

Article

Invention in Times of Global Challenges: A Text-Based Study of Remote Sensing and Global Public Goods

Ingrid Ott ^{1,*} and Simone Vannuccini ²¹ Chair of Economic Policy, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany² Groupe de Recherche en Droit, Économie et Gestion (GREDEG) Université Côte d'Azur, Sophia Antipolis Cedex, 06410 Biot, France; simone.vannuccini@univ-cotedazur.fr

* Correspondence: ingrid.ott@kit.edu

Abstract: We study whether remote sensing (RS), a set of technologies with global reach and a variety of applications, can be considered instrumental to the provision of global public goods (GPG). We exploit text information from patent data and apply structural topic modeling to identify topics related (or relevant) to GPG provision, and trace their participation in the evolution of remote sensing technology over time. We develop a new indicator of affinity to GPG (and other themes) using meta information from our dataset. We find that, first, RS displays features of a general-purpose technology. Second, while peripheral, GPG-relevant topics are present in the RS topic space, and in some cases overlap with topics with high affinity in AI and participation of public sector actors in invention. With our analysis, we contribute to a better understanding of the interplay between the dynamics of technology and (global) political economy, a field of research yet under-explored.

Keywords: remote sensing; global public goods; patents; unsupervised ML; structural topic modeling; text as data



Citation: Ott, Ingrid, and Simone Vannuccini. 2023. Invention in Times of Global Challenges: A Text-Based Study of Remote Sensing and Global Public Goods. *Economies* 11: 207. <https://doi.org/10.3390/economies11080207>

Academic Editor: Ralf Fendel

Received: 14 June 2023

Revised: 19 July 2023

Accepted: 21 July 2023

Published: 2 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It is not the first time in the history of humanity that societies are riddled with tensions and profound transformations. The very globalization of trade emerged at regular intervals throughout modern history (Arrighi 1994). However, the global scale of challenges that humankind faces nowadays is an absolute novelty: we are navigating a ‘polycrisis’ landscape, while, at the same time, being enveloped in a techno-economic paradigm shaped by information and communication technologies (Lombardi and Vannuccini 2022). The polycrisis we face is the convergence of several interconnected global challenges, ranging from climate change to (geo)political and economic vulnerabilities. The solution of such challenges requires global coordination amongst actors of different types (international institutions, nations, and transnational organizations), aimed at the provision of global public goods (GPG) and dealing with global commons. As classic public goods, GPGs are non-excludable and non-rival; however, they affect population at the world scale. While the coordination failures emerging when contributing to GPGs are relatively well studied (Buchholz and Sandler 2021), we know much less about how technology might facilitate the provision of GPGs. In other words, can a given technology be instrumental to the provision of GPG? In this paper, we explore a specific set of technologies—remote sensing (RS)—whose uses are multi-scale in nature and potentially directed at areas connected with public good provision.

RS technology is a type of data or information acquisition technology (Savona et al. 2022). It can be defined as “the acquisition of information about an object or phenomenon without making physical contact with the object, in contrast to in situ or on-site observation”.¹ In brief, the activity of RS employs sensor technology for detection at a distance. The global scope of RS applications makes it an interesting case to study the interplay of technology dynamics and GPGs. Furthermore, the tremendous cost decline of sensors during the

last two decades together with miniaturization have strongly fostered the diffusion of the technology across a wide array of domains, making RS a good candidate to be a general purpose technology (GPT).

Given its features and applications, we hypothesize that RS can play an instrumental role in the provision of GPGs. In fact, some of the technical challenges related to GPGs are strongly dependent on information acquisition from a distance. For example, protecting ecosystems, preserving biodiversity, monitoring climate change, addressing flows of refugees, or identifying terrorist threats, just to name a few, are activities that rely (or might rely) upon RS. The technology has the potential to be a mediator of incentives; hence, a better understanding of the interplay of RS development and GPGs can influence decision making, and direct attention and resources to develop or reinforce beneficial RS applications.

It remains an open question whether the potential of RS to intervene in and ease the provision of GPGs can be detected in the very patterns of evolution of the technology. To assess that, we conduct a series of exploratory exercises, using textual information from relevant patent abstracts. We employ structural topic modeling (STM) to draw a granular picture of RS nature and evolution. We identify a series of topics capturing essential features of RS technology, functions, and applications. We then map their evolution to identify structural shifts, and use features of STM to assess the affinity of RS topics to important themes—GPG in the first stance, but also artificial intelligence (AI), the role of public and private actors in RS development, and the international specialization of certain actors (i.e., China) on specific topics.

We find that the direction of inventive activities turns increasingly towards algorithms and the integration of modern methods of data/ image acquisition and analysis, and also that new applications emerge—making the case for looking at RS through the lenses of GPT. Most importantly, our results suggest that RS development features GPG-related elements, which appear in the topic space, even though they are yet peripheral compared to topics mostly including technical terms—an expected result, as patent information has a strong technical focus, rather than an application one. Still, themes related to GPGs play a role in RS technological development. Interestingly, top-ranking topics in what we label the GPG-affinity indicator are, in some cases, overlapping with topics with affinity to AI as well as to public actors, pointing at the fact that a share of technical progress in RS is dedicated to GPG-relevant trajectories. Therefore, we conclude that RS is a technology with general-purpose features and the capability to play a role as a tool in the provision of GPGs.

The paper offers two major contributions. The first is a novel angle of research: to our knowledge, this is the first study that combines a fine-grained analysis of technological evolution with a key theme of political economy, such as GPGs. By connecting these two worlds, we suggest that they can mutually gain: on the one hand, the reach of political economy of GPG can expand beyond the study of incentive structures to delve quantitatively into the role that technology can play to facilitate (or obstacle) the pursuit of global welfare. On the other hand, the economics of technology can begin looking beyond issues of production, adoption, and mere economic impact and explore the quasi-normative implications of the technological evolution on global, humanity-defining dynamics.

The second major contribution is methodological: exploiting features of STM in an innovative way, we generate metadata based on selected terms and use these to build covariates that are specific for the analysis. We label these covariates ‘affinities’ (in our case, to GPG and AI). This is a novel type of indicator, rather flexible and information-rich, that can be reused in further studies.

The paper proceeds as follow: in Section 2, we summarize the theoretical building blocks and the empirical approach we use in the study; namely, we introduce GPT, GPG, and STM. In Section 3, we set the groundwork for our analysis by presenting the dataset, offering a descriptive view on specific relevant terms and topics from our corpus, and setting up the topic modeling exercise. In Section 4, we discuss our findings. Section 5

concludes the paper. We thus contribute to a perspective that increasingly is taken into account in the context of technology analysis.

2. A Framework to Study Remote Sensing

Our goal is to understand the interplay of technology evolution and political economy using the case of RS. In order to conduct our analysis, we outline the theoretical and empirical building blocks we operate with. On the theoretical side, we place RS under the umbrella of the literature studying general purpose technology and global public goods, as both concepts offer interpretative angles to map our phenomenon of interest. On the empirical side, we introduce (structural) topic modeling as a technique to capture the nuances of technological evolution from text data.

2.1. Theoretical Building Blocks

General-purpose technology (GPT). GPTs are well identifiable, usually stand-alone technologies (or clusters of complementary technologies working in close synergy) that are used as an input or tool in a wide range of economic activities. Examples of GPTs are the steam engine, computer platforms, lasers, and especially integrated circuits (Cantner and Vannuccini 2012; Menz and Ott 2011). These technologies are potentially at the core of broader transformations in the whole economic system, and the most radical amongst them set the stage for successive technological eras (Knell and Vannuccini 2012). GPTs are identified through a series of characteristics that make them a peculiar family of breakthrough innovations. The literature discusses several of these—including the absence of technologies that can act as GPT substitutes, or their non-linear impact on productivity (Bekar et al. 2018)—but it tends to converge on three fundamental key features (Bresnahan and Trajtenberg 1995): general applicability (or pervasiveness), technological dynamism, and innovational complementarity (or spawning). General applicability is what allows the use of a technology at scale and across a variety of sectors: a GPT is pervasive because it does “nothing specific” (Simon 1987); in other words, it provides a generic function (i.e., computation, motion and control) to a wealth of user sectors. Technological dynamism refers to the steep learning curve of the technology, as, once realized, the demand for the GPT coming from different sources pushes down its production cost (and pushes up performance) drastically. Finally, innovation spawning captures the enabling capability of GPTs: the use of this technology lowers the barrier (or raises the returns) to conducting innovative activities and, in some cases, opens up new directions and approaches to innovate (i.e., digitization of business models favors firm entry into the app market).

How is the GPT concept relevant to RS? Thoma (2009) discussed the GPT nature of control technologies that enable process automation. In turn, control technologies are enabled by sensors, the same technological component at the core of RS technologies. This suggests that the RS nature might be well approximated by the GPT concept, and its development ‘read’ through this lenses. We posit that RS technologies have general-purpose features. First, they perform a generic function, that of information acquisition from a distance. Second, this generic function is put at work in a variety of rather heterogeneous application sectors that require measurement and surveying, from meteorology to intelligence and military uses. Third, even though the technology is not new *per se* (for instance, our analysis will cover fifty years of RS patenting), its dynamism increases as a function of the expansion of its uses. Fourth, like many other information-acquisition technologies, RS has the capability to be ‘enabling’, that is, to induce specific actions (i.e., innovation) by lowering uncertainty or expanding the choice set. The reason for this is that the use of RS produces, at lower costs and/or at a higher quality/coverage, an output (information) that feeds as an input into decision making. This feature of RS can potentially be welfare enhancing; in turn, given the wide reach of the technology, this welfare effect can be *global*, suggesting that next to generality of purpose, RS can play a role in the provision of GPGs.

Global public goods (GPG). GPGs “may correspond to pure or impure public goods that impact much of the world’s population.” (Buchholz and Sandler 2021, p. 488). They

include “identifying virulent pathogens, ameliorating global financial crises, adopting universal regulatory practices, protecting essential ecosystems, allocating geostationary orbits, diverting earthbound planetesimals, preserving cultural heritage, reversing ozone layer depletion, and curbing climate change. These and other GPGs (e.g., eradicating infectious diseases, developing disease treatment regimes, fostering cybersecurity, preserving biodiversity, reducing transnational terrorism, maintaining world peace, discovering scientific breakthroughs, and addressing refugee flows) represent some of the world’s most pressing problems” (Buchholz and Sandler 2021, p. 489). When non-excludable but rival, GPGs approximate global commons (i.e., worldwide reservoirs of resources, such as oceans). The well-known sustainable development goals (SDGs) pursued by the United Nations (UN) can be considered a subset of GPGs.

What makes GPGs a distinct family of public goods is their complexity, namely the “multi-actor, multi-sector, multilevel nature of their provision path” (Kaul 2012, p. 736), featuring high transaction costs, high risk of coordination failures, and the need to take into account issues of sovereignty influencing their production and provision. RS can play a supporting role in the provision of GPGs. In fact, Buchholz and Sandler (2021, p. 490) point out that “novel monitoring technologies allow humankind to spot some global public bads (GPBs) and GPGs (e.g., the accumulation of atmospheric greenhouse gases (GHGs), the melting of the planet’s icecaps, the health of the world’s forests, the state of the stratospheric ozone shield, and the spread of deserts).”. This can happen because, as we discuss in the next paragraph, RS provides an input (information) that feeds into decision making, expanding the knowledge set or reducing uncertainty. In other words, RS is a technical tool that can be used to facilitate coordination and that can influence the allocation of resources to the production of GPGs.

Outlook: the economics of RS. In economic terms, the use of RS technology produces an input (information) at lower cost or higher quality (i.e., better scale or resolution). In a growth theory framework, RS adoption can be seen as a production function shifter, that is, a form of capital (or labor, depending on the application) technical change. From this perspective, RS can be instrumental to the provision of GPG—*ceteris paribus* international cooperation incentives and mechanisms—as it affects the productivity of production factors and, indirectly, the shape of the choice set for decision makers. An alternative way to look at that is to consider RS as a push outwards to the production possibility frontier (PPF) of activities that rely on information acquired from a distance. Forney et al. (2012) illustrate this for the case of groundwater quality surveying using the RS system Landsat.

In sum, RS appears to be a general-purpose, enabling technology, used by both commercial and non-commercial actors, that can produce sizable social returns beyond private returns and shift outwards the possibility frontier of the applications adopting it. This is due to the fact that the information it provides allows for a better use of existing production factors. When used in applications that have global reach, this information represents a form of ‘Earth intelligence’, which can feed into the production of GPGs. The prominent ‘public’ role of RS is reflected in how the technology is described by actors pursuing missions with a global ‘flavor’. For example, NASA “observes Earth and other planetary bodies via remote sensors on satellites and aircraft that detect and record reflected or emitted energy. Remote sensors, which provide a global perspective and a wealth of data about Earth systems, enable data-informed decision making based on the current and future state of our planet”.² In the medium to long term, this information will also become important when it comes to proving war events and even war crimes.

2.2. Methodological Approach

A further dimension of our framework to study the nexus between technology and political economy through the case of RS is the methodological one. We trace the presence of GPG-related themes along the technological evolution of RS technology by exploiting information contained in patent documents. Patent data are the natural choice to study the evolution of technology. Despite the never-ending debate about their limitations for

economic analysis, patents represent a rich source of data when it comes to unpacking a given technology into its constituent components and techniques. Furthermore, patent data can be exploited to map technology and proximity spaces that capture the interconnections between different ‘quantums’ of knowledge (Alstott et al. 2016).

Patent data can be used as a source of structured and unstructured information. In our analysis, we focus on the exploitation of the more ‘fluid’, unstructured information by parsing the text of patents’ abstracts using structural topic modeling techniques. This approach allows us to cluster terms retrieved from abstracts into topics, to relate the topic with each others (i.e., using network methods) and to study and track them over time. For example, we can focus on topics (and terms) that are GPG relevant in order to understand to what extent this perspective has permeated RS-related inventions.

The scope of research enabled by the use of text as data cannot be understated. Thanks to advances in machine learning techniques and to digital availability of large corpora of text, textual information is increasingly used to address innovation economic questions. A comprehensive overview on text as data in innovation analyses can be found at Paunov et al. (2018). Gentzkow et al. (2019) provide a recent overview on text as data in economics, while Grimmer and Stewart (2013) address the opportunities and shortcomings of automated text analysis in study political questions. Related, recent overviews are provided by Ranaei et al. (2019) and Van Looy and Magerman (2019), who apply text analysis to study the relationship between science and technology based on papers and patents.

Over the last two decades, probabilistic topic models have become a prominent tool both for processing large amounts of text and for measuring latent variables. The most prominent method is latent Dirichlet allocation (LDA; Blei et al. 2003; Griffith and Steyvers 2004). Instead, in our analysis, we employ structural topic modeling (STM). Unsupervised modeling approaches of text classification imply that there is no need for strong a priori assumptions regarding the outcome. STM represents such an approach, which includes metadata in the classification procedure (compare Roberts et al. 2014, 2016b, 2016a).³

STM is a mixed membership model, which implies that the occurrence of topics within documents follows systematic patterns across the whole corpus (Blei and Lafferty 2007; Roberts et al. 2016a). In mixed membership models, documents are not assumed to belong to single topics but to belong simultaneously to several topics, and the topic distributions vary over documents. It also allows to endeavor to seek relationships between different topics across all documents. Generally speaking, STM—like other unsupervised machine learning (ML) algorithms—do not presuppose categories but infer contents from text (see Roberts et al. 2014, who apply STM to open-ended survey questions).

STM applies a so-called bag-of-words approach, i.e., for the analyzed text, the order of words is not considered. Instead, the distribution of words within documents—together with term weights—incorporate essential information on the content of the documents. Topics represent a distribution of words and thus are characterized by the frequent usage of the same vocabulary. The entire corpus can be split in K topics. Another feature of the STM approach is that each document comprises all K topics, though to varying shares.⁴ For a given corpus, the distribution of topics, words, and documents can be estimated using Bayesian statistics techniques. Compared to LDA, STM allows for topic distributions to depend upon covariates, which can be selected by the researcher: a feature we will exploit in our analysis.⁵

Aside from covariate information, another key decision to be made refers to the number of topics chosen K , which is not predetermined but must be chosen based on either prior knowledge of the research context or with the help of statistical indicators/diagnostics. In general, high topic numbers increase the validation need while at the same time introducing topics that are too fine and granular to be interpretable in a reasonable manner. Overall, the decision for the topic number faces a trade-off between the topics’ separability/exclusivity and their semantic coherence, the latter covering the co-occurrence of words (Mimno et al. 2011). Exclusivity, in contrast, measures how exclusive a term within a topic is compared to other topics (Airoldi and Bischof 2016).

The advantages of automatic content analysis come at the cost that it needs careful validation of the results, a process that cannot be performed automatically. Nelson (2003) argues that automated text analysis is complementary to human knowledge, a statement which is especially true when dealing with unstructured data and latent topics. However, this combination also opens up new research perspectives, such as the proposed link between technology development and GPGs.

3. Empirical Setup

3.1. Overview of the Dataset and RS Technological Profile

Our analysis is based on patent data. Relevant patents are identified by applying a truncated keyword search for the term ‘remote sens*’ in the PATSTAT 2021a database for the period 1963–2020. Patents are selected if the search term appears in the title OR in the abstract.⁶ We only consider patent cooperation treaty (PCT) filings and only the first member of a patent family; moreover, for patents with the first filing at the China Patent Office (SIPO), we restrict our selection only to those patents with a family size of at least 2.⁷ Our final sample includes 2.247 unique patent IDs and 2.189 unique abstracts from 24 authorities; we label these ‘international patents’. Table 1 summarizes descriptive information of our sample. The column ‘all patents’ lists all RS PCT patents independent of their family size. It includes 6.618 patents being filed at SIPO only, i.e., with family size 1; the column ‘international patents’ refers to all filings but only include Chinese patents if their family size exceeds one. For the analysis, we use the unique abstract texts of international patents and restrict the period to 1970–2018. This leaves us with 2.186 patents (i.e., we lose 3 patents due to restricting the period).

Table 1. Sample descriptives of the dataset.

	All Patents	International Patents
unique patent ids	8.865	2.247
unique abstracts	8.807	2.189
unique titles	8.739	2.159
period	1963–2020	1963–2020
authorities	24	24

To ground our analysis, we first explore some properties of the final dataset; in particular, we present filing dynamics and technology classes’ coverage. Concerning filing dynamics, Figure 1a shows the evolution of our international patents selection. Patenting in RS increases steadily over time—except for a slump between 2005 and 2010 followed by an even stronger increase—as the technology develops and diffuses. Overall, the US is the main patent authority represented in the sample. However, since 2012, Chinese international filings (labeled as ‘CN’) have been shaping the worldwide dynamics. This can be seen in Figure 1b, which tracks filings from a more continental perspective, distinguishing different ‘global regions’ and major authorities.⁸ Despite filings being dominated by the US, Asia and WO have been catching up, with a boost given by the more recent entry of CN into invention. Europe’s dynamics is characterized by steady growth until 2005; after a peak, it experiences a decline.

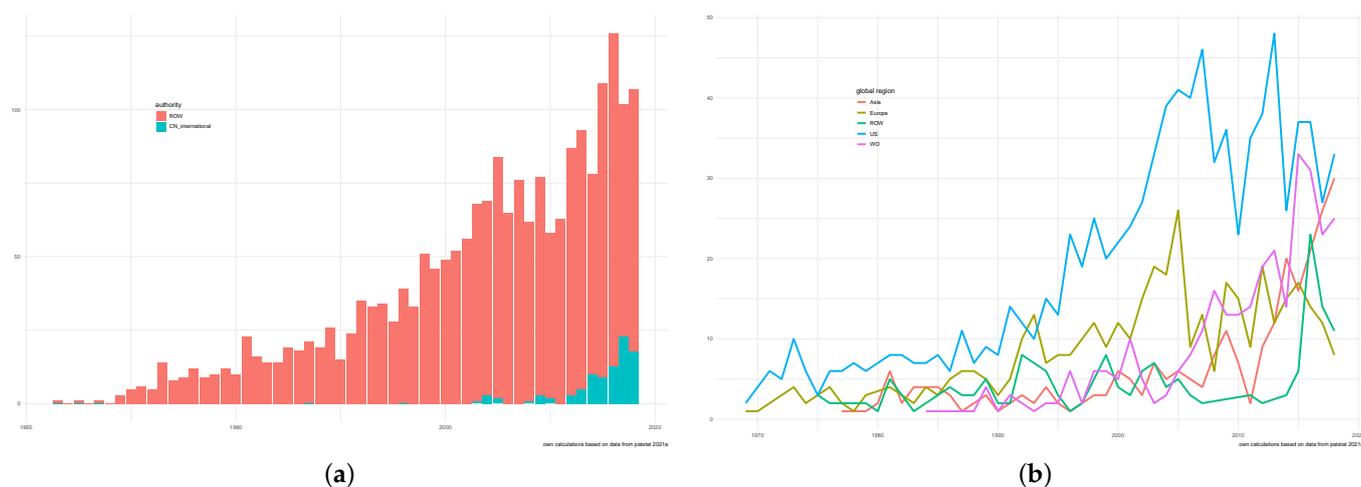


Figure 1. Dynamics of PCT filings world wide and by major global regions and major authorities. CN patents are only included if their family size is at least 2. Compare the appendix for details on the grouping of countries into ‘global regions’. (a) Evolution of international filings in remote sensing: blue color—CN international filings; red color—all other patents. (b) Evolution of PCT filings in remote sensing technologies seen from the perspective of ‘global regions’ and major authorities: US, Asia (including CN international patents), Europe, WO and ROW.

To capture the technological profile of RS, we can start by exploiting the structured part of patent data, and look at the CPC classes most listed in the patents. These are G06K (graphical data reading), G06T (image data processing or generation), and G01N (investigating or analyzing materials by determining their chemical or physical properties). The classes most cited in RS patents relate to technological components and functions of the technology. This is not surprising, as, to a larger extent, patents are meant to cover information on technical progress in RS. Hence, structured information might overlook applications that have GPG relevance.

Notwithstanding that, some preliminary insights related to our frameworks of analysis can already be drawn. For example, over time, RS-related patents cover an increasing variety of technology classes. This is one of the, admittedly rough, measures characterizing GPTs in the making (Hall and Trajtenberg 2006). Figure 2 depicts the diffusion dynamics of RS-related patents across (top 15 most relevant) CPC classes, making it clear how RS percolates through the technology space over time, gaining purposes and appearing in a wide array of inventions, though to a different extent (captured by the color intensity of the cells).

More important for our purposes is the fact that some of the mentioned classes, even if not dominant in terms of frequency, indicate the *direction* of RS evolution towards certain fields of application. This is the case, for example, of class Y02A (technologies for adaptation to climate change), which is strongly related to GPG themes.

In summary, from an inspection of patents’ structured information, we can extract two coarse results: first, RS diffusion across the technology space (approximated by patent classes) shows the general-purpose quality of the technology, rooted in its generic function of information acquisition from a distance. Second, it appears that an intersection of RS and GPG themes exists. While feeble in magnitude, this suggests that some of the uses of the technology have GPG relevance. In order to delve more in depth into this relationship, next, we resort to unstructured data analysis, as it allows for a much more granular assessment of the presence of GPG terms, shaping RS technical developments.

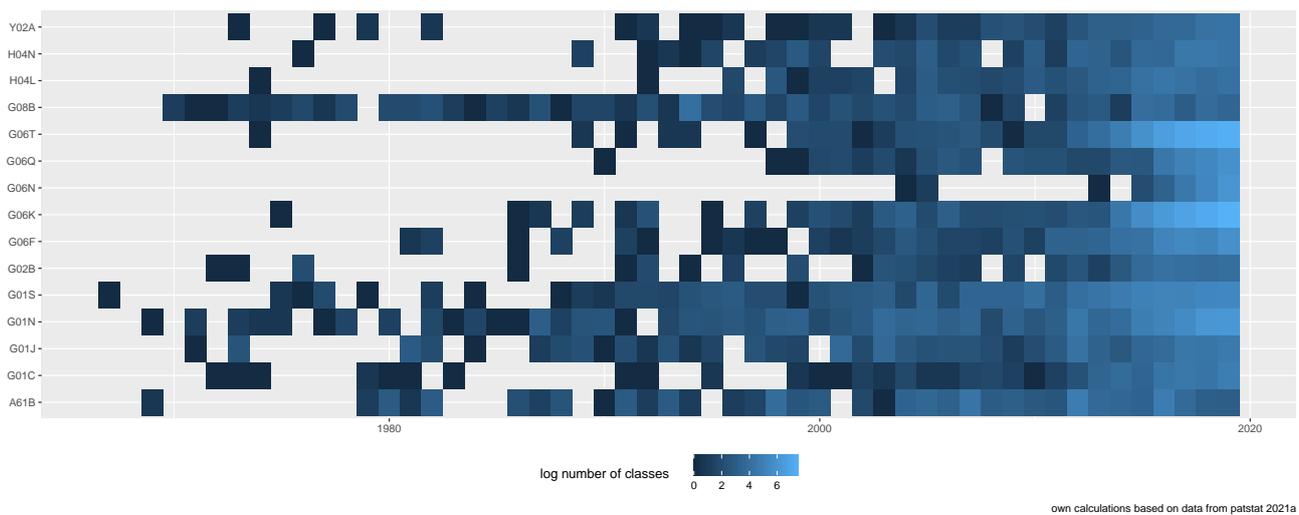


Figure 2. RS-related filing dynamics across the top 15 CPC classes for RS inventions; across time, filings take place in more and more CPC classes, including class Y02A, which is related to climate adaptation.

3.2. Text-Based Analysis: Terms

As a first exercise, we exploit text information by focusing on the dynamics of selected *terms* that we expect carry important signals regarding complementary technologies and domains to RS or fields of application. In Figure 3, we track the frequency of 12 terms over time. The terms refer both to hardware and software complements to RS (i.e., ‘drone’ and ‘neural’, respectively) or to specific domains of application of RS (i.e., ‘climate’ or ‘weather’) and to specific focus objects of the technology’s use (i.e., ‘crop’ or ‘water’).⁹ We can distinguish between terms that have appeared in the patents’ texts for a long time, such as ‘data’, ‘real time’, and ‘satellite’, and those that start being mentioned more recently, for instance ‘drone’ (‘climate’) being mentioned in only 7 (10) patents in our dataset. Figure A3 plots the diffusion curves to the topics in which these two terms have the highest impact, and it becomes obvious that both terms are parts in T3.

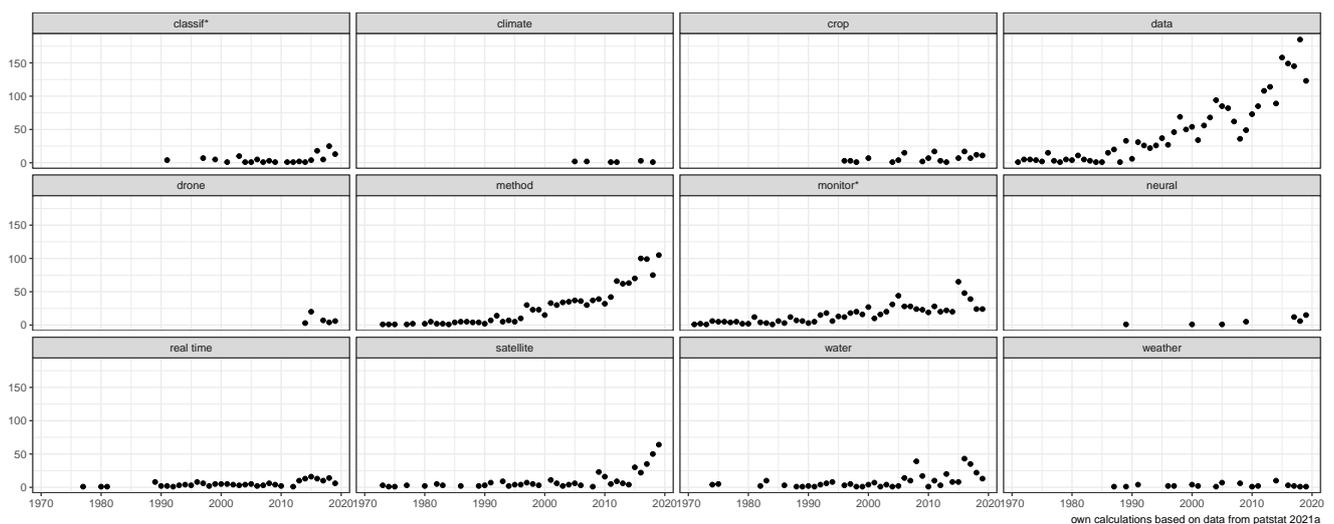


Figure 3. Evolution of selected terms.

Amongst the long-standing terms, we find those capturing the most essential features of the technology. Being a collection of data-acquisition technology, it is not surprising that ‘data’ appears in this group. What is more interesting is that the same term experiences a continuous increase in frequency, with a relatively recent acceleration. The driver of the dynamics might be related to a novel focus on RS data, now processed through artificial

intelligence (AI) algorithms. From this perspective, the appearance of terms such as neural (for neural network) and classif* (including terms such as classifier systems that relate to data-elaboration software), might indicate a growing interdependence between AI and RS technologies, meshing into more complex technology systems offering data collection and elaboration in real time (another term relevant in our sample).

Focusing on the terms that appeared more recently and that do not belong to methods and complementary technologies, our selection includes terms that are fairly relevant both for the diffusion of GPTs and for the provision of GPGs. On the one hand, the recent appearance of a term like drone illustrates the innovational complementarity feature characterizing RS that is embodied by devices (such as drones) that have their own technological trajectory and that can use RS as an ‘expansion’ of their capabilities. On the other hand, terms such as crop, water, and weather relate to changes in climate or to activities affected by global challenges. The terms water and crop indicate that RS technologies are employed in providing the Earth intelligence we already mentioned, which is instrumental in mapping environmental threats—in turn, a clear global public ‘bad’.

3.3. Text-Based Analysis: Structural Topic Modeling

Rationale. We run STM on our corpus. This gives us a granular picture of the technology at the level of words but organized within specific contexts (given by the topics). In other words, with STM, we are able to extract valuable signals from unstructured information. With the topics at hand, we can execute a series of interesting exercises. For example, we can identify shifts within the technology space by plotting topic diffusion curves, which allow to gauge structural shifts within the corpus. Additionally, we can trace the distribution of selected terms across topics—a feature leading back to the GPT issue of pervasiveness. Furthermore, with the use of STM, we can exploit metadata (i.e., covariates obtained from the dataset—a key novelty introduced by this paper—or constructed by the researcher(s)) to impose some structure on the topics and assess whether different issues of interest (i.e., GPG affinity) are relevant or not across the topics. Finally, we can represent information in network form, and apply network metrics to study the topic space. We discuss these exercises in sequence in Section 4.

For the analysis, we restrict our time period to after the year 1970, which reduces our dataset to 2.186 patents with unique abstracts. Concerning model selection, we opt for K (the number of distinct topics the model outputs) = 42.¹⁰

Data preprocessing. We take the texts from the patents’ abstracts and apply the following text preprocessing procedure covering these steps: (i) the identification of trigrams and bigrams—only those not including standard stopwords are kept; (ii) the removal of patent-specific stopwords (compare Table A1); and (iii) the running of the STM algorithm, which applies the removal of numbers and punctuation, custom stopwords, and frequent and rare words. The received work with patent data showed that in English texts, ngrams (i.e., conglomerates of n words) often represent technical terms or concepts. At the same time, there is patent-specific ‘jargon’ in the abstract texts. Since STM is based on text as data, it is crucial to think about how to deal with these specialities. Table A1 in the Appendix A.1 summarizes our selection.

Exploiting information from the covariates. Differently from other text-analysis techniques, such as LDA, STM allows to exploit meta information, which can be created from any variable of the dataset. We use two ‘types’ of covariates in order to offer a more elaborate analysis of the topics. First, we exploit covariates that are directly linked to variables in the dataset variables. These are as follows:

- Time:¹¹ This allows us to study the dynamics of the topics and the structural shifts within the corpus. It applies to all 2.186 patents.
- Authority and focus on CN international patents:¹² Patents filed at the Chinese patent authority (SIPO) and with a family size of 2 or more as discussed in Section 3.1. This allows to control for the recent filing boom in RS, and to assess whether there is a (macro)

geographical specialization in certain topics. The split of the 2.186 patents through this covariate shows that in our dataset, 113 patents (5.2%) are CN international patents.

- Sector assignment—private sector filings vs. non-private sector filings:¹³ The split of the 2.186 patents through this covariate shows that in our dataset, 1,561 patents (71.4%) may be assigned to the private sector (covering companies and individuals); 302 patents (13.8%) may be assigned to the non-private sector (covering non-profit and university), and for 323 patents (14.8%), no sector information is available. In the analysis, the latter are dropped.

Second, we use term-based covariates. These represent the key novelty of our analysis. In fact, through term-based covariates, we create meta information based on criteria that are relevant to our research question and that we define normatively. In particular, we build two families of covariates, one capturing the affinity of the reach topic to GPG, and the other to AI. Affinity is based on the inclusion in a patent abstract of at least one of the terms in Table 2. Our sample is split as follows along the covariates:

- GPG affinity 521 patents with GPG affinity (23.8%); 1.665 (76.2%) without GPG affinity.
- AI affinity given in 72 patents (3.3%); 2.114 patents (96.7%) without AI affinity.

Note that we can exploit this meta information even for rather small sub-samples of patents, e.g., compare the little number of 72 patents in the case of AI affinity.

Table 2. Term-based covariate creation for GPG affinity and AI affinity.

Covariate	Terms
GPG affinity	'agriculture', 'air', 'clean', 'climate', 'CO ₂ ', 'crop', 'crops', 'dioxide', 'ecological', 'ecology', 'fire', 'fires', 'flood', 'food', 'heat', 'nitrogen', 'sulfur', 'water', 'wildfire', 'wildfires'
AI affinity	'classified', 'classifier', 'classification', 'classify', 'classifying', 'misclassification', 'neural', 'preclassified', 'reclassification', 'supervised', 'unsupervised'

GPG affinity is the key dimension of interest for this paper. Introducing the covariate, we can estimate the closeness of each of the 42 topics to GPG-related terms to assess, for instance, how 'central' GPG-relevant topics are within the whole corpus. We decide to introduce further term-based covariates capturing AI to measure another characteristic of RS evolution: as discussed in Section 3, RS evolves as a technology system, in synergy with complementary technologies. AI is, among other things, a collection of software technologies for data processing (Vannuccini and Prytkova 2021), and data processing is the natural complement to the data acquisition function performed by RS. In fact, structural shifts in terms and topics relevance in RS can be driven by the increasing transition towards AI techniques used to manipulate remote-sensing-acquired data.

Outcomes of the K42 model. Figure 4 lists for the full period of analysis the topics extracted from our sample, ranked by expected topic proportions, and includes the 15 most important terms per topic. The top ranking topics are T22, T20, T28, T15, T33. As it could be expected, these relate to technical features of RS technologies, and represent the core of the invention direction in RS. Terms related to RS functions—in particular, data acquisition—also rank high, with T28 included in the top 3. However, when looking at the their dynamics, we see an increase in T28 (most important term: data) and a T15 (most important term: light), whereas the other big topics see a sharp decline. Table A2 in the Appendix A.2 presents the same information, ordered by topic number and for 17 terms.

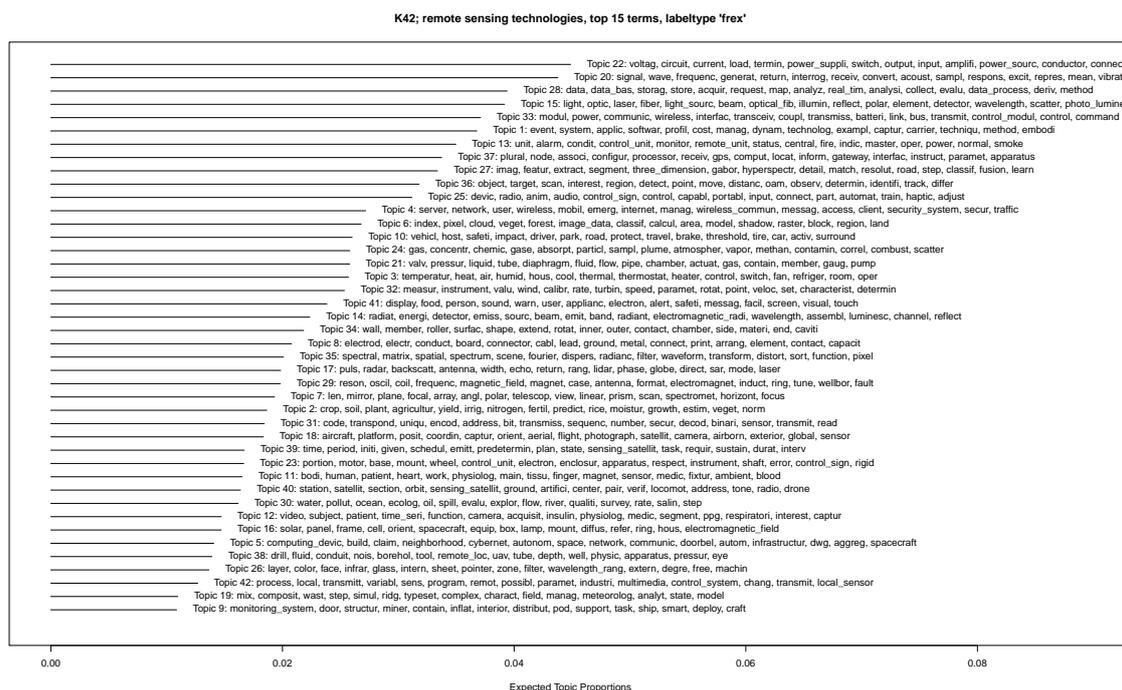


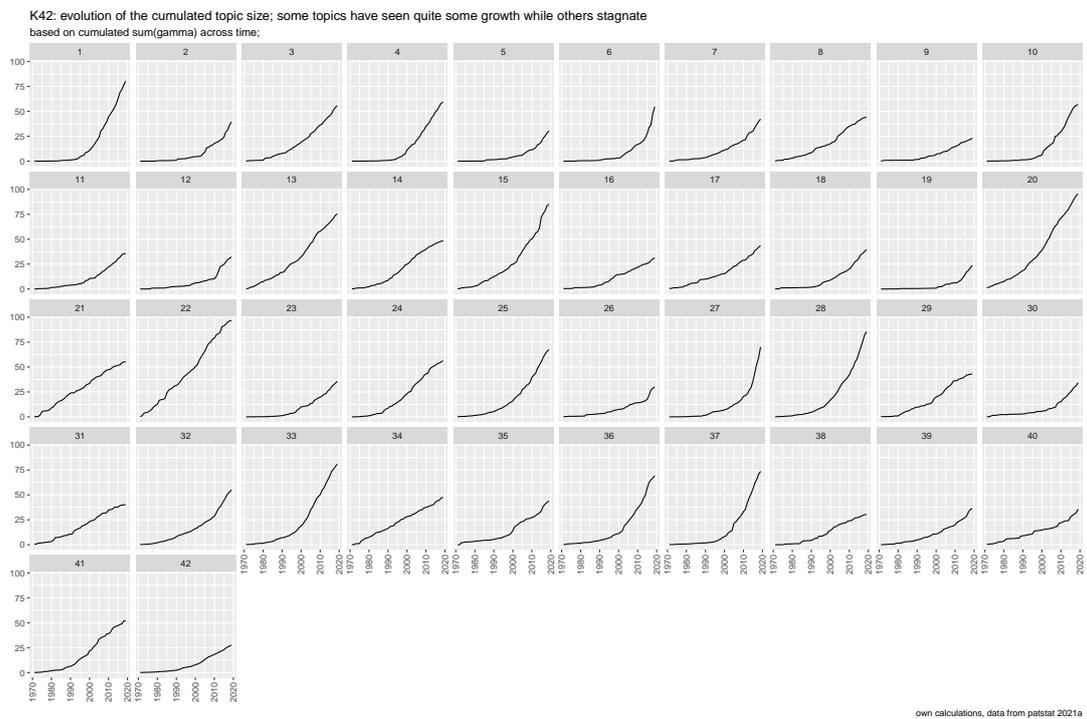
Figure 4. Expected topic proportion and the top 15 terms per topic for the K42 model; the expected topic proportion represents the importance of the topic in the corpus for the full period of analysis (it results from the γ matrix (per-document-per-topic) and is the sum of the γ values per topic divided by the sum of all γ values).

4. Results and Discussion

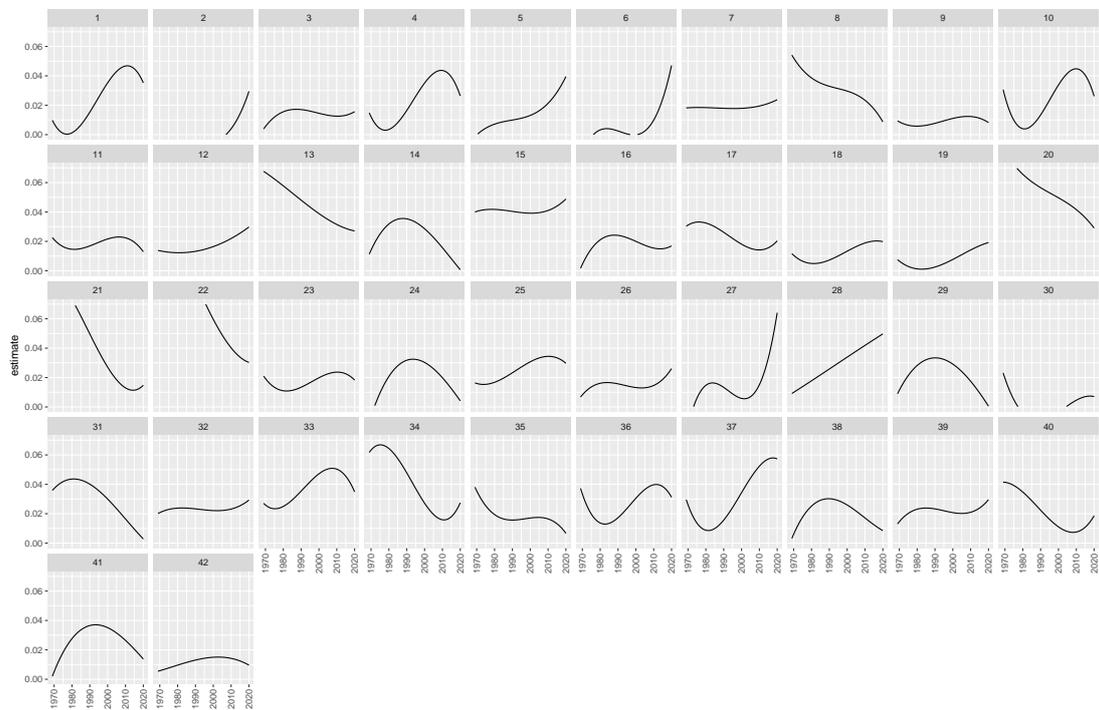
With some early insights on terms dynamics and the K42 model outcome at hand, we can now engage in a series of exercises to form a better picture of the RS nature, evolution, and propensity to be a medium for GPGs, amongst other things. First, we focus on the topic dynamics. Second, we look at the distribution of selected terms across topics. Third, we exploit our meta-information on covariates to study the affinities of specific themes to RS, including GPG and AI. Finally, we present the topic space as a network and compare different network metrics to identify, for instance, central and peripheral nodes.

Dynamics. In Figure 4, we can see the relative importance of the 42 topics (as represented by the topic proportion) as well as the most important terms per topic over the whole period. We now focus on their evolution. Figure 5 presents the trends of (cumulated) topic size in the top panel (Figure 5a), and topic dynamics in the bottom panel (Figure 5b). Concerning size, all topics display an increase that trails the very increase in the patent corpus. However, some topics' size growth tends to stagnate or to show an inflection point; this is the case, for instance, of T29 and T38. Both contain technical elements that might have already achieved a standardized configuration, and thus feature a slowdown in innovation. Alternatively, these topics could capture 'reverse salients' for the overall technology development. The structural shifts of topics over time become evident when inspecting the dynamics in Figure 5b, with certain topics relatively losing relevance and others gaining a more prominent role. This is the case for topics such as T6, T27, or T28 that cover issues related to image processing and data and have, thus, a more software-related nature. Their positive gradient already suggests that along the evolution of RS technology, the relative weight of physical vs. intangible components is shifted in favor of the latter, as the increasing diffusion of RS across different domains is grounded on the possibility to elaborate the data acquired by RS hardware. The relevance of the different topics is heterogeneous across global regions and major authorities. Figure 6 plots the spatial distribution of the topics evolution. Asia drives the dynamics in topics such as T6 and T27 that, as we

will see, are related to advances in image recognition (in turn, a key application field for AI systems).



(a)



(b)

Figure 5. Evolution of topics for the K42 model and the full period: cumulated size (corpus growth) and dynamics (corpus structure). (a) Cumulated topic size across time (the sum of γ -values per topic cumulated over time); (b) topic dynamics across time. For each year the expected topic proportions sum up to 1. These dynamics have to be interpreted in light of the steadily increasing number of patents.

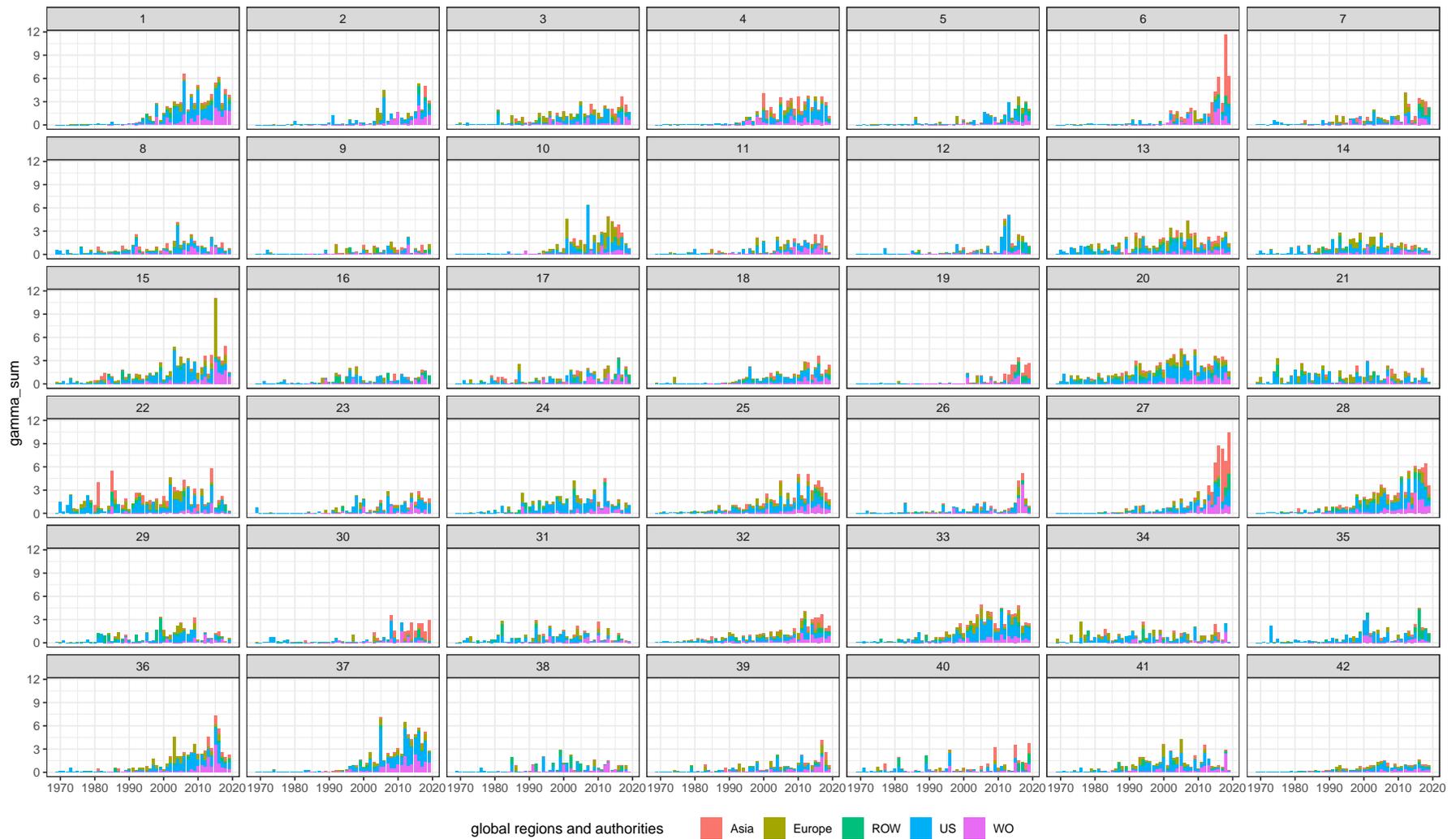


Figure 6. Spatial distribution related to the topic evolution, K42 model; for each topic, the up to three global regions or authorities which display the highest sum of γ -values are reported. Compare also the discussion related to CN affinity, e.g., in Figure 9.

Terms distribution across topics. In Figure 7, we plot the distribution of the terms' appearance across the topics for the full period. We focus on three selected terms, 'drone', 'real-time', and 'satellite'. The support of the figure is the 42 topics (for each term). This gives us some insight regarding the pervasiveness of some themes related to RS development. For example, the term drone is very 'localized' in a handful of topics. This might be related to the specific hardware-related synergies it develops with RS technology but also to the fact that it is a relatively novel feature in RS patents' texts as shown in Figure 3. We also see a strong embeddedness in T40—which thus bridges the terms 'drone' and 'satellite'. In contrast, real-time and satellite have a long-standing presence in the sample; however, they differ in spread across the topics. Real time is a pervasive term that appears in patents featuring diverse topics; this is certainly due to the fact that the speed of data acquisition has been and continues to be a key characteristic of RS, especially when considering intelligence and military uses, but also GPG-related uses, such as the monitoring of severe weather events, refugees flow, or cybersecurity threats. In contrast, satellite is a not as widely spread across topics—possibly because of its specific technological trajectory—but has experienced a sudden acceleration in frequency in recent periods (compare again Figure 3). This could be related to an increasing global focus of RS inventions, as well as to related trends in the private commercialization of Low-Earth-Orbit activities. In summary, through this exercise, we can obtain granular insights on the technological nature of RS—for example, emerging inventions driven by 'hardware' complementarities (drones), persistent and pervasive inventive activity based on the 'service' that RS technology delivers (real time), and growing inventions related to terms directly linked to the provision of Earth intelligence, monitoring and related services, and, thus, also to GPGs (satellite).

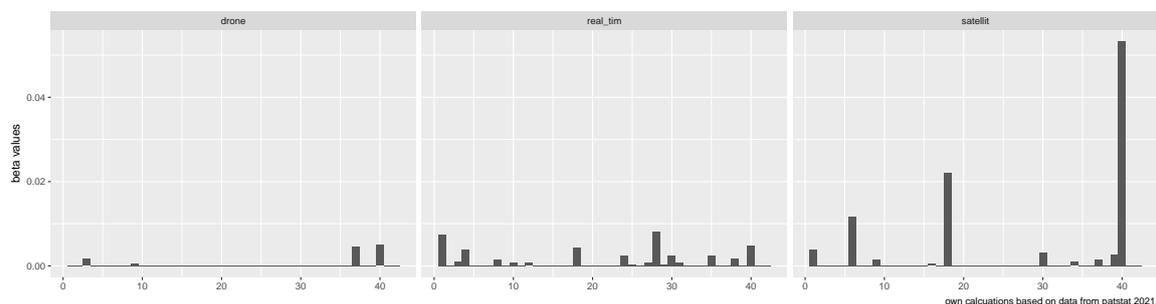


Figure 7. Term distribution across topics based on β -values; full period. Recall: The term 'drone' is recent and rare; 'satellite' is established and has seen quite some dynamics; 'real-time' is established and did not experience dynamics (compare also Figure 3).

Affinities. In Figure 8, we present the mean estimates of affinities for all topics and all four types of covariates: GPG (panel (a)), AI (panel (b)), public/private sector (panel (c)), and China origin (panel (d)). This allows us to compare topics relevant for each covariate type and to explore co-occurrences. We start with GPG covariates. The topics with the highest estimated mean affinities are T3, T2, T30, and T21. T2 and T30 are particularly relevant, as they include among their top terms several that are clearly related to global challenges, such as crops, soil, and agricultural activities (T2), or water, oil and ocean (T30). A visual depiction of the key terms for the structural shifts of the top four topics by GPG affinity is offered by the word clouds in Figure A1 and the diffusion curves in Figure A2, both in the Appendix A.3. Looking at the top topics in terms of AI affinity, we find T27, T6, T35, and T2. Here, key topics from a pure AI standpoint are T27 and T6, which include terms related to image processing (i.e., feature extraction)—the main AI capability used in RS. However, T6 also includes terms that have to do with specific uses of RS in the environmental domain and that can have potential global reach, such as forest and veget*.¹⁴ Importantly, T2 co-occurs also as a top GPG-relevant topic. This illustrates an important overlap: some inventions in RS that involve AI have GPG-related application,

showing technological synergies at work for global uses. When focusing on affinity to public actors (the left side of the support of panel (c); the right side captures affinity to private actors), the top topics are T2, T35, T27, and T6. These perfectly match the top topics for AI affinity, suggesting that—at least in the field of RS inventive activities—patents involving AI are also more related to public actors. This is a further piece of evidence to understand the RS-GPG nexus: public actors are often (even if not exclusively) the key actors involved in the production and provision of (global) public goods. In contrast, private actors’ affinity is higher on topics such as T22, T13, T10, and T4, which mostly refer to the specific hardware components of RS technology, and thus likely capture invention along the supplier chain of RS. Finally, we can offer some preliminary insight on whether there are signs of international specialization in RS invention by inspecting the mean estimates on China affinity. In this case, the top four topics are T27, T6, T19, and T40. We see that T27 and T6 are ranked top in affinity with AI covariates and non-private actors’ covariates. From a political economy perspective, China develops RS technology through the initiative of public actors, in directions strongly related to AI. RS appears again as a globally relevant technology, whose progresses are possibly pushed by a key emerging geopolitical actor. An additional support to this idea comes from another topic ranking high in China-affinity estimates, T40, which covers satellite technology—needless to say, a strategic and global one.

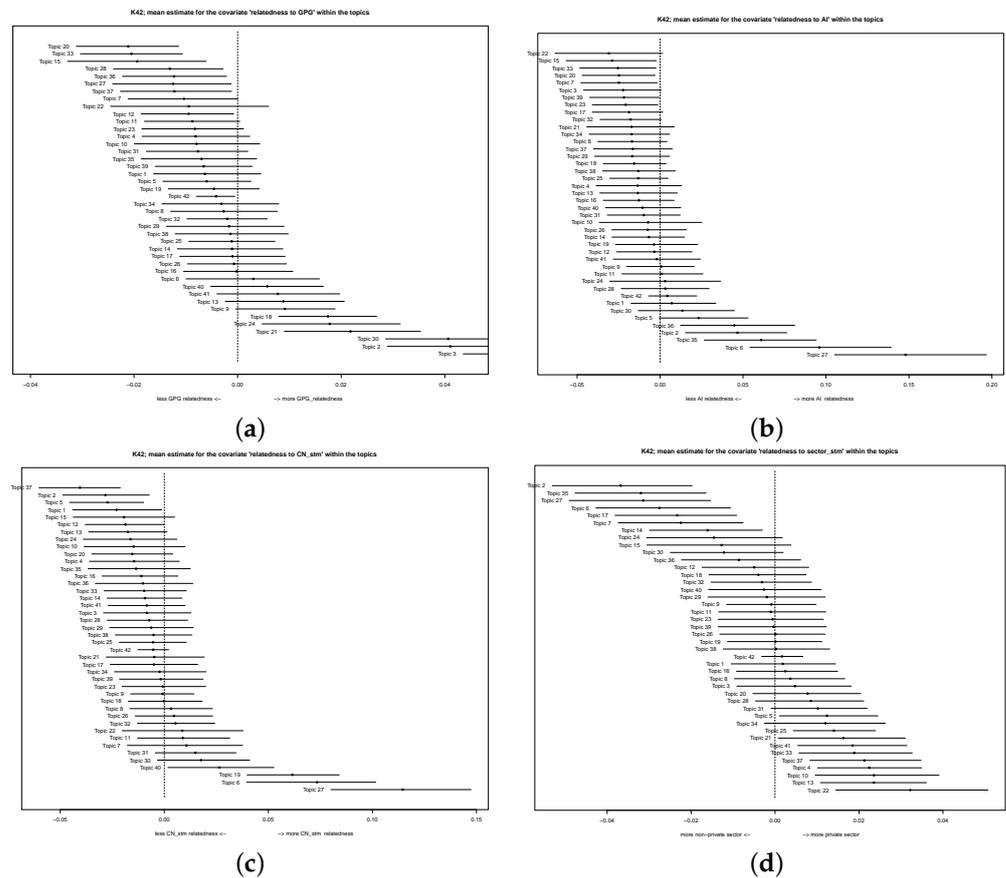


Figure 8. Identification of affinities. The mean estimate split allows to assign topics to selected affinities (vertical line depicted at the level 0). The dot represents the mean estimate value, and the horizontal line reflects the 95% confidence interval. The topic numbers mentioned in the subpanels are those showing a clear affinity to either perspective. They are used for the color-coded network in Figure 9. Related to sector affinity, the color code involves both perspectives, namely private sector affinity and non-private sector affinity. For the affinities in panels (a–c) only those with clear positive values are taken into account in the network. (a) GPG affinity: topics 3, 2, 30, 21, 24, 18, 9; (b) AI affinity: topics 27, 6, 35, 2, 36, 5; (c) CN affinity: topics 27, 6, 19, 40; (d) private sector affinity: private sector topics 22, 13, 10, 4, 37, 33, 41, 21, 25; non-private sector topics 2, 35, 27, 7, 17, 14.

Merging all perspectives. The insights we derived from the discussion of the affinity measures become more evident when topics are presented in network form. In Figure 9, we present the RS topic network, created by linking topic-sharing terms.¹⁵ We color code the network to highlight top ranking topics in terms of affinity indicators, as introduced in the previous paragraph. From visual inspection, we can confirm some overlap between topics showing high GPG, AI, public actors, and China affinity. As discussed, private actors seem to conduct inventive activities more focused on RS component technologies, while non-private inventors have high affinity with some topics that are also GPG and AI related, suggesting that RS can be instrumental to GPGs when invention is initiated by the classic providers of public goods. Furthermore, we could speculate that GPG-relevant invention trajectories might give rise to private coordination failures, and thus require initial public support; this argument goes in line with the failures characterizing the kick-off phase of GPT diffusion (Bresnahan and Trajtenberg 1995). Such a perspective can be supported by noting that GPG-related topics are yet rather peripheral in the topic network.

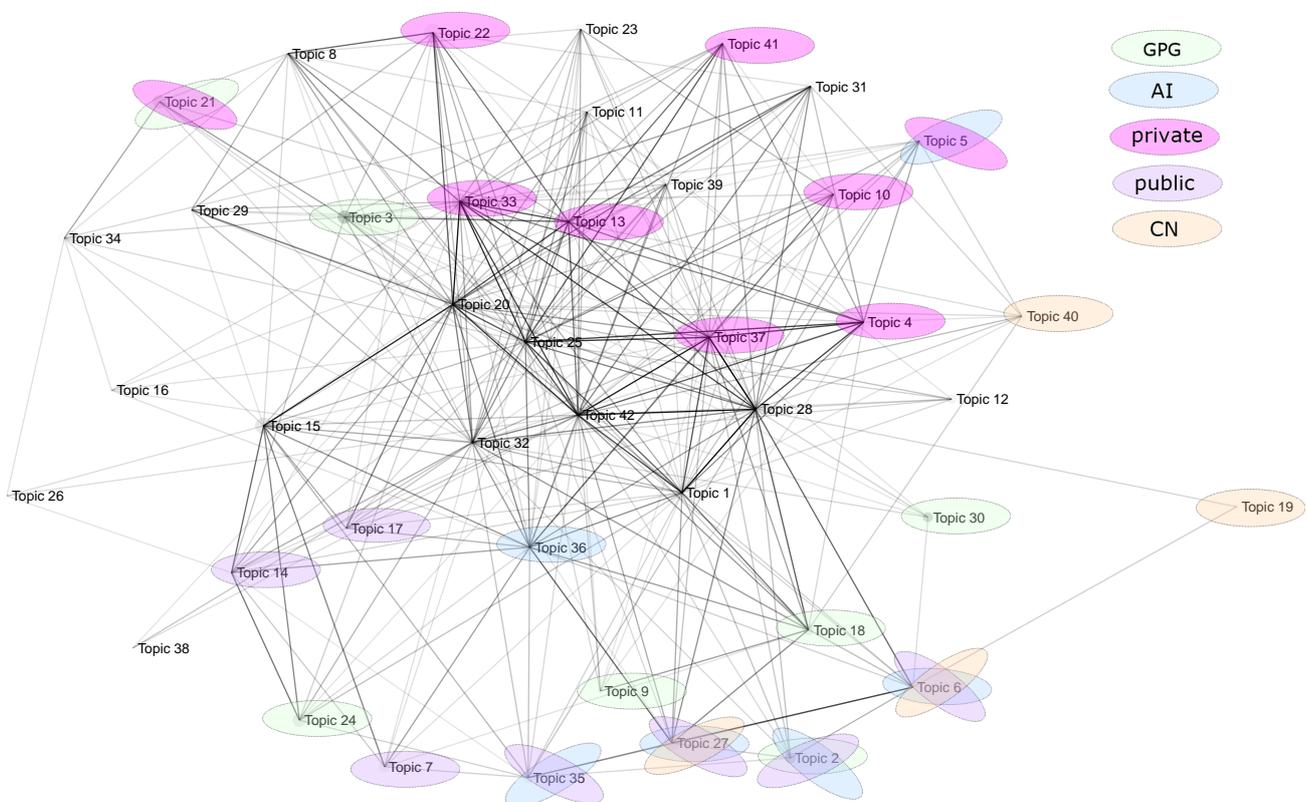


Figure 9. Topic network of the K42 model including all the dimensions of our analysis (the covariates), depicting (when present) the overlap of some topics across different perspectives; the color code is based on the mean estimates derived in Figure 8.

To provide more quantitative support to our claims, we can exploit network analysis and compute metrics on our topic space. Figure 10 presents a correlation plot of different network metrics (size, degree centrality), our affinity indexes, and topic dynamics. We confirm that the private sector inventors are negatively correlated with AI, which instead correlates positively with China patenting, making the case for some specialization in inventive activities taking shape internationally. GPG affinity is negatively correlated with degree, thus supporting the idea that GPG-related technological advances as captured by topics are yet a peripheral development. This is to be expected to a certain extent: in fact, GPG terms will relate with high likelihood to uses of RS, which are less frequently mentioned in patents' abstracts compared to the technological core of the patented inventions.

However, we stress that for a topic being peripheral in the topic space does not mean being irrelevant; on the contrary, our methodology allows to pick up emerging trends, with RS uses (including those that have GPG valence) increasingly appearing in the documents.

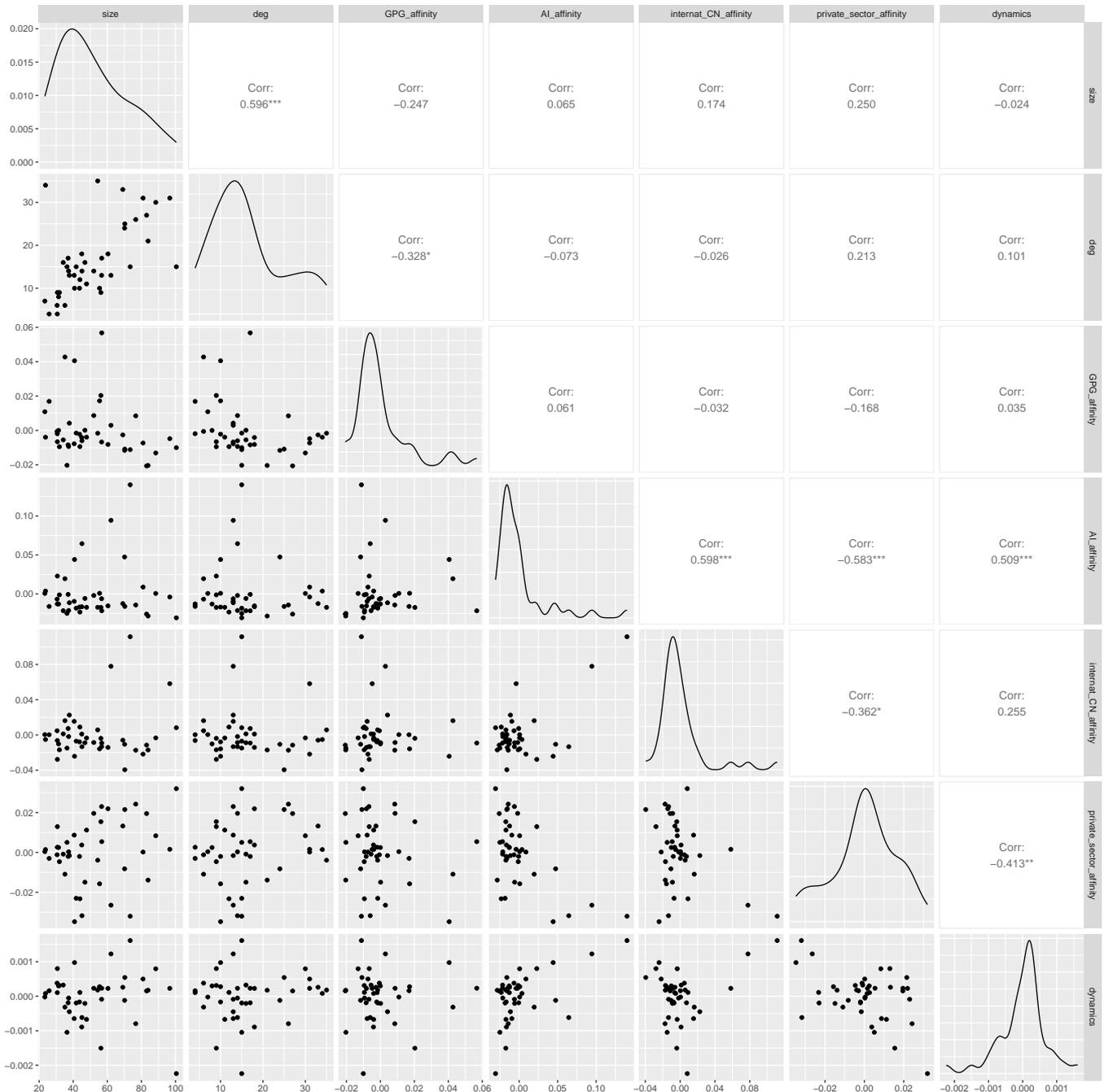


Figure 10. Correlation plot of topic network metrics: size, degree, GPG affinity, AI affinity, CN affinity, private sector affinity, and dynamics. *, **, *** indicate statistical significance levels at respectively 10%, 5% and 1%.

5. Conclusions and Outlook

In this paper, we studied the relationship between technological change and political economy issues using the case of remote sensing technology and the theme of global public goods. RS is a collection of information acquisition technologies, whose generic function can be applied to several goals, including some that have a global reach or impact. An example

is the monitoring of climate and environmental changes. RS is an important technology, as it displays the properties of a GPT—a type of innovation with transformative impact.

Our motivation for the study was to assess whether the fact that RS technology can influence the provision of GPG could already be detected from the direction of invention in RS, proxied by patenting. We employed structural topic modeling (STM) to exploit information from unstructured data (abstracts' text corpus) over an extensive time period. Our methodology has two main advantages: first, it provides us with granular information to explore the nexus between RS and GPG. Second, we could use metadata to build covariates that allow exploring specific aspects of RS invention dynamics: in particular, we estimated an indicator of topics' affinity with GPG and studied its relationship with other topics' features, such as affinity to AI technologies, public inventors, or inventive activities in a growing dynamic economy like China. We found that GPG-related topics are peripheral in the RS patents topic space, and yet they are emerging, suggesting that RS does play a role in how global public provision is deployed. It goes without saying that our approach has shortcomings too: for example, and despite our robustness checks, the choice of the topic number parameter has a degree of arbitrariness. In general, we are able to draw inference from a snapshot of terms distribution across topics, but we cannot identify specific mechanisms shaping the RS-GPG relationship.

We focused on tracing a theme of political economy into technical documents; hence, it is not our goal to address 'classic' issues related to public goods provision, such as the solution of coordination failure and the design of incentives. However, with our results, we shed light on the fact that a given technology can be instrumental in the pursuit of global goals. As societal challenges become increasingly pressing, this will provide fertile ground for continued growth in RS patent-filing activities, shaping its direction of invention.

The analysis we conducted is a first step to single out whether and how a given technology crosses aspects that are GPG relevant during its evolution. In this sense, this paper is a novel but also exploratory contribution. We hope to direct research and policy attention towards the fact that under the bridge that connects technology and GPGs, there are interesting issues worth further investigation.

As we focus on technical information and on the presence of GPG terms in it, we are not able to delve into the ethical and legal aspects of RS uses. However, an important aspect not to leave aside is if RS could be considered a dual-use technology (Forge 2010): in fact, the generic function of information acquisition from a distance can feed both positive and harmful uses. It goes without saying that the many purposes of RS range from humanitarian to military and intelligence ones. RS has the potential to contribute positively to the production of GPGs; however, scenarios in which its use entails (global) welfare reduction can also be imagined, from surveillance to espionage, terrorism and corporate data capture. While it is problematic to infer the desired uses of technology from patent documents, even at a granular level of analysis, this possibility makes even clearer the relevance of combining the technological and the political economy perspective in innovation research, and should inspire further studies along this trajectory.

Author Contributions: I.O.: conceptualisation; analysis and results interpretation; writing. S.V.: conceptualisation; analysis and results interpretation; writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The patent data has been taken from the patstat database 2021, spring version. The list of the patent IDs utilized for analysis can be provided upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Data Preprocessing

Table A1 collects all ngrams and specific stopwords used to preprocess our text corpus.

Table A1. K42 data preprocessing: ngrams, patent-specific and technology-specific stopwords.

Context	Terms
trigrams	'convolutional neural network', 'light emitting diode' 'local sensor data' 'synthetic aperture radar', 'unmanned aerial vehicle'
bigrams	'base station', 'central control', 'computing device', 'control circuit', 'control device', 'control module', 'control signal', 'control system', 'control unit', 'data base', 'data collection', 'data processing', 'database' 'earth's surface', 'electric power', 'electromagnetic radiation', 'electronic device', 'high resolution', 'image data', 'land use', 'light beam', 'light emitting', 'light source', 'local sensor', 'magnetic field', 'management system', 'measurement data', 'monitoring system', 'multi scale', 'multi-spectral', 'neural network', 'optical fiber', 'optical remote', 'output signal', 'output voltage', 'power source', 'power supply', 'processing system', 'processing unit', 'radio frequency', 'real time', 'remote location', 'remote unit', 'satellite data', 'security system', 'sensing apparatus', 'sensing satellite', 'signal processing', 'wavelength range', 'wireless communication'
custom stopwords	'along', 'also', 'and/or', 'art', 'can', 'claim', 'claiming', 'claims', 'comprise', 'comprised', 'comprises', 'comprising', 'contain', 'contained', 'containing', 'contains', 'correspond', 'corresponding', 'corresponds', 'describe', 'described', 'describes', 'describing', 'directed', 'disclose', 'disclosed', 'discloses', 'disclosing', 'first herein', 'include', 'included', 'includes', 'including', 'invention', 'least', 'like', 'may', 'obtain', 'obtained', 'obtaining', 'obtains', 'one', 'permit', 'permits', 'present', 'presented', 'presenting', 'presents', 'prior', 'provide', 'provided', 'provides', 'providing', 'relate', 'related', 'relates', 'relating', 'said', 'second', 'see', 'shown', 'thereof', 'two', 'wherein', 'within'
technology specific stopwords	'remote', 'sensing', 'sensor'

Appendix A.2. Topics, Terms and Prevalence Ordered by Topic Number

Table A2 lists the 42 topics by number, the topic prevalence and the top 17 terms for each topic.

Table A2. Topics, terms and prevalence, ordered by topic number (K42 model).

Topic	Topic Proportion	Terms
Topic 1	0.0368	system, method, use, detect, embodi, applic, base, util, event, exampl, perform, monitor, determin, manag, improv, particular, implement
Topic 2	0.0186	crop, field, use, plant, agricultur, yield, soil, irrig, predict, determin, estim, base, area, nitrogen, veget, applic, fertil
Topic 3	0.0257	air, hous, control, use, cool, fan, sensor, switch, instal, batteri, enclosur, mount, inlet, valu, assembl, respons, cooler
Topic 4	0.0272	network, wireless, server, user, sensor, receiv, transmit, local, monitor, transceiv, inform, devic, communic, mobil, internet, via, central
Topic 5	0.0141	system, communic, devic, build, computing_devic, network, claim, user, function, cybernet, applic, embodi, oper, level, report, util, adapt
Topic 6	0.0268	imag, area, method, pixel, model, base, cloud, surfac, index, veget, satellit, region, calcul, forest, classif, image_data, differ
Topic 7	0.0193	mirror, len, imag, plane, array, angl, platform, focal, surfac, element, optic, infrar, field, light, camera, secondari, spectromet
Topic 8	0.0208	electr, electro, ground, element, connect, conduct, plural, print, form, thermal, circuit, common, arrang, compon, lead, singl, head
Topic 9	0.0109	structur, monitoring_system, support, door, mount, use, system, deploy, inflat, task, interior, distribut, adapt, damag, attach, pod, caus
Topic 10	0.0260	vehicl, detect, determin, devic, impact, signal, driver, system, road, activ, inform, park, arrang, brake, respons, track, emiss
Topic 11	0.0165	bodi, sensor, patient, detect, use, analyt, posit, human, medic, blood, inform, magnet, physiolog, heart, user, fixtur, devic
Topic 12	0.0147	video, subject, use, method, determin, function, field, region, interest, loss, camera, reconstruct, time_seri, patient, electr, estim, acquisit
Topic 13	0.0350	unit, alarm, monitor, power, control_unit, oper, electr, control, central, sensor, condit, remote_unit, transmit, system, connect, suppli, master
Topic 14	0.0224	detector, radiat, energi, sourc, assembl, illumin, emit, reflect, detect, locat, beam, configur, region, mirror, infrar, filter, remot
Topic 15	0.0392	optic, light, element, fiber, optical_fib, receiv, reflect, light_sourc, polar, beam, environ, arrang, monitor, coupl, output, end, system

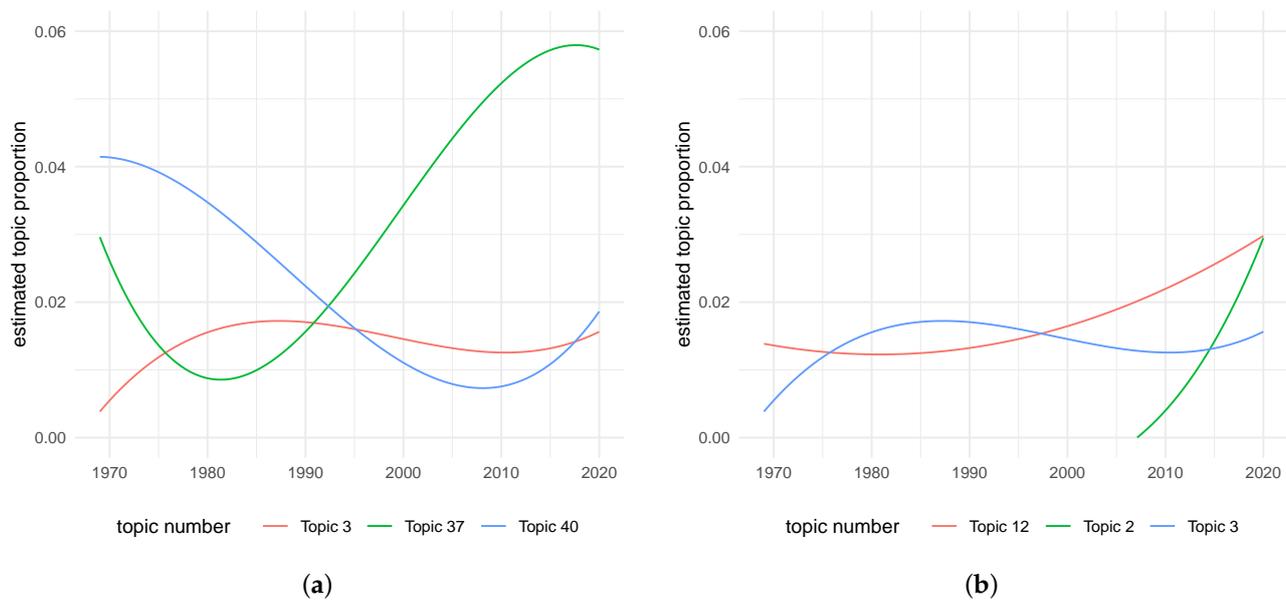


Figure A3. Diffusion curves of the topics which have the highest β values for the selected terms (note that the diffusion curve of T3 is colored in red in panel (a) and colored in blue in panel (b)). (a) 'drone': topics 40, 37, 3; (b) 'climate': topics 2, 12, 3.

Notes

¹ https://en.wikipedia.org/wiki/Remote_sensing, Wikipedia, accessed 13 March 2023.

² <https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing>, accessed on 13 March 2023.

³ Roberts et al. (2013) develop the STM which exploits document-level covariates affecting topical prevalence and/or topical content. The authors especially provide an R package (*stm*), which allows users to incorporate the specific structure of their corpus and thus to directly estimate the quantities of interest in applied problems. The approach to including the corpus structure intends to make inference about observed covariates rather than predicting covariate values in unseen text.

⁴ The generative process of each document can be understood as a procedure that (i) draws a document length, then (ii) word by word, draws a topic from the distribution of K topics, (iii) draws a word from the associated distribution, and (iv) proceeds with the following word. For a given set of documents, the underlying distributions can be estimated using Bayesian statistics techniques. Details on the generative process can be found in Blei et al. (2003).

⁵ The basic assumption is that the mean prevalence of a topic, i.e., its share in all documents at a given point of time, can be expressed by splines. A spline is a function defined piecewise by polynomials. In the STM package in R, the default is set to $d = 3$, i.e., piecewise third-degree polynomials allow for non-linear changes over time. This allows to avoid erratic behavior at the domain bounds.

⁶ While designing the concept of our technology breakdown, we also carried out cooperative patent classification (CPC) class search at ESPACENET based on the term 'remote sensing'. However, that did not provide additional information or insights. Compare EPO et al. (2022) for a CPC class-based technology breakdown to capture the technology field of space-borne sensing.

⁷ Overall, the term search resulted in 8.807 unique patents. We detect an exorbitant increase in filings after 2015, with SIPO filings overwhelmingly dominating the sample (6.618 patents out of 8.807). The literature argues that drivers of the huge increase in Chinese patents in almost any technology and not just remote sensing are strategic/political motives, rather than real innovations (e.g., EPO et al. 2022). Hence, we correct for this potential bias by imposing an additional restriction: we include Chinese patents into our dataset only if these patents have been filed at least at two authorities.

⁸ We aggregated the 24 patent authorities for which we have filings to what we call 'global regions': United States (US), WO (patents being filed directly at the World Intellectual Property Organization), Asia (including SIPO patents with family size of 2 and larger), Europe, and ROW (Rest Of World). See Appendix A.3 for a breakdown of the global regions by patent authorities.

⁹ We pay particular attention to 'polysemic' words, that is, terms carrying distinct meanings—for example, crop is a noun in the agriculture field but a verb in other domains, such as image cropping in computer graphics. We exclude polysemic words like 'forest', 'tree', or 'environment' from the term dynamic perspective since they may be either related to modern algorithms or the natural, technical or even urban environment. However, when applying STM, we are able to trace the importance of polysemic words via their embedding within topics and, for example, focusing on related topic dynamics.

¹⁰ The choice of K42 is dictated by the ease of elaboration and presentation of the results. We conduct robustness tests with K60 and K62 models, obtaining comparable results in terms of topics' clustering.

- 11 Based on the PATSTAT variable: earliest_filing_year.
- 12 Based on the PATSTAT variable: authority.
- 13 Based on the PATSTAT variable: psn_sector.
- 14 Here, the polysemic term ‘forest’ is related to the plants and not to algorithms.
- 15 Technically speaking, we link topics based on their cosine similarity with the edge size representing the level of similarity between two topics. We also apply a minimum threshold for edges to be shown.

References

- Airoldi, Edoardo M., and Jonathan M. Bischof. 2016. Improving and evaluating topic models and other models of text. *Journal of the American Statistical Association* 111: 1381–403. [\[CrossRef\]](#)
- Alstott, Jeff, Giorgio Triulzi, Bowen Yan, and Jianxi Luo. 2016. Mapping technology space by normalizing patent networks. *Scientometrics* 110: 443–79. [\[CrossRef\]](#)
- Arrighi, Giovanni. 1994. *The Long Twentieth Century: Money, Power, and the Origins of Our Times*. London: Verso.
- Bekar, Clifford, Kenneth Carlaw, and Richard Lipsey. 2018. General purpose technologies in theory, application and controversy: A review. *Journal of Evolutionary Economics* 28: 1005–33. [\[CrossRef\]](#)
- Blei, David M., and John D. Lafferty. 2007. A correlated topic model of science. *The Annals of Applied Statistics* 1: 17–35. [\[CrossRef\]](#)
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.
- Bresnahan, Timothy F., and Manuel Trajtenberg. 1995. General purpose technologies: ‘Engines of growth’? *Journal of Econometrics* 65: 83–108. [\[CrossRef\]](#)
- Buchholz, Wolfgang, and Todd Sandler. 2021. Global public goods: A survey. *Journal of Economic Literature* 59: 488–545. [\[CrossRef\]](#)
- Cantner, Uwe, and Simone Vannuccini. 2012. A new view of general purpose technologies. In *Empirische Makroökonomik und mehr—Festschrift zum 80. Geburtstag von Karl Heinrich Oppenländer*. Edited by Adolf Wagner and Ullrich Heilemann. Berlin: De Gruyter. pp. 71–96. [\[CrossRef\]](#)
- EPO, ESPI, and esa. 2022. *Space-Borne Sensing and Green Applications: Patent Insight Report*. Munich: European Patent Office.
- Forge, John. 2010. A note on the definition of “dual use”. *Science and Engineering Ethics* 16: 111–18. [\[CrossRef\]](#)
- Forney, William M., Ronald P. Raunikaar, Shruti Mishra, and Richard L. Bernknopf. 2012. An economic value of remote sensing information: Application to agricultural production and maintaining ground water quality. Paper presented at 2012 Socio-Economic Benefits Workshop: Defining, Measuring, and Communicating the Socio-Economic Benefits of Geospatial Information, Boulder, CO, USA, June 12–14; pp. 1–6.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. Text as data. *Journal of Economic Literature* 57: 535–74. [\[CrossRef\]](#)
- Griffith, Thomas L., and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America* 101: 5228–35. [\[CrossRef\]](#)
- Grimmer, Justin, and Brandon M. Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21: 267–97. [\[CrossRef\]](#)
- Hall, Bronwyn, and Manuel Trajtenberg. 2006. Uncovering GPTs with patent data. In *New Frontiers in the Economics of Innovation and New Technology. Essays in Honor of Paul A. David*. Edited by Cristiano Antonelli, Dominique Foray, Bronwyn Hall and W. Edward Steinmueller. Cheltenham: Edward Elgar, pp. 380–426.
- Kaul, Inge. 2012. Global public goods: explaining their underprovision. *Journal of International Economic Law* 15: 729–50. [\[CrossRef\]](#)
- Knell, Mark, and Simone Vannuccini. 2012. Tools and concepts for understanding disruptive technological change after Schumpeter. In *The Routledge Handbook of Smart Technologies*. Edited by Heinz D. Kurz, Marlies Schütz, Rita Strohmaier and Stella S. Zilian. Abingdon: Routledge, pp. 77–101.
- Lombardi, Mauro, and Simone Vannuccini. 2022. Understanding emerging patterns and dynamics through the lenses of the cyber-physical universe. *Patterns* 3: 100601. [\[CrossRef\]](#) [\[PubMed\]](#)
- Menz, Nina, and Ingrid Ott. 2011. *On the Role of General Purpose Technologies within the Marshall-Jacobs Controversy: The Case of Nanotechnologies*. No. 18. KIT Working Paper Series in Economics. Karlsruhe: arlsruher Institut für Technologie (KIT), Institut für Volkswirtschaftslehre (ECON).
- Mimno, David, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. Paper presented at 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, July 27–31; pp. 262–72.
- Nelson, Richard R. 2003. On the uneven evolution of human know-how. *Research Policy* 32: 909–22. [\[CrossRef\]](#)
- Paunov, Caroline, Sandra Planes Satorra, and Dominique Guellec. 2018. Semantic Analysis for Innovation Policy: Workshop Summary. Paper presented at Semantic Analysis for Innovation Policy, Paris, France, March 12–13.
- Ranaei, Samira, Arho Suominen, Alan Porter, and Tuomo Kässi. 2019. Application of text-analytics in quantitative study of science and technology. In *Springer Handbook of Science and Technology Indicators*. Edited by Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch and Mike Thelwall. Berlin/Heidelberg: Springer, pp. 957–82.
- Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi. 2016a. A model of text for experimentation in the social sciences. *Journal of the American Statistical Association* 111: 988–1003. [\[CrossRef\]](#)

- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2016b. Navigating the local modes of big data: The case of topic models. In *Computational Social Science*. Edited by R. Michael Alvarez. Cambridge: Cambridge University Press, pp. 51–97. [[CrossRef](#)]
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, and Edoardo M. Airoldi. 2013. The structural topic model and applied social science. *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation 4*: 1–20.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science* 58: 1064–82. [[CrossRef](#)]
- Savona, Maria, Tommaso Ciarli, Ed Steinmueller, and Simone Vannuccini. 2022. The design of digital automation technologies: Implications for the future of work. *CESifo Forum* 23: 4–10.
- Simon, Herbert A. 1987. The steam engine and the computer: What makes technology revolutionary. *Educom Bulletin* 22: 2–5.
- Thoma, Grid. 2009. Striving for a large market: evidence from a general purpose technology in action. *Industrial and Corporate Change* 18: 107–38. [[CrossRef](#)]
- Van Looy, Bart, and Tom Magerman. 2019. Text mining and science and technology studies. In *Springer Handbook of Science and Technology Indicators*. Edited by Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch and Mike Thelwall. Berlin/Heidelberg: Springer, pp. 929–56.
- Vannuccini, Simone, and Ekaterina Prytkova. 2021. Artificial intelligence's new clothes? From general purpose technology to large technical system. In *From General Purpose Technology to Large Technical System (April 7, 2021)*. Brighton: SWPS, vol. 2.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.