

Mapping the Evolution of Social Innovation in Scientific Publications: A Topic Modelling and Text Mining Approach

Uroš GODNOV¹, Jana HOJNIK², Simona KUSTEC²

¹ University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia, uros.godnov@famnit.upr.si

² University of Primorska, Faculty of Management, Koper, Slovenia, jana.hojnik@fm.upr.si, simona.kustec@fm.upr.si

Objective: To trace how academic discourse on social innovation has evolved from 2000 – mid-2024 in numbers and leading topics by applying a special topic modelling and text mining methodology.

Data & Sources: 4,703 full-text journal articles retrieved from Science Direct.

Methods: Literature review and PDF text extracted with PyPDF2 and pdfplumber; cleaned and tokenised in R; topic modelling performed with Latent Dirichlet Allocation (ldatuning-optimised); temporal and correlation analyses visualised via tidyverse.

Results: The number of publications increased significantly from 16 (in 2000) to 573 (in 2021), stabilizing thereafter. Seven dominant topics emerged: renewable energy, environmental/resource management, smart-city governance, sustainable food systems, corporate strategy, academic-method studies, and social-governance structures. "Social" and "innovation" became the top word pair after 2006; energy-related terms surged after 2016. Surprisingly, topics typically considered 'social' have not dominated the social innovation discourse in scientific communities compared to the aforementioned dominant topics.

Discussion: Our results largely confirm existing findings from literature reviews and affirm the interdisciplinary, vague, contested, and still intensively evolving nature of social innovation. Dominant social innovation topics in scientific papers reference to social innovation topics in global political and policy documents, notably from the EU (from 2013 onwards) and the 2015 UN SDGs agenda, also emphasising collaboration between scientific, business, political and non-governmental stakeholders, and can thus serve as scientific, evidence-based advocacy for other stakeholders involved in social innovation processes.

Conclusions: Social innovation research is now an established, systemic, and broadly interdisciplinary field of study, focusing on sustainability, emerging technologies, and governance topics. It is tightly connected with the political and policy agendas of leading international organisations, as well as business and non-governmental ones.

Implications: Findings guide scholars to under-explored social-related content and niches (such as governance and, especially, equity topics) and help policymakers and other stakeholders involved in social innovation processes locate evidence-based approaches and clusters when designing their socially innovative responses, interventions, solutions, and measures.

Keywords: Social innovation theories, Global policy agenda, Text mining, Topic modelling literature review

1 Introduction

Today, social innovation is understood within diverse yet interconnected contexts. Definitions of social innovation can thus be found, predictably, in academia, but also,

by its very nature, in broader everyday social and political discourse. Due to the very diverse nature of social innovation, there is no single, uniform definition. Moreover, as the concept has developed in both academic and broader socio-political discourse, definitions of social innovation

have intertwined, complemented each other, or reflected contemporary views shaped by political and wider user discourse (Ayob et al., 2016; Godin, 2016; Logue, 2020). This situation can, on one hand, bring confusion and diminish credibility within the scientific world and for scientific work, while on the other hand, it also helps in understanding the evolution of a newly developing field of scientific interest. Acknowledging the former while seeing potential in the latter.

This paper aims to trace the evolution of academic discourse on social innovation from 2000 to mid-2024 by employing advanced text mining techniques—including PDF text extraction, cleaning, tokenization, and topic modeling with Latent Dirichlet Allocation (LDA)—to analyse a comprehensive dataset of 4,703 full-text journal scientific papers retrieved from the ScienceDirect database.

The text analysis of the volume and prevailing topics in the analysed papers reveals a significant increase in publications on social innovation, from 16 in 2000 to 573 in 2021, followed by a period of stabilization. Seven dominant topics emerged, including renewable energy, environmental/resource management, smart-city governance, sustainable food systems, corporate strategy, academic and methodological studies, and social governance structures. The analysis highlights the interdisciplinary and contested nature of social innovation, mirroring global policy shifts such as the Sustainable Development Goals. The findings guide scholars to understand the prevailing scientific discourse on social innovation across different time periods, correlate this with the general evolution of international political and policy documents and agendas on social innovation, and identify potential for stronger advocacy-related relationships between science and real-world needs in social innovation policy design. The findings also help identify under-explored niches and assist policymakers in locating evidence clusters when designing socially innovative interventions.

The structure of this paper is as follows: Section 2 outlines the theoretical background, followed by Section 3 detailing the methodology. Section 4 presents the text mining analysis and discussion. Finally, Section 5 offers conclusions, limitations, and future research directions.

2 Theoretical background

Despite the fact that various types of innovation can be recognized, each type of innovation has its distinct meaning and scope. Social innovation, however, specifically focuses on developing new responses to pressing societal demands, which influence the process of social interactions. Its primary goal is to enhance human well-being by fostering solutions that are beneficial to both society and individuals, empowering citizens, civil society organizations, local communities, businesses, and public institu-

tions (Godin, 2016).

In the initial period of intensive development of the idea of social innovation in the academic world, there were supposedly 252 different definitions of social innovation, referred to in more than 2,300 scientific publications (Edwards-Schachter & Wallace, 2017), creating a large number of theoretical settings, as well as boundary conditions (Oeij et al., 2019; Van der Have & Rubalcaba, 2016). Summarizing an extensive review of definitions by Solis-Navarrete et al. (2021), the scientific definition of social innovation encompasses interdisciplinary, temporal, and target diversity frameworks, referring to new ideas, solutions, processes, or tools for solving social problems or meeting needs to achieve positive, deep, and long-lasting changes that improve and empower living conditions.

However, the fact that social innovation is contested, conceptually imprecise, and used in ways that may be seen as disagreeable should not dissuade us from engaging with the concept (Ayob et al., 2016). It should be the goal of research to make sense of complex concepts and to understand their evolution over time (Collier et al., 2006). According to an early outline by Gallie (1956), contested concepts, such as social innovation, have five main characteristics: 1) they are appraised as signifying a valued achievement; 2) they are internally complex and so variously interpreted; 3) they are variously describable by different actors; 4) they are open and amenable to change over time; and 5) they are recognized as contested by the stakeholders, meaning that each actor attempts to assert their own authority in defining the concept.

A few years after the intense development of the idea of social innovation in the academic world, the concept also started to receive an intense response in the wider public political sphere. Social innovation has been recognized as one of the central ways to recover from crises and a building block for social, political, and environmental resilience since the end of the global financial crisis after 2009. Sustainable social and environmental development topics were addressed through social innovation discourse in the UN's 2030 Agenda for Sustainable Development, adopted in 2015 (United Nations, 2024). The 2030 Agenda for Sustainable Development outlines specific goals, pinpoints the most pressing issues, and provides a shared blueprint for peace and prosperity for people and the planet, now and into the future. While addressing climate change and working on planet preservation, the Agenda recognizes that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and spur economic growth (United Nations, 2024). Among these, SDG 9 – Industry, innovation and infrastructure and SDG 12 – Responsible consumption and production, are particularly relevant to social innovation. Spadafora and Rapaccini (2024) explore the intersection of servitization and social innovation, proposing that servitization – a model emphasizing service-oriented

solutions – can reshape industrial societies and enhance well-being. Their findings suggest that servitization is not only an economic strategy but also a form of social innovation with the potential to drive societal progress.

Another important concept related to social innovation in the international policy community is global citizenship, which emphasizes shared values and shared responsibility, acknowledging that local events are significantly shaped by global events, and vice-versa. The World Economic Forum (2017) defines global citizens as individuals and organizations – ranging from corporations to civil society actors – who recognize their roles in shaping global developments. Segales et al. (2023) investigated the relationship between social innovation and global citizenship in the case of facilitating sustainable and democratic energy transitions in cities. They suggested five criteria stemming from the intersection of social innovation and global citizenship perspectives: 1) democratic governance; 2) civil empowerment and capacity building; 3) human rights approach; 4) diversity of actors; and 5) sustainability (Segales et al., 2023). Likewise, De Souza Joao-Roland and Granados (2023) postulated that the processes and structures of a collaborative and user-centred approach and participatory organizational culture are positively linked with social innovation performance. Their study encompassing 78 social enterprises from the UK highlights the importance of cooperation with diverse stakeholders, particularly the community, beneficiaries, and universities. Development of solutions based on the community's needs is guaranteed by employing creative tools such as design thinking, and the engagement of people inside the social enterprise who have an entrepreneurial mindset. These efforts result in a positive linear relationship with social innovation and thus successfully drive social innovation.

Against this backdrop, the European Commission as early as 2013 launched the Social Innovation agenda, defined as “the development and implementation of new ideas – whether in the form of products, services and/or models in order to meet social needs and create new social relationships or collaborations. Addressing neglected or inadequately met societal challenges – including environmental, social issues – is central to social innovation” (European Commission, 2013). The European Commission (2013) identifies four key components of the social innovation process: 1) Identification of new/unmet/inadequately met social needs; 2) Development of new solutions in response to these social needs; 3) Evaluation of the effectiveness of new solutions in meeting social needs; and 4) Scaling up of effective social innovations. Given the increasing demand for public services and limited resources, policymakers often promote ‘social innovation’ as a means to address these tensions and challenges (Purcell et al., 2025). On the other hand, some critics argue that social innovation is merely a ‘fashionable concept’ or ‘buzzword’ in public policy discourse, emphasizing the need for fur-

ther empirical research to improve our understanding of the actors and mechanisms that drive effective social innovations (Purcell et al., 2025).

The concept of social innovation has also recently been explored from different lenses, including open innovation and R&D innovation perspectives. Arvaniti et al. (2024) define open social innovation as a new fast-growing discipline within Open Innovation, where they (re)emphasize the importance of co-creation, collaboration, and co-working under the auspices and shared efforts to tackle societal challenges. Today, many organizations and universities adopting social innovation principles employ open innovation strategies to generate tangible societal benefits (Arvaniti et al.).

By bridging public and market sectors, social innovation also enables the creation of products and services that align with both individual and collective aspirations. However, despite its potential, social-innovation-oriented companies often encounter resource constraints, which can significantly impact their economic or social performance (Lu & Wang, 2024). Lu and Wang (2024) analysed data from 598 social-innovation-oriented firms (period from 2010 to 2020) and found that non-R&D innovation complements R&D innovation, ultimately enhancing both economic and social performance. Their findings also suggest that the long-term impact of social innovation is more significant than its immediate impacts. Conversely, Jeannerat and Lavanchy (2024) pinpoint that traditional innovation policies historically prioritized technological advancements (e.g., traditional and technological innovation), while social innovation has only recently gained recognition, largely due to global challenges such as climate change and social inequalities. Further historical analysis conducted by Hu et al. (2024) traces social innovation back to 1642, demonstrating its role as a mediator between technological advance and economic expansion. They also highlight the Internet-of-Things as a powerful mechanism of social innovation in the sixth Kondratieff wave. As a result, a transformative policy paradigm is emerging, shifting from market-driven competition to a model that acknowledges social innovation as a crucial driver of systemic change (Jeannerat & Lavanchy, 2024). This evolving framework moves beyond the traditional ‘triple helix’ model—focused on interactions between research institutions, industry, and government. In addition, it involves and engages NGOs, civil society organizations, and individual citizens as key stakeholders in social innovation (Jeannerat & Lavanchy, 2024).

In this section, we have outlined the key dimensions of social innovation, highlighting the link between the academic and the actual social, political, and economic/business user environment, as currently reflected through the aspects of social, sustainability, and more recently digitalisation topics, as well as global governance and citizenship discourse. Social innovation is intertwined with all of the

abovementioned terms and is also an active player when addressing not only wider social challenges but also environmental, and broadly speaking, sustainable ones.

One of the most provocative questions relates to the creation of the social innovation discourse: who created what context of social innovation, and for whom? Has the scientific community shaped a broader social understanding of social innovation via its publications, or has the everyday social, environmental, economic, and political community steered science's view on social innovation? What topics and issues constitute social innovation? In seeking answers to these dilemmas, the purpose of this article is to examine how scientific publications have addressed social innovation over time, and how the field has evolved from a research perspective as well as in relation to the broader social agenda of the time. This is addressed by analysing the volume and variety of prevailing topics in leading academic work and international political documents from the pivotal year 2010 up to mid-2024. Based on existing findings regarding the evolution of social innovation in scientific and wider public discourse, the analytical part of this paper will focus on the following research questions:

RQ1: Has the number of scientific publications on social innovation risen in recent decades, corresponding to increasing interest in the field?

RQ2: Do the main themes in scientific papers on social innovation relate to social, environmental, and sustainability issues, systems of governance, and digitalisation?

RQ3: Does text analysis of scientific papers confirm a linkage between scientific, political, and business discourse on social innovation?

3 Methodology

We focused on scientific research papers and employed Science Direct as a source. We intentionally neglected book chapters, etc., as complete data for these could not be gathered. We focused on both review and research papers published in English. On July 24th, a search for “social innovation” yielded 4,703 papers. We downloaded all of them in their full-text version; six people simultaneously downloaded their respective shares of papers. We covered the time period from 2000 until mid-2024 (published until July 24, 2024). After that, we employed a combination of Python and R programming languages, leveraging their respective strengths in data extraction and statistical analysis.

3.1 Data Collection and Extraction

To efficiently extract textual content from the PDFs, we developed a custom Python script in Python 3.11.9 that

utilized the PyPDF2 (Fenniak et al., 2024) and pdfplumber (Singer-Vine et al., 2024) libraries. The script was designed to attempt text extraction using PyPDF2 first, due to its speed and efficiency in handling well-structured PDFs. If PyPDF2 failed to extract text—often the case with scanned documents or those with intricate formatting—the script automatically switched to pdfplumber, which is more robust in handling complex layouts but at the cost of increased processing time. This two-step approach maximized our ability to retrieve textual data while optimizing performance.

After successful extraction, the text from each document was cleaned to remove non-textual elements such as images, tables, and metadata. The cleaned text was then saved into CSV files, with each row representing a single document. This structured format facilitated seamless importation into R for subsequent analysis.

Data Preprocessing in R

Upon importing the extracted data into R version 4.4.1 using RStudio 2024.09.0, we undertook extensive preprocessing to prepare the text for analysis. The preprocessing steps included:

- **Tokenization:** Splitting the text into individual words or tokens using the tidytext package (Silge & Robinson, 2016).
- **Normalization:** Converting all text to lowercase to ensure consistency.
- **Stop Words Removal:** Eliminating common stop words (e.g., “and,” “the,” “of”) that do not contribute meaningful information to the analysis. Besides including stopwords from package, we defined the list of our own stopwords.
- **Stemming and Lemmatization:** Reducing words to their root forms using the textstem package (Rinker, 2018), which helps in grouping similar terms together.
- **Removing Punctuation and Numbers:** Excluding non-alphabetic characters to focus on meaningful text.
- **Handling Sparse Terms:** Removing infrequent terms that appear in a minimal number of documents to reduce dimensionality.

These preprocessing steps resulted in a clean and manageable dataset, from which we constructed a Document-Term Matrix (DTM) using the tm package (Feinerer, Hornik, & Meyer, 2008). The DTM is a mathematical matrix that describes the frequency of terms appearing in a collection of documents, serving as the foundation for text mining and topic modelling.

Parallel Processing

Given the large size of the dataset—comprising [over 400 000 words]—computational efficiency was a priority. We utilized parallel processing techniques to expedite data processing and analysis. By leveraging the doParallel (Microsoft Corporation & Weston, 2022) package, we

were able to distribute tasks across 11 cores of a multicore processor. This approach significantly reduced processing times, enabling us to perform complex computations that would otherwise be time-prohibitive.

Topic Modelling

To identify latent themes and patterns within the text data, we employed topic modelling techniques using the topicmodels package (Grün & Hornik, 2011). Specifically, we implemented Latent Dirichlet Allocation (LDA), a generative probabilistic model that allows sets of observations to be explained by unobserved groups. The number of topics was determined based on model perplexity and coherence scores, optimizing for interpretability and statistical validity.

We assessed the quality of the topics generated by examining the top terms associated with each topic and their distribution across documents. This analysis provided insights into prevalent research themes and how they have evolved over time.

Data Visualization and Analysis

For data manipulation and visualization, we utilized the tidyverse ecosystem (Wickham et al., 2019), including the dplyr and ggplot2 packages. dplyr facilitated efficient data frame operations such as filtering, summarizing, and joining datasets. ggplot2 allowed us to create high-quality visualizations to represent term frequencies, topic distributions, and temporal trends.

Results Interpretation

The combination of advanced text extraction, comprehensive preprocessing, and robust statistical modelling enabled us to uncover meaningful insights from academic literature. The analysis revealed trends across different periods, contributing to a deeper understanding of the academic landscape.

4 Text mining analysis with discussion

In this section, we will present the number of journal articles and words through years (4.1), keywords by year (4.2), word correlations (4.3), and topic modelling (4.4).

4.1 Number of journal articles and words through years

The bar chart (Figure 1) displays the number of publications per year from 2000 to 2024, highlighting a clear trend of growth over time. In the initial years between 2000 and 2010, the number of publications remained relatively low and stable, fluctuating between single digits and around 30 publications annually. Specifically, there was a slight peak in 2003 with 24 publications, followed by a dip to 9 publications in 2004, and modest growth afterward.

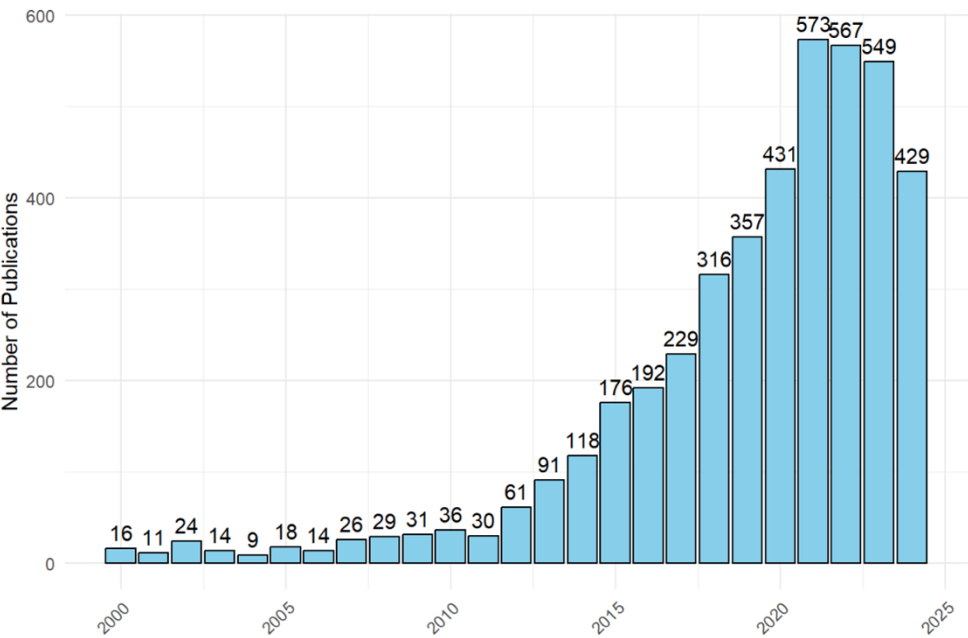


Figure 1: Number of publications per year from 2000 to 2024

From 2011 onward, a steady upward trend emerged, signalling increasing scholarly interest. The growth becomes more pronounced starting in 2015, where the publication count rises sharply from 61 publications in 2012 to 176 in 2015. This upward trajectory continues into the subsequent years, reaching 431 publications by 2020, marking a significant acceleration. We hypothesize that this growth from 2015 onwards is closely related to the 2030 Agenda for Sustainable Development, adopted in 2015 by all United Nations Member States. Especially considering sustainability topics and related pressing issues such as intensive consumption and resource use, social innovation can be seen as part of the solution pertaining to sustainable development goals (SDGs; United Nations, 2024). The period from 2021 to 2023 showcases a peak in publication activity. The highest number of publications occurred in

2021, with 573 publications, followed closely by 567 in 2022. These years demonstrate a sharp increase compared to previous years, possibly reflecting heightened interest or global factors influencing research output.

However, a slight decline is observed from 2023. In 2023, publications dropped to 549, and by mid-2024 (data collected until July 24), the count was 429. Although these numbers remain significantly higher than in earlier years, the 2023 figure suggests a possible stabilization, while the 2024 count reflects partial year data. Overall, the chart reveals a long-term growth trend in the number of publications, particularly accelerating from 2010 onward, peaking around 2021-2022, followed by a minor decline in 2023 and partial data for 2024. This pattern indicates a maturing research field with sustained, albeit with a recent plateau or slight decrease in fully completed years', scholarly output.

Table 1: Total number of words in publications annually between 2000 and 2024

Year	Total Words
2000	55,316
2001	87,017
2002	89,126
2003	57,211
2004	40,446
2005	62,028
2006	62,444
2007	128,238
2008	104,558
2009	151,812
2010	157,590
2011	141,783
2012	227,858
2013	447,530
2014	524,918
2015	759,884
2016	906,218
2017	1,216,992
2018	1,666,712
2019	1,764,112
2020	2,270,012
2021	2,878,006
2022	3,060,713
2023	2,718,392
2024	2,083,199

The dataset reveals a significant growth trend in the total number of words published annually between 2000 and mid-2024 (Table 1). In the early years (2000–2004), word counts remained relatively low, with a minimum of 40,446 words recorded in 2004. This period reflects limited research activity or fewer contributions to the field. From 2007 onward, the total word count increased steadily, surpassing 100,000 words and marking a transition to a more productive phase. By 2012, word counts surged to 227,858, illustrating a pivotal period of growth. The upward trend became exponential after 2015, where the total word count nearly tripled over five years, reaching its peak of 3,060,713 words in 2022. The years 2020–2022 represent the most productive period, coinciding with a global focus on research, possibly influenced by worldwide events like the COVID-19 pandemic that may have spurred academic activity. Post-2022, a slight decline is observed, with 2024 (partial year) recording 2,083,199

words. Despite the reduction, the output remains significantly higher than earlier years. Overall, this progression highlights a maturing research landscape, with a sharp increase in scholarly activity from 2012 onward, solidifying the growing importance and attention toward research topics during this period, and so also confirming our first research question expectation.

4.2 Keywords by year

From the table 2 related to the top 5 most frequent words by year (2000–2024), some key patterns emerge:

- Dominance of “social” and “innovation”: Starting from 2006, the word “social” becomes consistently dominant, highlighting its importance in academic discourse. It continues to remain the most frequent word, reflecting a focus on societal themes in research.

Table 2: The top 5 most frequent words by year (2000–2024)

Year	Word 1	Freq 1	Word 2	Freq 2	Word 3	Freq 3	Word 4	Freq 4	Word 5	Freq 5
2000	scenario	419	change	397	energy	366	future	354	environmental	353
2001	future	645	change	500	world	477	development	452	technology	446
2002	innovation	868	future	830	system	821	social	576	change	493
2003	change	624	social	490	process	456	system	369	technology	315
2004	science	330	innovation	267	service	263	social	257	system	228
2005	network	675	future	467	system	426	research	399	study	377
2006	social	695	change	388	future	373	system	371	study	351
2007	university	955	change	797	regional	682	project	660	development	656
2008	social	876	policy	661	change	630	innovation	630	development	622
2009	social	1633	innovation	1061	system	854	research	794	process	769
2010	social	1146	change	964	study	770	policy	718	system	696
2011	social	1359	network	874	research	839	city	791	process	720
2012	innovation	2768	social	1891	research	1553	system	1467	change	1458
2013	innovation	4234	social	3050	change	2846	policy	2747	research	2344
2014	social	5282	innovation	4656	change	3054	research	2941	study	2731
2015	social	7706	innovation	5760	research	3818	system	3798	development	3513
2016	social	9027	innovation	6849	energy	5585	research	4720	system	4572
2017	social	11233	innovation	9590	energy	6863	research	6329	study	5779
2018	social	15753	innovation	12710	energy	10093	development	8622	policy	8618
2019	social	17430	innovation	14943	research	10472	study	9105	system	8251
2020	social	19169	innovation	16683	research	14288	energy	12002	study	11461
2021	social	28943	innovation	22836	research	18937	study	15440	system	12707
2022	social	28171	innovation	26479	research	19894	energy	19752	study	17266
2023	innovation	24097	social	23166	research	18397	study	16562	development	12907
2024	social	17794	innovation	17243	research	13923	study	13235	energy	9622

- Rise of “innovation” and “research”: The word “innovation” gains prominence around 2012, marking its peak in 2013 and maintaining its high frequency thereafter. Similarly, “research” emerges as a frequently recurring term, showcasing an emphasis on scholarly investigation.
- Energy and Climate Themes: Words like “energy” and “change” appear repeatedly, particularly from 2016 onward, which could be attributed to rising concerns around sustainability, renewable energy, and climate change.
- Growth in “development” and “system”: Terms such as “development” and “system” remain consistent across the years, suggesting their central role in discussions related to technological, economic, or societal advancements.

Overall, the shift in word frequency over time reflects evolving priorities in research, from environmental concerns in the early 2000s to innovation, sustainability, and societal themes dominating recent years. Interestingly, this part of the analysis also does not support our expectations related to the increased presence of social-related contents, contents related to different forms of governance, includ-

ing also non-governmental and civil society communities and, also not more recently, digitalisation-related topics.

4.3 Words correlation

The correlation chart above (Figure 2) visually represents the relationships between words that frequently co-occur across all journal articles. Since the number of articles at the beginning of the examined period was small, the analysis was conducted on the entire dataset to reveal stronger word associations over time. The nodes (blue dots) represent individual words, and the edges (lines) connecting them indicate the level of correlation, where thicker lines signify stronger associations. Several distinct thematic clusters can be observed, reflecting different areas of academic focus in the dataset.

Major Observations and Interpretation

• Energy and Sustainability Cluster

A prominent cluster revolves around the term “energy,” which has strong connections to related words like “renewable,” “solar,” “electricity,” and “development.” This cluster highlights a significant focus on renewable

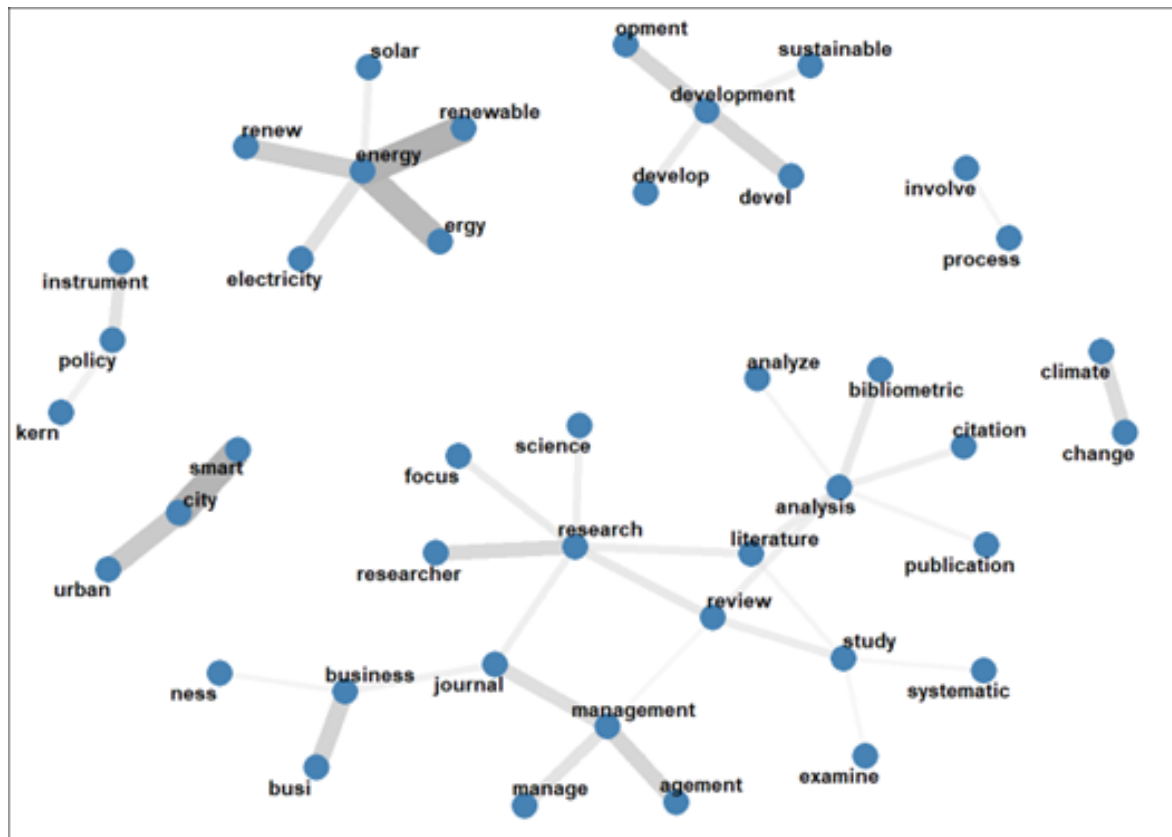


Figure 2: The relationships between words that frequently co-occur across all journal articles

energy and sustainable energy systems, key themes in research during the studied period. The appearance of terms like “solar” and “renewable” suggests increasing interest in clean energy technologies and climate action, aligning with global sustainability goals. Chen et al. (2025) note that sustainability-driven forces are key drivers for social innovations in microbusiness clusters.

- **Research and Analysis Cluster**

At the centre of the chart is the word “research,” which connects to words such as “science,” “focus,” “review,” “literature,” and “analysis.” This central position indicates the overarching importance of research-related activities in academic discourse. Terms like “bibliometric,” “citation,” and “publication” reflect a meta-analysis approach, where research on academic output and scholarly citations has become a subject of study itself. There are many studies conducted including term social innovation, however Daniel and Jenner (2022) highlight that social innovation still raises interest among scholars and policymakers, as a potential panacea for social disenfranchisement and civic dysfunction. What is troubling, according to researchers, is diverse perspectives of social innovation abound, creating inconsistencies in methodological approaches which confound theory development.

- **Urban Development and Smart Cities**

Another noticeable cluster includes terms such as “smart,” “city,” and “urban,” forming a coherent group. The co-occurrence of these words highlights the growing research focus on urbanization, smart city technologies, and sustainable urban development. This trend likely reflects the increasing importance of cities in global policy discussions around sustainability and infrastructure development. Researchers in this topic emphasize that there are present several challenges such as high implementation costs, slow technological adoption, and social equity issues, highlighting the complexity of achieving inclusive and sustainable urban evolution (Oyadeyi & Oyadeyi, 2025). This is related to the social dimension of social innovation, enhancing and providing well-being and equity to all and form of cooperation in resolving and addressing properly these challenges.

- **Management and Business Studies**

A distinct group focuses on “business,” “management,” and related words like “journal” and “busi” (likely stemming from truncation during preprocessing). The inclusion of “manage” and “agement” emphasizes themes surrounding business operations, management processes, and entrepreneurship. This cluster underscores a consistent academic interest in organizational management and corporate strategies over the years. Related to social innovation and entrepreneurship the results of literature review conducted by Grilo and Moreira (2022) reveal that the connection between social innovation and social entrepreneurship is in its take-off phase, but it still is a fragmented field with a huge lack of consensus.

- **Climate Change and Environmental Focus**

A small yet significant cluster features the terms “climate” and “change,” which strongly correlate. This reflects the consistent attention given to climate change research and its implications. The standalone yet connected nature of this cluster suggests that climate change remains a central yet independent topic, tying into broader themes like energy and sustainability. This is in line with the literature review conducted by Kouam and Asongu (2022), who as well emphasized the role of social innovation in achieving sustainable development. From a sustainable development perspective, social innovation plays a crucial role in addressing economic, social, and environmental challenges. In more detail, social innovation drives economic growth by creating jobs, fostering entrepreneurship, promoting sustainable agriculture, and encouraging innovation. It supports the transition to more inclusive and resilient economic systems. From the social dimension point of view, it promotes social cohesion, reduces inequalities, and expands opportunities for all, including women, youth, people with disabilities and the most vulnerable. By fostering inclusive participation, social innovation strengthens communities and promotes equitable development. From the environmental dimension, social innovation tackles environmental challenges by introducing new climate change adaptation and mitigation technologies. It also promotes sustainable consumption patterns and measures resilience, which all together support long-term environmental sustainability. (United Nations, 2024)

- **Process and Involvement Themes**

Words such as “process” and “involve” cluster together, indicating that studies often emphasize processes (methodological, operational, or organizational) and the involvement of stakeholders. This may signify a research focus on collaboration, engagement, or participatory frameworks in various contexts. Li and Bacete (2022) note technology’s role in driving social innovation through direct adoption or indirect engagement in co-design processes.

The correlation analysis reveals that academic research within this dataset has consistently focused on themes of energy sustainability, climate change, urban development, and management processes. The prominence of clusters such as energy-renewable-electricity and research-analysis-bibliometric reflects both domain-specific and methodological concerns. As global priorities shifted toward sustainability and urbanization, these themes have become intertwined with technological and policy-oriented research. Moreover, the interconnectedness of words like “research,” “analysis,” and “publication” signals an increasing interest in bibliometric studies and knowledge dissemination. Meanwhile, the growing focus on smart cities and renewable energy highlights a response to global challenges like climate change and rapid urbanization. This chart (Figure 2) effectively demonstrates how thematic priorities in academic articles are interconnected,

reflecting the focus of research in this field. Above all, it clearly supports our research question expectations regarding prevailing topics and the interconnectedness between scientific, political, and business/entrepreneurial social innovation agendas.

4.4 Topic modelling

The initial Document-Term Matrix (DTM) for topic modelling contained 4,347 documents (after removing documents with insufficient text for analysis, if applicable, or state this number is for topic modelling specifically) and

455,065 unique terms. It had 6,230,699 non-sparse entries. The sparsity level was very high, reflecting the common phenomenon of many terms appearing rarely across the document set. The high sparsity was addressed by filtering out words that appeared in fewer than 40 documents, reducing the matrix to 13,428 terms, which significantly improved the efficiency of topic modelling (Table 3).
To determine the optimal number of topics, the log-likelihood curve and multiple topic coherence metrics were evaluated (Figure 3). The ldatuning package (see Figure 4) provided metrics like Griffiths (2004), CaoJuan (2009), Arun (2010), and Deveaud (2014).

Table 3: Document-Term Matrix (DTM) Comparison: Initial vs. Filtered

Matrix Details	Initial DTM	Filtered DTM
Number of Documents	4347	4347
Number of Terms	455,065	13,428
Non-/Sparse Entries	6,230,699/1.97B	5,049,933/53.3M
Sparsity	100%	91%
Maximal Term Length	19	19
Weighting	Term Frequency	Term Frequency

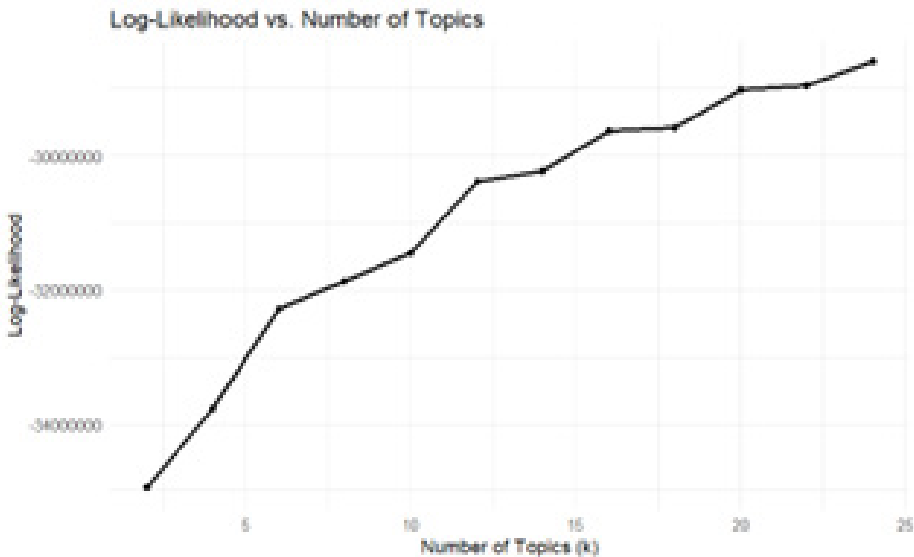


Figure 3: Determining the Optimal Number of Topics: Log-Likelihood vs. Number of Topics in LDA

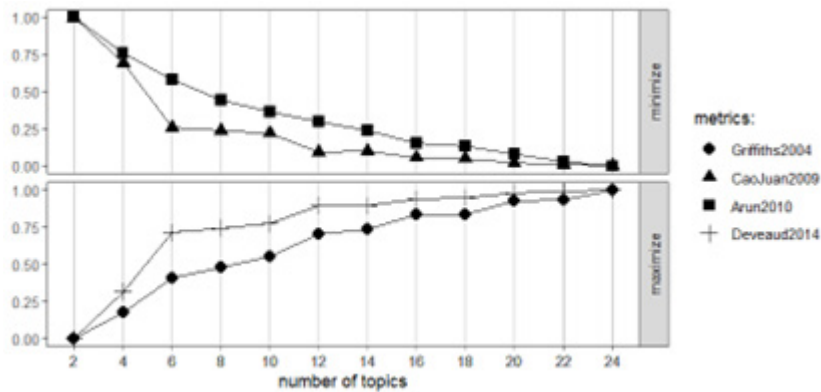


Figure 4: Comparison of Topic Coherence Metrics for LDA

Table 4: Identified Topics with Keywords and Suggested Names

Topic	Top Terms	Proposed Topic Name	Addressed in Policy Documents Discourse
1	supply, water, emission, price, consumption, gas, trade	Environmental and Resource Management	UN, EU, WEF
2	policy, system, community, local, public, project	Governance and Social Structures	UN, EU, WEF
3	research, literature, review, study, publication, analysis	Academic Research and Study Methods	UN, EU, WEF
4	business, management, product, service, innovation	Corporate Strategy and Entrepreneurship	UN, EU, WEF
5	energy, renewable, electricity, power, transition	Renewable Energy and Power	UN, EU, WEF
6	city, urban, smart, sustainable, citizen, development	Smart Cities and Urban Sustainability	UN, EU, WEF
7	food, waste, sustainable, chain, production, consumer	Sustainable Food Systems and Production	UN, EU, WEF

Note: UN = United Nations; EU = European Union; WEF = World Economic Forum

- The log-likelihood graph showed significant improvement in model fit as the number of topics increased from 2 to 7. Beyond 7 topics, the improvement plateaued, suggesting diminishing returns with additional topics.
- The coherence metrics also pointed to 7 topics as the optimal balance between clarity and complexity.

The Latent Dirichlet Allocation (LDA) and NMF (Non-negative Matrix Factorization) models generated 7 topics (see Table 4), each characterized by a set of words and their weights (β values).

Pertaining to Table 4, the Environmental and Resource Management topic focuses on themes related to resource management, emissions, energy consumption, and waste management. Terms like supply, water, emission, and trade

suggest a strong emphasis on sustainability and resource efficiency.

The Governance and Social Structures topic centres around governance, social structures, and public projects, featuring terms such as policy, system, community, and local. It highlights the critical role of governance in societal development and community initiatives. In Academic Research and Study Methods, terms like research, study, literature, review, and analysis indicate a focus on academic methodologies, scholarly publications, and systematic reviews. The Corporate Strategy and Entrepreneurship topic emphasizes business management, entrepreneurship, and product innovation, as shown by words like business, management, service, and product. The Renewable Energy and Power topic highlights renewable energy and electrification, with terms such as energy, renewable, electricity, and power, reflecting global sustainability trends and energy transitions. Addressing urban planning and smart city development, the Smart Cities and Urban Sustainability topic includes terms like city, urban, smart, sustainable, and citizen, reflecting research on sustainable infrastructure and urban innovations. Finally, the Sustainable Food Systems and Production topic focuses on food systems, waste

management, and production chains, with terms like food, waste, chain, production, and consumer, emphasizing increasing concerns about sustainable consumption and food security.

Above we can see that social innovation is present in different topics and addresses different dimensions of sustainability. Social innovation is perceived as an independent innovation type but also is seen in inter-dependence with other innovation forms (e.g., technological, product, service, organizational, business, and design-driven innovations). It stems from multi-stakeholder and cross-sectoral cooperation and results in either development or adoption of social innovation between the public and private actors, in collaboration with civil society (Edwards-Schachter, 2018). Moreover, also The Organization for Economic Co-operation and Development (Organisation for Economic Co-operation and Development [OECD], 2025) has extensively addressed social innovation in its policy documents, recognizing its pivotal role in addressing social and environmental challenges. Social innovation is defined by the OECD as the development and implementation of new solutions that improve the quality of life for individuals and communities.

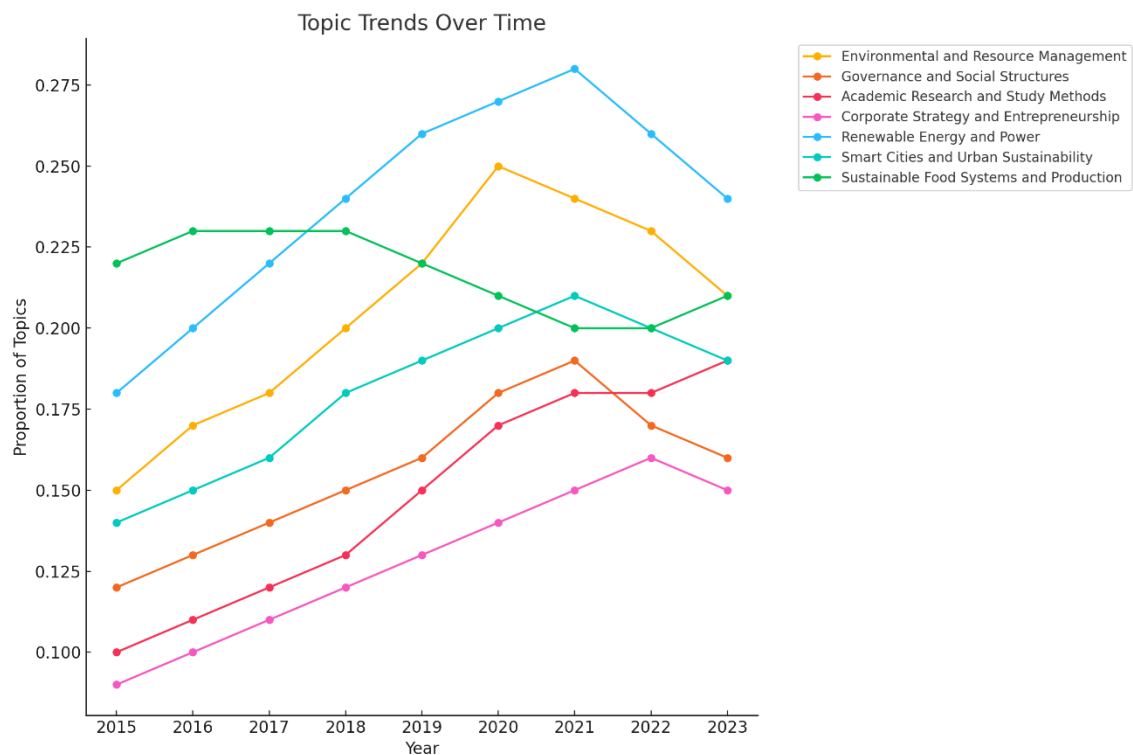


Figure 5: Evolution of Topic Proportions from 2015 to 2023

The visualisation (Figure 5) illustrates the trends of seven key topics identified through topic modelling over the period 2015–2023, with each line representing the proportion of articles addressing a specific topic in a given year. Renewable Energy and Power emerges as the most dominant topic, steadily rising from 2015 and peaking around 2021 before slightly declining in 2022–2023, reflecting growing global interest in sustainability and renewable energy transitions. Similarly, Environmental and Resource Management shows a notable upward trend, particularly between 2015 and 2020, as its share increased from approximately 15% to over 25%, driven by discussions on resource efficiency, emissions control, and energy management. Sustainable Food Systems and Production maintains a consistent presence throughout the years, hovering around 22%–23%, underscoring ongoing research interest in sustainable consumption, food security, and production chains. Meanwhile, Smart Cities and Urban Sustainability exhibits steady growth, peaking at approximately 21% in 2021, highlighting the focus on smart infrastructure and urban innovation. Academic Research and Study Methods demonstrates gradual but steady growth from approximately 10% in 2015 to around 18% in 2023, reflecting the importance of scholarly methodologies and systematic reviews. Corporate Strategy and Entrepreneurship begins at a lower level but steadily increases, reaching its peak around 2021, signalling a growing emphasis on innovation, entrepreneurship, and business management strategies. Although Governance and Social Structures remains among the lower proportion topics overall, it experiences noticeable growth until 2021, emphasizing the role of governance and community-oriented initiatives. Collectively, these trends reveal evolving research priorities over these years, with a pronounced focus on renewable energy, sustainability, and innovation, aligning with global movements toward environmental consciousness and strategic advancements. This part of the text analysis and its interpretation particularly supports our research question expectations regarding the tight interconnectivity between the academic agenda and global political and policy agendas.

5 Conclusion

This paper examined a set of questions related to social innovation content within the scientific community over time. By analysing a large sample of scientific research papers (4,703 articles) from 2000 to mid-2024 and employing a combination of Python and R programming languages, we arrived at several conclusions.

The evolution of scientific interest in social innovation topics has been rising both in volume and content variety within the scientific community over time. Our main operational findings are as follows. First, while in 2000 there

were 16 published papers including the term “social innovation,” in the first seven months of 2024 alone (data until July 24), 429 papers were published on the topic. Especially since 2011, the results demonstrate increasing interest in social innovation among researchers, with a steady annual increase in publications on the subject, closely followed by the emergence of strategic political and policy documents on social innovation from the EU, UN, World Bank, and OECD.

Second, our findings indicate that “social” and “innovation” have been among the top five most frequent words by year since 2006 (being the core search terms). In 2012, there was a rise in the use of “research” (alongside the search terms), while from 2016 onward, words like “energy” and “change” began to appear more frequently. Additionally, “development” and “system” have remained central terms across all years, as many publications address technological, economic, or social issues.

Third, regarding word correlations, our text mining analysis identified six major clusters: 1) Energy and sustainability, 2) Research and analysis, 3) Urban development and smart cities, 4) Management and business studies, 5) Process and involvement themes, and 6) Governance and Policy.

Fourth, our topic modelling identified seven key topics based on their top terms: 1) Environmental and Resource Management, 2) Governance and Social Structures, 3) Academic Research and Study Methods, 4) Corporate Strategy and Entrepreneurship, 5) Renewable Energy and Power, 6) Smart Cities and Urban Sustainability, and 7) Sustainable Food Systems and Production. These topics highlight the broad scope of social innovation and its potential to bring benefits and improvements across many areas and various sectors of everyday life, fostering better modes of mutual cooperation among diverse stakeholders in society. This part of the analysis has also shown and reinforced that, despite its diversity, social innovation is developing as an important part of academic research and methodology. A major and unexpected finding from the content analysis is the relatively low incidence of specific ‘social content’ terms (beyond the core search terms ‘social’ and ‘innovation’) compared to other thematic contents. This underscores a critical need for the scientific community to more explicitly address diverse social-related aspects within social innovation in future research, not only for academic purposes but also for practical applications and policy impact.

Referring back to the theoretical introduction, we can generally confirm our research questions, which were based on previously gained conclusions about the heterogeneous, diverse, evolving, and intertwined understandings of social innovation within academic and actual everyday social and political discourse. Gallie’s (1956) parameters of contested concepts have not only been confirmed but also further reinforced by the assessment that it

is these differences that can lead to progressive competition and better quality of arguments over time.

The dominant central themes in academic publications also clearly coincide with actual social, economic, environmental, and political conditions and their global exposure within the context of social innovation. The discourse of social innovation in academia appears to align closely with everyday policy-making and business-run agendas.

Regarding the limitations of our study, one limitation stems from the fact that we examined only scientific research papers. Another one pertains to the method used in this paper, where strengths include a large full-text dataset analysis and robust computational techniques for identifying trends, while weaknesses involve reliance on a single database and the inherent simplifications of automated text processing and topic interpretation. Future research could extend this analysis to other types of literature, such as books, and, in particular, conduct a detailed analysis of leading political and policy documents. Additionally, our study focused on the ScienceDirect database, which covers a significant portion of relevant scientific journal articles; however, future research could incorporate other databases to provide a more comprehensive view and complement these recent findings.

Hence, future research directions could include replicating our study using different publication types or employing different databases. Future research could also explore specific aspects of social innovation or analyse its impact within particular domains. Given the increasing trend in publications pertaining to social innovation and the growing practical interest in the topic, we can conclude that social innovation is gaining well-deserved recognition and merit. This rise may stem from the broader sustainability movement and the acknowledgment that social innovation can be a response to environmental, social, and economic challenges.

Concluding with the thoughts of Minguijon et al. (2024), who emphasize that social services function as a protective system designed to support individuals throughout their lives, while innovation plays a crucial role in addressing contemporary societal challenges and adapting to an ever-evolving world. Ultimately, social innovation plays a crucial role in fostering responsibility and driving positive change in our environment.

Literature

- Arvaniti, E. N., Stylios, C. D., & Dima, A. (2024). What do European stakeholders think about open social innovation? *IFAC-PapersOnLine*, 58(3), 55–60. <https://doi.org/10.1016/j.ifacol.2024.07.124>
- Ayob, N., Teasdale, S., & Fagan, K. (2016). How social innovation ‘came to be’: Tracing the evolution of a contested concept. *Journal of Social Policy*, 45(4), 635–

653. <https://doi.org/10.1017/S004727941600009X>
- Chen, W. H., Chen, S., Lin, Y. C., Chen, C. L., Ng, W. K., Ding, Y., Shih, Y. S., & Chen, H. N. (2025). Achieving social innovation in the digital age: A case study of microbusiness. *International Journal of Innovation Studies*, 9(1), 16–29. <https://doi.org/10.1016/j.ijis.2024.12.003>
- Collier, D., Hidalgo, F. D., & Maciuceanu, A. O. (2006). Essentially contested concepts: Debates and applications. *Journal of Political Ideologies*, 11(3), 211–246.
- Daniel, L. J., & Jenner, P. (2022). Another look at social innovation: From community – for community. *International Journal of Innovation Studies*, 6(2), 92–101. <https://doi.org/10.1016/j.ijis.2022.04.001>
- de Souza João-Roland, I., & Granados, M. L. (2023). Towards social innovation strategy: An analysis of UK social enterprises. *Technological Forecasting and Social Change*, 187, 122189. <https://doi.org/10.1016/j.techfore.2022.122189>
- Edwards-Schachter, M. (2018). The nature and variety of innovation. *International Journal of Innovation Studies*, 2(2), 65–79. <https://doi.org/10.1016/j.ijis.2018.08.004>
- Edwards-Schachter, M., & Wallace, M. L. (2017). ‘Shaken, but not stirred’: Sixty years of defining social innovation. *Technological Forecasting and Social Change*, 119, 64–79.
- European Commission. (2013 / 2025). *Guide to social innovation*. Retrieved February 19, 2025, from <https://european-social-fund-plus.ec.europa.eu/en/social-innovation>
- Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54. <https://doi.org/10.18637/jss.v025.i05>
- Fenniak, M., Stamy, M., pubpub-zz, Thoma, M., Peveler, M., exiledkingcc, ... pypdf Contributors. (2024). *pypdf* (Version 4.0.0) [Computer software]. <https://pypi.org/project/pypdf/>
- Gallie, W. B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56, 167–198.
- Godin, B. (2016). Making sense of innovation: From weapon to instrument to buzzword. *Quaderni*, 90, 1–?? <https://doi.org/10.4000/quaderni.977>
- Grilo, R., & Moreira, A. C. (2022). The social as the heart of social innovation and social entrepreneurship: An emerging area or an old crossroads? *International Journal of Innovation Studies*, 6(2), 53–66. <https://doi.org/10.1016/j.ijis.2022.03.001>
- Grün, B., & Hornik, K. (2011). Topicmodels: an R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30. <https://doi.org/10.18637/jss.v040.i13>
- Hu, M. C., Phillips, F., & Wu, C. Y. (2024). Social innovations mediate technology and economy in the Kondratieff waves. *Technological Forecasting and Social Change*, 201, 123262. <https://doi.org/10.1016/j.tech>

- fore.2024.123262
- Jeannerat, H., & Lavanchy, P. (2024). Transformative social innovation in, of and by the city: Beyond mission-driven policy rationales. *Environmental Innovation and Societal Transitions*, 52, 100890. <https://doi.org/10.1016/j.eist.2024.100890>
- Kouam, J. C., & Asongu, S. A. (2022). Effects of taxation on social innovation and implications for achieving sustainable development goals in developing countries: A literature review. *International Journal of Innovation Studies*, 6(4), 259–275. <https://doi.org/10.1016/j.ijis.2022.08.002>
- Li, C., & Bacete, G. (2022). Mapping the technology footprint in design for social innovation. *International Journal of Innovation Studies*, 6(3), 216–227. <https://doi.org/10.1016/j.ijis.2022.07.003>
- Logue, D. (2020). *Theories of social innovation*. Edward Elgar. <https://doi.org/10.4337/9781786436894>
- Lu, X., & Wang, J. (2024). Is innovation strategy a catalyst to solve social problems? The impact of R&D and non-R&D innovation strategies on the performance of social innovation-oriented firms. *Technological Forecasting and Social Change*, 199, 123020. <https://doi.org/10.1016/j.techfore.2023.123020>
- Microsoft Corporation, & Weston, S. (2022). *doParallel: Foreach parallel adaptor for the “parallel” package* (Version 1.0-17) [R package]. <https://CRAN.R-project.org/package=doParallel>
- Minguijón, J., Serrano-Martínez, C., & Raya-Díez, E. (2024). Reliability and validity of the public innovation hexagon (PIH) as applied to the Spanish social services. *International Journal of Innovation Studies*, 8(3), 237–248. <https://doi.org/10.1016/j.ijis.2024.02.004>
- Murzintcev, N. (2020). *ldatuning: Tuning of the latent Dirichlet allocation models parameters* (Version 1.0-2) [R package]. <https://CRAN.R-project.org/package=ldatuning>
- Organisation for Economic Co-operation and Development. (2025). *Better policies for better lives*. Retrieved February 26, 2025, from <https://www.oecd.org/>
- Oeij, P. R., Van der Torre, W., Vaas, F., & Dhondt, S. (2019). Understanding social innovation as an innovation process: Applying the innovation journey model. *Journal of Business Research*, 101, 243–254.
- Oyadeyi, O. A., & Oyadeyi, O. O. (2025). Towards inclusive and sustainable strategies in smart cities: A comparative analysis of Zurich, Oslo, and Copenhagen. *Research in Globalization*, 10, 100271. <https://doi.org/10.1016/j.resglo.2025.100271>
- Pedersen, T. L. (2024). *ggraph: An implementation of grammar of graphics for graphs and networks* (Version 2.2-1) [R package]. <https://CRAN.R-project.org/package=ggraph>
- Purcell, C., Manthorpe, J., & Malley, J. (2025). Understanding the role of internal governance units in the process of social innovation: The case of Shared Lives Plus in England. *Research Policy*, 54(1), 105152. <https://doi.org/10.1016/j.respol.2024.105152>
- Rinker, T. W. (2018). *textstem: Tools for stemming and lemmatizing text* (Version 0.1-4) [R package]. <https://github.com/trinker/textstem>
- Segales, M., Hewitt, R. J., & Slee, B. (2023). Social innovation and global citizenship: Guiding principles for sustainable, just and democratic energy transition in cities. *Energy Research & Social Science*, 106, 103295. <https://doi.org/10.1016/j.erss.2023.103295>
- Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, 1(3), Article 37. <https://doi.org/10.21105/joss.00037>
- Singer-Vine, J., & The pdfplumber contributors. (2024). *pdfplumber* (Version 0.11-4) [Computer software]. <https://github.com/jsvine/pdfplumber>
- Solis-Navarrete, J. A., Bucio-Mendoza, S., & Paneque-Gálvez, J. (2021). What is not social innovation. *Technological Forecasting and Social Change*, 173, 121190.
- Spadafora, M., & Rapaccini, M. (2024). Bridging the gap between servitization and social innovation. *Journal of Cleaner Production*, 452, 142178. <https://doi.org/10.1016/j.jclepro.2024.142178>
- United Nations. (2024). *The 17 goals*. Retrieved December 17, 2024, from <https://sdgs.un.org/goals>
- Van der Have, R. P., & Rubalcaba, L. (2016). Social innovation research: An emerging area of innovation studies? *Research Policy*, 45(9), 1923–1935.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- World Economic Forum. (2017). *What is global citizenship?* Retrieved December 17, 2024, from <https://www.weforum.org/stories/2017/11/what-is-global-citizenship/>

dr. **Uroš Godnov** is an assistant professor in the Department of Computer Science at the University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies. He studies data science, text mining, and artificial intelligence. He designs courses that teach students to turn raw data into clear insights and covers state-of-the-art principles in Python and R.

Assoc. prof. dr. **Jana Hojnik**, Faculty of Management, University of Primorska, Koper, Slovenia. She works as an associate professor at the Department of Entrepreneurship at Faculty of Management. Her research interests include sustainability, circular economy, eco-innovation, social innovation and knowledge transfer. Jana Hojnik is also a member of the Centre for Development and Transfer of Knowledge of the University of Primorska and editor-in-chief of the scientific journal *Managing Global Transitions* (About the Journal | *Managing Global Transitions*).

Prof. dr. **Simona Kustec**, Faculty of Management, University of Primorska, Koper, Slovenia. She works as a full professor at the Department of Political Science, Sociology, Communication, Ethics and Cultural Studies in Management. Her research interests include public policy analysis, governance and management, innovation and knowledge policies, human rights and democratic practices. Simona Kustec is also a head of the Centre for Development and Transfer of Knowledge of the University of Primorska and editor-in-chief of the University of Primorska Press.

Razvoj družbenih inovacij v znanstvenih publikacijah: pristop tematskega modeliranja in besedilnega rudarjenja

Cilj: S pomočjo posebne metodologije tematskega modeliranja in besedilnega rudarjenja slediti razvoju akademskega diskurza o družbenih inovacijah od leta 2000 do sredine leta 2024 z vidika obsega objav in izpostavljenih vsebinskih vidikov oz. vodilnih tematik.

Podatki in viri: 4.703 polnih besedil znanstvenih člankov, pridobljenih iz zbirke Science Direct.

Metode: Pregled literature; besedilo iz datotek PDF, ekstrahirano s PyPDF2 in pdfplumber; očiščeno in tokenizirano v programskem okolju R; tematsko modeliranje, izvedeno z Latentno Dirichletovo alokacijo (optimizirano z ldatuning); časovne in korelacijske analize, vizualizirane s paketom tidyverse.

Rezultati: Število publikacij se je znatno povečalo s 16 (leta 2000) na 573 (leta 2021), nato pa se je ustalilo. Pojavilo se je sedem prevladujočih tematik: obnovljiva energija, okoljsko upravljanje/upravljanje z viri, upravljanje pametnih mest, trajnostni prehranski sistemi, korporativna strategija, akademsko-metodološke študije in strukture družbenega upravljanja. Besedni par »družbeno« in »inovacija« je postal najpogostejši po letu 2006; izrazi, povezani z energijo so močno narasli po letu 2016. Presenetljivo je, da tematike, ki se običajno štejejo za 'družbene', v znanstvenih skupnostih v diskurzu o družbeni inovaciji v primerjavi z zgoraj navedenimi prevladujočimi tematikami niso prevladovali.

Razprava: Naši rezultati v veliki meri potrjujejo obstoječe ugotovitve iz pregledov literature ter interdisciplinarno, pa tudi nejasno, sporno in še vedno intenzivno razvijajočo se naravo družbenih inovacij v znanstvenem objavljanju. Prevladujoče tematike družbenih inovacij v znanstvenih člankih sovpadajo z diskurzom družbenih inovacij v globalnih političnih in strateških dokumentih, zlasti EU (od leta 2013 naprej) in agende Ciljev trajnostnega razvoja ZN iz leta 2015, pri čemer poudarjajo tudi sodelovanje med znanstvenimi, poslovnimi, političnimi in nevladnimi deležniki. Kot takšne, znanstvene objave o družbenih inovacijah lahko služijo tudi kot znanstveno, na dokazih temelječe zagovornišvo za druge deležnike, vključene v procese, povezane z vsebinami družbenih inovacij.

Zaključki: Raziskovanje družbenih inovacij je danes uveljavljeno, sistemsko in široko interdisciplinarno področje preučevanja, osredotočeno na trajnost, nove tehnologije in teme upravljanja. Tesno je povezano s političnimi in strateškimi agendami vodilnih mednarodnih organizacij, pa tudi poslovnega sveta in nevladnih organizacij.

Priporočila: Ugotovitve usmerjajo raziskovalce k manj raziskanim družbeno relevantnim vsebinam in nišam (kot so upravljanje in zlasti teme pravičnosti) ter pomagajo oblikovalcem politik in drugim deležnikom, vključenim v procese družbenih inovacij, pri iskanju in dokazih temelječih pristopov pri oblikovanju njihovih družbeno inovativnih odzivov, intervencij, rešitev in ukrepov.

Ključne besede: *Teorije družbenih inovacij, Globalna politična agenda, Besedilno rudarjenje, Tematsko modeliranje, Pregled literature*