, D., Godfrey Ochieng, E., Oteke Omwenga, V., 2021. Optimising ICT infrastructure performance in developing countries: Kenyan viewpoint. Technol. Forecast. Soc. Chang. 169, 120844.

Pasaoglu, G., Harrison, G., Jones, L., Hill, A., Beuadet, A., Thiel, C., 2016. A systems dynamics based market agent model simulating future powertrain technology transition: scenarios in the EU light duty vehicle road transport sector. Technol. Forecast. Soc. Chang. 104, 133–146.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., Rand, D.G., 2014. Structural topic models for open-ended survey responses. Am. J. Polit. Sci. 58 (4), 1064–1082.

Roberts, M.E., Stewart, B.M., Tingley, D., 2019. STM: an R package for structural topic models. J. Stat. Softw. 91 (2), 1–40. https://doi.org/10.18637/jss.v091.i02.

Saracco, F., DiClemente, R., Gabrielli, A., Squartini, T., 2015. Randomizing bipartite networks: the case of the world trade web. Scientific Reports 5, p. 10595.

Savin, I., 2021. On optimal regimes of knowledge exchange: a model of recombinant growth and firm networks. J. Econ. Interac. Coord. 16, 497–527.

Savin, I., Teplyakov, N., 2022a. Using computational linguistics to analyse main research directions in Economy of Regions. Econ. Reg. 18 (2) https://doi.org/10.17059/ekon.reg.2022-2-3.

Savin, I., Teplyakov, N., 2022b. Topics of the nationwide phone-ins with Vladimir Putin and their role for public support and Russian economy. Inf. Process. Manag. 59 (5), 103043.

Savin, I., van den Bergh, J., 2021. Main topics in EIST during its first decade: a computational linguistic analysis. Environ. Innov. Soc. Transit. 41, 10–17.

Savin, I., Drews, S., Mestre Andrés, S., van den Bergh, J., 2020. Public views on carbon taxation and its fairness: a computational linguistics analysis. Clim. Chang. 162, 2107–2138.

Savin, I., Drews, S., van den Bergh, J., 2021. Free associations of citizens and scientists with (green) economic growth: a computational linguistics analysis. Ecol. Econ. 180, 106878 https://doi.org/10.1016/j.ecolecon.2020.106878.

Savin, I., Chukavina, K., Pushkarev, A., 2023. Topic-based classification and identification of global trends for startup companies. Small Bus. Econ. 60, 659–689. https://doi.org/10.1007/s11187-022-00609-6.

Savin, I., Ott, I., Konop, C., 2022b. Tracing the evolution of service robotics: insights from a topic modeling approach. Technol. Forecast. Soc. Chang. 174, 121280.

Schoenmakers, W., Duysters, G., 2010. The technological origins of radical inventions. Res. Policy 39 (8), 1051–1059.

Shibata, N., Kajikawa, A., Sakata, I., 2011. Measuring relatedness between communities in a citation network. J. Assoc. Inf. Sci. Technol. 62, 1360–1369.

Simon, J.L., 1981. Resources, population, environment: an oversupply of false bad news. Technol. Forecast. Soc. Chang. 19 (3), 207–225.

Speier, W., Ong, M.K., Arnold, C.W., 2016. Using phrases and document metadata to improve topic modeling of clinical reports. J. Biomed. Inform. 61, 260–266.

Tvinnereim, E., Fløttum, K., 2015. Explaining topic prevalence in answers to open-ended survey questions about climate change. Nat. Clim. Chang. 5, 744–747.

Uglanova, I., Gius, E., 2020. The order of things. A study on topic modelling of literary texts. In: CHR 2020: Workshop on Computational Humanities Research, pp. 57–76. Amsterdam, The Netherlands.

Voutilainen, A., 2003. Part-of-speech tagging. In: Mitkov, R. (Ed.), The Oxford Handbook of Computational Linguistics. Oxford University Press Inc, New York, pp. 219–232.

Wang, Y., Kung, L.A., Byrd, T.A., 2018. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technol. Forecast. Soc. Chang. 126, 3–13.

Youn, H., Strumsky, D., Bettencourt, L., Lobo, J., 2015. Invention as a combinatorial process: evidence from US patents. J. R. Soc. Interface 12 (20150), 272.

Zhu, L., Cunningham, S.W., 2022. Unveiling the knowledge structure of technological forecasting and social change (1969–2020) through an NMF-based hierarchical topic model. Technol. Forecast. Soc. Chang. 174, 121277.

Ivan Savin is an Associate Professor of Business Analytics at the ESCP Business School, Madrid campus. Between 2018 and 2023 he served as a postdoctoral research fellow within the ERC project “Behavioral-evolutionary analysis of climate policy: Bounded rationality, markets and social interactions” (EVOCLIM) at ICTA-UAB, and earlier also within the KIT-BETA project devoted to general purpose technologies, creativity and sustainability (2015-2018), and was a postdoc for four years at the Friedrich Schiller University and the Graduate College "The Economics of Innovative Change" (2011-2015). He received his PhD from the Department of Economics of the Justus Liebig University of Giessen (Chair of Econometrics, supervisor Prof. Dr. Peter Winker) in 2011. In 2017 he habilitated at KIT with the thesis titled "Applications of Evolutionary Optimization and Modelling in Economics". Given all that, his research interest naturally include agent-based models, economics of innovation and climate policy analysis.