Taís Maria Nunes Carvalho[1], Francisco De Assis De Souza Filho[2], and Mariana Madruga de Brito[3]

[1]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Universität Leipzig
[2]Hydraulics and Environmental Engineering Department, Federal University of Ceara
[3]Helmholtz Center for Environmental Research

December 21, 2023

Unveiling Water Allocation Dynamics: A Text Analysis of 25 Years of Stakeholder Meetings

Taís Maria Nunes Carvalho[1,2], Francisco de Assis de Souza Filho[3] and Mariana Madruga de Brito[2]

[1] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Universität Leipzig, Leipzig, Germany
[2] Helmholtz Center for Environmental Research, Leipzig, Germany
[3] Hydraulics and Environmental Engineering Department, Federal University of Ceara, Fortaleza, Brazil

E-mail: tais.carvalho@informatik.uni-leipzig.de

Abstract

Managing water resources in regions with high climate variability and frequent extreme weather events poses challenges for policymakers. To facilitate water allocation in these cases, participatory and collaborative decision-making approaches have become common. However, the evaluation of these approaches is hindered by the lack of structured methods and data to understand them. To address this knowledge gap, we propose a novel methodology that leverages text data to identify key topics, conflicts, and influential actors that shape water allocation dynamics. Our methodology is tested using records of 1020 water basin committee meetings held between 1997 and 2021 across twelve basin committees in Ceará, Brazil – a region known for its extensive history of droughts that have impacted water governance. To uncover key water management issues discussed during these meetings, we employed a three-step topic modeling framework: (1) sentence embedding, (2) dimensionality reduction, and (3) sentence clustering. Furthermore, we used entity recognition, dependency parsing, and network graphs to identify powerful actors influencing these meetings and, ultimately, the decisions taken. Our findings revealed stakeholders' heightened concern for urban water supply over agricultural demand during droughts. We found that "reservoir operation" was the most recurring topic, especially in basins where the strategic reservoirs are located. Discussions related to "climate information" became significantly more important over time, which indicates that water allocation decisions are increasingly based on the seasonal forecast and data on oceanic indices provided by the meteorology agency. Despite the presence of local users in the committees, governmental representatives dominated the discussions and were central in all river basins. In conclusion, our proposed approach harnesses existing text data to uncover spatiotemporal patterns related to participatory water allocation. This study opens new avenues for investigating water governance using text-based analysis.

Keywords: water allocation, text mining, participatory management

1. Introduction

Managing water resources in scenarios where their availability is uncertain and their uses are conflicting requires a coordinated, integrated response from stakeholders and decision-makers [1]. Water management is a complex and political task, so its social context cannot be disengaged [2,3]. Acknowledging this need, frameworks such as Integrated Water Resources Management (IWRM) [4] and Adaptive Management [5,6] have been developed. In the IWRM framework [7], the focus is on integrating management across multiple scales and resources (i.e. water and land) while attending to the needs of multiple users. Adaptive management, on the other hand, incorporates uncertainty assessment by promoting a flexible, continuous learning approach [8,9].

In recent decades, collaborative and participatory approaches have been widely applied to promote IWRM and adaptive management. While these methodologies have gained significant traction in the United States and Europe [16], experiences in the Global South are relatively limited [14,15], with notable examples in São Paulo and Ceará, Brazil [17,18]. The appeal of these approaches lies in their effectiveness when implemented within a participatory framework. Specifically, such frameworks have proven to be effective in managing conflict [10], reaching agreed targets [11], integrating local knowledge [12], and improving water planning [13].

The advantage of participatory water management approaches is rooted in their transparency, thus fostering trust and promoting social learning [19]. Furthermore, as they are done in a joint process with relevant actors, they can also incorporate the perspectives of vulnerable and marginalized groups – a dimension often overlooked in model-driven approaches. Nevertheless, participatory water management also entails risks related to power imbalances [20] and the prevalence of hierarchical approaches [16]. Additionally, the lack of systematic analysis of the process poses a risk to the effectiveness of participatory approaches [21]. Therefore, there is a pressing need to develop evaluation tools to better understand the impact of participatory water management initiatives.

The evaluation of participatory water management is, however, often hindered by the lack of structured methods and data. The analysis of water management approaches is typically done using inductive methods based on document analysis, cursory analysis (e.g., [22] and interviews (e.g. [23,24]. For instance, [25] conducted 25 interviews with water managers in Canada to evaluate the perceived collaborative water management outcomes. Minutes of meetings have also been analyzed qualitatively to provide insights into the effectiveness of water management policies [26,27]. Within this context, [28] investigated the minutes of 47 meetings dealing with water negotiations between Israel and Palestine. Despite

providing in-depth details, these methods are limited in terms of their comprehensiveness, as only a few documents or interviews can be analyzed.

In the wake of digitalization, much of the information about water management decisions has moved to virtual platforms, including databases of meeting records (e.g., European Council [29] and UN meetings) and legislative and policy documents (e.g., the FAOLEX database: http://faolex.fao.org/faolex/index.htm). This digital landscape provides vast amounts of text on an unprecedented scale, breadth, and depth. At the same time, advancements in the fields of machine learning (ML) and natural language processing (NLP) have opened previously unforeseen opportunities for using the unprecedented abundance of texts to understand water management practices.

Here, we seek to exploit these new developments systematically by proposing an automated approach to investigate participatory water management practices based on NLP, and social network analysis. More specifically, we consider the minutes of 1017 water basin committee meetings held in Ceará, northeast Brazil, to provide insights into the participatory IWRM activities conducted in the past 20 years. The goal was threefold: (1) to identify the water management topics that concern stakeholders in the decision-making process, (2) to map their spatiotemporal dynamics and trajectories, and (3) to identify influential actors involved in water allocation dynamics, their linkages (or disconnections) and how they are related to the water management topics. To the best of our knowledge, this approach is the first to perform an automated text analysis of water committee meeting records as well as an automated assessment of stakeholder networks based on text data.

2. Methods

2.1 Case study area

To illustrate how text data can be used to assess participatory water management, we considered the case of Ceará, northeast Brazil. This state, predominantly located (>90%) in the Brazilian semi-arid region, is frequently affected by severe and long-lasting meteorological and hydrological droughts [30,31]. The state is divided into twelve hydrographic basins (Figure 1), with extensive water infrastructure, including 155 monitored artificial reservoirs with a storage capacity of 18.67 billion m³.

Water availability in Ceará is not only seriously affected by frequent and recurrent drought events but is also subject to a high temporal and spatial climate variability – one of the highest in the world [32]. To address these challenges, water governance in Ceará has been regulated since the early 90s through national laws and state legislation. Notably, the establishment of river basin committees and negotiated water allocation strategies have emerged as pivotal mechanisms for fostering public participation. The first committee was established in 1992, whereas the others were implemented in

subsequent years, according to the need for urgent collective decision-making on water allocation (Figure 1a).

The committees meet regularly to discuss water management issues. After the rainy season, they also hold annual water allocation meetings, which work as a negotiation forum in single or multiple reservoir water systems, allowing for participatory

reservoir operation. Currently, the river basin committees are composed of representatives of governmental and non-governmental institutions, with the following distribution and percentage of participation: local users (30%); civil society (30%); municipal public power (20%), and state/federal public power (20%) [33].
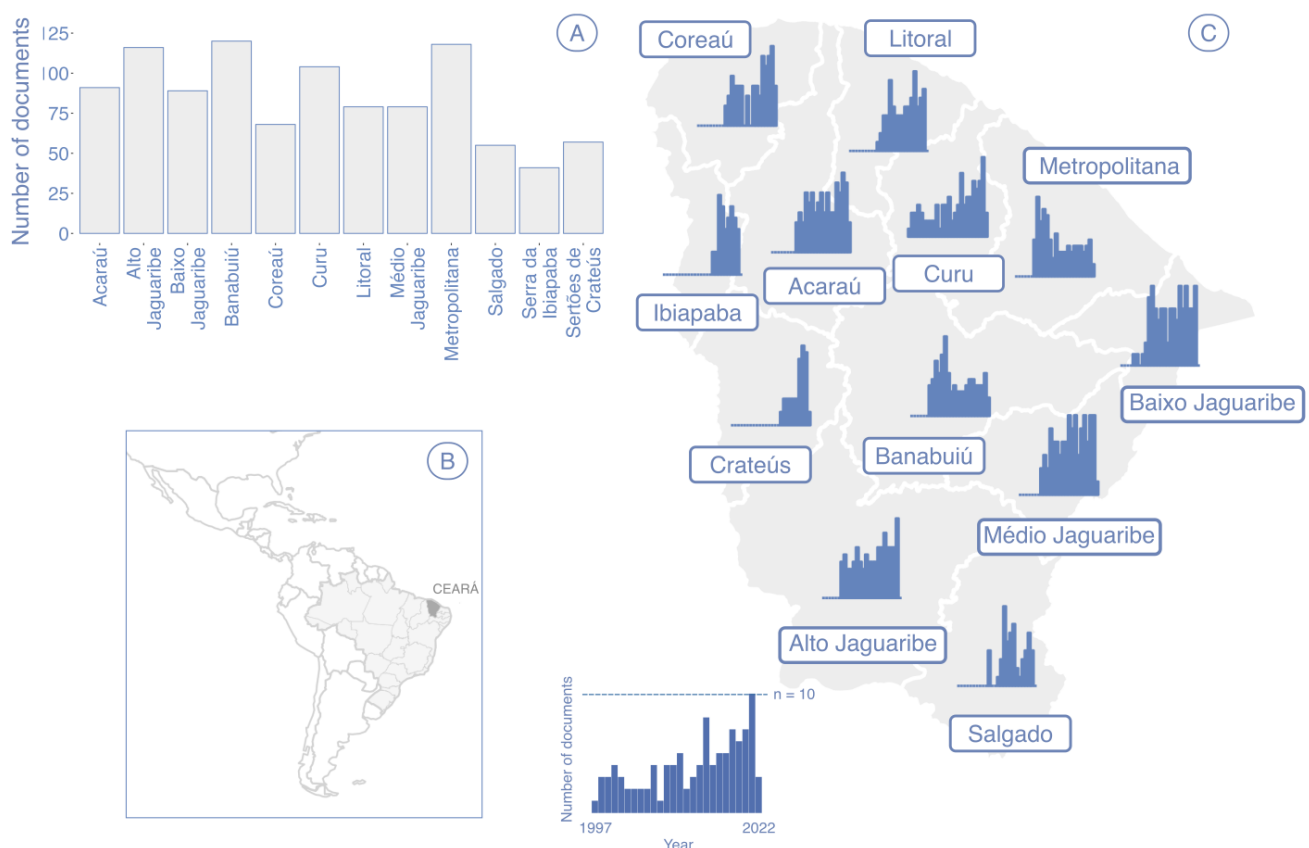

Figure 1. Location of the study area and distribution of the meeting documents in time and space. (A) Total number of documents per basin. Since the committees were created at different periods, there is an uneven frequency of meetings across them (B) Ceará (dark gray) location in northeast Brazil (C) Water basins of Ceará and histograms with the yearly number of documents available in each basin.

2.2 Text data collection and pre-processing

Minutes from committee meetings held between 1997 and 2022 in Ceará were scrapped from the water basins' websites (https://portal.cogerh.com.br/comites-de-bacias-

hidrograficas). The minutes are publicly available and are uploaded after being approved by the committee members. A total of 1020 documents were obtained and converted into plain text. They were then classified according to (1) the

corresponding water basin committee (Figure 1b) and (2) the date the meeting was held.

We pre-processed the text data by removing stop words, numbers, and special characters. Furthermore, to avoid introducing noise in the identification of key topics discussed, we also removed proper nouns (e.g., names of persons or locations) and structural elements of the minutes. For example, most of the minutes start with a list of participants' names and an outline of the meeting agenda. To automatically remove this information, we used regex to detect in which part of the text the minute starts (e.g., "the meeting started with" or "the meeting was opened by"). Municipality names and people's names detected with the Named Entity Recognition (NER) tool from SpaCy [37] were also removed from the texts. Finally, we split the text into sentences using the sentence tokenizer module from the Python package nltk [38]. We further removed sentences composed of less than five words, as we considered these to be uninformative. To validate this assumption, we randomly selected a sample of ten sentences with less than five words for each committee and read them. As an outcome, we obtained a clean corpus of 49,111 sentences identified by the committee name and the corresponding meeting date.

2.3 Topic modelling

To detect the main topics discussed in the water basin meetings, we created a topic modeling approach which (1) transfers the sentences into high-dimensional vectors, i.e., represents sentences as vectors (sentence embedding), (2) reduces the dimensionality of these embeddings, and (3) cluster the resulting dimensionality reductions. We opted to work at the sentence level to better capture the context and the overall message of the discussions. An unsupervised deep-learning framework based on the sentence-level BERT (sBERT) algorithm [39] was used to perform sentence embedding. Unlike previous state-of-the-art text topic models such as word2vec [40], sBERT allows word representations to be enriched with contextual information.

In the second step, we reduced the dimensionality of these embeddings to decrease the computational cost of clustering them while keeping their characteristics, i.e. sentences with similar contexts would have similar vector representations. Dimensionality reduction was performed with Uniform Manifold Approximation and Projection (UMAP) [41], which is commonly applied for text embeddings [42]. We also tested Principal Components Analysis (PCA) and t-SNE for this task, but the topics obtained with these methods were not as informative and representative. Possible explanations for this are that while PCA performs a linear transformation and preserves global structure in data [43], UMAP can capture non-linear relationships and preserve local and global structures in data [42]. Furthermore, while t-SNE can deal with nonlinear data [44], it is less effective in maintaining global structure when using random

initialization [45]. Hence, as maintaining local and global structures is relevant in the case of sentence embeddings, we opted to use UMAP.

In the final step, we clustered the reduced-embedded sentences using the mini-batch k-means model – a faster implementation of k-means that updates cluster centers with mini-batches. Compared to other clustering methods, such as HDBSCAN and hierarchical clustering, k-means produced more coherent and well-distributed topics. We used batches of size 100 and a maximum of 10 iterations over the dataset. As a result, we obtained 40 clusters corresponding to different topics discussed in the water committee meetings.

2.4 Evaluation of the topic modeling clusters

The quality of the clusters obtained with the topic modeling was evaluated quantitatively and qualitatively. For the quantitative evaluation of how sensitive the clusters are to parameter choices, we used the Silhouette coefficient [46]. This coefficient measures how similar each sentence is to its cluster compared to other clusters. Hence, it quantifies how well each sample was classified. To obtain an overall measure of cluster quality, we calculated the average silhouette coefficient for all clusters.

For hyperparameter tuning, we used a grid search approach. We varied the number of neighbors in UMAP (5 to 20, varying in windows of 5) and the number of clusters in k-means (30 to 80, varying in windows of 10) and calculated the average silhouette coefficient for all parameter combinations. The analysis revealed little variation in the clustering quality (the silhouette coefficient varied between 0.25 and 0.32), and we chose the parameter combination that provided the best topic assessment following a qualitative analysis (i.e., clusters representing coherent and informative topics).

2.5 Labeling of the topic modeling clusters

To label the topics and reduce the subjectivity, which is inherent in this task [47], we followed a hierarchical clustering approach, combining both quantitative and qualitative analyses. First, we inspected the 40 topics generated by the mini-batch k-means model, considering the ten most frequently occurring words in each of them. We further examined the ten sentences closest to the cluster's centroids. Throughout the labeling process, certain topics were found to be irrelevant to our focus and thus were removed from the analysis. For example, some topics only listed meeting participants or were mainly composed of unimportant information, such as metadata or the approval of previous meeting minutes. Topics that embedded multiple distinct topics were further partitioned into two or more clusters by rerunning k-means (see Table S2 for a full list of labeled topics and the divisive approach). To evaluate the newly created clusters, we used the cosine similarity to measure the degree of similarity among the clustered sentences. Furthermore, in some cases, the topics identified by our model were qualitatively similar enough that they were given identical labels and merged. As a result of this

analysis, a total of 35 topics were selected for spatiotemporal analyses.

2.6 Social network analysis

In addition to investigating the key topics discussed, we also examined influential actors, water management instruments, and how they are related (or not) to each other. To assess participation and analyze the social networks that pertain to the water allocation process, we used NER to extract the actors mentioned in each document in the corpus that did not undergo proper noun removal (see Section 2.2). Here, we define actors as entities of people with one common goal, i.e., organizations, companies, associations, government bodies, and public, private, and third-sector entities. Besides, we also considered management instruments, such as water permits and river basin plans. While these do not constitute actors per se, we assume them as nonhuman entities that can relate to other actors – and which often hold account for water management [48].

After automatically extracting actors' names, we mapped them into 18 groups (Table 1) using a regular expression rule-based search [49]. Details on the criteria used to classify the actors into organizations are described in the supplementary material (Table S3). This analysis did not include people's names mentioned in the meetings, as accurately attributing them to their respective actor group is challenging given that the minutes

typically provide only first names (e.g., 'José' could either refer to 'José Medeiros' or 'José Xavier').

After identifying the stakeholders and constructing their co-occurrence network, we computed different centrality measures (namely, degree centrality, betweenness, and closeness).

2.6 Statistical analyses

To assess spatial differences in topic frequency across water basins, we applied the Kruskal-Wallis test [50]. This nonparametric test compares groups of independent observations of a variable to check for differences between them. In our case, the groups were the topics identified in the meeting minutes, and the observations were the frequency of topic mentions over time.

To investigate how each actor is related to the six topics related to agriculture, water quality, and climate, we used a Linear Mixed-Effects regression model. In this model, the response variable was the topic frequency, and the frequency of actors' mentions in the meetings was treated as the independent variable. For this analysis, we regarded the meeting dates as a unit of time. Additionally, we considered the 12 water basin committees as a grouping variable, treating the data from each committee as independent. This allowed us to explore the connections between topics and actors while considering the unique characteristics of each committee.

Table 1. Classification criteria of the actors identified in the document.

Actor class	Entities
State Water Company	Water Resources Management Company (COGERH)
Water basin committee	Water basin committee, allocation commission, collegiate, committee chairwoman, committee chairman
State power	Civil defense, state secretariats, water companies, state police, legislative assembly, state institutions
Federal power	Federal institutions (e.g., national water company), national army, public ministry, public prosecutor
Municipal power	City hall, mayors, city councils, firefighters, municipal secretariat, municipal water management service
Media	Radio, newspapers
Agriculture users	Agricultural associations, farmers
Industry users	Industry companies and associations
Aquaculture users	Aquaculture companies and associations, shrimp farms
Management instrument	Plans, programs, policies, water permits
Working groups	Working groups, articulation groups
Civil organizations	Labor unions, community associations, study groups, charity associations
Banks	Bank or financial institution
Educational institutions	Schools, university, federal education institutes, colleges
Consulting companies	Consulting and/or engineering companies
Religious institutions	Churches, catholic church
International organizations	International institutions (e.g., WHO), European Union
Other institutions	Water companies from other states, technical assistance companies

3. Results

3.1 Dynamics of water management topics over time and space

Using topic modeling, we mapped key discussions related to water supply, water use, agriculture, water quality, climate, and the participatory water management process in each of the basins in Ceará state to 35 topics (see Figure 2). Overall, topics related to 'participatory water management', 'water supply and use', and 'water allocation' themes dominated the discussions over the last 20 years (31%, 18%, and 16% of all sentences, respectively). Indeed, 'participatory water management' was identified as a key topic for all basins investigated (Figure 3). Conversely, topics such as 'pesticide use and cattle raising' and 'agriculture water use', which are crucial for the management of irrigation water demands – the second highest in the state, were not as widely discussed by stakeholders. Likewise, we found that few discussions were centered on environmental problems.
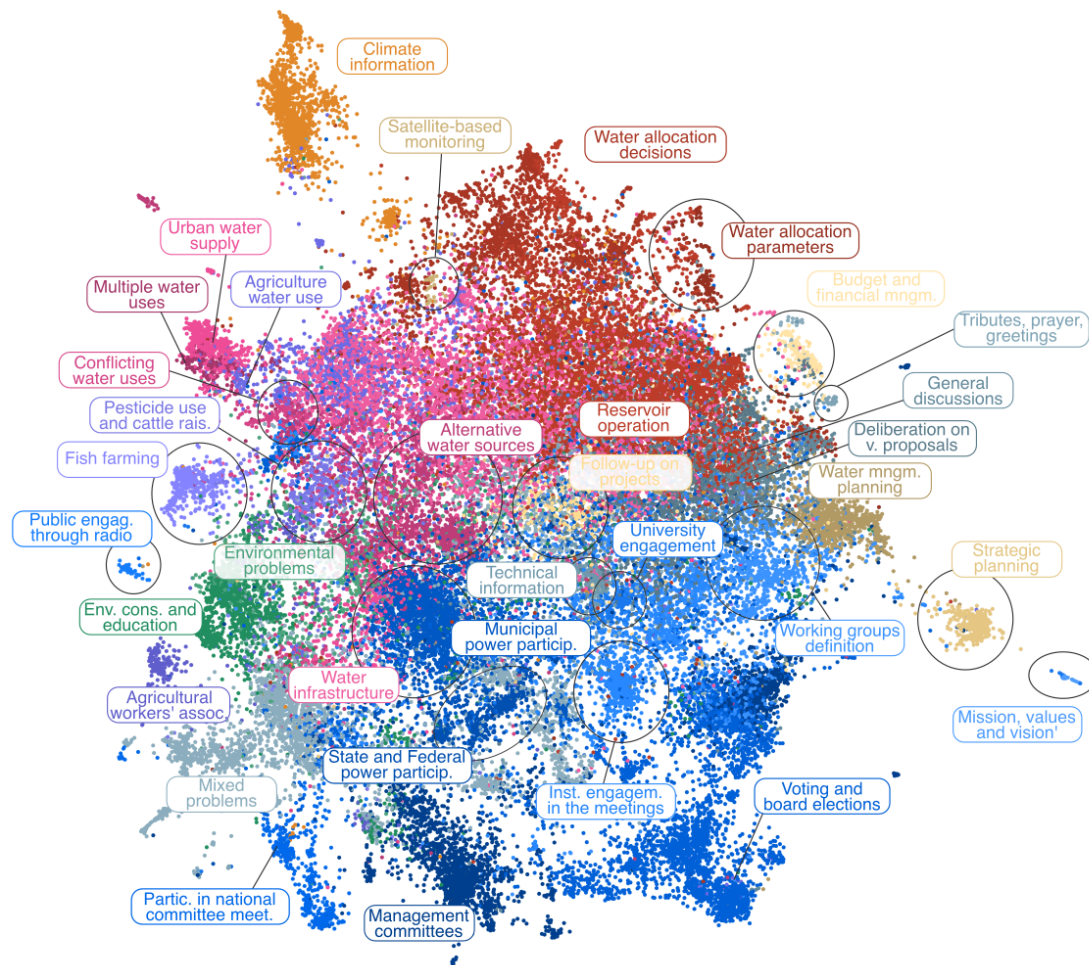
Figure 2. Visualization of the topics in the minute's corpora. Each dot corresponds to a sentence from the meeting minutes, the colors represent the associated topic. Similar sentences are close together, and dissimilar are apart.
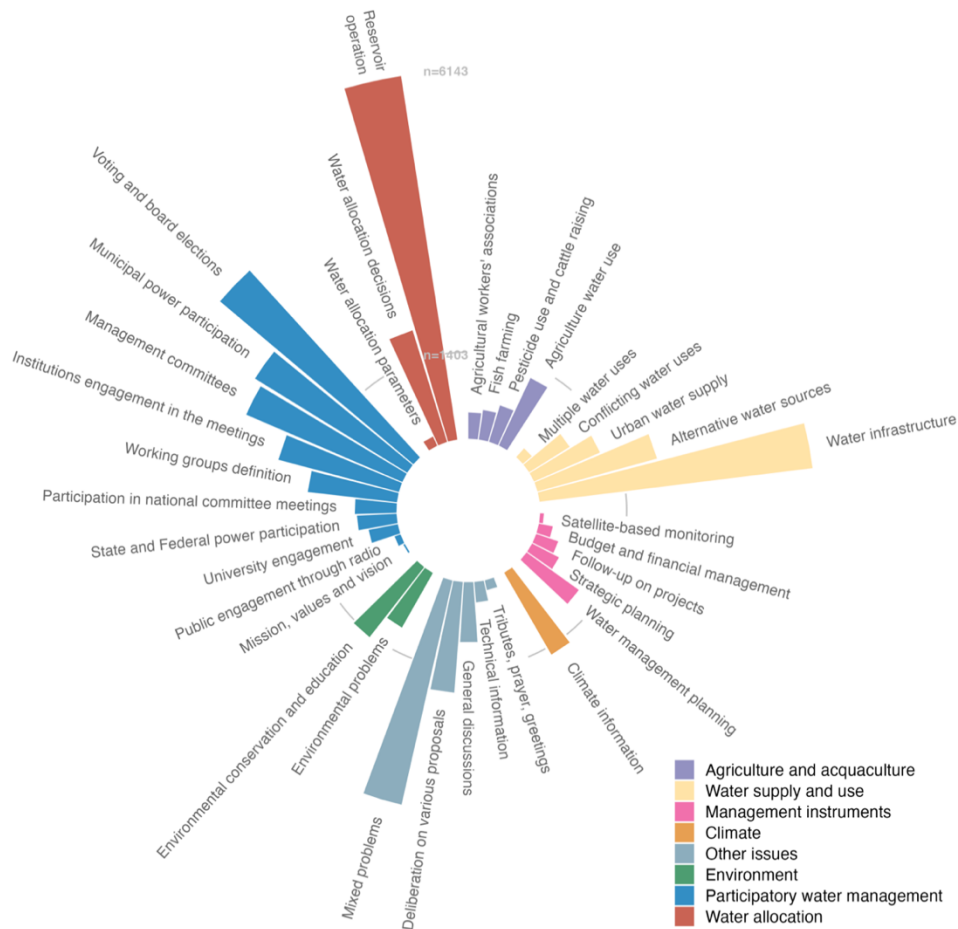
Figure 3. Total frequency of topic mentions in the corpus of meeting minutes, grouped by their main theme. The gray lines represent the maximum (n = 6,143) and the average (n = 1,403) number of mentions for each topic.

We found a significant spatial difference regarding the distribution of some topics across the basins (Figure 4). This indicates that stakeholders have different concerns depending on the basin they are acting on, reinforcing the need for adaptive management practices. For example, the topics 'water infrastructure', 'municipal power participation', and 'water allocation decisions' are distributed differently across the basins. Since 2013, a substantial infrastructure project called Cinturão das Águas has been implemented. The project will add over 1,300 km of channels, tunnels,

and siphons in Ceará so that the entire state can be supplied by surface reservoirs. The committees with a high frequency of mentions associated with 'water infrastructure' (i.e. Alto and Médio Jaguaribe) coincide with the water basins where most of the construction has been initiated or completed. Conversely, topics associated with management instruments have an almost homogeneous distribution across all the committees.

Regarding the temporal distribution, a noticeable upward trend in the frequency of topics such as

'alternative water sources' can be observed (Figure 5). This indicates a possible increased search for capacity expansion of water supply. Indeed, when qualitatively examining these sentences, we found that potable water trucks are frequently mentioned (n=102 word mentions), besides wells (n=302 mentions) and cisterns (n=20 mentions). Desalination (n=12) is also mentioned as an alternative to surface reservoirs.
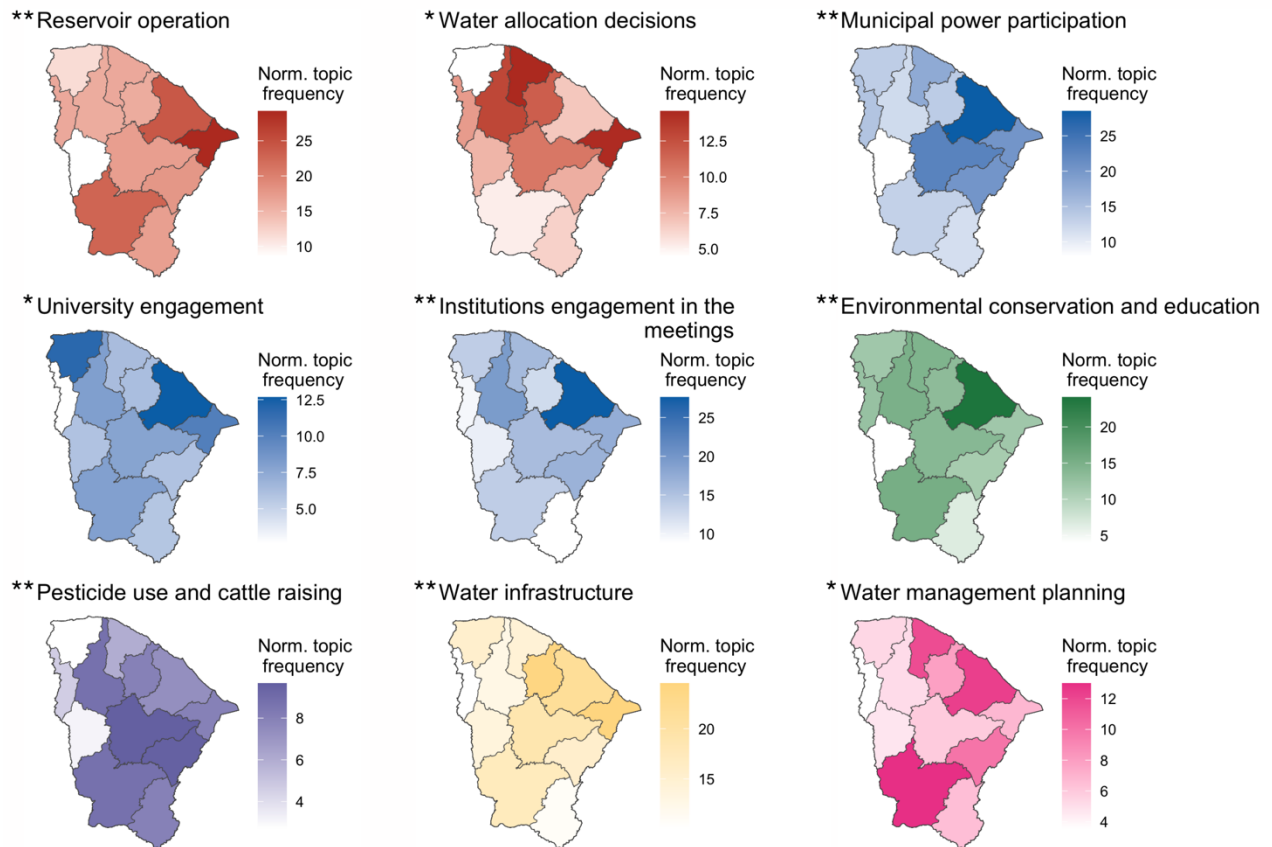

Figure 4. Normalized topic frequency per water basin. Here, we represent only the topics for which there is a statistically significantly different topic frequency distribution across water basins (* indicates $p < 0.01$ and ** indicates $p < 0.05$).

Also, discussions regarding infrastructure frequently mention wells, surface reservoirs, and channels (Table S2), indicating the significant role that decentralized water sources play in the supply system. Interestingly, water management and allocation seem to be increasingly based on 'climate information' - or at least stakeholders have discussed more about it in the past years ($p<0.01$). 'Water allocation decisions' and 'water allocation parameters' gained special attention over time ($p<0.01$), especially after the 2012-2018 drought started. This reinforces the hypothesis that the occurrence and frequency of hydrological extremes also drive the management mechanisms in Ceará. As expected, topics related to the creation and administration of the committees, including

'participation in national committee meetings' and 'management committees', became less relevant over time (p<0.01).
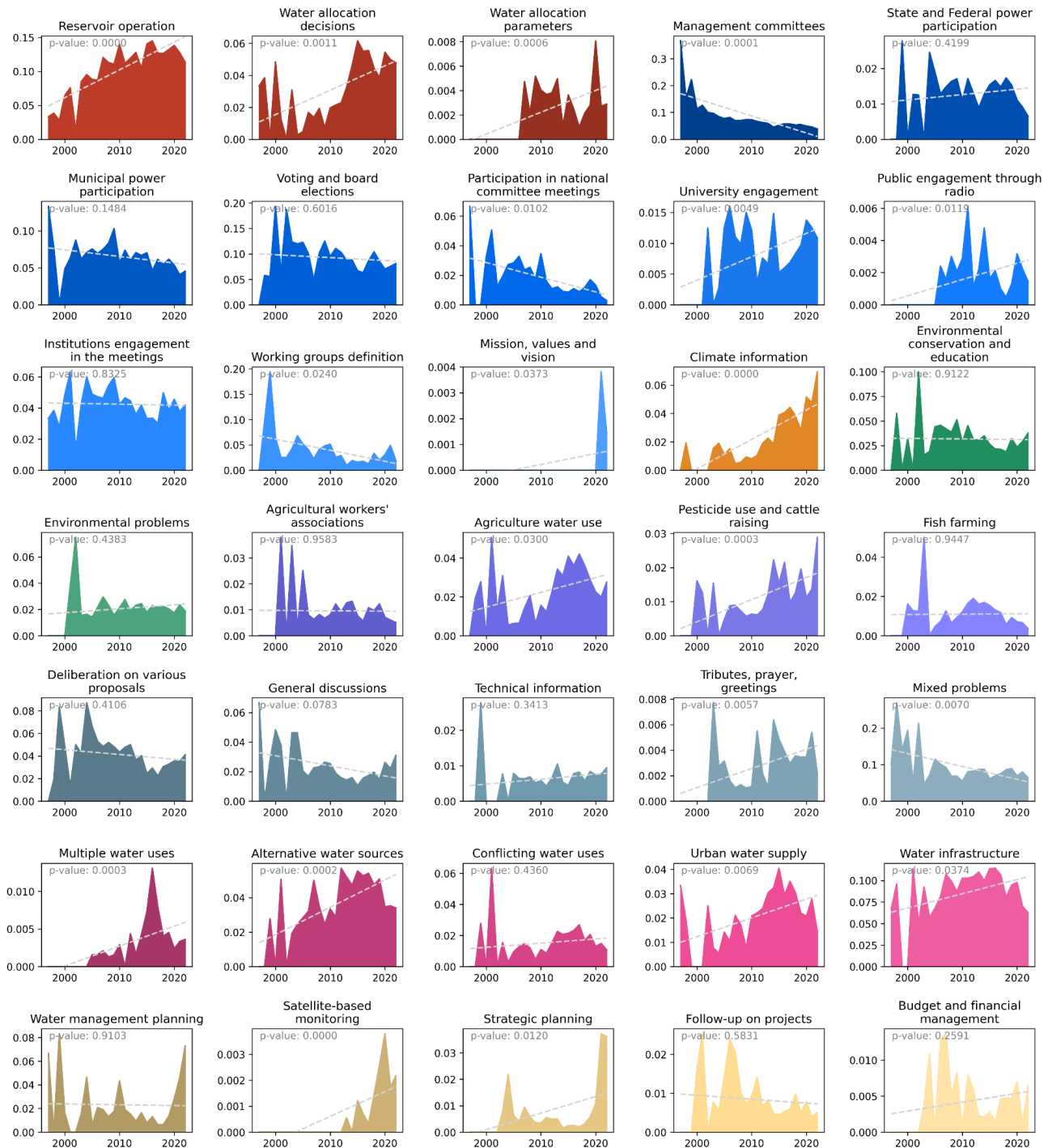

Figure 5. Normalized topic frequency over time for all committees. The area charts represent topic frequency over time. For each time series, we fitted a regression model to identify the overall trend (dashed gray line).

3.2 Actors' involvement and social network analysis

To investigate the role actors played in each of the different water allocation periods, we performed a social network analysis.
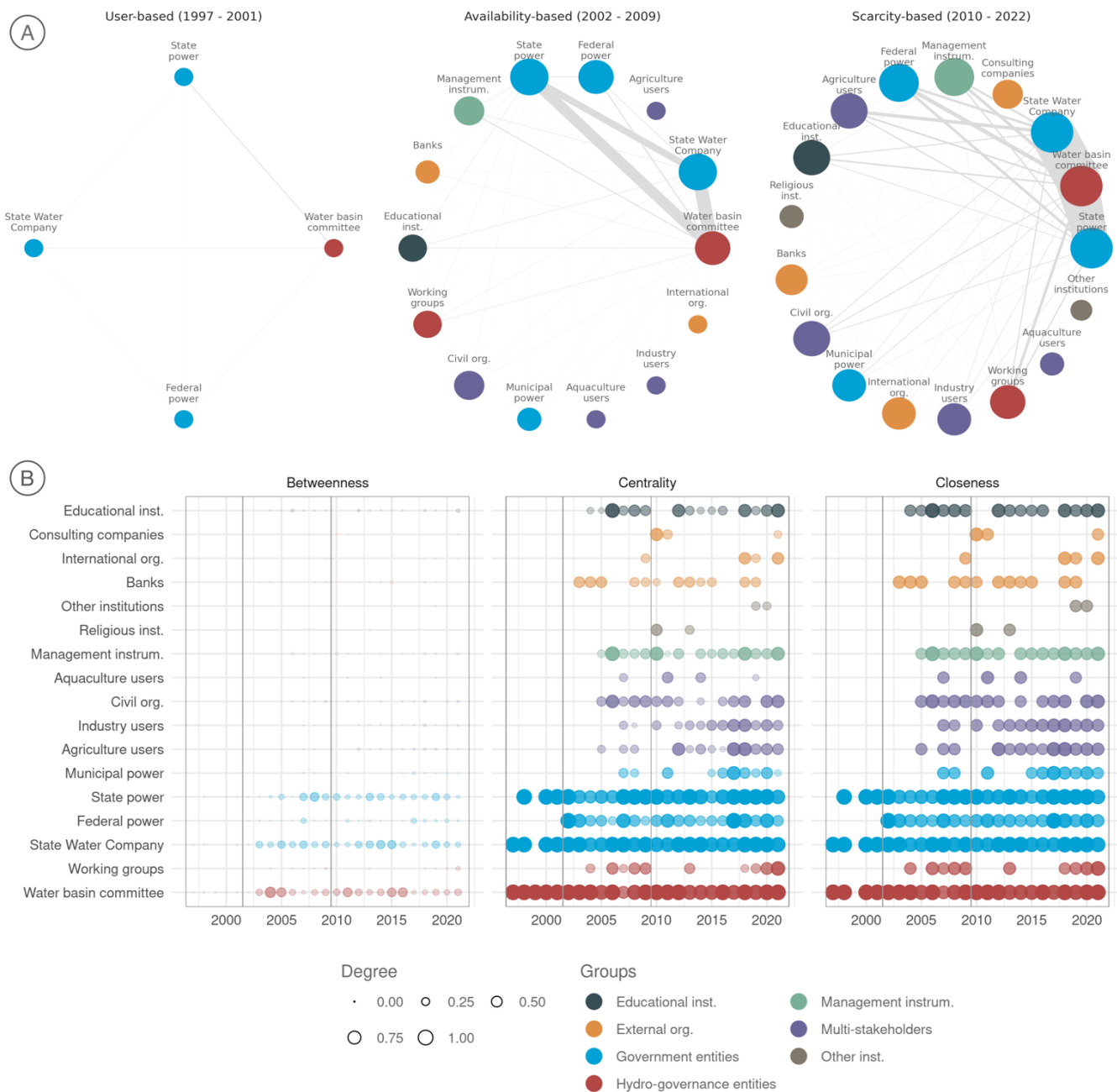
Figure 6. (A) Actors' network graphs across different allocation stages (demand-based, availability-based, and scarcity-based) The size of the nodes is proportional to its degree. (B) Network centrality measures per year for each organization group. Empty spaces indicate that the corresponding organization group was not identified in the documents in that period. Circle size and color intensity are proportional to the measured value, i.e., organizations with darker, bigger circles have a higher influence on information flow (betweenness), are more well connected to other actors (centrality), and take longer to communicate and spread information (closeness).

In the user-based water allocation period (1997-2001), state water companies, state power, and the water basin committees dominated the discourse, and institutions such as agriculture users were not represented. In fact, up until 2006, most of the actor groups were not even mentioned in the meetings. This lack of representation can be explained by the prevailing governance approach at that time; if water is easily available, fewer people tend to be interested in contributing to its management. Alternatively, this can also be explained by the limited number of minutes between 1997 and 2001, as not all committees had been established during that time (0.5% of all sentences).

From 2001 to 2002, the State Water Company initiated an experimental program, that among other measures, enabled small agriculture producers to exchange water rights with large producers during the incurring drought [51]. This program opened the possibility of a water market-like governance mode. In 2005, the second version of the State Water Plan was elaborated by representatives of the State Water Company, the Water Resources Secretariat, and other water-related state agencies. According to the legislation, the plan required approval from the water basin committees – and this can explain the increase in participation and interconnectedness between actors from 2005 onwards.

During the scarcity-based stage, especially after the 2012-2018 drought, agriculture and industry users became more relevant. One interesting aspect of Ceará's water management policy is that water availability is conditioned to the negotiated water allocation, i.e., the decisions taken during the meetings can restrict water withdrawals. Consequently, when water quantity diminishes, water users are more likely to be interested in participating in the meetings.

Over the entire period, statutory entities were the actors that could most easily reach all others in the network, as depicted by the betweenness metric in Figure 6b. While the closeness measure is homogeneous for all actors, actors representing the state power and the water basin committee are the only to have a significant betweenness value. This indicates that although participation has increased in the past years, government entities still control the information flow and strongly influence decision-making during these meetings.

Previous studies in Brazil revealed that committee members perceive an unequal power distribution in water basins where techno-scientific knowledge is more strongly used in decision-making [17,52,53]. On the other hand, stakeholders appreciate expert support, indicating a successful trust-building process [18,54].

3.3 Relationship between actors and different agriculture, water quality, and climate topics

The analysis of the relationships between actors and the topics (Figure 2) showed several significant relationships (Table 2). Specifically, topics related to environmental problems showed a strong connection with actors from educational institutions

(p<0.01). This was expected as researchers affiliated with local universities and research institutes have a heightened concern about water quality in semi-arid reservoirs, as shown by several studies published in the past years [55–57].

International organizations were frequently mentioned together with topics on agriculture water use (p<0.001) and pesticides (p<0.001). These associations may be attributed to the stringent requirements imposed by the European Union, Brazil's largest market, to import agricultural products [58].

Finally, climate information seems to be a concern not only to municipal power (p<0.05) but also to consulting companies (p<0.05). This does not necessarily mean that they provide this information but that their associated tasks (e.g., water supply projects and environmental diagnosis) are often contextualized with seasonal forecasts and climate projections.

4. Discussion

The water resources system of Ceará has been historically affected by severe and frequent droughts [31], which encouraged the implementation of important management approaches and instruments, e.g., public participation in water management decisions through basin committees and the negotiated water allocation. In this study, we systematically investigate the shift from IWRM towards adaptive management practices in Ceará by analyzing the minutes of 1017 meetings.

Using an unsupervised deep-learning approach, we show that discussions on water management seem to be transitioning to a climate-informed approach, but still focused on quantity rather than quality. This is reflected by the highly frequent, uniform mentions of topics associated with infrastructure and urban supply and the reduced frequency of topics related to environmental problems (which entails water quality issues). Indeed, water quality concerns such as 'pesticide use' seem only to be discussed in times of crisis (e.g. after the increase in algae bloom episodes in 2015 [60], Figure S3). This topic, alongside 'agriculture water use' shows an increased frequency during drought periods (e.g. the 2012-2018 drought).

The increased frequency of 'climate information' in water allocation meetings can not only be associated with the development of climate services in Brazil and the effort by the local meteorology agency to produce seasonal forecasts and monitor drought in the past decades [60] but also with an increasingly level of averseness to risk adopted by water managers [61].

We found that topic distribution across water basins is not uniform – mirroring the needs of each basin. For example, 'reservoir operation' has a stronger focus in the basins where the state's strategic reservoirs are located (i.e., the largest

ones). On the other hand, 'environmental conservation and education' is prominent in the 'Metropolitana' basin, which holds the state's capital and is almost entirely urbanized [62]. This underscores the need for a flexible, collaborative water management approach.

Furthermore, we learned that although the participation of multiple actors has increased in the past decade, technical knowledge and information flow are still mainly dominated by government entities and representatives of the water basin committees (Figure 6b). In all committees, independent of the water allocation stage, state actors had higher connectedness and could communicate more easily than other actors. Previous empirical studies have indicated that this is indeed the case: multi-stakeholders can vote on the reservoir operation strategy, but the water availability and allocation scenarios are usually prepared in advance by the water management company and some committee representatives [54]. Risk tolerance and the expected future user behavior are decisions that rarely are agreed upon during the water allocation process. To improve the adaptive capacity of water systems, all organizations and actors must be able to interpret and use technical information [6,17].

It is also worth noting that organizations beyond the water sector (e.g., energy, environment, health) seem to be missing from the participative process. A cross-sectoral approach to water management is important to ensure water security and reduce the impacts of climate hazards and the uncertainties associated with the dependence of decisions made in other sectors [63]. A detachment from other sectors was also observed by the stakeholders of the water basin committees from another Brazilian state [23]. To improve collaboration across sectors and align their interests, it is necessary to increase institutional capacity so that actors from different governance levels and sectors work together to develop strategies and measures [63].

The proposed NLP-based approach allows for a low-cost (in terms of computer and human resources) and efficient assessment of unstructured data. Furthermore, it provides a strategy to extract longitudinal information on participatory processes that would only be possible if panel surveys had been performed. However, leveraging information from meeting registers has some drawbacks, such as the intrinsic subjectivity of the minute taker and the inconsistency of organizations' acronyms. However, it still provides reliable information on the meeting content, as it is usually approved by the participants upon publication.

Table 2. Results of the mixed linear regression model fitted to the topics related to agriculture, water quality, and climate. $p < 0.05$: *, $p < 0.01$: **.

	Climate information			Environm. conservation and education			Environmental problems			Agriculture water use			Pesticide use and cattle raising			Fish farming		
	Coef	Std.Er	P>|z|	Coef	Std.Er	P>|z|	Coef	Std.Er	P>|z|	Coef	Std.Er	P>|z|	Coef	Std.Er	P>|z|	Coef	Std.Er	P>|z|
Intercept	-148,85	27,77	**0,00	17,52	16,54	0,29	-16,68	13,86	0,23	-69,75	18,38	**0,00	-41,70	20,03	*0,04	11,29	6,20	0,07
Year	0,07	0,01	**0,00	-0,01	0,01	0,30	0,01	0,01	0,22	0,04	0,01	**0,00	0,02	0,01	*0,04	-0,01	0,00	*0,01
Industry users	0,09	0,08	0,24	0,02	0,06	0,81	-0,03	0,05	0,53	0,19	0,06	**0,00	0,21	0,06	**0,00	0,14	0,08	0,07
Int. Org.	0,07	0,12	0,56	0,25	0,10	*0,01	0,04	0,07	0,56	0,33	0,09	**0,00	0,44	0,10	**0,00	-	-	-
Educational inst.	-0,03	0,05	0,54	0,02	0,03	0,58	0,07	0,03	*0,01	-0,07	0,04	0,08	0,00	0,04	0,97	-0,02	0,04	0,68
State power	0,00	0,01	0,73	0,03	0,01	*0,00	0,03	0,01	**0,00	0,01	0,01	0,35	0,03	0,01	**0,00	0,01	0,01	0,48
Municipal power	-0,35	0,15	*0,02	0,05	0,08	0,52	0,04	0,07	0,58	-0,15	0,11	0,18	-0,16	0,10	0,12	0,08	0,12	0,52
Federal power	-0,01	0,03	0,71	0,04	0,03	0,11	0,06	0,02	**0,00	0,00	0,03	0,90	0,02	0,04	0,52	0,12	0,03	**0,00
Water basin committee	0,01	0,01	0,07	0,01	0,00	*0,02	-0,01	0,00	*0,03	-0,01	0,00	0,14	0,00	0,00	0,45	-0,01	0,00	0,11
Aquaculture users	-0,13	0,17	0,43	-	-	-	0,02	0,11	0,87	-0,11	0,19	0,55	0,27	0,15	0,07	0,04	0,13	0,77
Management instrument	-0,04	0,04	0,34	0,02	0,02	0,42	0,00	0,02	0,84	-0,02	0,03	0,42	-0,01	0,03	0,80	0,03	0,04	0,44
Civil org.	-0,02	0,08	0,82	0,07	0,06	0,20	0,08	0,05	0,14	0,04	0,07	0,57	0,04	0,07	0,53	0,03	0,08	0,75
Agriculture users	-0,05	0,03	0,16	0,00	0,03	0,93	-0,03	0,02	0,23	0,07	0,03	*0,01	0,06	0,03	**0,03	-0,03	0,04	0,35
State Company	0,02	0,01	*0,02	-0,01	0,01	0,22	0,01	0,01	*0,02	0,03	0,01	**0,00	-0,01	0,01	0,26	0,02	0,01	0,10
Religious Inst.	-0,40	0,45	0,37	0,06	0,26	0,82	0,33	0,22	0,14	-0,05	0,34	0,89	-	-	-	-0,17	0,29	0,56
Consulting companies	0,62	0,30	*0,04	0,01	0,09	0,87	0,03	0,11	0,78	-0,05	0,12	0,68	-	-	-	0,17	0,19	0,38
Other	0,36	0,33	0,27	-	-	-	0,16	0,28	0,58	-0,26	0,24	0,28	0,02	0,21	0,95	-0,13	0,32	0,67
Working groups	0,00	0,03	0,94	0,03	0,02	0,23	0,01	0,02	0,49	0,03	0,02	0,20	0,02	0,02	0,31	0,07	0,03	*0,05

5. Conclusion

A key element for the success of water management strategies lies in the consistent evaluation and systematic adaptation of ongoing practices. This should preferably be done systematically rather than organically to ensure that external factors such as political change will not discontinue the process. However, obtaining information on this process is a difficult task, as continuous surveys or interviews are not readily available or feasible for a long-term assessment spanning multiple years or decades. Valuable information on decision-making, task distribution, and actor engagement are often hidden in minutes of meetings, which nowadays are usually publicly available. Hence, automated analysis strategies such as the one presented here represent a way forward to evaluate participatory approaches and learn about water management strategies.

By considering the case of the water systems in Ceará, Brazil, we evaluated how power networks developed and changed with time and what the stakeholders are worried about. We learned that the management process has become more participative and diverse in Ceará, but it still fails to incorporate actors from sectors other than the water resources and that practical information still mainly originates from the state power and the water basin committee. In this context, moving to a polycentric governance approach might be a way to deal better with the complexity of the water resources system and climate and social uncertainties.

Our methodology presents an opportunity for stakeholders from the water sector to evaluate the dynamics of the participatory process and to verify which areas or actors deserve closer attention or to be better incorporated into it. Ultimately, it provides a strategy to improve our understanding regarding the implementation of water management strategies.

Data statement

All data is available in the following repository: github.com/taiscarvalho/water-meet-mining.

References

[1] Kapetas L, Kazakis N, Voudouris K and McNicholl D 2019 Water allocation and governance in multi-stakeholder environments: Insight from Axios Delta, Greece *Sci. Total Environ.* **695** 133831

[2] Anderson E P, Jackson S, Tharme R E, Douglas M, Flotemersch J E, Zwarteveen M, Lokgariwar C, Montoya M, Wali A, Tipa G T, Jardine T D, Olden J D, Cheng L, Conallin J, Cosens B, Dickens C, Garrick D, Groenfeldt D, Kabogo J, Roux D J, Ruhi A and Arthington A H 2019 Understanding rivers and their social relations: A critical step to advance environmental water management *WIREs Water* **6** e1381

[3] Walker W E, Loucks D P and Carr G 2015 Social Responses to Water Management Decisions *Environ. Process.* **2** 485–509

[4] United Nations 1992 The Dublin statement on water and sustainable development *International Conference on Water and the Environment, Dublin, Ireland*

[5] Hjorth P and Madani K 2023 Adaptive Water Management: On the Need for Using the Post-WWII

Science in Water Governance *Water Resour. Manag.* **37** 2247–70

[6] Pahl-Wostl C 2006 Transitions towards adaptive management of water facing climate and global change *Water Resour. Manag.* **21** 49–62

[7] Biswas A K 2008 Integrated Water Resources Management: Is It Working? *Int. J. Water Resour. Dev.* **24** 5–22

[8] Walters C J and Hilborn R 1978 Ecological Optimization and Adaptive Management *Annu. Rev. Ecol. Syst.* **9** 157–88

[9] Pahl-Wostl C 2008 Requirements for Adaptive Water Management *Adaptive and Integrated Water Management* ed C Pahl-Wostl, P Kabat and J Möltgen (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 1–22

[10] Bijlsma R M, Bots P W, Wolters H A and Hoekstra A Y 2011 An empirical analysis of stakeholders' influence on policy development: the role of uncertainty handling *Ecol. Soc.* **16**

[11] Vollmer D, Abell R, Bezerra M, Harrison I, Hauck S, Shaad K and Souter N 2023 A watershed moment for healthy watersheds *Nat. Sustain.* **6** 233–5

[12] Budryte P, Heldt S and Denecke M 2018 Foundations of the participatory approach in the Mekong River basin management *Sci. Total Environ.* **622–623** 349–61

[13] Graversgaard M, Jacobsen B, Kjeldsen C and Dalgaard T 2017 Stakeholder Engagement and Knowledge Co-Creation in Water Planning: Can Public Participation Increase Cost-Effectiveness? *Water* **9** 191

[14] Yasmin T, Farrelly M, Rogers B, Krause S and Lynch I 2023 A tale of two cities: evidence from the Global South on established versus emerging cities' approaches to adaptive and sustainable water governance *Ecol. Soc.* **28** art38

[15] Brown J 2011 Assuming too much? Participatory water resource governance in South Africa *Geogr. J.* **177** 171–85

[16] Lukat E, Pahl-Wostl C and Lenschow A 2022 Deficits in implementing integrated water resources management in South Africa: The role of institutional interplay *Environ. Sci. Policy* **136** 304–13

[17] Lemos M C, Puga B P, Formiga-Johnsson R M and Seigerman C K 2020 Building on adaptive capacity to extreme events in Brazil: water reform, participation, and climate information across four river basins *Reg. Environ. Change* **20** 53

[18] de Assis Souza Filho F, de Carvalho Studart T M, Filho J D P, Martins E S P R, Ayrimoraes S R, Pessoa C A P, Rolim L Z R, de Araujo Junior L M, da Silva S M O, Carvalho T M N and de Aquino S H S 2023 Integrated proactive drought management in hydrosystems and cities: building a nine-step participatory planning methodology *Nat. Hazards* **115** 2179–204

[19] Lumosi C K, Pahl-Wostl C and Scholz G 2020 Evaluating trust and shared group identities in emergent social learning processes in the Zambezi river basin *Humanit. Soc. Sci. Commun.* **7** 172

[20] Rana M M P and Piracha A 2020 Supplying water to the urban poor: Government's roles and challenges of participatory water governance *Cities* **106** 102881

[21] Morinville C and Harris L M 2014 Participation, politics, and panaceas: exploring the possibilities and limits of participatory urban water governance in Accra, Ghana *Ecol. Soc.* **19**

[22] Vos J, Boelens R, Venot J-P and Kuper M 2020 Rooted water collectives: Towards an analytical framework *Ecol. Econ.* **173** 106651

[23] Barbosa M C, Mushtaq S and Alam K 2017 Integrated water resources management: Are river basin committees in Brazil enabling effective stakeholder interaction? *Environ. Sci. Policy* **76** 1–11

[24] Ben-Daoud M, Mahrad B E, Elhassnaoui I, Moumen A, Sayad A, ELbouhadioui M, Moroşanu G A, Mezouary L E, Essahlaoui A and Eljaafari S 2021 Integrated water resources management: An indicator framework for water management system assessment in the R'Dom Sub-basin, Morocco *Environ. Chall.* **3** 100062

[25] Ferreyra C and Beard P 2007 Participatory evaluation of collaborative and integrated water

management: Insights from the field *J. Environ. Plan. Manag.* **50** 271–96

[26] Andrew McGuire S and Ehlinger T 2022 Restoration as social-ecological transformation: Emergence in the Pike River Watershed *J. Gt. Lakes Res.* **48** 1505–16

[27] Kunz N C, Moran C J and Kastelle T 2013 Implementing an integrated approach to water management by matching problem complexity with management responses: a case study of a mine site water committee *J. Clean. Prod.* **52** 362–73

[28] Fischhendler I and Katz D 2013 The impact of uncertainties on cooperation over transboundary water: The case of Israeli–Palestinian negotiations *Geoforum* **50** 200–10

[29] Hayes-Renshaw F, Van Aken W and Wallace H 2006 When and Why the EU Council of Ministers Votes Explicitly* *JCMS J. Common Mark. Stud.* **44** 161–94

[30] Brito S S B, Cunha A P M A, Cunningham C C, Alvalá R C, Marengo J A and Carvalho M A 2018 Frequency, duration and severity of drought in the Semiarid Northeast Brazil region *Int. J. Climatol.* **38** 517–29

[31] Marengo J A, Torres R R and Alves L M 2017 Drought in Northeast Brazil—past, present, and future *Theor. Appl. Climatol.* **129** 1189–200

[32] Gutiérrez A P A, Engle N L, De Nys E, Molejón C and Martins E S 2014 Drought preparedness in Brazil *Weather Clim. Extrem.* **3** 95–106

[33] Ceará 2010 Lei Nº 14.844

[34] Campos J N B and de Carvalho Studart T M 2008 Drought and water policies in Northeast Brazil: backgrounds and rationale *Water Policy* **10** 425–38

[35] Studart T M de C, Campos J N B, Souza Filho F A de, Pinheiro M I T and Barros L S 2021 Turbulent waters in Northeast Brazil: A typology of water governance-related conflicts *Environ. Sci. Policy* **126** 99–110

[36] Pontes Filho J D, Souza Filho F de A, Martins E S P R and Studart T M de C 2020 Copula-Based Multivariate Frequency Analysis of the 2012–2018 Drought in Northeast Brazil *Water* **12** 834

[37] Ines Montani, Matthew Honnibal, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem and Henning Peters 2023 explosion/spaCy: v3.7.2: Fixes for APIs and requirements

[38] Bird S, Klein E and Loper E 2009 *Natural language processing with Python: analyzing text with the natural language toolkit* (O'Reilly Media, Inc.)

[39] Reimers N and Gurevych I 2019 Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

[40] Mikolov T, Chen K, Corrado G and Dean J 2013 Efficient Estimation of Word Representations in Vector Space

[41] McInnes L, Healy J and Melville J 2018 UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction

[42] Allaoui M, Kherfi M L and Cheriet A 2020 Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study *Image and Signal Processing* Lecture Notes in Computer Science vol 12119, ed A El Moataz, D Mammass, A Mansouri and F Nouboud (Cham: Springer International Publishing) pp 317–25

[43] Hotelling H 1933 Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24** 417

[44] Hinton G E and Roweis S 2002 Stochastic neighbor embedding *Adv. Neural Inf. Process. Syst.* **15**

[45] Kobak D and Linderman G C 2021 Initialization is critical for preserving global data structure in both t-SNE and UMAP *Nat. Biotechnol.* **39** 156–7

[46] Rousseeuw P J 1987 Silhouettes: A graphical aid to the interpretation and validation of cluster analysis *J. Comput. Appl. Math.* **20** 53–65

[47] Hsu A and Rauber R 2021 Diverse climate actors show limited coordination in a large-scale text analysis of strategy documents *Commun. Earth Environ.* **2** 30

[48] Latour B 1996 On actor-network theory: A few clarifications *Soz. Welt* **47** 369–81

[49] Mansouri A, Affendey L S and Mamat A 2008 Named entity recognition approaches *Int. J. Comput. Sci. Netw. Secur.* **8** 339–44

[50] Moore D S, McCabe G P and Craig B A 2012 *Introduction to the practice of statistics* (New York: Freeman [u.a.])

[51] Formiga-Johnsson R M and Kemper K 2005 Institutional and Policy Analysis of River Basin Management: The Jaguaribe River Basin, Ceara, Brazil *World Bank Policy Res. Work. Pap. No 3649*

[52] Abers R N, Formiga-Johnsson R M, Frank B, Keck M E and Lemos M C 2009 Inclusão, deliberação e controle: três dimensões dedemocracia nos comitês e consórcios de bacias hidrográficas no Brasil *Ambiente Soc.* **12** 115–32

[53] Engle N L, Johns O R, Lemos M C and Nelson D R 2011 Integrated and Adaptive Management of Water Resources *Ecol. Soc.* **16**

[54] Formiga-Johnsson R M, Kumler L and Lemos M C 2007 The politics of bulk water pricing in Brazil: lessons from the Paraíba do Sul basin *Water Policy* **9** 87–104

[55] Mesquita J B de F, Neto I E L, Raabe A and Araújo J C de 2020 The influence of hydroclimatic conditions and water quality on evaporation rates of a tropical lake *J. Hydrol.* **590** 125456

[56] Raulino J B S, Silveira C S and Neto I E L 2021 Assessment of climate change impacts on hydrology and water quality of large semi-arid reservoirs in Brazil *Hydrol. Sci. J.* **66** 1321–36

[57] Lopes F B, Andrade E M D, Meireles A C M, Becker H and Batista A A 2014 Assessment of the water quality in a large reservoir in semiarid region of Brazil *Rev. Bras. Eng. Agríc. E Ambient.* **18** 437–45

[58] Esteves R C, Do Amaral Vendramini A L and Accioly F 2021 A qualitative meta-synthesis study of the convergence between organic crop regulations in the United States, Brazil, and Europe *Trends Food Sci. Technol.* **107** 343–57

[59] Laurent A, Godoy R F B, Elias M C, Franco G, Carril A F, Trevisan E and Flombaum P 2023 Extremes in water quality parameters modulated by weather and climate conditions in Ceara State, Brazil *Ecohydrol. Hydrobiol.* S1642359323000617

[60] Escada P, Coelho C A S, Taddei R, Dessai S, Cavalcanti I F A, Donato R, Kayano M T, Martins E S P R, Miguel J C H, Monteiro M and Moscati M C L 2021 Climate services in Brazil: Past, present, and future perspectives *Clim. Serv.* **24** 100276

[61] Kirchhoff C J, Lemos M C and Engle N L 2013 What influences climate information use in water management? The role of boundary organizations and governance regimes in Brazil and the U.S. *Environ. Sci. Policy* **26** 6–18

[62] Ceará 2009 Caderno regional das bacias Metropolitanas

[63] Pahl-Wostl C and Knieper C 2023 Pathways towards improved water governance: The role of polycentric governance systems and vertical and horizontal coordination *Environ. Sci. Policy* **144** 151–61

We built an interactive web visualization tool to make it easier to verify the content of the topics identified with our methodology for topic modeling: https://taiscarvalho.github.io/water-meet-mining/.

Figure S1. Absolute topic frequency per water basin (i.e., how often a topic was detected in the entire corpus). Topics for which there is a statistically significantly different topic frequency distribution across water basins are marked with a * (p < 0.01) or ** (p < 0.05).

Figure S2. Topic frequency across months. Topic frequency was normalized by the topic with higher occurrence every month.
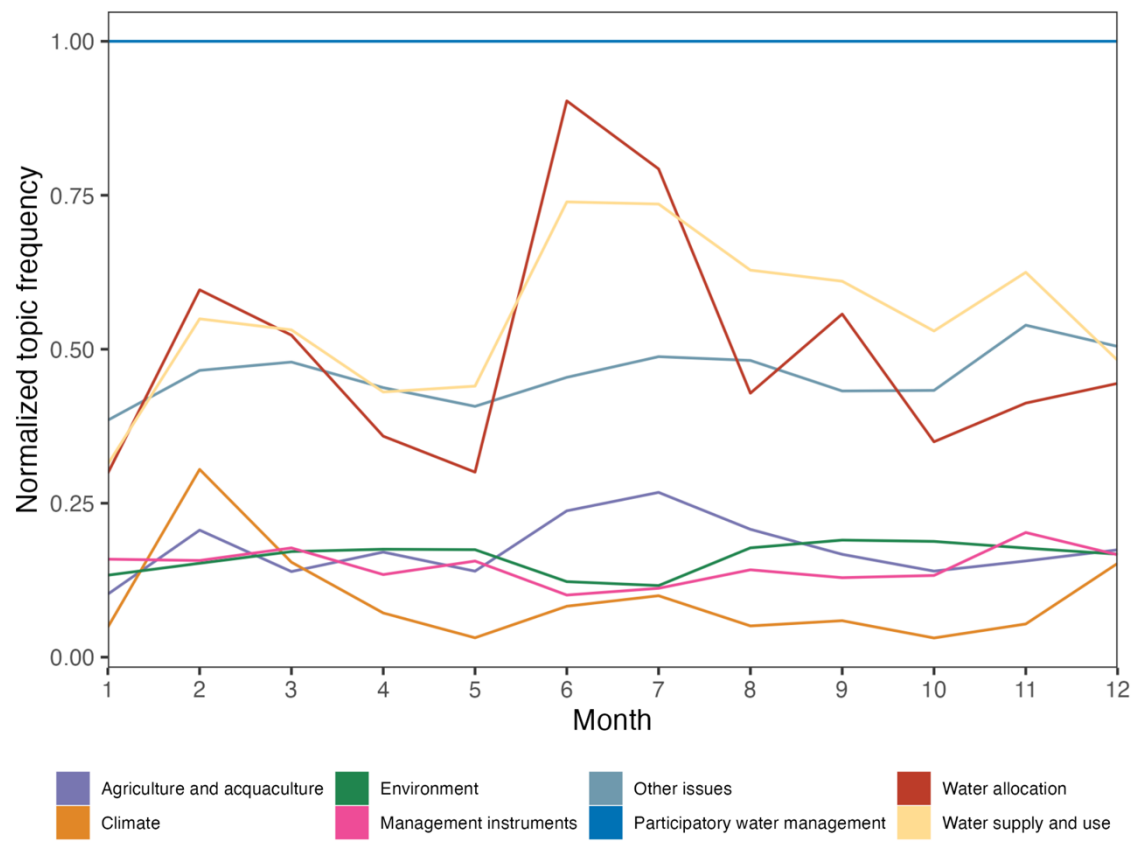
Figure S3. Temporal distribution of topic frequency, normalized by the maximum annual frequency of each topic. Color intensity is proportional to topic frequency.
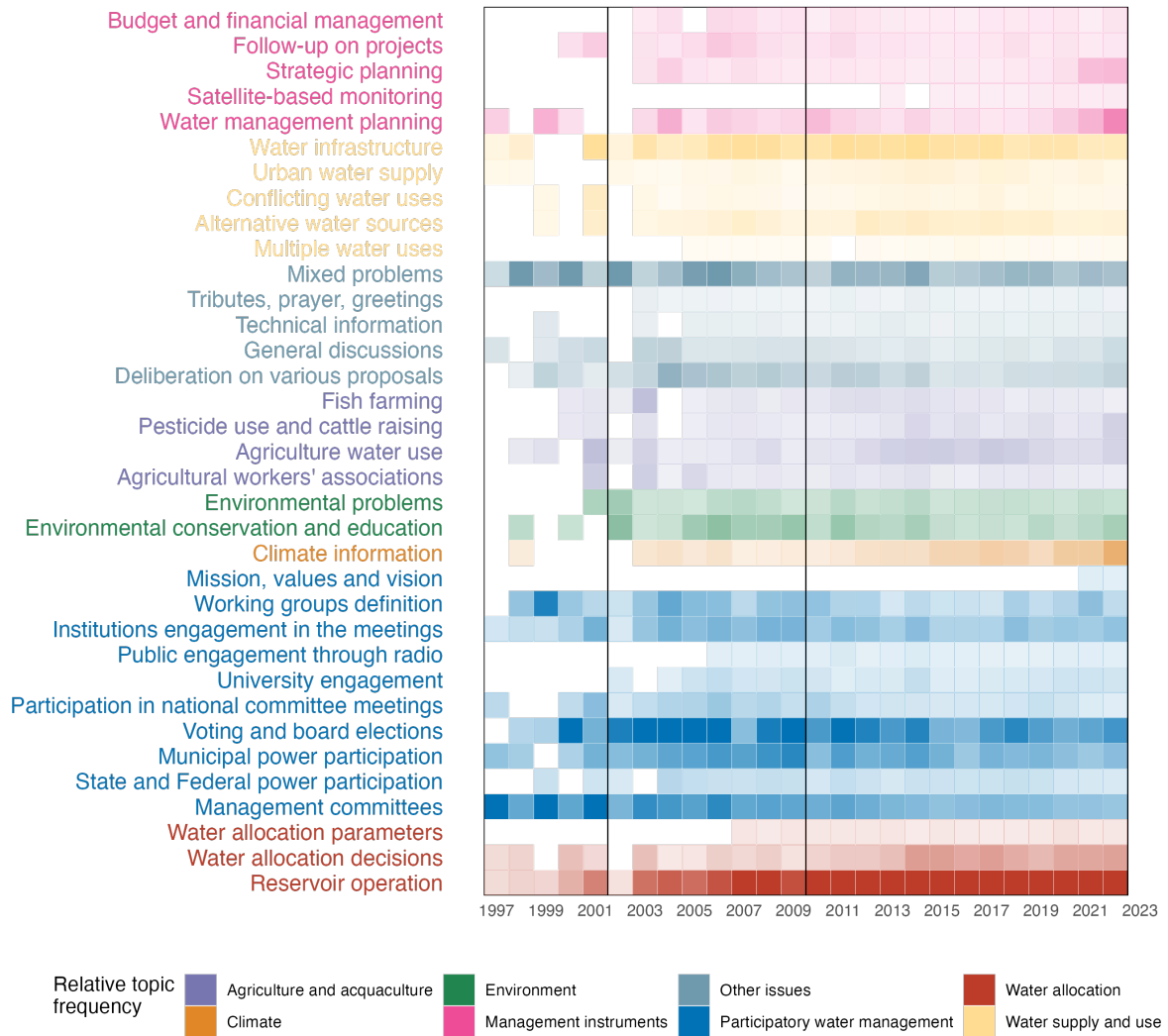
Table S1. Description and representative sentences of the topics detected in the corpus of meeting minutes.

Topic	Description
Reservoir operation	Decisions on reservoir operation.
Water allocation decisions	Presentation of allocation and water demand scenarios and assessment of reservoir levels.
Water allocation parameters	Definition of water allocation parameters (i.e., maximum and minimum reservoir outflows).
Management committees	Reservoirs monitored by the water management company have their own management committee. This topic is associated with their activities and specific tasks.
State and Federal power participation	Discussion of issues that shall be solved by federal or state organizations, collaboration with the Ministry of National Integration.
Municipal power participation	Municipal power participation into the meetings and its responsibilities; announcements from municipal organizations.
Voting and board elections	Decisions regarding the election of committee members, changes in the election process, nomination for awards or commendations.
Participation in national committee meetings	Planning and assessment of the participation of the committee representatives into the annual national meeting of water basin committees.
University engagement	Involvement of researchers associated with the university in the committee activities, water quality assessment, drought planning and water management.
Public engagement through radio	Information dissemination through radio, be it about water allocation decisions or environmental awareness campaigns.
Institutions engagement in the meetings	Discussions on how to ensure the engagement of institutions that should be participating into the meetings, announcements of institutions interested in coming to the meetings.
Working groups definition	Working groups are constituted to discuss and propose solutions for specific challenges, such as water allocation, project studies, communication, and internal regulations. Discussions on project proposals and decision-making with the groups.
Mission, values, and vision	Definition of the mission, values, and vision of the water basin committees.
Climate information	Seasonal forecast and climate information presented in the meetings (usually by the meteorological organization), including precipitation forecast and ocean conditions.
Environmental conservation and education	Discussions regarding programs for environmental conservation. Environmental education campaigns and initiatives in schools.

Environmental problems	Discussions regarding inappropriate waste and wastewater disposal, fires, technical information on water pollution. Description of episodes of fish mortality in reservoirs (causes, monitoring, possible impacts) and other problems associated with fish farming.
Agricultural workers' associations	Support, concerns, and interests of farmers unions.
Agriculture water use	Issues associated with irrigation; adaptation measures taken by farmers. Debates on organic production (agriculture products, advantages of the system, etc.)
Pesticide use and cattle raising	Aspects associated with farming, cultivation, use of pesticides and sustainable practices.
Fish farming	Implications of limited water supply to fish farming, unauthorized fish farming.
Deliberation on various proposals	Transitory sentences on the context of collective deliberation of projects, presentations, and documentation.
General discussions	Discussions about generic aspects related to the meeting, including debates, presentations, opinions, and voting procedures.
Technical information	Presentation or request of technical information on infrastructure building, water quality and water supply.
Tributes, prayer, greetings	Greetings, appreciation of birthdays and commendations, prayers.
Mixed problems	Discussions associated with diverse problems that could not be classified into separate clusters because are not recurrent (e.g., conservation units, reservoir bathymetry campaigns, etc.). Sentences that are too long to be classified into one single topic.
Multiple water uses	Discussions on reservoirs that serve multiple uses, how to address them, how to allocate water between multiple users.
Alternative water sources	Usage of groundwater and sources such as water trucks to supply water.
Conflicting water uses	Problems and discussions regarding reservoirs with multiple water uses.
Urban water supply	Water demand projections, alternative urban supply sources, strategies to ensure urban supply in appropriate quantity and quality.
Water infrastructure	Discussions on the construction of pipelines and wells, the installation of hydrometers, replacement of pumps, etc.
Water management planning	Goal setting, planning of activities associated with water resources management and drought response.
Satellite-based monitoring	Discussions about the implementation of a satellite-based monitoring for water use.
Strategic planning	Development of strategic water resources and/or drought planning.

Follow-up on projects	Presentation of goals and results of new/ongoing projects and generic sentences regarding regulatory/administrative issues associated with them.
Budget and financial management	Decisions on how to allocate the committee budget and the usage of money collected from granted water permits.

1 **Table S2.** Top ten most frequent words mentioned per topic, excluding stopwords. The number between parentheses refers to the count of mentions
2 each word had in the corresponding topic.

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
Urban water supply	0,63	humano (338), abastecimento (327), demanda (89), volume (69), irrigação (67), consumo (57), liberação (40), média (38), perenização (38), usos (38)	Urban water supply	0,63	-
Reservoir operation	0,58	perguntou (200), alocação (150), após (147), açudes (138), pediu (119), demanda (119), momento (115), fala (115), sugeriu (106), pode (96)	Reservoir operation	0,58	-
Water infrastructure	0,55	problema (141), construção (105), região (97), problemas (89), poços (89), cagece (67), usuários (63), estrutura (60), abastecimento (59), dnocs (59)	Water infrastructure	0,55	-
Climate information	0,60	chuvas (630), média (345), chuvosa (304), quadra (278), estado (230), funceme (228), normal (206), período (187), acima (157), abaixo (156)	Climate information	0,60	-
Voting and board elections	0,61	comissão (842), comitês (528), gestora (351), renovação (261), comissões (240), diretoria (191), processo (185), gestoras (176), resolução (132), conerh (127)	Voting and board elections	0,61	-
Agriculture water use	0,57	irrigação (190), abastecimento (138), humano (109), hídrica (85), área (80), usuários (77), segurança (73), consumo (67), demanda (66), sistema (62)	Agriculture water use	0,57	-
Fish farming	0,57	volume (284), operação (246), açudes (193), pesca (179), peixes (127), capacidade (118),	Fish farming	0,54	pesca (179), peixes (127), peixe (84), pescadores (66), piscicultura

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
		apresentou (111), alocação (105), simulado (102), final (94)			(61), colônia (57), açudes (54), associação (45), qualidade (44), mortandade (41)
			Reservoir operation	0,57	volume (270), operação (241), açudes (139), capacidade (114), apresentou (105), simulado (102), alocação (99), final (88), simulação (83), parâmetros (74)
General planning	0,59	plano (762), atividades (273), planejamento (242), ações (229), após (189), discussão (176), aprovação (149), trabalho (144), cada (132), câmara (126)	Water management planning	0,59	plano (698), atividades (254), planejamento (235), ações (199), trabalho (113), planos (98), aprovação (85), capacitação (78), apresentou (77), ação (76)
			General discussions	0,55	discussão (163), após (138), resolução (115), discussões (102), decisão (71), próxima (67), discutir (64), plano (64), aprovação (64), câmara (63)
Alternative water sources	0,61	comunidade (419), comunidades (377), poços (250), abastecimento (223), águas (185), onde (162), projeto (161), situação (151), estado (141), adutora (135)	Alternative water sources	0,61	-
Organic farming	0,59	sindicato (214), rurais (205), trabalhadores (185), associação (97), ambiental (96), agricultura (94), agricultores (89), rural (73), projeto (70), familiares (59)	Environmental conservation and education	0,55	ambiental (86), ambiente (39), ambientais (33), conservação (27), produção (26), área (24), solo (23), gestão (21), pessoas (20), orgânicos (20)

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
			Agricultural workers' associations	0,59	sindicato (212), rurais (203), trabalhadores (185), agricultura (86), agricultores (85), associação (80), familiares (59), rural (53), projeto (41), familiar (36)
			Public engagement through radio	0,57	rádio (54), rádios (44), divulgação (28), comunicação (17), spot (14), plano (12), programa (12), municípios (10), projeto (10), jornalista (9)
Conflicting water uses	0,59	irrigação (156), segurança (153), abastecimento (136), região (100), hídrica (98), barragem (95), barragens (91), águas (84), onde (83), humano (83)	Conflicting water uses	0,59	-
Deliberation on various proposals	0,57	diretoria (138), informes (130), participação (118), gestão (108), evento (103), após (92), proposta (91), sugeriu (91), trabalho (90), site (87)	Deliberation on various proposals	0,57	-
Management committees	0,65	secretaria (1013), secretário (559), executiva (417), gestão (346), assinada (212), presente (204), estado (193), diretoria (190), oliveira (182), núcleo (174)	Management committees	0,65	-
Water infrastructure	0,58	águas (314), hídrica (200), situação (200), poços (175), açudes (167), canal (155), estado (124), barragem (123), qualidade (120), apresentou (113)	Water infrastructure	0,58	-
Pesticide use and cattle raising	0,56	produção (68), agricultores (57), agricultura (56), hectares (54), empresa (54), área (51),	Pesticide use and cattle raising	0,56	-

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
		sementes (48), beckman (44), milho (39), agrotóxicos (37)			
Municipal power participation	0,59	prefeitura (397), municípios (381), municipal (257), cidade (161), projeto (161), abastecimento (110), cagece (109), comunidades (101), comunidade (88), onde (86)	Municipal power participation	0,59	-
Water allocation decisions	0,63	volume (1122), capacidade (555), média (425), reservatório (317), operação (301), apresentou (262), açudes (253), período (229), reservatórios (212), aporte (199)	Water allocation decisions	0,63	-
Negotiated water allocation	0,56	irrigação (148), açudes (144), seca (123), hídrica (92), situação (91), volume (89), águas (89), castanhão (87), média (86), seco (86)	Agriculture water use	0,56	irrigação (141), hídrica (88), águas (87), açudes (81), lima (79), campos (79), média (74), castanhão (70), riacho (59), volume (57)
			Water allocation decisions	0,52	seca (121), seco (82), açudes (61), secos (46), poços (45), situação (34), volume (29), aporte (25), período (25), apresentou (21)
			Satellite-based monitoring	0,50	satélite (42), imagens (42), fiscalização (24), drone (20), monitoramento (18), imagem (17), drones (11), áreas (8), onde (8), empresa (7)
Voting and board elections	0,58	votação (450), votos (369), eleição (285), eleitoral (233), diretoria (193), comissão (156), proposta (151), processo (149), após (140), voto (106)	Voting and board elections	0,58	-

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
Committees strengthening	0,58	comitês (786), nacional (455), comissão (374), encontro (329), brasil (132), participação (127), águas (120), hidrográficas (117), estado (113), programa (104)	Participation in national committee meetings	0,58	nacional (452), comitês (330), encontro (303), brasil (129), águas (120), hidrográficas (114), participação (84), agência (76), encob (73), realizado (72)
			Management committees	0,55	comitês (456), comissão (360), discussão (62), estado (57), plano (56), gestora (55), acompanhamento (52), comissões (46), operação (43), outros (43)
Involvement with other institutions	0,56	orçamento (109), trabalho (107), informações (87), usuários (84), visita (76), projeto (71), despesas (64), dnocs (57), apoio (56), empresa (53)	Budget and financial management	0,49	orçamento (109), despesas (63), faturamento (39), planejamento (33), arrecadação (33), apresentou (30), custo (21), projeto (20), onde (19), reais (19)
			Water infrastructure	0,56	trabalho (98), usuários (81), informações (77), visita (74), dnocs (57), apoio (55), projeto (51), empresa (49), semace (47), demanda (47)
Climate information	0,62	oceano (204), atlântico (172), pacífico (137), niña (74), temperatura (69), niño (69), chuvosa (69), condição (67), chuvas (62), condições (61)	Climate information	0,62	-
Urban water supply	0,65	humano (398), abastecimento (391), animal (131), dessedentação (116), poços (109), usos (66), captação (63), consumo (56), atender (52), atual (52)	Multiple water uses	0,61	animal (128), humano (126), dessedentação (114), abastecimento (109), usos (51), animais (40), irrigação (40),

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
					indústria (37), consumo (29), prioridade (24)
			Urban water supply	0,66	humano (271), abastecimento (250), apenas (35), demanda (33), atender (33), consumo (26), superfície (23), liberação (19), trecho (18), atendimento (18)
			Alternative water sources	0,63	poços (82), captação (50), atual (44), abastecimento (32), ideal (31), reais (27), operação (23), tubulares (21), igações (21), adutora (20)
Environmental problems	0,59	abastecimento (157), projeto (137), ambiental (130), poços (116), famílias (115), situação (107), área (104), qualidade (104), ambientais (99), ambiente (97)	Environmental problems	0,59	-
Water allocation parameters	0,53	parâmetros (147), alocação (120), açudes (84), resolução (78), definição (58), parâmetro (55), decisão (51), operação (51), mínimo (51), máximo (45)	Water allocation decisions	0,49	resolução (70), decisão (50), decisões (33), alocação (30), açudes (24), conerh (20), informações (19), tomada (16), momento (13), período (13)
			Tributes, prayer, greetings	0,45	oração (28), feliz (28), natal (28), desejou (24), cenários (23), próxima (20), momento (19), participação (17), constar (16), padre (15)

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
			Water allocation parameters	0,55	parâmetros (142), alocação (87), açudes (57), parâmetro (53), definição (48), mínimo (46), máximo (45), operação (42), definir (33), após (25)
Follow-up on multiple projects	0,56	projeto (359), técnico (152), regional (135), técnicos (125), região (125), trabalho (78), gerência (78), gerente (77), projetos (75), dados (58)	Follow-up on projects	0,53	projeto (335), projetos (66), região (27), plano (23), ações (21), construção (21), execução (16), feito (15), recuperação (15), empresa (15)
			Mixed problems	0,52	regional (119), região (91), gerente (66), gerência (60), trabalho (54), arrecadação (51), faturamento (45), transporte (43), despesas (43), cagece (37)
			Technical information	0,53	técnico (152), técnicos (122), dados (44), coordenador (27), núcleo (26), projeto (23), engenheiro (22), operação (21), tecnologias (20), trabalho (18)
Mixed problems	0,65	associação (1073), prefeitura (930), municipal (925), oliveira (808), sousa (583), alves (573), comunitária (503), lima (477), rurais (463), santos (456)	Mixed problems	0,65	associação (1073), prefeitura (927), municipal (923), oliveira (807), sousa (582), alves (572), comunitária (503), lima (477), rurais (463), santos (455)
			Strategic planning	0,63	planejamento (368), estratégico (354), plano (152), estratégicos (149), trabalho (82), objetivos (81),

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
					gestão (66), rossana (66), conceitos (66), visão (64)
Working groups definition	0,58	grupo (464), trabalho (296), plano (229), regimento (207), comissão (146), diretoria (144), interno (142), comunicação (141), capacitação (137), grupos (135)	Working groups definition	0,58	-
Reservoir operation	0,57	situação (183), operação (160), açudes (158), cagece (137), sistema (121), poços (121), qualidade (116), usuários (106), apresentou (105), fiscalização (104)	Reservoir operation	0,57	-
Municipal power participation	0,59	governo (209), municípios (195), prefeitura (188), municipal (173), estado (170), projeto (124), público (121), regional (105), saneamento (90), câmara (88)	Municipal power participation	0,59	-
Institutions engagement in the meetings	0,59	instituições (515), gestão (460), estado (165), participação (145), instituição (137), núcleo (132), comitês (116), representantes (113), processo (113), trabalho (111)	Institutions engagement in the meetings	0,59	-
Urban water supply	0,55	municípios (142), prefeitura (101), comunidades (81), municipal (70), abastecimento (66), senador (65), poços (57), situação (57), estado (55), onde (52)	Urban water supply	0,55	-
Mixed problems	0,65	ambiental (825), ambiente (508), projeto (478), ações (430), comissão (372), gestão (360), educação (338), secretaria (335), trabalho (319), plano (314)	Environmental conservation and education	0,66	ambiental (629), ambiente (323), educação (263), ambientais (180), projeto (142), ações (113), trabalho (109), escolas (91), municípios (90), semace (89)

Topic	Average cosine similarity	Ten most frequent words (word frequency)	Topic after reassessment	Average cosine similarity	Ten most frequent words (word frequency)
			Mixed problems	0,62	projeto (336), ações (317), comissão (300), gestão (275), comitês (268), plano (266), secretaria (263), águas (241), estado (239), alocação (228)
			University engagement	0,54	professor (128), sociedade (84), professora (68), participação (65), civil (64), universidade (60), gestão (50), comissão (44), representantes (41), federal (40)
Collaboration with the federal power	0,61	gestão (479), norte (455), público (364), limoeiro (308), núcleo (288), poder (254), sociedade (252), federal (250), civil (229), ministério (204)	Mixed problems	0,63	norte (442), gestão (387), limoeiro (304), núcleo (282), gerência (135), regional (122), coordenador (112), analista (97), técnico (93), gerente (82)
			State and Federal power participation	0,61	público (337), poder (224), federal (206), ministério (186), sociedade (167), civil (161), estadual (151), segmento (135), usuários (110), municipal (98)
Voting and board elections	0,58	duas (411), diretoria (273), eleitoral (237), eleição (227), votos (208), secretário (202), processo (173), instituições (165), votação (163), proposta (159)	Voting and board elections	0,58	-
Mission, values and vision	0,74	sustentabilidade (22), espaço (21), gestão (18), visão (17), reconhecido (16), principal (16), promoção (16), participativa (16), próximos (14), participação (10)	Mission, values and vision	0,74	-

Table S3. Search criteria used to classify the organizations identified with NER into classes. We used three dictionaries: (1) Match, which contains patterns to be detected in the organizations (either completely or partially); (2) Joint match, which contains pairs or triplets of strings that must be jointly detected in the entity name; and (3) Exact match, with strings to be exactly matched among the organization's vectors.

Organization class	Dictionaries of strings
State Water Company	Match: "cogerh", "nucleo" Joint match: ["conselho", "cogerh"]
Water basin committee	Match: "comissao", "comite", "cbh", "csbh", "camara tecnica de", "cbh", "comites" Joint match: ["comite", "bacia"] Exact match: ['csbhs baixo', 'camara tecnica de o plano de bacia', 'cbh', 'comite', 'comite de bacia', 'diretoria de o comite', 'comissao gestora', 'comissao gestora de o acude acarau mirim', 'comissoes gestoras', 'comissao de operacao de o vale de o acarau', 'comite gestor', 'comissao gestora de o farias de souzacbh-rmf', 'colegiado coordenador de o forum nacional de o comites de bacia hidrografica', 'diretorias de o comites de o estado', 'comites', 'comissao', 'cbh-coreau', 'cogerh', 'comites de o ceara', 'comites', 'comite', 'comite de bacia', 'comite de bacias', 'comites de bacia', 'comite de bacia hidrografica de o banabuiu', 'comite de bacia', 'secretaria executiva de o cbhs acarau', 'comites de bacia', 'comite de bacias', 'presidente de o cbh-acarau', 'cbh-acarau', 'coordenador de o nucleo de gestao de o bacias de o acarau', 'cbh-acarau de o sra', 'cbhs', 'camara de capacitacao e comunicacao', 'camara de compensacao ambiental', 'camara tecnica de educacao ambiental e mobilizacao social', 'camaras tecnicas e grupos de trabalho', 'camara tecnica', 'colegiado', 'colegiado coordenador', 'colegiado de assessoramento e representacao interinstitucional de o recursos hidricos carirh', 'comissao de alocacao', 'coordenador de o comissao eleitoral', 'coordenador de o nucleo de gestao', 'coordenadora de o junta eleitoral', 'coordenadora de o nucleo de gestao de o gerencia de o bacias metropolitanas', 'fccbh']
State power	Match: "defesa civil" Joint match: ["conselho", "estad"] Exact match: ['nucleo tecnico de o dnocs/ cest - ce', 'diretoria de o dnocs', 'dnocs', 'apremace', 'sinprece', 'empresa de assistencia tecnica e extensao rural ematerce', 'DNOCS', 'ibama-ce', 'cras', 'cedec', 'assembleia legislativa', 'fdid','seuma', 'conselho de altos estudos', 'defesa civil', 'sda', 'srh', 'promotor de justica', 'camara', 'secretario adjunto de recursos hidricos', 'secretaria de recursos hidricos', 'secretario de recursos hidricos', 'secretaria de desenvolvimento agrario', 'sohidra', 'secretaria de obras', 'cagece', 'camara tecnica de meio ambiente', 'secretario adjunto de o srh', 'conerh', 'conselho estadual de recursos hidricos', 'camaras de

	vereadores', 'ematerce', 'semace', 'secretaria de meio ambiente', 'secretarios de o recursos hidricos', 'grupo de altos estudos de o assembleia legislativa', 'secretaria de cultura', 'ce', 'conpam', 'defesa civil de o estado', 'shr', 'pm', 'assuntos estrategicos de o assembleia legislativa', 'caeae', 'funceme', 'srh', 'defesa civil', 'a celula de o regional de o secretaria estadual de saude-ce', 'adagre', 'adagri', 'aneel', 'arce', 'cap', 'ce', 'ceara', 'ciat', 'cmb', 'cnrh', 'coaf', 'codevasf', 'coelce', 'coema', 'comite semace', 'companhia de agua e esgoto do ceara/cagece', 'conerh', 'conpam', 'conpan', 'copam', 'conselho de altos estudos', 'conselho de altos estudos de o assembleia legislativa', 'crea', 'crea-ce', 'ematece', 'ematerce', 'ematerce titular', 'empresa de assistencia tecnica e extensao rural ematerce', 'enel', 'fiec', 'funcap', 'funceme', 'fundacao cis', 'fundacao cis francisca', 'geman', 'orientadora de o celula de articulacao de o srh', 'parlamento cearense', 'poder publico estadual', 'sda', 'seagri', 'seapa', 'sebrae', 'sedec', 'sedet', 'sema', 'semace', 'sesa', 'soihbra', 'sohidra', 'srh', 'srh/cogerh', 'sudene', 'tecnicos agricolas de nivel medio de o estado de o ceara', 'tce', 'unbme', 'cogerh']
Federal power	Joint match: ["conselho", "nacio"] Exact match: ['sesai', 'ana', 'ibama', 'funai', 'cprm', 'proagua', 'banco de o nordeste', 'banco mundial', 'bnb', 'projeto rondon', 'abes', 'abrasco', 'abrh', 'anvisa', 'aba', 'chefe de o poder executivo', 'colegiado nacional', 'comando de o exercito', 'conab', 'cprm', 'dnpm', 'exercito', 'exercito brasileiro', 'funasa', 'funai', 'fundacao brasil cidadao', 'ibama', 'ibama titular fernando cela pinto', 'icmbio', 'marinha de o brasil', 'mp', 'mpf', 'procurador de o republica', 'servico geologico do brasil'],
Municipal power	Exact match: ['camara de miraima', 'prefeitura', 'sisar-bac', 'sisar', 'secretaria de agricultura de forquilha', 'secretaria de meio ambiente de nova russas', 'saae', 'condema', 'amma', 'cmds', 'poder publico municipal de pacuja', 'secretaria de acao social', 'secretaria de agricultura', 'secretarias de saude e educacao', 'ceat', 'sec de agricultura de forquilha', 'prefeito', 'diretora de o escola profissionalizante antonio mota filho', 'aabb-sobral', 'casa cultura e arte solidaria', 'secretarias municipal', 'ama', 'camaras municipais', 'comite camara municipal de quixada', 'condema', 'corpo de bombeiros', 'icmbio de ubajara', 'poder publico municipal', 'prefeitos', 'aprece']
Media	Match: "radio" Exact match: ['jornal diario do nordeste', 'diario de o nordeste', 'reporter assembleia']
Agriculture users	Match: "agropolos", "agrovale", "associacao agro", "associacao de agri", "associacao de o agri" Joint match: ["associacao", "comunitaria", "agric"], ["associacao", "irrig"], ["associacao", "irrig"], ["associacao", "campo"], ["associacao", "produtor"], ["associac", "agric"] Exact match: 'esperanca agropecuaria', 'distar', 'promovale', 'dibau', 'conselho de o associacoes de o perimetro', 'coif', 'cooperativa de o irrigantes de o forquilha', 'perimetros irrigados', 'camara tecnica de

	irrigacao', 'dipan', 'dibau', 'cooperativa agroorganica', 'cooperativa de o irrigantes de forquilha', 'fapija', 'fetraece', 'nutrivale', 'organizacoes de produtores de o perimetros publicos de irrigacao fapid', 'produtores de o agricultura familiar ipueirense', 'trabalhadores rurais', 'trabalhadoras rurais de solonopole', 'univale', 'ascos associacao de fomento a caprino ovinocultura de gado de leite de sao joao do jaguaribe'
Industry users	Match: "carbomil", "cipp" Exact match: 'petrobras', 'industrial', 'ducoco', 'asterussas', 'ambev', 'associacao de empresas do complexo industrial e portuario de o pecem', 'companhia siderurgica de o pecem', 'construtora maciel', 'heineken', 'mineradora globest', 'votorantin', 'beckman sementes'
Aquaculture users	Match: "aquicultores", "associacao de carcin", "associacao de aqui", "aceaq", "bomar", "carcinicultura gaviao ltda", Joint match: ["associacao", "pescador"] Exact match: ['aquabras aquicultura ltda', 'aceaq']
Management instrument	Exact match: 'singerh', 'outorga', 'prodam', 'progerirh', 'funorh', 'prodham', 'singreh', 'politica de recursos hidricos', 'plano estrategico - pacto', 'plano', 'plandesva', 'plandesva', 'plano em o bacia', 'ppa de o governo', 'plano plurianual', 'plano de aproveitamento de o recursos hidricos', 'plano de aproveitamento de o recursos hidricos de santa quiteria', 'plano de o bacia', 'fiscalizacao e monitoramento de o recursos hidricos em o regularizacao de o usos', 'programa pro-agua', 'planosplano de bacia', 'plano de bacia de o acarau', 'cobranca de agua bruta', 'enquadramento', 'ferramentas de gestao participativa', 'gestao participativa de o recursos hidricos', 'instrumentos de gestao', 'pae', 'paerh', 'perh', 'plano', 'plano nacional de recursos hidricos', 'planos municipais', 'planos municipais de saneamento basico', 'pmsb', 'pnrh', 'politica nacional de recursos hidricos', 'politica nacional de seguranca de barragens', 'proagua', 'procomite', 'prodhamprogerirh', 'progestao', 'programa agua doce', 'programa nacional de fortalecimento de o comites de bacias hidrograficas', 'programa pro-comite', 'promovalereuniao de alocacao negociada', 'sistema integrado de gestao de recursos hidricos - sigerh'
Working groups	Match: "grupo de trabalho" Exact match: 'camara tecnica de plano de bacia', 'ct', 'grupo articuladores', 'grupo de cobranca', 'camaras tecnicas', 'ct de plano de bacia', 'ct de plano de bacia : camara tecnica de o bacia do acarau instituicao membro e-mail', 'ct de sistemas hidricos', 'camara tecnica', 'grupo de contingencia de o seca', 'gt', 'gti', 'gts'
Civil organizations	Match: "social", "acap", "associacao comunitari", "associacao beneficente", "associacao aratu", "associacao caatinga" "comunidade" Joint match: ["associacao", "comunitaria", "morador"], ["associacao", "comunitaria", "trabalhador"], ["associacao", "comunitaria", "quilomb"], ["associacao", "comunitaria", "usu"], ["associacao", "morador"], ["associacao", "idoso"], ["associacao", "cultura"], ["associacao",

"saude"], ["associacao", "vazant"], ["associacao", "trabal"],
["associacao", "profis"], ["federacao", "associac"], ["federac", "civ"],
["associacao", "beneficente"]

Exact match: 'movimento de mulheres de o serra de o ibiapaba mim',
'trabalhadoras rurais', 'sttr pereiro', 'agema', 'sttr de itapaje', 'cetra',
'movimento de mulheres', 'comunidade de curral velho', 'comite sindicato
de o trabalhadores rurais de quixeramobim', 'fambaci', 'famuva', 'grupo
de articulacao de o semi-arido', 'grupo de articuladores', 'grupo de
interesse ambiental', 'grupo gia', 'instituto mangara', 'instituto terramar',
'movimento de mulheres de o serra de o ibiapaba mim', 'movimento de o
atingidos', 'movimento por o soberania popular em o mineracao',
'movimento potiguar-tapuia', 'oliveira associacao dos professores de
independencia- aprofi', 'oscip', 'senge-ce', 'sindicato de o professores de
crateus', 'sindicatos de o trabalhadores rurais', 'terramar', 'comunidade de
o acude sao vicente', 'associacao santanense de promocao social',
'comunitaria taboense', 'conselho de o povo potyguara de o serras de o
matas - mons', 'associacao de o pescadores e marisqueiros', 'associacao
de o marisqueiros', 'associacao de o marisqueiros de curral velho',
'federacao de o associacoes comunitarias de santa quiteria', 'acaap', 'acao
social de campos sales', 'acpa assoc', 'adel', 'agente jovem ambiental',
'amuvale', 'amab', 'comunidade', 'atasa associacao de o trabalhadores de
acarau', 'segmento sociedade civil'

Banks	Exact match: "banco", "banco de o nordeste", "banco mundial", "bird", "bnb", "bndes"
Educational institutions	Exact match: 'unifor', 'ceja', 'urca crajubar', 'urca', 'ufca', 'facedi', 'ifce acarau', 'cefet', 'fafidam', 'ifce', 'centec', 'eafi', 'embrapa', 'escola familia agricola', 'faec', 'programa cientista chefe', 'uece', 'ufc', 'ufc campus russas', 'uva'
Consulting companies	Exact match: ['kl engenharia', 'consultoria ibi', 'ibi', 'ibi consultoria', 'ibi engenharia', 'ibi - relatorio fase apresentacao a o plenaria']
Religious institutions	Match: 'igreja', 'igreja catolica'
International organizations	Exact match: ['caritas', 'caritas brasileira', 'oms', 'onu', 'uniao europeia', 'ue']
Other institutions	Exact match: 'cagepa', 'geosolos', 'empresa de assistencia tecnica'