CS-E4875 - Research Project in Machine Learning, Data Science and Artificial Intelligence D

Comparison of Offline and Online Interaction of Climate Change Actors

Isabella Rositi

May 17, 2024

Abstract

Nowadays, climate and sustainability are crucial aspects of any policy-making process in many countries, and this project aims to investigate the intricate dynamics among climate change players across five countries worldwide, exploring seven distinct networks with each nation's corporate landscape. These networks reveal diverse climate-related relationships among high-profile organizations and highlight nuanced decision-making processes regarding climate policies. Leveraging a dataset of tweets sourced from official accounts and employees of the companies in question, the behaviour of both the networks and the tweet content is been analysed. By performing topic modeling on the tweets, this research examines whether the presence of a direct connection significantly impacts the number of shared topics between two companies, which are ultimately found to be influenced by the connectivity of the networks and the primary interests of their entities.

1 Introduction

In today's rapidly evolving society, it is crucial to address climate change issues for organizations, even when their core business is not directly related to sustainability. Indeed, by fostering economic and social progress, organizations can have great influence in shaping policies that mitigate environmental impact. With companies' chronically online presence, special attention needs to be places on the content shared, in order to assure its alignment with the company values. This dual representation of companies beliefs provides a unique opportunity to analyse the differences between the organizations' offline dynamics, such as their relationships with other companies in their ecosystem, and their online dialog. By exploring whether organizations that self-declare themselves as "close" in regards to climate change policies, also show ideological proximity and topical similarity in their online conversations, this project aims to uncover deeper insights into the connections between organizational relationships, climate change policies, and online discussions.

1.1 Problem Statement

Are the relationships established among organizations offline reflected in the ideologies and conversational topics they engage with online? By investigating this question, this project seeks to analyse the extent to which organizations' online conversations mirror their offline interactions, providing valuable insights into the dynamics of organizational behavior in the context of climate change policies.

2 Related Works

2.1 Topic Modeling: BERTopic

The topic modeling performed in this project is based on BERTopic, Grootendorst's work in [7]. Egger and Yu [4] performed an in-depth comparison of four different techniques on Twitter data: LDA, NMF, BERTopic, and Top2Vec. Latent Dirichlet Allocation (LDA) [2] is a generative probabilistic model, Nonnegative Matrix Factorization (NMF) [5] uses a linear algebra approach for topic extraction, and both BERTopic and Top2Vec [1] use an embedding approach.

Their research shows that both BERTopic and NMF can discern well-separated topics, yet NMF's topics tend to be more general. The authors attribute BERTopic's success to its leveraging of pre-trained embedding models, however, while Top2Vec also employs such models, its results often overlap between diverse concepts. Lastly, LDA also produces overly broad and uninformative topics [4]. Besides the lack of insightful concepts identified, both LDA and NMF require a pre-defined amount of topics and struggle to scale efficiently to large datasets, making these methods unsuitable for this analysis.

2.2 The COMPON Project

This research revolves around data provided by Compon.org, a project focusing on *Comparing Climate Change Policy Networks* (COMPON), aiming to shed light on the complexities behind variations in national climate change policies across different countries. The core idea of the COMPON project is an exploration of meso-level policy networks and media discussion networks, with the intent to highlight the actors and coalitions shaping climate policy decisions.

The gathering and analysis of policy network data is part of the COMPON project, which involves conducting surveys among the most influential organizations within each country's climate change policy domain. These surveys focused on mapping inter-organizational collaborations, information exchanges, and attitudes towards climate change, including adherence to UN-established emission reduction norms [18].

Over the years several studies have emerged on climate policy networks, mostly focusing on individual countries. For instance, Yun et al. [19] found two conflicting clusters in South Korea's political landscape, with the stronger group prioritizing economic growth and the weaker emphasizing environmental concerns. Similarly, in Finland, Gronow and Ylä-Anttila [6] identified a powerful coalition favoring economic growth over environmental preservation, which can influence important climate policies also on a national level. In the United States, Jasny et al. [10] observed distinctive patterns in communication networks among politicians and scientists, which show that the information shared reinforces existing beliefs and it contributes to political polarization and policy immobility at a national level.

Research by Tindal et al. [17] reckons the scarcity of climate politics analyses that integrates media discourse with policy network analysis and highlights the importance of media in shaping public perception and understanding of climate change and explain how important players use their visibility to communicate their policy ideologies through the media. Building upon this observation, Kotkaniemi et al. delve into the dynamics between offline and online policy influence in Finland [11]. Their findings reveal that actors with perceived authority in offline policy-making also show significant influence online, independently from their institutional role. However, individuals in influential offline roles without strong perceived credibility do not demonstrate the same level of online influence.

3 Methodology

3.1 Proposed Solution

Very few work has been researched in regards to a combined analysis of offline and online behaviour of the climate actors present in the COMPON survey, so far only covering the Finnish ecosystem, hence this project aims to provide additional insights in this direction. The aim is to navigate whether there is an alignment between the self-declared organizations' relationships and their shared values and topic of discussions on social media, namely on Twitter. This investigation will be carried out on five different countries around the globe: Australia (AU), Canada (CA), Finland (FI), Ireland (IE), and Sweden (SE).

3.2 Policy Networks

3.2.1 Data

For every country (Australia, Canada, Finland, Ireland, and Sweden) between 40 and 89 organizations have been surveyed and a number of questions have been asked regarding their relationships to other organizations present in their same country related to climate change. Table 1 shows the questions the companies were asked in order to be able to create national policy networks.

Network	Description
N1	Influential in domestic climate change politics
N2	Source of expert scientific information
N3	Regular collaboration
N4	Reliable advice about climate policy
N5	Provide a public forum to exchange ideas
N6	*Not answered by the countries taken into analysis*
N7	Provide with scientific information
N8	*Not answered by the countries taken into analysis*
N9	Which policy actors have a strong influence upon the organizational stance of your organization?

Table 1: Networks description

However, not all countries replied to all of the questions, and as can be seen in Table 2, the networks available for every nation are "N1 - Influential in climate change politics" and "N3 - Regular collaboration".

Network	AU	CA	FI	IE	SE
N1	●	●	●	●	●
N2	●	○	●	●	●
N3	●	●	●	●	●
N4	○	○	○	●	○
N5	○	○	○	●	○
N6	○	○	○	○	○
N7	○	○	○	●	○
N8	○	○	○	○	○
N9	○	○	○	●	○

Table 2: Networks available for every country

3.2.2 Distances

In order to assess the closeness of the organizations within the networks, the shortest path length is calculated employing the Dijkstra algorithm [3] implemented in the Python library NetworkX.

3.3 Tweets

3.3.1 Data

For each of the country, tweets about climate have been scraped in the span of 5 years (2017-2023) whose authors are either official accounts of the organizations present in the networks or identifiable employees. For each country between 50.000 and 111.000 tweets have been retrieved.

3.3.2 Topic Modeling

The topic modelling system builds on several steps feeding into each other, using the following order: pre-processing, generating embeddings, dimensionality reduction, clustering, and cluster naming.

1. **Pre-processing**: all text is lemmatized, and words are broken down to their root.

2. **Embeddings**: instead of using BERT, the embedding model proposed in BERTopic, in this project is implemented All-MPNet-Base-v2, as it has shown strong performance [15] and builds on the work of [16]. It processes each tweet generating a 768-dimensional output.

3. **Dimensionality Reduction**: the dimensionality reduction of the embeddings is performed with the UMAP (Uniform Manifold Approximation and Projection) algorithm. The choice of UMAP is motivated by its exceptional performance in terms of both speed and accuracy, as demonstrated by Mcinnes et al. in their study comparing dimensionality reduction methods [13]. Moreover, they also show that UMAP preserves the local and global structures of the original data.

   The algorithm consists of multiple hyper-parameters, and in this project the parameters `n_neighbors` and `n_components` were selected for finetuning. `n_neighbors` determines the number of neighbors considered when constructing a smaller graph used for the reduction and `n_components` specifies the number of dimensions the data is being reduced to.

4. **Clustering**: the clustering is performed by HDBSCAN [12], a density-based algorithm. One of the most implemented clustering approaches is hierarchical clustering [8], however, it is sensitive to noise and outliers, it struggles to handle clusters of varying densities and complex structures, and it necessitates the number of clusters to be known apriori. These constraints prevent the implementation of this method, in favour of a less common approach known as density-based clustering [12], which groups data points based on their density in the feature space. HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) combines the strengths of both hierarchical and density-based clustering, allowing for the automatic discovery of clusters of varying densities and robust handling of noise and outliers.

In this project the fine-tuning concerns the parameters `min_cluster_size` and `min_samples`, which control respectively the minimum size of a cluster and the minimum number of samples that need to be in a point's neighborhood in order to form a cluster.

5. **Cluster Naming**: the cluster naming uses c-TF-IDF as suggested in the BERTopic documentation [7], which is a class based version of the popular TF-IDF technique. Once the clusters have been finalised, all of the documents within the same cluster are concatenated into a single document (a class). Class-based TF-IDF models the importance of words in clusters instead of individual documents and [7] shows that by analysing the words with high TF-IDF scores, the most significant terms and concepts in a cluster can be pinpointed.

In order to identify the best combination of UMAP and HDBSCAN hyper-parameters, HDBSCAN provides the relative_validity_ function, which is a relative measure of goodness of clustering used to compare results across different choices of hyper-parameters. Table 3 shows the values tested for each parameter, and in bold can be observed the best combination identified.

Algorithm	**Hyper-parameter**	**Values Tested**
UMAP	`n_components`	[2, **5**, 15, 30]
UMAP	`n_neighbours`	[**2**, 5, 15, 30]
HDBSCAN	`min_samples`	[**1**, 5, 15, 30, 50]
HDBSCAN	`min_cluster_size`	[**5**, 15, 30, 50, 100]

Table 3: Hyper-parameters tested, and in bold the values for best performing model

Determining the accuracy of the topic modeling is not trivial, as there is not a correct answer the results can be compared to. Therefore, a limited set of items (200 per country) has been manually given a "topic" so that these can serve as a validation set. The accuracy can then be determined by manually reviewing how similar the topics identified by the model are to those of the validation set.

3.4 Networks-Tweets Comparison

The number of shared topics is determined for each pair of companies. Afterwards, each pair is categorised based on their shortest path length, which ranges from 1 to 6, based on the country and the network under analysis. Consider μ_i and μ_j the means of the distributions of shared topics for all pairs of companies belonging to categories i and j, where $i < j$, the hypothesis tested are:

$$\begin{cases} H_0 : \mu_i = \mu_j \\ H_1 : \mu_i > \mu_j \end{cases} \tag{1}$$

The test is conducted using t-testing [9] for distributions that met the assumptions and Mann-Whitney U test[14] for those that do not. Assumptions include normality for category distributions with at least 10 observations and equality of variance among distributions with at least 10 observations. The means of the distributions are considered different when the null hypothesis is rejected with a confidence level of 95%, hence when the p-value obtained is less than 0.05.

4 Results

Table 4 shows descriptive statistics of the policy networks, and it can be observed that Finland is the country with the highest density, revealing the high interconnection that distinguishes the Finnish culture. On the other hand, Canada shows very low densities, highlighting the fragmented nature of the country. Further in-depth descriptive analysis of these networks can be found in Table 9 in the Appendix A.

Network	**Country**	**Directed**	**Nodes**	**Edges**	**Density %**
N1	AU	True	39	361	24.36
	CA	True	80	476	7.53
	FI	True	89	2309	29.48
	IE	True	53	627	22.75
	SE	True	85	1529	21.41
N2	AU	True	35	104	8.74
	FI	True	87	1004	13.42
	IE	True	52	264	9.95
	SE	True	75	695	12.52
N3	AU	False	33	109	20.64
	CA	False	72	183	7.16
	FI	False	89	1159	29.60
	IE	False	53	358	25.98
	SE	False	79	369	11.98
N4	IE	True	52	261	9.84
N5	IE	True	51	207	8.12
N7	IE	True	54	212	7.41
N9	IE	True	53	328	11.90

Table 4: Statistics on policy networks

Table 5, on the other hand, displays the statistics of the tweets retrieved, showing a high number of clusters for each country. These results stem from a clustering approach that categorises topics with high resolution, favouring the creation of very detailed, and small sized clusters.

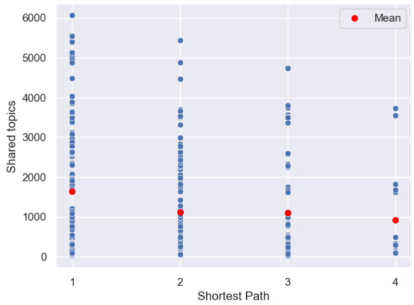
Country	Companies	Tweets	Clusters
AU	39	76.989	7.598
CA	82	51.095	5.219
FI	89	111.364	10.834
IE	54	53.537	4.597
SE	85	50.175	5.094

Table 5: Statistics on Tweets

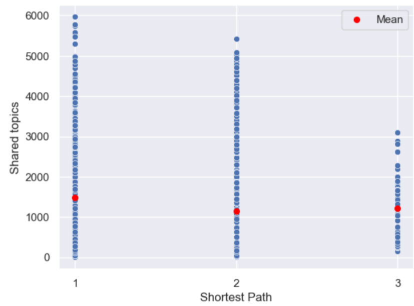
4.1 N1 - Influential in domestic climate change politics

For Australia, Finland and Sweden, shown in Figure 1, the mean of shared topics between companies that receive direct influence in domestic climate change politics (shortest path length = 1) is higher than the mean of shared topics between companies without direct influence (shortest path length > 1) with a level of confidence of 95%.

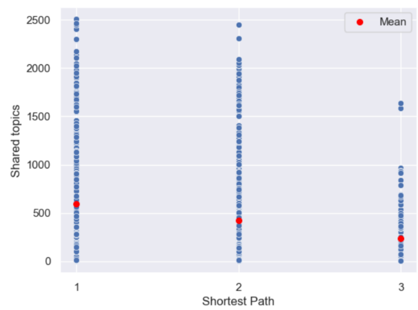
For Canada and Ireland, on the other hand, there is no sufficient evidence to reject the null hypothesis: direct influence does not reflect in a higher number of shared topics. Significant p-values can be found in Table 8 in Appendix A.



(a) Australia



(b) Finland

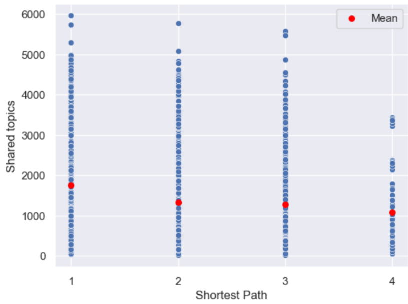


(c) Sweden

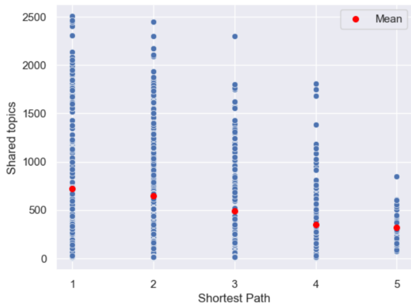
Figure 1: Distributions of topics shared for shortest path length for network "N1 - Influential in domestic climate change politics" for Australia, Finland and Sweden. Mean is shown for t-testing and median for Mann-Whitney U test

4.2 N2 - Source of expert scientific information

For Finland and Sweden, shown in Figure 2, the mean of shared topics between companies that share direct expert scientific information (shortest path length = 1) is higher than the mean of shared topics between companies without direct share of such expertise (shortest path length > 1) with a level of confidence of 95%.



(a) Finland

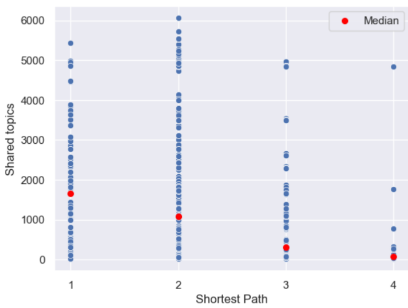


(b) Sweden

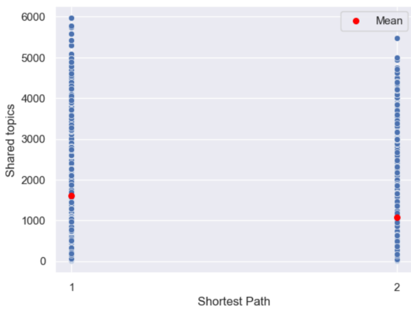
Figure 2: Distributions of topics shared for shortest path length for network "N2 - Source of expert scientific information" for Finland and Sweden. Mean is shown for t-testing and median for Mann-Whitney U test

For Australia and Ireland, on the other hand, there is no sufficient evidence to reject the null hypothesis: direct scientific exchange does not reflect in a higher number of shared topics. Significant p-values can be found in Table 8 in Appendix A.

4.3 N3 - Regular Collaboration



(a) Australia



(b) Finland

Figure 3: Distributions of topics shared for shortest path length for network "N3 - Regular Collaboration" for Australia and Finland. Mean is shown for t-testing and median for Mann-Whitney U test

For Australia and Finland, shown in Figure 3, the mean of shared topics between companies that regularly collaborate (shortest path length = 1) is higher than the mean of shared topics between companies without direct collaboration (shortest path length > 1) with a level of confidence of 95%.

For Canada, Ireland and Sweden, on the other hand, there is no sufficient evidence to reject the null hypothesis: regular collaboration does not reflect in a higher number of shared topics. Significant p-values can be found in Table 8 in Appendix A.

4.4 Other Networks

Ireland has additional 4 networks, but none of them shows sufficient evidence to reject the null hypothesis with a confidence level of 95%: direct interaction does not reflect in a higher number of shared topics. Significant p-values can be found in Table 8 in Appendix A.

5 Discussion

Table 6 summarizes the results, highlighting the networks that where direct interaction is found to have an impact on the number of shared topics. As observed, for Finland direct interactions always reveal a higher number of topics shared online, whereas for Canada and Ireland, this is never the case.

Country	**N1**	**N2**	**N3**	**N4**	**N5**	**N7**	**N9**
AU	●	○	●	X	X	X	X
CA	○	X	○	X	X	X	X
FI	●	●	●	X	X	X	X
IE	○	○	○	○	○	○	○
SE	●	●	○	X	X	X	X

Table 6: Overview of the results of hypothesis testing

Legend:
●: direct interaction shows more shared topics
○: direct interaction does not show more shared topics
X: network not available

By analysing the structure of the networks in Table 9 in the Appendix A, it stands out that Canada policy networks have very low reciprocity and density, and a very high number of components and maximum path length compared to the other countries. These statistics suggest that Canadian landscape is highly fragmented, with smaller ecosystems hardly interconnected.

Finland, on the other hand, has very high reciprocity and density and low values of maximum path lengths in all policy networks, compared to the other countries. For N1, the Finnish landscape shows 30% of total possible connections, against Canada showing only 7%. Similarly, Finnish companies are at most 3 steps away from each other, whereas Canadians go as far as 6. This hyper-connection hints at an ecosystem highly communicative, which is mirrored also online.

Moreover, investigating the three most common topics for every country's tweets offers deeper insights into why direct connections may not influence dialogue across all networks homogeneously. As previously discussed, in Australia, Finland and Sweden the first-hand connections show an impact on online discourse within network "N1 - Influential in domestic climate change politics". Notably, Table 7 illustrates that political themes are among the most prominent in these countries ("auspol", "the Greens", "elections"), whereas in Canada and Ireland they are not.

Similarly, direct interactions within "N2 - Source of expert scientific information" are relevant for online conversations only in Finland and Sweden, both of which show keywords related to science in their most common topics ("reports", "climate science" and "researchers").

Country	**Topic Keywords**	**Size**
AU	coal	48
	decarbonised reduce emissions	38
	auspol	35
CA	climateanxiaty	30
	climateambition cop26	29
	fossilfuel emissions	28
FI	coal (*hiilen*) the Greens (*vihreät*)	112
	globalwarming anomaly visualizations (*dataviz*)	53
	reports (*ipccraportti*) climate science (*ilmastotiede*)	40
IE	climatemergency climateaction	122
	reduce carbon emissions 2025	60
	fossilfree	49
SE	elections (*val*) gender equality (*jämställdhet*)	44
	researchers (*forskare*) sustainability (*hållbarhet*)	28
	climatesecurity conflict	28

Table 7: Topic keywords for the three most common topics in every country. In *italics* the original keywords which have been translated to English.

5.1 Challenges and Future Work

One clear challenge of performing topic modeling without a predefined set of themes or without a large set of manually labeled text is the lack of clear assessment of the goodness of the results. Even with large amount of labeled tweets, the process would be biased because of the human deciding the "topic" of the text, making it hard to be able to generalize the goodness of the performance.

Future work could focus on further investigate suitable performance measures for unsupervised learning tasks with text-based unlabeled ground truth. Additionally, analysing one specific country and leveraging in-depth knowledge about its ecosystem could provide more accurate insights into the dynamics of climate change discourse within that context. Lastly, there is an opportunity to critically assess the definition of *topic* within this framework and determine the optimal resolution for categorizing topics within the dataset.

6 Conclusions

This project intended to investigate whether the self-declared relationships among companies active in climate change policies are reflected in their online discourses on Twitter. Through a topic analysis and consequent hypothesis testing, it emerged that it cannot be stated that overall the offline relationships are reflected in the themes discussed online. This relation depends highly both on the structure of the national corporate ecosystem and the main interests of the entities of the networks. High interconnection among players favours the similarity of online and offline behaviours and high fragmentation discourages it. However, the singular cultures and landscape attributes at national level should be investigates and taken into account to better understand these connections.

A Appendix

Network	Country	Alternative Hypothesis H_1	P-Value
N1	AU	$1 > 2$	0.000
	AU	$1 > 3$	0.003
	AU	$1 > 4$	0.027
	FI	$1 > 2$	0.000
	FI	$1 > 3$	0.039
	SE	$1 > 2$	0.000
	SE	$1 > 3$	0.000
	SE	$2 > 3$	0.000
N2	AU	$2 > 3$	0.034
	FI	$1 > 2$	0.000
	FI	$1 > 3$	0.000
	FI	$1 > 4$	0.000
	FI	$2 > 4$	0.037
	SE	$1 > 2$	0.044
	SE	$1 > 3$	0.000
	SE	$1 > 4$	0.000
	SE	$1 > 5$	0.000
	SE	$2 > 3$	0.000
	SE	$2 > 4$	0.000
	SE	$2 > 5$	0.000
	SE	$3 > 4$	0.000
	SE	$3 > 5$	0.007

Continues in next page

Network	**Country**	**Alternative Hypothesis H_1**	**P-Value**
N3	AU	$1 > 2$	0.004
	AU	$1 > 3$	0.000
	AU	$1 > 4$	0.000
	AU	$2 > 3$	0.000
	AU	$2 > 4$	0.000
	AU	$3 > 4$	0.000
	FI	$1 > 2$	0.000
	SE	$1 > 3$	0.007
	SE	$2 > 3$	0.007
N4	IE	$2 > 4$	0.003
	IE	$2 > 5$	0.006
	IE	$3 > 4$	0.018
	IE	$3 > 5$	0.017

Table 8: Overview of all significant p-values < 0.05 for the hypothesis testing

Network	Country	Directed	Nodes	Edges	Reciprocity (%)	Density (%)	Strong Comp.	% Nodes in Big Comp.	% Edges in Big Comp.	Average Degree	Clustering Coefficient	Average Path	Max Path	Reachable Network (%)
N1	AU	True	39	361	18.84	24.36	16	61.54	52.91	18.51	0.34	1.83	4	82.05
	CA	True	80	476	4.20	7.53	67	16.25	8.82	11.90	0.18	2.28	6	63.75
	FI	True	89	2309	29.71	29.48	17	82.02	82.63	51.89	0.44	1.67	3	98.88
	IE	True	53	627	29.98	22.75	15	73.58	72.41	23.66	0.48	1.75	3	96.23
	SE	True	85	1529	23.68	21.41	31	64.71	62.85	35.98	0.39	1.73	4	98.82
N2	AU	True	35	104	3.85	8.74	33	8.57	4.81	5.94	0.21	1.74	4	71.43
	FI	True	87	1004	12.35	13.42	40	54.02	51.99	23.08	0.37	2.12	4	65.52
	IE	True	52	264	12.88	9.95	23	57.69	53.41	10.15	0.28	2.41	5	73.08
	SE	True	75	695	6.91	12.52	49	36.00	30.94	18.53	0.28	2.23	5	53.33
N3	AU	False	33	109	-	20.64	1	100.00	100.00	6.61	0.55	2.17	5	-
	CA	False	72	183	-	7.16	1	100.00	100.00	5.08	0.21	2.62	6	-
	FI	False	89	1159	-	29.60	1	100.00	100.00	26.04	0.59	1.70	2	-
	IE	False	53	358	-	25.98	1	100.00	100.00	13.51	0.61	1.87	4	-
	SE	False	79	369	-	11.98	1	100.00	100.00	9.34	0.41	2.28	5	-
N4	IE	True	52	261	23.75	9.84	16	71.15	67.05	10.04	0.32	2.53	7	94.23
N5	IE	True	51	207	15.46	8.12	26	50.98	52.17	8.12	0.25	2.55	6	84.31
N7	IE	True	54	212	22.64	7.41	31	44.44	46.23	7.85	0.52	2.16	4	98.15
N9	IE	True	53	328	26.83	11.90	15	73.58	76.52	12.38	0.35	2.35	5	96.23

Table 9: Descriptive analysis of every national network

References

[1] Dimo Angelov. Top2Vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*, 2020.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[3] Jing-Chao Chen. Dijkstra's shortest path algorithm. *Journal of formalized mathematics*, 15(9):237–247, 2003.

[4] Roman Egger and Joanne Yu. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in sociology*, 7:886498, 2022.

[5] Nicolas Gillis and Stephen A Vavasis. Fast and robust recursive algorithms for separable nonnegative matrix factorization. *IEEE transactions on pattern analysis and machine intelligence*, 36(4):698–714, 2013.

[6] Antti Gronow and Tuomas Ylä-Anttila. Cooptation of ENGOs or treadmill of production? Advocacy coalitions and climate change policy in Finland. *Policy Studies Journal*, 47(4):860–881, 2019.

[7] Maarten Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[8] Trevor Hastie, Jerome Friedman, and Robert Tibshirani. *Unsupervised Learning*, pages 437–508. Springer New York, 2001.

[9] Henry Hsu and Peter A Lachenbruch. Paired t test. *Wiley StatsRef: statistics reference online*, 2014.

[10] Lorien Jasny, Joseph Waggle, and Dana R Fisher. An empirical examination of echo chambers in US climate policy networks. *Nature Climate Change*, 5(8):782–786, 2015.

[11] Anniina Kotkaniemi, Tuomas Ylä-Anttila, and Ted Hsuan Yun Chen. Policy influence and influencers online and off. *Policy Studies Journal*, 2023.

[12] Leland McInnes, John Healy, Steve Astels, et al. HDBSCAN: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205, 2017.

[13] Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[14] Patrick E McKnight and Julius Najab. Mann-Whitney U Test. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[15] Nils Reimers. Sentence-Trasformers: Pretrained Models. `/https://www.sbert.net/docs/pretrained_models.html`. [Accessed: May 7, 2024].

[16] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[17] David B Tindall, Mark CJ Stoddart, and Candis Callison. The relationships between climate change news coverage, policy debate, and societal decisions. In *Oxford Research Encyclopedia of Climate Science*. 2018.

[18] Tuomas Ylä-Anttila, Antti Gronow, Mark CJ Stoddart, Jeffrey Broadbent, Volker Schneider, and David B Tindall. Climate change policy networks: Why and how to compare them across countries. *Energy Research & Social Science*, 45:258–265, 2018.

[19] Sun-Jin Yun, Dowan Ku, and Jin-Yi Han. Climate policy networks in South Korea: Alliances and conflicts. *Climate policy*, 14(2):283–301, 2014.