Following Socio-Environmental Conflict Narratives About Energy Transition in Chile: A Spatio-Temporal Analysis Using Dynamic Topic Modeling

Jonas Rieger[1,*], Felipe Muñoz[2], Lars Grönberg[1], Kai-Robin Lange[1], Iván Ojeda-Pereira[3], Darío Briceño[2], Christian Nass[2], Carsten Stahl[1], José Cassola[4], Carolina Rojas-Córdova[5], Brian Keith-Norambuena[4], Marcelo Lufin[6], Fernando Campos-Medina[2] and Sebastián Herrera-León[7]

[1]*Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany*
[2]*Department of Sociology, Universidad de Chile, Capitán Ignacio Carrera Pinto 1045, Ñuñoa, Chile*
[3]*Department of Social Sciences, LUT University, Yliopistonkatu 34, Lappeenranta, Finland*
[4]*Department of Computing and Systems Engineering, Universidad Católica del Norte, Angamos 0610, Antofagasta, Chile*
[5]*Department of Industrial Engineering, Universidad Católica del Norte, Angamos 0610, Antofagasta, Chile*
[6]*Department of Economics, Universidad Católica del Norte, Angamos 0610, Antofagasta, Chile*
[7]*Institute of Engineering Sciences, Universidad de O'Higgins, Avenida Libertador Bernardo O'Higgins 611 Rancagua, Chile*

Abstract

Understanding the construction of socio-environmental narratives at a national scale is a complex challenge, particularly when research remains fragmented across disconnected case studies. In Chile, the energy transition has generated territorial disputes as extractive industries and renewable energy projects expand, yet large-scale systematic analyses of how these conflicts are represented in public discourse remain scarce. This paper addresses this gap by applying a spatio-temporal topic modelling framework to a corpus of 1,996 validated news articles covering conflicts related to the energy transition in Chile from 2011 to 2025. Using RollingLDA, a dynamic adaptation of latent Dirichlet allocation that prevents information leakage from future documents, we identify twelve topics that provide insights into the public narratives surrounding socio-environmental conflicts. Our analysis reveals how specific conflicts, such as the HidroAysén dam project, the Dominga mining controversy, and pollution in sacrifice zones such as Quintero-Puchuncaví, have evolved over time, with some narratives declining while others, including green hydrogen development and lithium extraction, have emerged as central concerns. We complement this temporal analysis with a spatial dimension by mapping the prevalence of topics across Chilean regions through an interactive dashboard. By combining established methods, our work offers a reproducible framework that can be adapted to topic modelling results incorporating spatial and temporal dimensions, enabling the tracking of how socio-environmental narratives emerge, evolve, and fade over time. Please also refer to the GitHub repository at https://github.com/JonasRieger/t2s2026.

Keywords

Dynamic Topic Modeling, RollingLDA, Socio-Environmental Conflicts, Energy Transition, Narrative Extraction

1. Introduction

Socio-environmental conflicts have consolidated as a structural phenomenon in countries of the Global South, particularly in contexts where economic development strategies have consistently been based

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'26 Workshop, Delft (The Netherlands), 29-March-2026

*Corresponding author.
rieger@statistik.tu-dortmund.de (J. Rieger); felipe.munoz.5@ug.uchile.cl (F. Muñoz); lars.groenberg@tu-dortmund.de (L. Grönberg); kalange@statistik.tu-dortmund.de (K. Lange); ivan.ojeda-pereira@lut.fi (I. Ojeda-Pereira); dario.briceno@ug.uchile.cl (D. Briceño); christian.concha@ug.uchile.cl (C. Nass); carsten.stahl@tu-dortmund.de (C. Stahl); jose.cassola@alumnos.ucn.cl (J. Cassola); carojas@ucn.cl (C. Rojas-Córdova); brian.keith@ucn.cl (B. Keith-Norambuena); mlufin@ucn.cl (M. Lufin); fernandocampos@uchile.cl (F. Campos-Medina); sebastian.herrera@uoh.cl (S. Herrera-León)
0000-0002-0007-4478 (J. Rieger); 0009-0003-5305-5988 (F. Muñoz); 0009-0002-0639-1121 (L. Grönberg); 0000-0003-1172-9414 (K. Lange); 0000-0002-5146-0002 (I. Ojeda-Pereira); 0000-0002-1029-1548 (J. Cassola); 0000-0001-5734-8962 (B. Keith-Norambuena); 0000-0002-6856-6401 (M. Lufin); 0000-0002-7772-3544 (F. Campos-Medina); 0000-0003-3315-9272 (S. Herrera-León)

on extractive industries and large-scale infrastructures [1]. In these settings, the expansion of economic activities has generated persistent disputes over access to and use of natural resources, environmental and health impacts, and the legitimacy of decision-making processes, commonly framed under the concept of extractivism [2, 3, 4]. Chile represents a particularly relevant case in this context, mainly due to the central role that mining plays in the national production matrix [5, 6]. After the return to democracy in the 1990s, Chile embraced the so-called commodities consensus in Latin America [3, 7, 8], a cross-party agreement that framed natural resource extraction and its fiscal revenues as drivers of poverty reduction and economic growth. However, the reduction of extreme poverty has been accompanied by a proliferation of socio-environmental controversies that have gained increasing public visibility, articulated both in territorial settings and in the media sphere [9, 10, 11, 12, 13].

Although Chile is a regional leader in the transition to clean energy, researchers warn that the local territorial dynamics associated with the global energy transition often reproduce extractive logics [14, 15, 16]. First, energy technologies are materially dependent on minerals, which has intensified global demand and further expanded mining activities that were already generating significant social and environmental tensions [17, 18]. Second, the development of new energy production and transport infrastructures has also generated territorial conflicts, particularly in contexts where these projects intersect with Indigenous territories and other local economic activities [19, 20]. From a critical analytical perspective, the concept of the extractive-energy nexus highlights how the energy transition itself may deepen extractivism, its frictions, and its socio-environmental impacts [21].

Academic research on extractivism and related conflicts has intensified in recent years, consolidating a dynamic and diverse field while also revealing recurring limitations. An analysis of indexed publications from 2015 to 2020 addressing conflict, territory, and extractivism in the Chilean case identifies several problematic patterns: (i) the routinisation of extractivism as a concept, frequently used as an adjective or contextual descriptor rather than as an analytical category; (ii) the treatment of conflict as a largely taken-for-granted notion, with limited conceptual problematisation; (iii) a simplified understanding of territory; (iv) a recurrent classification of actors along a state, company, and civil society axis; (v) a predominant focus on the local scale; and (vi) a clear dominance of case studies as the main methodology [22]. In response, the authors argue that renewing the field requires, among other elements, methodological innovation and a more explicit discussion of the key concepts structuring these studies [22].

This paper engages directly with the identified limitations. In particular, we adopt a critical stance towards the strong localist perspective that characterises large parts of the research on socio-environmental conflicts in Chile, both in the classic literature on extractivism and in more recent studies focused on energy-related conflicts. The predominance of case study-based approaches limits the capacity to identify general patterns, recurring dynamics, and shared features across conflicts, especially when these unfold over discontinuous or intermittent time spans, or across territorially dispersed settings.

Building on this diagnosis, we propose a shift in the analysis to the national scale, using news data to identify narrative patterns that reveal trends and temporal dynamics in an industrial landscape continuously reconfigured by the introduction of new technologies. Accordingly, we approach socio-environmental conflicts as configurations that are constructed, circulated, and transformed within public narratives. This perspective enables us to trace how different controversies are narratively articulated over time and how shared discursive patterns emerge across conflicts that might otherwise appear disconnected. Beyond specific topics, meaning-based groupings or clusters are identified through human interpretative analysis, capturing conflict dynamics linked to different industrial transformations, such as those associated with fossil fuels or new renewable energy technologies, while simultaneously foregrounding institutional dimensions and impacts on communities and fragile ecosystems.

1.1. Related Quantitative Approaches

This shift in the scale of analysis, from qualitative case-based research to quantitative national-level analysis, is not uncommon, especially in narrative research, which has greatly benefited from the recent

development of powerful natural language processing (NLP) models. In particular, models based on the Transformer architecture [23] have demonstrated a high level of language understanding, enabling them to extract narratives from documents without having to rely on rigid, manually defined linguistic rules. NLP pipelines such as RELATIO [24] and its adaptations [25] utilise classic NLP tasks such as part-of-speech tagging, named entity recognition, and coreference resolution to extract economic narratives from a corpus. The Narrative Trails methodology [26] identifies relevant related documents based on a given start and end document and arranges them along a timeline. More recently, studies have shown that decoder-only models provide more stable results than the previously mentioned pipelines by performing narrative extraction in a single step through accurate prompting of the model. [27, 28, 29, 30].

While the aforementioned studies show the theoretically optimal performance that could be achieved by applying an LLM such as GPT [31] to an entire corpus, they also highlight the significant costs of such an endeavour. Prompting an LLM for all documents in a large corpus results in substantial financial or computational costs for the user. Another option, not within the scope of this paper, is to use small or distilled language models, which often maintain a large part of the performance of larger models through different approaches, such as parameter-efficient algorithms or teacher-student learning [32], making them significantly less complex. Other studies have focused on developing appropriate narrative extraction tools using traditional low-resource NLP methods. For instance, [33, 34] use a temporal adaptation [35] of the classic topic model latent Dirichlet allocation [LDA, 36] to detect shifts in the content of newspaper articles during Covid-19 as well as shifts in political discussions in the German Bundestag. This approach was extended in [37] to explain these shifts and annotate them according to the Narrative Policy Framework [38, 39, 40].

These studies on narrative extraction propose methods that focus not only on extracting narratives from a corpus, but also on performing this extraction in a low-resource environment and integrating a temporal component into their research. However, they do not take possible non-temporal factors into account in their analysis.

1.2. Contribution

The general objective of this article is to analyse socio-environmental conflicts associated with the energy transition in Chile through a quantitative narrative extraction approach applied to news media data. Specifically, the article aims to: (i) design and implement a methodological pipeline oriented towards the identification and structuring of conflict narratives in the media; and (ii) identify and characterise dominant and topic-specific socio-environmental conflict narratives at different spatial levels, considering their distribution and temporal dynamics.

This paper is positioned at the intersection of socio-environmental conflict research, social studies of energy transitions, and narrative analysis based on text mining methods. By conceptualising socio-environmental conflicts as emerging narrative configurations within news corpora, the paper proposes a methodological framework consisting of already existing methods, thereby contributing to the combination and development of computational approaches for narrative identification, structuring, and comparison in complex public debates. In this sense, the contribution not only extends the application domain of narrative extraction methods, but also provides a challenging empirical test case, namely nationwide socio-environmental conflicts related to the energy transition in Chile, to assess the ability of these methods to capture multi-scalar, temporal, and thematic dynamics in non-European and non-English-language contexts.

The approaches mentioned in Section 1.1 largely demonstrate that automating narrative extraction alone does not (yet) work reliably. This is one reason why we use a hybrid approach in this work, combining NLP methods with human judgement and avoiding black box automation where appropriate. Consequently, from a broader perspective, our work lies within the newly proposed field of interactive narrative analytics [41].

2. Data

To analyse socio-environmental conflict narratives at the national level, we constructed a corpus of digital news articles that captures the public visibility of these disputes in the Chilean media ecosystem. The following subsections detail the data collection process and the quality assurance measures.

2.1. Composition

The data collection was based on digital news articles published in Chile, using Google News (Chile edition) as the primary source. To this end, a web scraping system was developed specifically to index national news outlets, configured with Google's standard URL parameters for territorial and linguistic filtering (gl=CL; hl=es-419) to ensure the retrieval of content produced and circulated within the Chilean media space. The search process relied on a broad and systematic set of keyword combinations aimed at capturing socio-environmental conflicts related to the energy transition, including references to types of projects, specific territories, and terms associated with controversies, complaints, and contentious actions. In total, the system operated with 651 Boolean search combinations, resulting from two successive scraping phases that expanded and refined the coverage of the corpus.

The scraping process led to an initial collection of 24,051 news articles. This dataset was subsequently subjected to a filtering and classification process based on automated systems, complemented by manual quality control checks. Inclusion criteria were restricted to news articles that reported disputes, controversies, or collective actions related to socio-environmental impacts, the use or management of natural resources, or energy and extractive infrastructures, and that displayed some form of public opposition, conflict, or dispute between actors. Exclusion criteria comprised duplicated items, purely descriptive or informational content lacking any reference to conflict, corporate announcements, institutional press releases without evidence of controversy, and articles not situated within the Chilean context. As a result of this process, 20,783 records were excluded, yielding a corpus of 3,268 validated news articles relevant to the topic of interest, that is, containing coverage of socio-environmental conflicts. An additional filtering step removed articles with missing text content or publication dates, resulting in a final corpus of 1,996 articles that constitute the empirical basis of the analysis. Thus, after applying the filter according to the required features (text content and publication date), 61% of the articles previously identified as relevant could be used.

The final dataset comprises 30 structured variables, including information on media source, publication date, territorial location, type of conflict, actors involved, and characteristics of contentious action. The systematic documentation of both included and excluded content, together with the explicit recording of exclusion criteria, ensures methodological transparency and traceability. This corpus construction process[1] enables the analysis of socio-environmental conflict as a narrative phenomenon at the national scale, grounded in explicit selection criteria and rigorous control over inclusion and exclusion decisions.

2.2. Integrity

The filtering process described in Section 2.1 positively affected the integrity and quality of the data. The filtered dataset contains substantially fewer missing values than the initial scraped dataset. After filtering, the share of missing values in the data columns referring to spatial localisation (*region*, *province*, *communes*) decreased from 61% to $< 1\%$, 66% to 15%, and 66% to 13%, respectively, allowing precise spatial analysis. With respect to the temporal dimension, the filtering resulted in only 0.2% missing date values, compared with 6.4% in the initial scraped dataset. Although the filtering improved the data quality and allows for spatio-temporal modelling, it cannot solve inherent source problems originating from the outlets. The newspapers "La Izquierda Diario Chile", "Ex-Ante", and "latribuna.cl" show 100% missing text data and therefore indicate total scraping failure. Furthermore, several other outlets, namely "Elquiglobal", "País Circular", and "BioBioChile", are missing a large number of full-text articles,

[1]For more detailed information, please check `construction_of_database.txt` at https://github.com/JonasRieger/t2s2026.

ranging from 90% to 34%. Since text data are essential for the analysis, the described filtering reduces the number of articles from these outlets to a small number of observations in the filtered dataset. As a result, for our analysis we were able to use at least 61% of the relevant articles.

3. Spatio-Temporal Topic Modeling Framework

For our analysis, we developed a generic spatio-temporal framework that is applicable to all topic modelling results, requiring only that each document include both a spatial and a temporal dimension.

3.1. Preprocessing

For preprocessing the data, we use the language-specific spaCy preprocessing pipeline (cf. Section A in the Appendix) followed by a few dataset-specific human adjustments that make the preprocessing more tailored. For our case (cf. Section 4), these adjustments include fixing a tokenisation artefact for the company Albemarle, which was lemmatised as “albemar” and “albemar él”. In addition, we exclude tokens containing digits and remove punctuation. For stopword removal, we use the internal stopword list of spaCy, supplemented with additional words and connectors[2] provided by domain experts.

3.2. Dynamic Topic Modeling

As a topic model, we propose using a dynamic topic model. Since latent Dirichlet allocation [LDA, 36] is convenient and still performs robustly compared with neural topic models [42, 43], we use RollingLDA [35] as temporal variant of LDA. Unlike other temporal adaptations of LDA[44, 45, 46, 47, 48] or neural topic models [e.g., 49], RollingLDA never utilises documents from future temporal components when modelling topics. This prevents the word and topic distributions from being influenced by information that is only available in the future, that is, a form of information leakage. Rather, it allows new data to be easily added, as the modelling approach can simply be continued, with only the new data requiring further modelling.

The topic assignments of classic LDA are based on a latent probability model using two distributions. The document-topic distribution θ_m and the word-topic distribution ϕ_k both follow a Dirichlet distribution with priors α and η, defining the mixture of topics. Utilising the underlying probability model, for a document $m = 1, \dots, M$, a topic vector $\boldsymbol{T}^{(m)} = \left(T_1^{(m)}, \dots, T_{N^{(m)}}^{(m)}\right)$ with $T_n^{(m)} \in \boldsymbol{T} = \{T_1, \dots, T_K\}$ is sampled following the generative model [50], which can be summarised as

$$W_n^{(m)} \mid T_n^{(m)}, \phi_k \sim \text{Discr}(\phi_k), \quad \phi_k \sim \text{Dir}(\eta), \quad T_n^{(m)} \mid \theta_m \sim \text{Discr}(\theta_m), \quad \theta_m \sim \text{Dir}(\alpha).$$

For a given parameter set $\{K, \alpha, \eta\}$, LDA assigns one of the K topics to each token. An estimator for the distributions $\theta_m = (\theta_{m,1}, \dots, \theta_{m,K})^T \in (0,1)^K$ (document-topic) and $\phi_k = (\phi_{k,1}, \dots, \phi_{k,V})^T \in (0,1)^V$ (word-topic) can be derived from the Gibbs iterations [50] via

$$\hat{\theta}_{m,k} = \frac{n_k^{m(\bullet)} + \alpha}{N^{(m)} + K\alpha}, \quad \hat{\phi}_{k,v} = \frac{n_k^{(\bullet v)} + \eta}{n_k^{(\bullet\bullet)} + V\eta}.$$

Since a single LDA is not trained deterministically, we stabilise this process and ensure the reliability of the results by using LDAPrototype [51], selecting the medoid, that is, the most representative LDA from a set of replicated candidate LDAs. We combine this method with the aforementioned RollingLDA [35] for dynamic modelling. The method models documents sequentially for each temporal component, using a number of previous temporal components specified by the user as a form of *seeding*. This rolling window with fading memory approach translates into the model *remembering* recent content as part of its memory while *forgetting* content further in the past that lies outside the memory window. As a result,

[2]For more detailed information, please check `preprocessing.py` at https://github.com/JonasRieger/t2s2026.

it allows gradual changes in topics over time while maintaining a consistent overall interpretation of each topic.

To interpret how topics change over time, we make use of leave-one-out word impacts [33, 34], which determine the most influential words driving gradual changes in topics. We calculate cosine similarities between the word count vectors of the selected topic k for two consecutive temporal components t_1 and t_2, first using the complete vocabulary V and second excluding the respective word ($V \setminus v$). We repeat this process for all words in the vocabulary and present those with the highest values. The score for word $v = 1, \ldots, V$ is given by

$$\cos(t_1, t_2, k, V \setminus v) - \cos(t_1, t_2, k, V).$$

3.3. Postprocessing and Visualization

Based on the metadata (time and location) of the dataset, we process the results of the dynamic topic model to enable spatio-temporal analyses. This metadata can be provided at different levels of granularity; for instance, the location can be specified as continents, countries, states, or regions, among others. We use the JavaScript library Leaflet with OpenStreetMap to visualise the spatial components. The resulting dashboard can be seen in Figure 2 in the Appendix. On the right-hand side, several settings are available, the most important of which are explained below. We distinguish between three modes: *Corpus*, *Topics*, and *Clusters*, each of which leads to corresponding analyses and visualisations in the left panel. At the same time, the map displays the document counts (*Corpus*) or the prevalence of the selected topic or cluster, restricted to the selected time period and normalised according to the selected spatial granularity. The *Clusters* option allows users to define a second level of hierarchy by consolidating sets of topics. The dashboard includes, but is not limited to, the following features:

- Top words per topic (and location) over time.
- Overall topic prevalence over time.
- Gini coefficient showing the inequality of topic prevalence across locations over time.
- Number of documents across locations and over time.
- Cluster prevalence across locations and over time.
- Interactive filtering by time period and geographic granularity.
- Option to download individual graphics and tables.

4. Results

To perform the analysis, we used the Python implementation of LDAPrototype and RollingLDA, available in the library "ttta: Tools for Temporal Text Analysis" [52]. The code for all analysis steps is available in the GitHub repository https://github.com/JonasRieger/t2s2026. For legal reasons, the data cannot be made publicly available, but may be shared upon specific request.

4.1. Study design

In our temporal modelling using RollingLDA, we use yearly time chunks, with the first ten years serving as a warm-up period modelled using LDAPrototype with 100 candidate models. To capture recent developments in environmental narratives, we set a memory window of three years for each consecutively modelled time period. In an iterative process involving human domain experts (cf. Section 4.2), we set the number of modelled topics to $K = 12$ and used a document-topic prior of $\alpha = 0.05$ for the final analysis.

We restricted the analysis to articles published from 2011 onwards, resulting in news coverage spanning from 14 February 2011 to 13 October 2025. Three different levels of granularity for the spatial component, namely regions, provinces, and municipalities, allow us to identify both broad structural patterns and local heterogeneity.

4.2. Parameter choice

In topic model analyses, automated proxy measures for human judgement are often used for model and parameter selection. Due to their negative correlation with human judgement [53], likelihood-based measures are less commonly used today, while coherence-based measures have gained considerable popularity. One advantage of these measures is that they can be calculated automatically and quickly, making it easy to evaluate and compare a large number of potential models. However, one disadvantage is that these measures are not suitable for reliably assessing traditional against neural topic models [42, 43, 54], which can lead to an incoherent selection of the best models [54] based on, e.g., normalized pointwise mutual information (NPMI) [55], the most established measure. Additionally, there are numerous variations of aggregating topic coherences into a metric for model selection, some of which differ substantially in the ranking of potential models [42]. For the reasons mentioned above, and with a view to human interaction, we follow the concept of interactive narrative analysis [41] and implement an iterative workflow with human input to select the most suitable parameters K and α.

Within this iterative process, we first computed five models with $K = 10, 15, 20, 25, 30$, and $\alpha = 1/K$. In a second step, we examined all 25 combinations of $K = 10, 11, 12, 13, 14, 15$ with $\alpha = 0.01, 0.05, 0.25, 0.5, 1$ in greater detail. After testing all models by domain experts in a non-standardised way, verifying the quality and coherence of the topics of a non-dynamic LDAPrototype model trained on the entire corpus using subjective criteria such as interpretability and plausibility, we chose a total of $K = 12$ topics and a document-topic prior of $\alpha = 0.05$ for the final analysis. The final number of topics was established based on their explanatory capacity, that is, their ability to account for both specific socio-environmental conflicts and territorially extended conflictualities linked to particular industries, whether dependent on the closure of fossil fuel sources or associated with renewable energies, as well as tensions related to the functioning of environmental institutions.

4.3. Findings

Table 1 presents the 12 distinct topics that capture the main narrative threads in Chilean socio-environmental conflict discourse, together with their top representative words and our interpretive labels, derived from domain expert analysis. Table 2 shows a higher level of abstraction in which the raw topics have been grouped into topical clusters based on further domain expert analysis.

For this purpose, we re-analysed the topics through human interpretation based on their terminological and semantic proximity, as identified by the computational process, evaluating their associative capacity to form narrative units that express specific forms of socio-environmental conflictuality. From this interpretive exercise, the topics were organised into five meaning clusters, which allow us to observe how these narratives evolve over time and across space. In particular, the temporal and territorial dynamics of these clusters, analysed over 15 years and across all 16 political-administrative regions of the country, reveal differentiated patterns of growth, stabilisation, or decline in socio-environmental conflictuality. This finding reinforces the idea that these conflicts are not static but rather respond to industrial, institutional, and territorial transformations that persist over time.

4.3.1. Spatio-Temporal Cluster Analysis

In Figure 1, we observe the dashboard organised by clusters. In this case, we selected Cluster 2 *Environmental Impacts & Energy Projects*, which groups Topics 1, 3, and 7. The map allows regional visualisation and the possibility to explore greater detail at smaller territorial scales, such as provinces and municipalities. At the same time, it allows the selection of specific years, as illustrated in Figure 3 in the Appendix, where we highlight three maps from different years.

The first map corresponds to 2012 and shows a high prevalence of news coverage in the Aysén region due to opposition to the HidroAysén project. The second map shows the situation in 2016, when the prevalence of news about conflicts shifted towards the south-central and central parts of the country. This shift is explained by conflicts linked to small run-of-river hydroelectric plants in Mapuche territory and to the Alto Maipo mega-hydroelectric project near the capital. The final map corresponds to 2025,

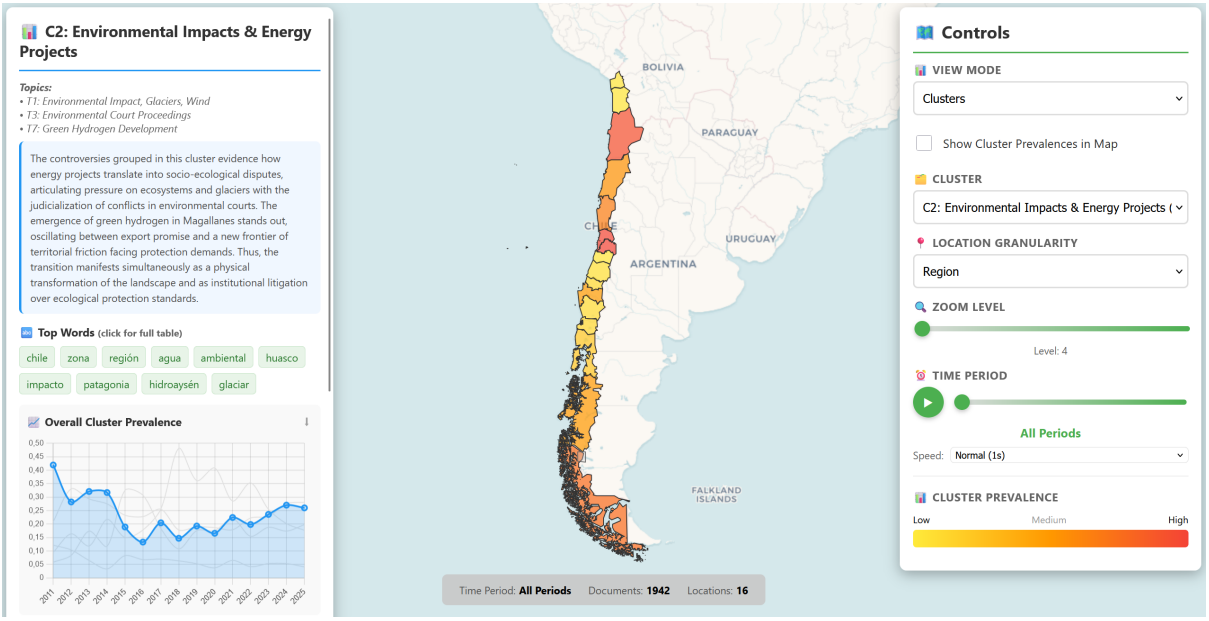

Figure 1: Screenshot of the interactive dashboard illustrating Cluster 2 *Environmental Impacts & Energy Projects* (T1, T3, T7). **Click the Figure to open the dashboard** in your browser or download it via our GitHub repository.

Table 1
The 12 topics identified by RollingLDA with top-5 representative words and interpretive labels.

Topic	Label	Top-5 Words
1	Environmental Impact, Glaciers & Wind	glaciar, impacto, zona, agua, ambiental
2	Hydroelectric & Indigenous Rights	maipo, río, mapuche, agua, comunidad
3	Environmental Court Proceedings	dominga, huasco, comité, ambiental, tribunal
4	Coal & Decarbonization	energía, carbón, central, descarbonización, eléctrico
5	Local Impacts & Communities	gente, pueblo, cielo, lumínico, comunidad
6	Industrial Pollution & Sacrifice Zones	quintero, contaminación, puchuncaví, zona, salud
7	Green Hydrogen Development	hidrógeno, hidroaysén, verde, patagonia, magallanes
8	Environmental Impact Assessment	ambiental, resolución, tribunal, evaluación, sma
9	Government & Environmental Legislation	ley, gobierno, comisión, derecho, política
10	General Word & Concepts	artículo, resumen, autor, investigación, texto
11	Lithium Mining	litio, salar, sqm, atacama, comunidad
12	Copper Mining & Water	minera, pelambres, agua, minero, cobre

Table 2
Topic clusters grouping thematically related topics for higher-level analysis. Clusters were defined through domain expert interpretation, grouping algorithmically identified topics into thematically coherent categories that reflect the structure of Chilean socio-environmental conflicts.

Cluster	Label	Topics Included
1	Water Resources & Mining Operations	T2 (Hydroelectric & Indigenous Rights) T11 (Lithium Mining) T12 (Copper Mining & Water)
2	Environmental Impacts & Energy Projects	T1 (Environmental Impact, Glaciers & Wind) T3 (Environmental Court Proceedings) T7 (Green Hydrogen Development)
3	Industrial Pollution & Decarbonization	T4 (Coal & Decarbonization) T6 (Industrial Pollution & Sacrifice Zones)
4	Environmental Institutions	T8 (Environmental Impact Assessment) T9 (Government & Environmental Legislation)
5	Local Impacts & Communities	T5 (Local Impacts & Communities)

where we observe a high prevalence across Chile, with the Antofagasta and Magallanes regions standing out due to large renewable energy projects.

In summary, the clustering of topics and their illustration through dynamic maps, both temporally and spatially, allow for the organisation of large volumes of information related to socio-environmental conflicts stemming from the energy transition in Chile, while also enabling flexible interaction in terms of scope and level of detail for different users. The validation of this system and its results by an academic team with knowledge of the social and historical context of these conflicts ensures that different social actors, stakeholders, and policymakers can interact with the information and generate new outputs.

5. Discussion

The results demonstrate how our spatio-temporal topic modelling framework overcomes the identified limitation of "methodological localism" [22], which has fragmented knowledge about these conflicts Our results show, for example, how narratives about lithium and green hydrogen cannot be understood as isolated cases, but rather as dynamic expressions of the same process: the materialisation of the extractive-energy nexus in Chile's energy transition. This contradiction challenges the sustainability of the energy transition, as decarbonisation narratives often obscure the socio-territorial frictions and local impacts inherent in the expansion of the extractive-energy nexus [16, 18].

By combining low-resource methods that require only modest computational resources, our framework maintains a low barrier to real-world application. Thanks to its flexibility, the proposed framework can be easily applied to other suitable datasets with spatial and temporal components, requiring only minor adjustments and a small number of human decisions regarding the model parameters. Additionally, the integration of LLMs allows for the automated generation of labels and descriptions for topics and clusters. Further extensions of the framework could build on this capability by using LLMs to explain broader narrative threads derived from the spatio-temporal topic modelling results.

This work contributes to the social sciences by demonstrating how computational methods can combine qualitative and quantitative approaches to study complex phenomena. By treating socio-environmental conflicts as narrative configurations circulating in the public sphere, our approach uncovers both the relevant topics and how their meanings and territorial dimensions evolve over time. For instance, the lexical evolution of Topic 2 *Hydroelectric & Indigenous Rights*, from *HidroAysén* and *dam* toward *hydrogen* and *Magallanes*, shows how public narratives are reconfigured in response to new projects and forms of resistance. This methodology allows for the identification of inflection points and unexpected convergences between apparently distant conflicts, offering the social sciences a tool to map the complexity of socio-environmental conflicts without losing sight of their specific contextual dimensions.

Our findings offer insights into how energy transition conflicts are represented in public narratives. On the one hand, some conflicts show long-term media visibility. This is the case of the sacrifice zones discourse around Quintero-Puchuncaví, which has remained present in public debate over extended periods. On the other hand, older controversies, such as HidroAysén, have gradually faded from media attention, while new sites of contestation are emerging. For instance, green hydrogen has become a prominent narrative theme since 2022, particularly in media coverage of the Magallanes region. This suggests that Chile's energy transition is not only reshaping the energy matrix but also reconfiguring the geography of socio-environmental conflicts. The appearance of green hydrogen discourse in Magallanes, following patterns observed in earlier hydropower debates in Aysén, indicates that renewable energy projects may not be exempt from the territorial contestation that characterised fossil fuel and hydropower developments. Whether this parallel extends beyond media framing, at the level of local mobilisation or policy responses, remains an open question for future research.

6. Conclusion & Limitations

This paper presents a spatio-temporal topic modelling framework for analysing socio-environmental conflict narratives in news corpora. Applied to 1,996 Chilean news articles spanning 2011–2025, our RollingLDA-based approach, with humans in the loop, identifies 12 topics that provide insight into the underlying narrative themes structuring public discourse on energy transition conflicts. The temporal analysis reveals how conflict narratives evolve, for instance from the HidroAysén controversy to green hydrogen debates, while the spatial dimension maps these narratives onto Chile's diverse territorial landscape.

We used Google News as our primary source, which introduces a selection bias: smaller regional media outlets are underrepresented, and the number of articles increases over time, partly due to the growing availability of digital news rather than solely to the intensity of the conflicts. Furthermore, our data capture public discourse as shaped by the media, but not necessarily the perspectives of local communities. In future work, we intend to compare the usefulness of the data presented here with high-quality data from a selection of Chilean newspapers[3], accessed directly through the official distributor of the digital archives of these newspapers. In the corresponding analysis, we aim to leverage the advantages of both data types: the diversity of sources in the scraped dataset and the completeness of the directly retrieved dataset.

While we were able to utilise postprocessing to integrate spatial factors into our analysis, the results could be improved by directly incorporating spatial factors into the modelling process of the topic model itself. In future work, we plan to develop further methodological improvements, including a metadata-aware topic model that co-models both temporal and non-temporal (e.g. spatial) factors. This would allow the model to allocate parameters in order to deliberately identify differences in spatial topic content and prevalence.

We also plan to examine in future work the extent to which our proposed topic modelling framework can be complemented with LLM components or replaced in certain parts. To this end, we intend to focus on distilled LLMs and neural topic models, comparing their quality through experiments with human coders as well as automated measures [42, 43, 54, cf.]. Moreover, we aim to develop further postprocessing and presentation tools to illustrate even more clearly the evolution of narratives.

Acknowledgments

This work was partially funded by the German Federal Ministry for Economic Affairs and Climate Action (Bundesministerium für Wirtschaft und Klimaschutz, BMWK) as part of the joint project *Social (in)justice in the energy transition - from the digital debate to the living world* (project no. 03EI5267B). It was also partially funded by the National Agency for Research and Development of the Chilean Government (Agencia Nacional de Investigación y Desarrollo, ANID) through the project Anillos ATE240017 *Sociotechnologies of Lithium*, the project FONDECYT Regular N°1231353 *Energy Transition and Critical Minerals*, and the project FONDEF ID25I10072 *Narrative Panopticon: Intelligent Platform For Mapping And Monitoring Information Narratives From Multi-Source Data Streams*.

Declaration on Generative AI

During the preparation of this work, we used Generative AI tools in a limited and clearly defined manner. ChatGPT was used exclusively to support grammar and spelling checks and to improve linguistic clarity. In addition, GitHub Copilot was used within the methodological workflow to assist in the automated classification and review of news articles during the data filtering process. Claude Code was used to assist in the programming of the dashboard. All analytical decisions, methodological design, interpretation of results, and the final wording of the manuscript were carried out and fully reviewed by the authors. We take full responsibility for the content of this publication.

[3]Currently, it is planned to include the newspapers 24 Horas, Diario Financiero, El Mercurio, La Nación, La Tercera, Pulso.

References

[1] F. Campos-Medina, I. Ojeda-Pereira, The invisible risks of extractive infrastructure networks and mining: Assemblages and forgotten landscapes in Chile, Visual Anthropology Review (2025) e70014. doi:10.1111/var.70014.

[2] E. Gudynas, et al., Diez tesis urgentes sobre el nuevo extractivismo, Extractivismo, política y sociedad 187 (2009) 187–225.

[3] E. Mantovani, M. Svampa, En las fronteras del cambio de época. escenarios de una nueva fase del extractivismo en América Latina, in: K. Gabbert, M. Lang (Eds.), ¿Cómo se sostiene la vida en América Latina? Feminismos y re-existencias en tiempos de oscuridad, Abya-Yala, 2019, pp. 169–217.

[4] M. Svampa, Las fronteras del neoextractivismo en América Latina. Conflictos socioambientales, giro ecoterritorial y nuevas dependencias, CALAS and Bielefeld University Press, Guadalajara, Mexico, 2019. URL: http://calas.lat/sites/default/files/svampa_neoextractivismo.pdf, online book.

[5] H. De Solminihac, L. E. Gonzales, R. Cerda, Copper mining productivity: Lessons from chile, Journal of Policy Modeling 40 (2018) 182–193. doi:10.1016/j.jpolmod.2017.09.001.

[6] J. I. Guzmán, A. Karpunin, C. Araya, P. Fúndez, M. Bocchetto, R. Camacho, D. Desormeaux, J. Galaz, I. Garcés, W. Kracht, G. Lagos, I. Marshall, V. Pérez, J. Silva, I. Toro, A. Vial, A. Wood, Chile: On the road to global sustainable mining, Resources Policy 83 (2023) 103686. doi:10.1016/j.resourpol.2023.103686.

[7] B. Bringel, M. Svampa, Del "consenso de los commodities" al "consenso de la descarbonización", Nueva Sociedad 306 (2023) 51–70. URL: https://nuso.org/articulo/306-del-consenso-de-los-commodities-al-consenso-de-la-descarbonizacion/, accessed January 29, 2026.

[8] M. N. Svampa, Consenso de los commodities y lenguajes de valoración en américa latina, Nueva Sociedad 244 (2013). URL: https://nuso.org/articulo/consenso-de-los-commodities-y-lenguajes-de-valoracion-en-america-latina/, accessed January 29, 2026.

[9] M. Allain, Conflictos y protestas socio-ambientales en chile: Reflexiones metodológicas y resultados, Revista de Sociología 34 (2019). doi:10.5354/0719-529X.2019.54271.

[10] G. Delamaza, A. Maillet, C. Martinez Neira, Socio-territorial conflicts in Chile: Configuration and politicization (2005–2014), European Review of Latin American and Caribbean Studies (2017). doi:10.18352/erlacs.10173.

[11] F. Campos-Medina, I. Ojeda-Pereira, P. Ponce Labra, Extractivismo en Chile: Tres claves para comprender la lógica de aglomeración de los proyectos industriales y proponer alternativas de resistencia, in: C. Alister, X. Cuadra, D. Julián-Vejar, B. Pantel, C. Once (Eds.), Cuestionamientos al Modelo Extractivista Neoliberal desde el Sur. Capitalismo, Territorios y Resistencias, Ariadna, Santiago, Chile, 2021, pp. 57–85. doi:10.2139/ssrn.5187287.

[12] F. Irarrázabal, Social protest at mining territories: Examining contentious politics at mining districts in chile, Resources Policy 78 (2022) 102787. doi:10.1016/j.resourpol.2022.102787.

[13] National Institute of Human Rights of Chile, Mapa de conflictos socioambientales, 2026. URL: https://mapaconflictos.indh.cl/#/, accessed January 29, 2026.

[14] E. Banks, E. Benton Hite, S. Schwartz, Green extractivisms and alternative transitions: A critical review of Latin America's renewable energy landscape, Environment and Society 16 (2025) 13–30. doi:10.3167/ares.2025.160102.

[15] F. Vergara-Perucich, M. Arias-Loyola, Locked in extraction: unveiling the path dependence of Chile's neoextractivist economy in a global transition era, Cambridge Journal of Regions, Economy and Society (2025) rsaf041. doi:10.1093/cjres/rsaf041.

[16] J. Barandiarán, The exhaustion of people and environment in an energy transition, Environment and Society 16 (2025) 31–49. doi:10.3167/ares.2025.160103.

[17] M. Akchurin, Mining, water conflicts, and climate change in Chile's Atacama Desert, Current History 124 (2025) 61–67. doi:10.1525/curh.2025.124.859.61.

[18] M. M. Ciftci, X. Lemaire, Deciphering the impacts of "green" energy transition on socio-environmental lithium conflicts: Evidence from Argentina and Chile, The Extractive Industries and Society 16 (2023) 101373. doi:10.1016/j.exis.2023.101373.
[19] S. Huneeus, S. Toro, J. P. Luna, D. Sazo, A. Cruz, D. Alcatruz, B. Castillo, C. Bertranu, J. Cisterna, Delayed and approved: A quantitative study of conflicts and the environmental impact assessments of energy projects in Chile 2012–2017, Sustainability 13 (2021) 6986. doi:10.3390/su13136986.
[20] F. Campos-Medina, E. Korneeva, R. Krayneva, I. Ojeda-Pereira, Challenge to sustainability: The territorial agglomeration of renewable energies projects in central-south Chile, in: Proceedings of the Second Conference on Sustainable Development: Industrial Future of Territories (IFT 2021), Atlantis Press, 2021, pp. 32–39. doi:10.2991/aebmr.k.211118.007.
[21] N. Bainton, D. Kemp, E. Lèbre, J. R. Owen, G. Marston, The energy-extractives nexus and the just transition, Sustainable Development 29 (2021) 624–634. doi:10.1002/sd.2163.
[22] A. Maillet, M. Allain, G. Delamaza, F. Irarrázabal, R. Rivas, C. Stamm, K. Viveros, Conflicto, territorio y extractivismo en Chile: Aportes y límites de la producción académica reciente, Revista de Geografía Norte Grande (2021). doi:10.4067/S0718-34022021000300059.
[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2023. doi:10.48550/arXiv.1706.03762, arXiv.
[24] E. Ash, G. Gauthier, P. Widmer, Relatio: Text semantics capture political and economic narratives, Political Analysis 32 (2024) 115–132. doi:10.1017/pan.2023.8.
[25] K.-R. Lange, M. Reccius, T. Schmidt, H. Müller, M. Roos, C. Jentsch, Towards extracting collective economic narratives from texts, Ruhr Economic Papers 963 (2022). doi:10.4419/96973127.
[26] F. German, B. Keith, C. North, Narrative trails: A method for coherent storyline extraction via maximum capacity path optimization, in: Proceedings of the Text2Story'25 Workshop, volume 3964 of *CEUR-WS*, 2025, pp. 15–27. URL: https://ceur-ws.org/Vol-3964/.
[27] T. Schmidt, K.-R. Lange, M. Reccius, H. Müller, M. Roos, C. Jentsch, Identifying economic narratives in large text corpora – an integrated approach using large language models, 2025. doi:10.48550/arXiv.2506.15041, arXiv.
[28] F. Gilardi, M. Alizadeh, M. Kubli, ChatGPT outperforms crowd workers for text-annotation tasks, Proceedings of the National Academy of Sciences 120 (2023). doi:10.1073/pnas.2305016120.
[29] A. Gueta, A. Feder, Z. Gekhman, A. Goldstein, R. Reichart, Can LLMs learn macroeconomic narratives from social media?, in: Findings Proceedings of the 2025 NAACL-Conference, Association for Computational Linguistics, 2025, pp. 57–78. doi:10.18653/v1/2025.findings-naacl.4.
[30] T. Schmidt, Narrating inflation: How German economic journalists explain post-covid price rises, DoCMA Working Paper 14, 2025. doi:10.17877/DE290R-25380.
[31] OpenAI, A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, ..., GPT-4o system card, 2024. doi:10.48550/arXiv.2410.21276, arXiv.
[32] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, 2019. doi:10.48550/arXiv.1910.01108, arXiv.
[33] J. Rieger, K.-R. Lange, J. Flossdorf, C. Jentsch, Dynamic change detection in topics based on rolling LDAs, in: Proceedings of the Text2Story'22 Workshop, volume 3117 of *CEUR-WS*, 2022, pp. 5–13. URL: http://ceur-ws.org/Vol-3117/.
[34] K.-R. Lange, J. Rieger, N. Benner, C. Jentsch, Zeitenwenden: Detecting changes in the German political discourse, in: Proceedings of the 2nd Workshop on Computational Linguistics for the Political and Social Sciences, 2022, pp. 47–53. URL: https://old.gscl.org/media/pages/arbeitskreise/cpss/cpss-2022/workshop-proceedings-2022/254133848-1662996909/cpss-2022-proceedings.pdf.
[35] J. Rieger, C. Jentsch, J. Rahnenführer, RollingLDA: An update algorithm of latent Dirichlet allocation to construct consistent time series from textual data, in: Findings Proceedings of the 2021 EMNLP-Conference, ACL, 2021, pp. 2337–2347. doi:10.18653/v1/2021.findings-emnlp.201.
[36] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation, Journal of Machine Learning Research 3 (2003) 993–1022. doi:10.1162/jmlr.2003.3.4-5.993.
[37] K.-R. Lange, T. Schmidt, M. Reccius, H. Müller, M. Roos, C. Jentsch, Narrative shift detection: A hybrid approach of dynamic topic models and large language models, in: Proceedings of the

Text2Story'25 Workshop, volume 3964 of *CEUR-WS*, 2025, pp. 67–80. URL: https://ceur-ws.org/Vol-3964/.

[38] M. D. Jones, M. K. McBeth, A narrative policy framework: Clear enough to be wrong?, Policy Studies Journal 38 (2010) 329–353. doi:10.1111/j.1541-0072.2010.00364.x.

[39] E. A. Shanahan, M. D. Jones, M. K. McBeth, C. M. Radaelli, The narrative policy framework, in: Theories of the policy process, Routledge, 2018, pp. 173–213. doi:10.4324/9780429494284.

[40] C. Schlaufer, J. Kuenzler, M. D. Jones, E. A. Shanahan, The narrative policy framework: a traveler's guide to policy stories, Politische Vierteljahresschrift 63 (2022) 249–273. doi:10.1007/s11615-022-00379-6.

[41] B. Keith, Interactive narrative analytics: Bridging computational narrative extraction and human sensemaking, IEEE Access 14 (2026) 2268–2284. doi:10.1109/ACCESS.2025.3650352.

[42] I. Harrando, P. Lisena, R. Troncy, Apples to apples: A systematic evaluation of topic models, in: Proceedings of the 2021 RANLP-Conference, 2021, pp. 483–493. URL: https://aclanthology.org/2021.ranlp-1.55/.

[43] A. M. Hoyle, R. Sarkar, P. Goel, P. Resnik, Are neural topic models broken?, in: Findings Proceedings of the 2022 EMNLP-Conference, ACL, 2022, pp. 5321–5344. doi:10.18653/v1/2022.findings-emnlp.390.

[44] X. Song, C.-Y. Lin, B. L. Tseng, M.-T. Sun, Modeling and predicting personal information dissemination behavior, in: Proceedings of the 11th SIGKDD-Conference, ACM, 2005, pp. 479–488. doi:10.1145/1081870.1081925.

[45] D. M. Blei, T. L. Griffiths, M. I. Jordan, J. B. Tenenbaum, Hierarchical topic models and the nested Chinese restaurant process, in: Advances in Neural Information Processing Systems, volume 16, MIT Press, 2003, pp. 17–24. URL: https://proceedings.neurips.cc/paper/2003/hash/7b41bfa5085806dfa24b8c9de0ce567f-Abstract.html.

[46] X. Wang, A. McCallum, Topics over time: A non-Markov continuous-time model of topical trends, in: Proceedings of the 12th SIGKDD-Conference, ACM, 2006, pp. 424–433. doi:10.1145/1150402.1150450.

[47] D. M. Blei, J. D. Lafferty, Dynamic topic models, in: Proceedings of the 23rd ICML-Conference, ACM, 2006, pp. 113–120. doi:10.1145/1143844.1143859.

[48] C. Wang, D. M. Blei, D. Heckerman, Continuous time dynamic topic models, in: Proceedings of the 24th UAI-Conference, AUAI Press, 2008, pp. 579–586. URL: https://dl.acm.org/doi/10.5555/3023476.3023545.

[49] A. Boutaleb, J. Picault, G. Grosjean, BERTrend: Neural topic modeling for emerging trends detection, in: Proceedings of the Workshop on the Future of Event Detection (FuturED), ACL, 2024, pp. 1–17. doi:10.18653/v1/2024.futured-1.1.

[50] T. L. Griffiths, M. Steyvers, Finding scientific topics, Proceedings of the National Academy of Sciences 101 (2004) 5228–5235. doi:10.1073/pnas.0307752101.

[51] J. Rieger, C. Jentsch, J. Rahnenführer, LDAPrototype: A model selection algorithm to improve reliability of latent Dirichlet allocation, PeerJ Computer Science 10 (2024) e2279. doi:10.7717/peerj-cs.2279.

[52] K.-R. Lange, N. Benner, L. Grönberg, A. Hachcham, I. Kolli, J. Rieger, C. Jentsch, ttta: Tools for temporal text analysis, 2025. doi:10.48550/arXiv.2503.02625, arXiv.

[53] J. Chang, J. Boyd-Graber, S. Gerrish, C. Wang, D. M. Blei, Reading tea leaves: How humans interpret topic models, in: NIPS: Advances in Neural Information Processing Systems, volume 22, Curran Associates Inc., 2009, pp. 288–296. URL: https://papers.nips.cc/paper/2009/hash/f92586a25bb3145facd64ab20fd554ff-Abstract.html.

[54] A. Hoyle, P. Goel, A. Hian-Cheong, D. Peskov, J. L. Boyd-Graber, P. Resnik, Is automated topic model evaluation broken? The incoherence of coherence, in: NeurIPS: Advances in Neural Information Processing Systems, 2021. URL: https://openreview.net/forum?id=tjdHCnPqoo.

[55] M. Röder, A. Both, A. Hinneburg, Exploring the space of topic coherence measures, in: Proceedings of the 8th WSDM Conference, ACM, 2015, pp. 399–408. doi:10.1145/2684822.2685324.

A. Preprocessing

For general text cleaning and preprocessing, we apply the Spanish text processing pipeline *es_core_news_lg* (version 3.8.0) from spaCy. We choose this pipeline over the small and medium counterparts due to its higher lemmatization accuracy (97% vs. 96%) reported in the package documentation. Within the pipeline, each text passes through a tokenization embedding layer. This layer tokenizes the text and projects the tokens into a vector space. These vector representations are then passed to a morphologizer, which adds part-of-speech tags and morphological features used for the subsequent lemmatization step. The lemmatization stage leverages the token vectors together with the morphological and positional features from the previous layer to produce lemmatized tokens. As a final step, the token vector representations are passed through a named entity recognition pipeline to assign entity tags, which are useful for the spatial dimension and subsequent token filtering. The attributes generated by the spaCy pipeline are then fed into a postprocessing pipeline. The full stopword list will be made available in our GitHub repository upon publication of this paper.

B. Dashboard Screenshots

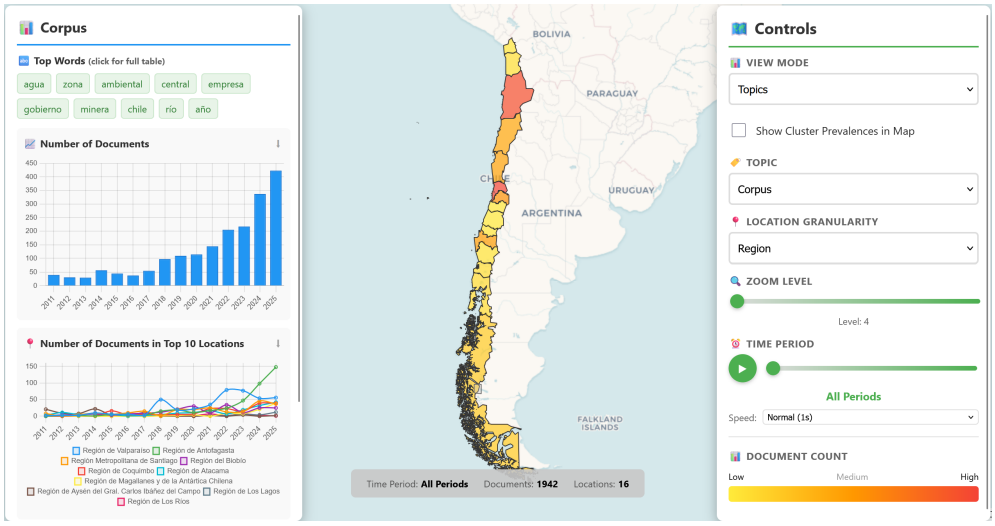

Figure 2: Screenshot of the interactive spatio-temporal dashboard showing topic distribution across Chilean regions. **Click the Figure to open the dashboard** in your browser or download it via our GitHub repository.

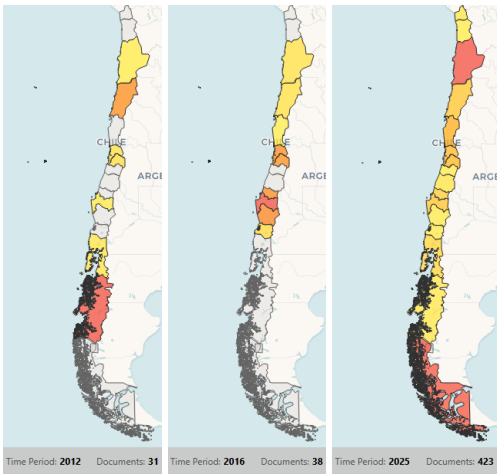

Figure 3: Dashboard of regional distribution of Cluster 2 *Environmental Impacts & Energy Projects* in 2012, 2016, and 2025 (from left to right).