PROCEEDINGS OF THE
INTERNATIONAL CARBON NEUTRALITY TRAINEESHIP PROGRAM
Volume.01, Number.1, 2023, 37-49

Research Topics and Trends in European Union Energy Policy: A Structural Topic Model

Yeye LI

Hainan University, Haikou City, Hainan Province, China, 570228

E-mail: liyeye0320@126.com

***Corresponding author**

Abstract

Energy policy is considered a key factor in achieving carbon neutrality in the EU. In recent decades, a substantial number of academic articles have been published on EU energy policy. This study aims to provide a comprehensive review of the development of energy policy research in the EU. We utilized Structural Topic Modeling (STM) to analyze research topics and trends in EU energy policy literature. STM is an unsupervised machine learning method that facilitates large-scale unstructured text mining and reveals research topics and evolutionary trends. We collected and analyzed 1777 articles published between 1975 and 2022. Our findings indicate that the primary academic focus of EU energy policy is related to port management, container operations, and liner shipping management. Furthermore, our analysis reveals that in the early days, researchers focused on energy performance and alternative energy sources such as wind and bioenergy. Later, research shifted towards broader topics such as renewable energy, climate change, and energy efficiency. More recently, CO_2 emissions, sustainability, energy management, energy consumption, carbon pricing, decarbonization, energy poverty, and energy equity are hot topics of research. The major research topics and emerging trends identified from STM can assist researchers, funding communities, and policymakers in identifying contemporary research issues and making more informed decisions.

Keywords: European union; energy policy; text mining; structural topic model; research trends.

1. Introduction

Climate change is one of the most pressing global issues of our time, and the energy sector plays a significant role in contributing to carbon dioxide emissions that have a major impact on it. Therefore, improving energy efficiency and reducing energy demand are crucial tools for mitigating climate change (Bertoldi, P., 2018). The European Union (EU) has been actively pursuing a climate policy for the past few decades, recognizing that energy production and use account for 80% of all greenhouse gas emissions in the region (Economidou et al., 2020). Thus, the development of effective energy policies is crucial to the EU's energy transition and to achieving the goal of carbon neutrality. A comprehensive understanding of current market conditions and developments is necessary to ensure the effectiveness of energy policies.

In recent years, EU energy policy has attracted significant attention from researchers, and efforts have been made to review research content and trends in this area. For example, Economidou et al. (2020) reviewed 50 years of EU policies on building energy efficiency and provided insights and recommendations for achieving complete decarbonization of buildings in the future. Kanellakis et al. (2013) outlined the historical evolution and current status of EU energy policy from 1950 to 2012. However, due to the coexistence of different perspectives, the scientific opinion on EU energy policy remains highly fragmented. Traditional analytical methods are limited by a small number of studies and are subject to researcher subjectivity and selective bias. To overcome these limitations, this study aims to gain an overview of the research profile in the field of EU energy policy and to predict future research trends using artificial intelligence technology and text mining techniques.

The Structural Topic Model (STM) is used to analyze research topics and trends in EU energy policy literature by collecting and examining metadata from publications in the 27 EU member states. This study seeks to provide insights and recommendations to support researchers, policymakers, and entrepreneurs. The remainder of this study is organized as follows: Section 2 describes the research methodology, Section 3 presents the results of the structural topic model analysis, including visualization and discussion of the findings, and the paper concludes with a summary and discussion of the key findings.

2. Methodology

2.1. Structural Topic Model (STM)

Structural Topic Modeling (STM) is an unsupervised machine learning approach that enables the discovery of the latent topic structure within a given corpus of text data. It has become a popular method for exploring the key themes and trends within various research fields due to its efficiency and ability to handle large volumes of unstructured data. For instance, scholars such as Farrell (2016) and Sietsma et al. (2021) have used STM to investigate topics related to corporate finance and ideological polarization about climate change, and scientific literature on climate change adaptation, respectively. Surprisingly, however, STM has yet to be employed in the realm of energy policy research.

2.2. Research outline

The research outline presented in Figure 1 illustrates the various steps involved in the research process. The first step is data collection, which forms the foundation for all subsequent research activities. Following data collection, the collected data is pre-processed, including case conversion, punctuation removal, stemming, among other techniques. The aim of this step is to ensure that the modeling results are accurate. Finally, the research requires determining the number of topics (K) and gaining insights from the model results. Each step will be elaborated upon in the following sections.

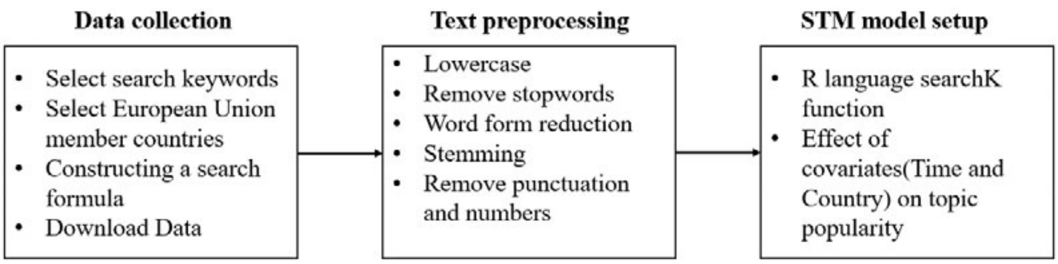


Fig.1 Research outline

2.3. Data collection

The research literature on energy policy in the 27 EU member countries was retrieved by conducting a search on the Web of Science (WOS). The search was conducted using the following search formula: (TI=ENERGY AND TI=POLICY) AND (DT==("ARTICLE") AND SILOID==("WOS") AND CU==("AUSTRIA" OR "BELGIUM" OR "BULGARIA" OR "CROATIA" OR "CZECH REPUBLIC" OR "DENMARK" OR "ESTONIA" OR "FINLAND" OR "FRANCE" OR "GERMANY" OR "GREECE" OR "HUNGARY" OR "IRELAND" OR "ITALY" OR "LATVIA" OR "LITHUANIA" OR "LUXEMBOURG" OR "MALTA" OR "NETHERLANDS" OR "POLAND" OR "PORTUGAL" OR "ROMANIA" OR "SLOVAKIA" OR "SLOVENIA" OR "SPAIN" OR "SWEDEN" OR "CYPRUS")). The search yielded a total of 1777 research papers. The number of research papers retrieved for each EU member country is presented in Table 1. It is important to note that, due to cooperation between some countries, an article may have been counted for more than one country.

Table 1. Number of articles by Country

Country	Number of Articles	Country	Number of Articles
GERMANY	382	LITHUANIA	36
ITALY	236	HUNGARY	30
NETHERLANDS	203	CZECH REPUBLIC	27
FRANCE	153	ROMANIA	27
SPAIN	153	LATVIA	22
SWEDEN	140	CROATIA	20
FINLAND	97	CYPRUS	17
AUSTRIA	96	SLOVENIA	14
GREECE	91	LUXEMBOURG	11
POLAND	90	SLOVAKIA	11
DENMARK	89	ESTONIA	10
BELGIUM	86	MALTA	4
PORTUGAL	56	BULGARIA	3
IRELAND	47		

2.4. Data preprocessing

We utilized the "stm" package in the R language to perform data pre-processing tasks as outlined in Roberts et al. (2019). Data pre-processing is a crucial step that involves cleaning the text data to make it more normalized. This involves converting all capital letters to lowercase (e.g., Energy to energy), removing stopwords (e.g., a, an, the, of and other words that are often found in the text but do not carry any substantive meaning), reducing word forms to their root form (e.g., reducing plural or morphologically different words to their prototype form such as energies to energy), stemming (i.e., removing suffixes from words to reduce them to their root form), and eliminating punctuation and numbers.

2.5. STM model setup

One of the most critical tasks in constructing an STM model is to determine the appropriate number of topics, denoted as K. This is because the choice of K is linked to the degree of interpretability of the model. Nevertheless, there is no definitive method for selecting the optimal number of topics in the STM algorithm. Therefore, we will employ the R language "searchK" function to assist us in determining the number of topics.

The STM model enables us to examine the impact of covariates on topic prevalence by utilizing a generalized linear model. In this study, we contend that researchers' attention towards different topics varies over time, while the degree of attention given to topics differs among countries. As such, we have defined two covariates, Time and Country, to represent the year of publication and the country where the article was published, respectively. Equation (1) depicts the relationship between topic prevalence and these two covariates. Additionally, we have considered the interaction of time and country, hypothesizing that the pattern of change in attention towards topics varies across countries over time.

$$Prevalence = g\,(Time, Country, Time \times Country) \qquad (1)$$

3. Results

3.1. Descriptive analysis

Table 2 presents the details of the literature data, including the year of publication, the source of the literature, and the average number of publications per year. Figure 1 shows the annual distribution of papers, indicating that the first article was published in 1975, and the number of papers has been slowly increasing. However, the number of papers started to rapidly increase in 2002, and by 2021, the number of papers published has exceeded 150, suggesting an increasing attention to research on EU energy policy. Table 3 provides information on the top 10 cited papers. Figure 3 illustrates the distribution of papers by journal, where *Energy Policy* is the top journal with 336 publications, followed by *Energie* and *Renewable&Sustainable Energy Reviews* with 86 and 68 articles, respectively. Figure 4 shows the word cloud of author keywords, where the font size represents the frequency of occurrence. It can be observed that keywords such as "energy transition", "climate policy", "climate change", and "renewable energy sources" are widely mentioned and important research topics in this field. Figure 5 presents the keyword dynamics over time, revealing the evolution of research in this field. Initially, the research mainly focused on energy performance and alternative energy sources such as wind and bioenergy. Later, the research expanded to broader topics such as renewable energy, climate change, and energy efficiency. In recent times, hot research topics include CO_2 emissions, sustainable development, energy management, energy consumption, carbon pricing, decarbonization, and social issues such as energy poverty and energy equity.

Table 2. Article statistics

Description	Results
Timespan	1975:2023
Sources (Journals, Books, etc)	527
Documents	1754
Document Average Age	7.02
Average citations per doc	22.11
References	71569
Keywords Plus (ID)	2100
Author's Keywords (DE)	4041
Authors	4075
Authors of single-authored docs	339
Single-authored docs	385
Co-Authors per Doc	2.94
International co-authorships %	40.82

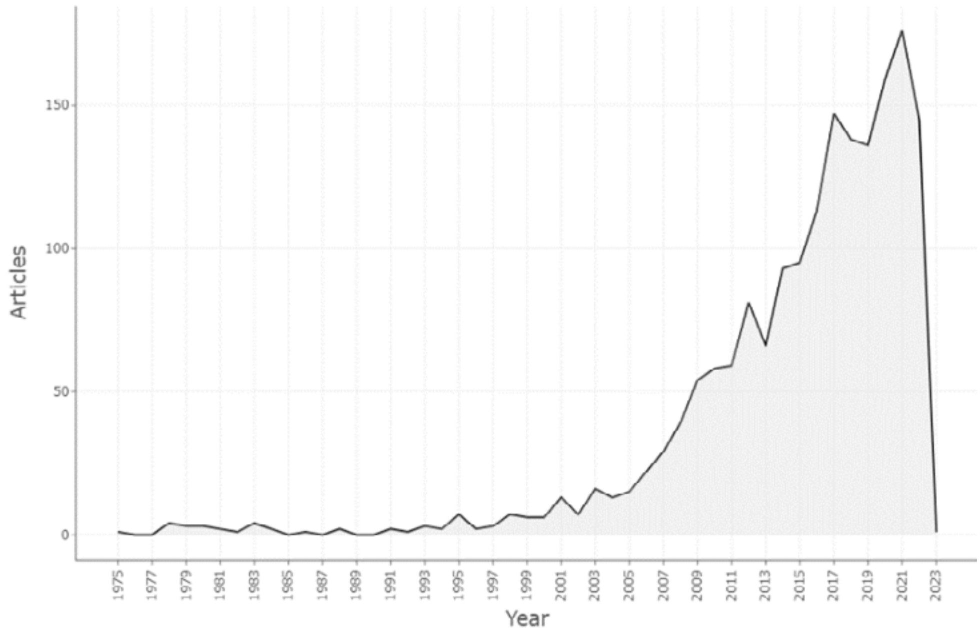


Fig.2 Annual distribution of Articles

Table 3 Most cited Articles

Article	DOI	Total Citations	TC per Year	Normalized TC
JOHNSTONE N, 2010, ENVIRON RESOUR ECON	10.1007/s10640-009-9309-1	797	61.31	14.45
JACOBSSON S, 2006, ENERG POLICY	10.1016/j.enpol.2004.08.029	611	35.94	9.60
STEG L, 2005, J ENVIRON PSYCHOL	10.1016/j.jenvp.2005.08.003	597	33.17	7.71
MENANTEAU P, 2003, ENERG POLICY	10.1016/S0301-4215(02)00133-7	465	23.25	9.38
LINDEN AL, 2006, ENERG POLICY	10.1016/j.enpol.2005.01.015	228	13.41	3.58
DE GROOT HLF, 2001, ENERG ECON	10.1016/S0140-9883(01)00083-4	218	9.91	5.41
FOUQUET D, 2008, ENERG POLICY	10.1016/j.enpol.2008.06.023	207	13.80	5.12
POLZIN F, 2015, ENERG POLICY	10.1016/j.enpol.2015.01.026	204	25.50	8.02
NESTA L, 2014, J ENVIRON ECON MANAG	10.1016/j.jeem.2014.01.001	203	22.56	7.12
SCARLAT N, 2015, RENEW SUST ENERG REV	10.1016/j.rser.2015.06.062	197	24.63	7.74

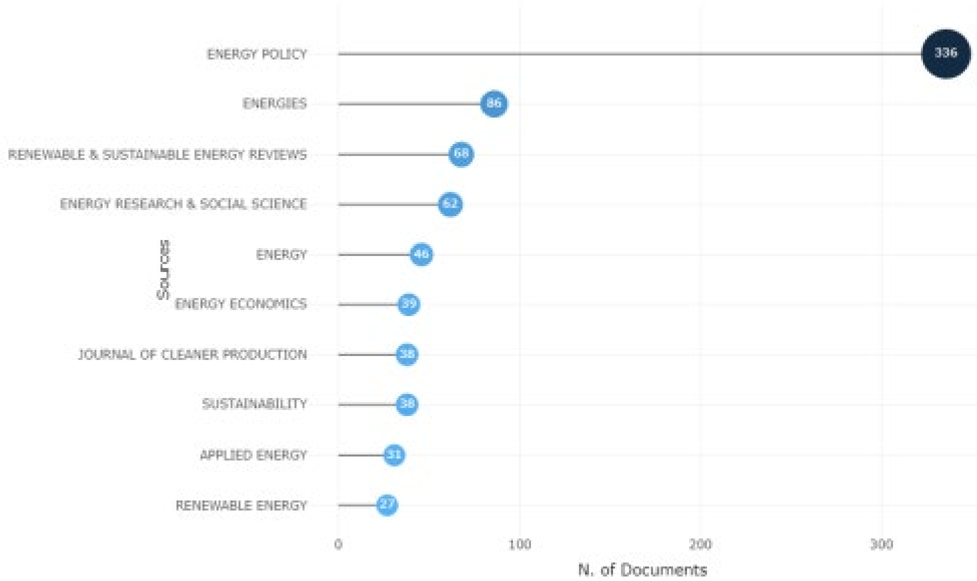


Fig.3 Journal distribution of Articles

Fig.4 Keyword word cloud

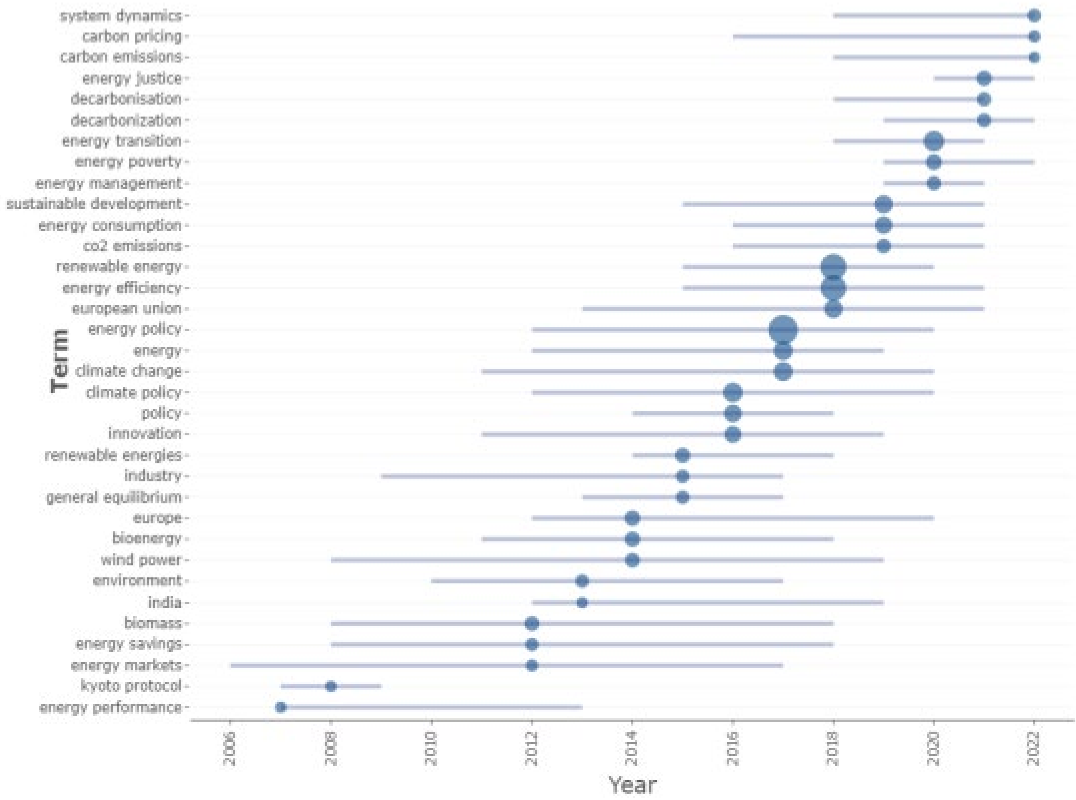
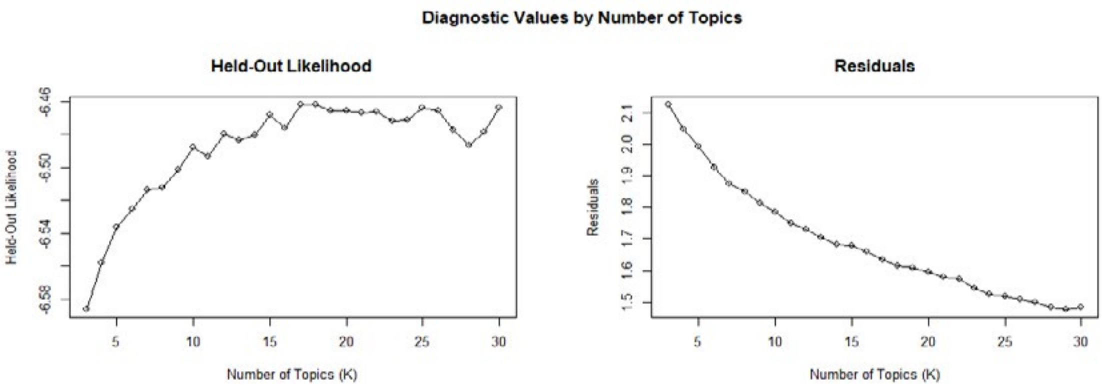


Fig.5 Keyword dynamics

3.2. Structural Topic Model (STM) Results

3.2.1. Choosing the number of topics

Figure 6 displays the diagnostic values for topic count. The top-left plot, "Held-out Likelihood", is used to estimate the final performance of the model after training and validation on the dataset. The bottom-left plot, "Semantic Coherence", is used to assess the coherence of the extracted topics. The bottom-right plot, "Lower Bound", represents the lower bound of the model convergence. It can be observed that both "Held-out Likelihood" and "Semantic Coherence" exhibit an inflection point at K=10, indicating that the optimal number of topics is 10.



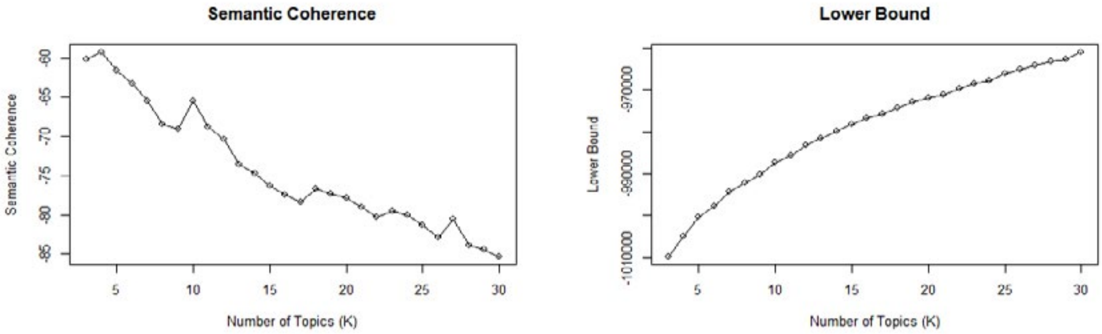


Fig.6 Topic number diagnostic value

3.2.2. Topic summary and validation

Table 4 shows the results of the STM model. The second column shows the top 10 words for each topic and the STM package provides four topic word outputs: Highest Prob is a high frequency word; FREX is a word with a high interpretability based on the probability and exclusivity ratio; Lift is a word with a high probability and high exclusivity with high discriminability; Score is the log frequency of the word in the topic divided by the log frequency of the word in other topics. Please refer to Roberts et al.(2019) for the exact calculation. The first column shows the names of the topics grouped according to keywords.

Table 4 displays the results of the STM model. The second column lists the top 10 words for each topic, and the STM package provides four different topic word outputs:

"Highest Prob" represents a high-frequency word.

"FREX" represents a word with high interpretability based on the probability and exclusivity ratio.

"Lift" represents a word with both high probability and high exclusivity and high discriminability.

"Score" represents the log frequency of the word in the topic divided by the log frequency of the word in other topics.

The exact calculation can be found in Roberts et al. (2019). The first column shows the names of the topics, which are grouped according to their respective keywords.

Topic 1 explores the impact of energy policy on both the environment and the economy. Topic 2 is primarily concerned with energy efficiency. Topic 3 examines public opinion on energy policy. Topic 4 focuses on nuclear energy. Topic 5 explores the energy consumption associated with computing, algorithms, and big data. Topic 6 discusses the development of renewable energy policy. Topic 7 is centered on climate change. Topic 8 examines the renewable energy markets. Topic 9 analyzes policies related to energy. Topic 10 primarily deals with energy consumption in buildings, house renovation, and similar topics.

The significance of each topic varies, and Figure 7 depicts the distribution of topic percentages. The analysis reveals that Topic 9 receives the highest research attention, followed by Topic 6 and Topic 3. In contrast, Topic 10 is the least explored area in the field.

Figure 8 presents the interrelationships between the ten topics, indicating that Topic 5 and Topic 6 are relatively independent. Topic 9 exhibits strong linkages with Topics 3 and 4, as it pertains to policy analysis that must consider public opinion, which is particularly crucial in the context of nuclear energy. Additionally, Topic 2 is linked to Topic 10, where energy-efficient design, house renovation, and other measures are essential for reducing energy consumption in buildings. Furthermore, Topic 7 is linked to Topic 8 and Topic 1, which primarily address climate, environmental, and economic issues.

The preceding analysis demonstrates that our model possesses a high degree of interpretability.

Table 4 Topic summary

Topic	Keywords
Topic 1:The economic and environmental impact of energy policy	Highest Prob: energi, effect, polici, use, result, product, impact, countri, environment, econom

Topic 1:The economic and environmental impact of energy policy	FREX: incom, tax, subsidi, panel, bioga, posit, regress, effect, elast, relationship
	Lift: irrig, ordinari, farmer, cointegr, autoregress, cge, asymmetr, short-run, capita, crop
	Score: irrig, household, tax, crop, elast, price, bioga, subsidi, agricultur, incom
Topic 2：Energy efficiency	Highest Prob: energi, effici, polici, measur, industri, paper, effect, save, use, improv
	FREX: effici, save, barrier, measur, industri, residenti, behaviour, standard, improv, end-us
	Lift: properti, smes, rebound, voluntari, applianc, white, dutch, effici, end-us, money
	Score: properti, effici, residenti, save, rebound, industri, smes, applianc, barrier, audit
Topic 3:Citizens' views	Highest Prob: energi, transit, research, polici, develop, innov, sustain, social, studi, technolog
	FREX: transit, citizen, research, social, societi, communiti, innov, attitud, expert, dimens
	Lift: postal, everyday, gender, sdg, media, conscious, finnish, justic, democraci, synthesi
	Score: gender, transit, citizen, democraci, justic, everyday, poverti, innov, narrat, communiti
Topic 4:Nuclear Energy	Highest Prob: energi, govern, state, region, polit, polici, articl, power, intern, european
	FREX: german, cooper, germani, nuclear, chines, govern, russia, reform, intern, coalit
	Lift: fukushima, ideolog, lobbi, vote, bilater, began, accid, salienc, state-own, amend
	Score: lobbi, nuclear, polit, german, govern, chines, germani, coalit, russia, feder
Topic 5:Data calculation and energy consumption	Highest Prob: model, energi, polici, system, optim, use, propos, network, result, can
	FREX: optim, model, simul, comput, batteri, machin, algorithm, paramet, network, stochast
	Lift: idl, radio, throughput, traffic, algorithm, determinist, energy-awar, markov, outperform, sensor
	Score: idl, optim, algorithm, model, sensor, batteri, machin, wireless, simul, traffic
Topic 6:Renewable energy policy developments	Highest Prob: energi, renew, develop, european, countri, polici, sourc, union, nation, sustain

Topic 6:Renewable energy policy developments	FREX: res, union, poland, european, lithuania, sourc, member, situat, countri, india
	Lift: lithuania, lithuanian, inland, polish, lignit, eurostat, poland, czech, disclos, ghana
	Score: inland, european, lithuania, renew, res, union, poland, lithuanian, countri, sourc
Topic 7:Energy and climate change	Highest Prob: emiss, energi, climat, global, fuel, scenario, carbon, gas, sector, reduct
	FREX: emiss, global, mitig, greenhous, fuel, fossil, ghg, carbon, transport, gas
	Lift: safeti, pledg, emit, sequestr, food, co-benefit, ghg, passeng, greenhous, net-zero
	Score: emiss, safeti, fuel, carbon, scenario, ghg, fossil, transport, climat, food
Topic 8:Renewable energy market	Highest Prob: electr, renew, technolog, cost, market, invest, power, system, wind, support
	FREX: electr, wind, tariff, solar, invest, feed-, deploy, plant, risk, investor
	Lift: cost-effici, fleet, motor, turbin, onshor, premium, investor, tariff, bid, tradabl
	Score: motor, wind, electr, renew, cost, feed-, power, solar, tariff, price
Topic 9:Policy analysis	Highest Prob: polici, climat, instrument, differ, approach, chang, process, analysi, integr, framework
	FREX: interact, prefer, mix, instrument, integr, outcom, coher, nexus, coordin, qualit
	Lift: contest, incoher, fossil-bas, intermediari, nich, cross-sector, horizont, uncov, enact, coher
	Score: contest, instrument, polici, climat, nexus, conflict, actor, prefer, intermediari, nich
Topic 10:Building renovation	Highest Prob: build, energi, heat, local, urban, citi, perform, studi, use, hous
	FREX: build, hous, citi, urban, retrofit, heat, municip, renov, stock, local
	Lift: metropolitan, retrofit, epc, hous, renov, mayor, epbd, coven, citi, municip
	Score: epc, build, heat, hous, citi, urban, retrofit, renov, municip, dwell

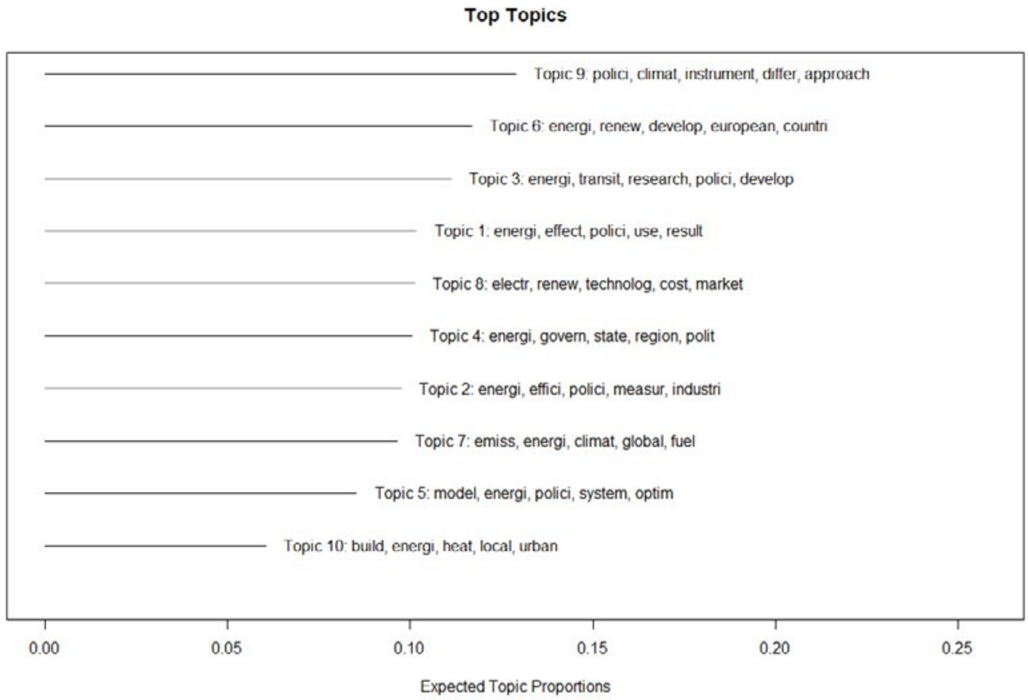


Fig.7 Topic distribution

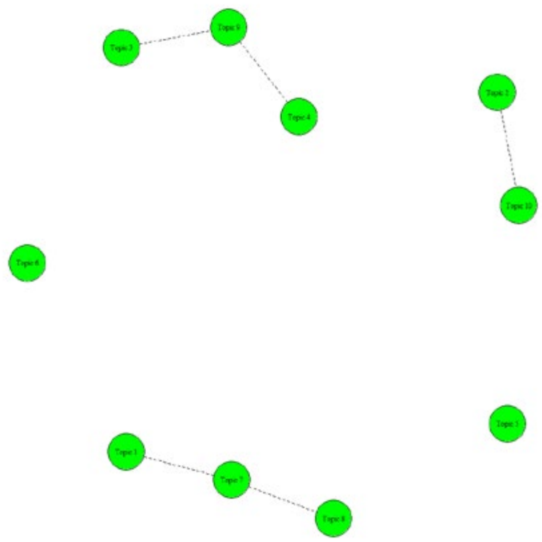


Fig.8 Thematic relationship diagram

3.2.3. Effect of covariates (Time and Country) on topic popularity

We have conducted a temporal analysis of the trends in topics, as presented in Figure 8. Our findings indicate that Topic 2 and Topic 4 have exhibited a declining trend, whereas Topic 5, Topic 7, and Topic 10 have shown an increasing trend. Additionally, Topic 3 and Topic 9 have trended upwards, but have experienced a significant decline in recent times. Other topics have displayed fluctuating changes. These changes may be related to policy enactments and significant social events, which warrant further investigation in future studies.

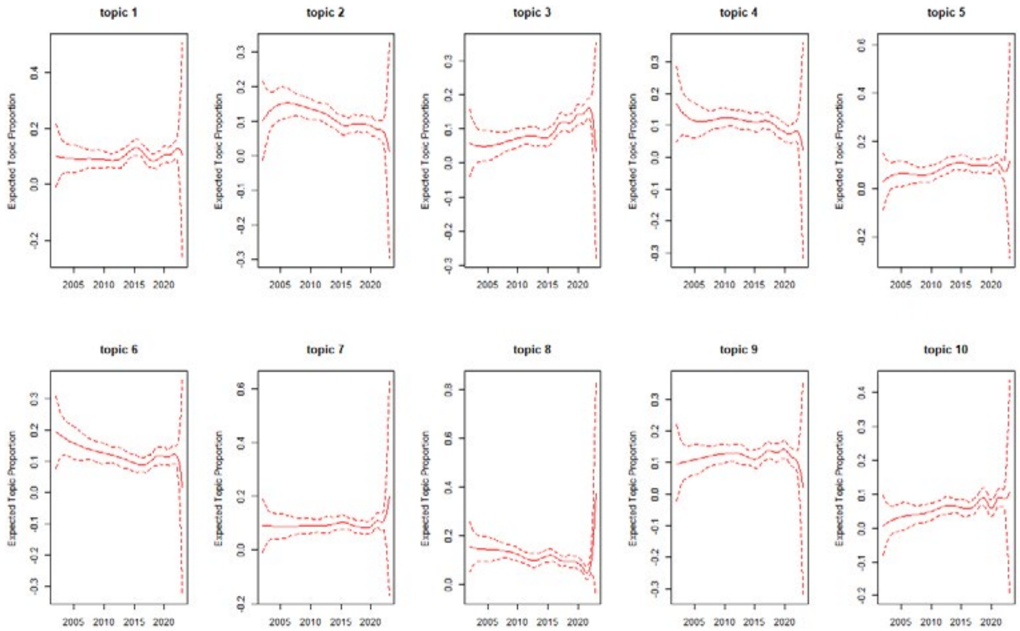


Fig.9 Trend of topics over time

The research focus varies across different countries. In our STM model, we added countries as covariates and compared the differences between Germany and Italy, as illustrated in Figure 10. Germany and Italy were selected for comparison because they have the highest number of publications in the field. Nonetheless, it is also possible to compare any two countries using the R language STM package. Our analysis shows both similarities and differences in the research focus of these two countries. Specifically, there is little difference in their focus on Topic 3 and Topic 4, while significant differences exist in their focus on Topic 5, Topic 10, Topic 4, and Topic 9, as well as in Topic 1, Topic 2, Topic 7, and Topic 8.

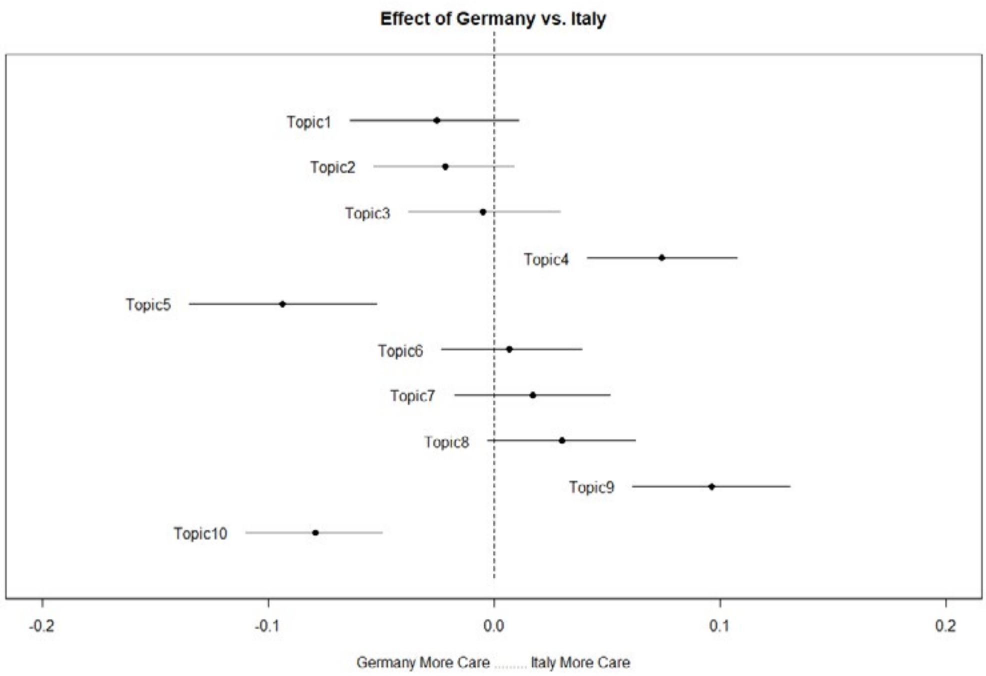


Fig.10 Similarities and Differences in Country Topics Focus

4. Conclusions

We conducted unsupervised machine learning using STM topic modeling on the abstract sections of 1777 articles related to energy policy published in 27 EU member states. We identified ten topics based on diagnostic values, which are detailed in Section 3.2.2. Through visual presentation of the topic distribution and interrelationships, we not only gained a better understanding of the model results but also confirmed the model's interpretability. Moreover, we analyzed how the topics evolved over time and compared the similarities and differences in research focus among different countries. We provided a detailed account of our data collection and processing procedures, the R language packages employed, and the key functions used in our study.

To our knowledge, our research is the first to apply STM topic modeling to the field of energy policy. The major research topics and emerging trends we identified through STM analysis can assist researchers, funding agencies, and policy makers in identifying current research issues and making well-informed decisions. However, our study has some limitations. For example, further investigation is needed to determine the factors contributing to the growth or decline of certain topics.

Reference

[1] Bertoldi, P. (2018). The Paris agreement 1.5 ◦C goal: What it does mean for energy efficiency. Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings; ACEEE: Washington, DC, USA.
[2] Economidou, M., Todeschi, V., Bertoldi, P., D'Agostino, D., Zangheri, P., & Castellazzi, L. (2020). Review of 50 years of EU energy efficiency policies for buildings. Energy and Buildings, 225, 110322.
[3] Farrell, J. (2016). Corporate funding and ideological polarization about climate change. Proceedings of the National Academy of Sciences, 113(1), 92-97.
[4] Kanellakis, M., Martinopoulos, G., & Zachariadis, T. (2013). European energy policy—A review. Energy Policy, 62, 1020-1030.
[5] Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. Journal of Statistical Software, 91, 1-40.
[6] Sietsma, A. J., Ford, J. D., Callaghan, M. W., & Minx, J. C. (2021). Progress in climate change adaptation research. Environmental Research Letters, 16(5), 054038.