Note: Group 4 is going to work in English

MAIN KEYWORDS (shared)

- Energy Transition
- Greenhouse Effect
- Biodiversity
- Extreme weather events
- CO2
- Emissions
- Global Warming
- Melting Glaciers
- Renewable energy
- Fake News

OUR WORDS:

1. Ecosystem
2. Fossil Fuels
3. Energy Consumption
4. Normatives
5. Deforestation
6. Flooding
7. Tesla
8. Green Policies
9. Rain
10. Electric Vehicles

EXTRAS:

11. Natural Disaster
12. Clean Energy
13. Net zero
14. Tesla
15. Heatwaves

TO FIND:

degree distribution
average degree
strength distribution
average strength
calculate betweenness centrality
calculate closeness centrality
local measured
global graph density
global clustering
graph diameter
community modularity
clusters

```
KEYWORDS = [
   "Ecosystem", "Fossil Fuels", "Energy Consumption", "Normatives", "Deforestation",
   "Floodings", "Heatwaves", "Green Policies", "Rain", "Electric Vehicles"]
```

LOCALITA’

Norvegia,
California
Quebec
Thailand

Code

https://colab.research.google.com/drive/1YCgFFEoYF0_aKGQUhLZa4QS-DHM7P2UI?usp=sharing

https://github.com/francescosilvano/complex-systems

OBSERVATION WITH DIFFERENT COUNTRIES AND 25 KEYWORDS

General Analysis

The results of this study offer a detailed view of how climate-related and energy-related concepts are discussed on Bluesky. By collecting posts that included a set of twenty-five selected keywords, divided in shared keywords and 15 our keywords, and analysing how these terms appeared together, we were able to construct a weighted network that reflects the structure of the online conversation. In this network, each keyword represents a node, while the edges describe how many times two terms were mentioned in the same post.

The analysis clearly shows that the discourse surrounding climate change is highly interconnected. An interesting finding is the dominant presence of the keyword AI, which appears with almost all other concepts at unusually high frequencies. For instance, AI co-occurs 260 times with Rain, 207 times with Heatwaves, 188 times with Floodings, and 201 times with Extreme weather events. It also forms strong links with topics such as Glaciers, Natural Disaster, and Normatives. These patterns suggest that users frequently discuss environmental events together with technological themes, especially artificial intelligence, which is often associated with prediction, data analysis, misinformation, or innovation in climate solutions.

Alongside AI, several environmental keywords also play central roles in the discourse. Rain, Heatwaves, Emissions, CO2, and Floodings show consistently high co-occurrence values with many other terms, confirming that users often relate extreme weather phenomena to pollution, emissions, and climate impacts. For example, Rain appears 69 times with Floodings and 48 times with Heatwaves, while Emissions is strongly linked to the Greenhouse Effect, Heatwaves, and Net Zero. These frequent associations underline the multidimensional nature of public discussions on climate issues, where natural events, scientific explanations, and policy measures are often mentioned together.

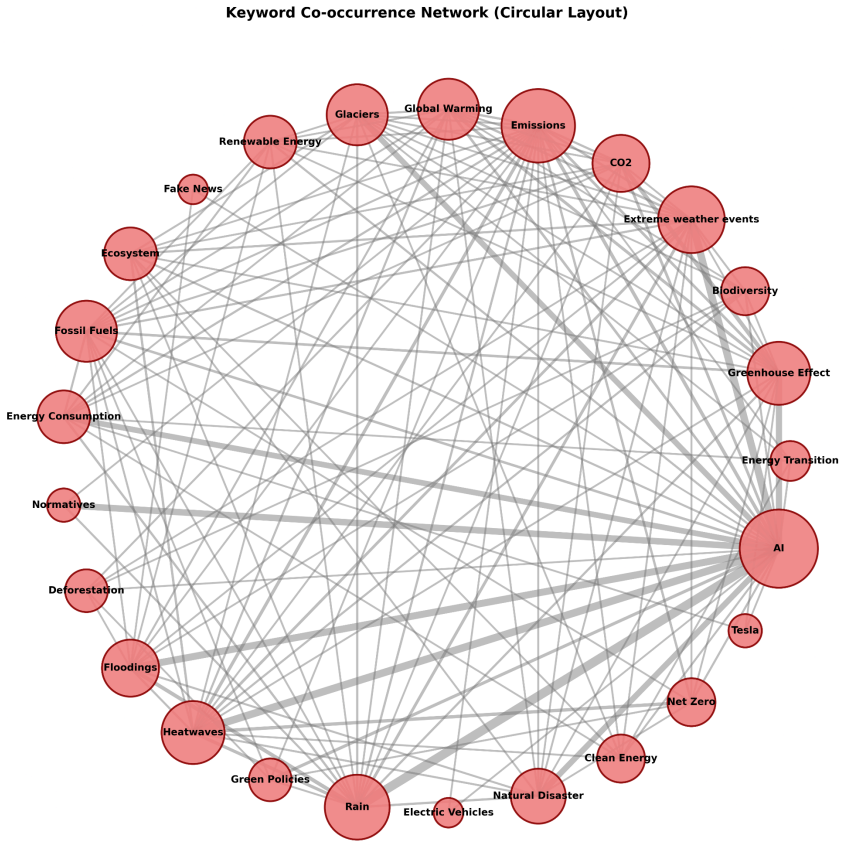
The network metrics help to interpret the structure of this discourse. The high density of the graph, combined with elevated clustering coefficients for most nodes, reveals a cohesive network where concepts tend to appear in recurring groups. Although a small number of nodes have very high betweenness centrality—especially AI, Emissions, and Heatwaves—the majority of terms are closely connected. This means that the overall conversation is not separated into isolated subtopics but instead forms a unified structure in which various themes are regularly integrated with one another.

Sentiment analysis further enriches this interpretation. Most posts containing the keywords express a neutral tone, which indicates that the discussion is often descriptive or factual. Positive sentiment appears mainly in relation to topics such as clean energy or technological progress, while negative sentiment is associated with

environmental threats or extreme weather conditions. This emotional balance supports the idea that the climate conversation on Bluesky is largely informative rather than polarised.

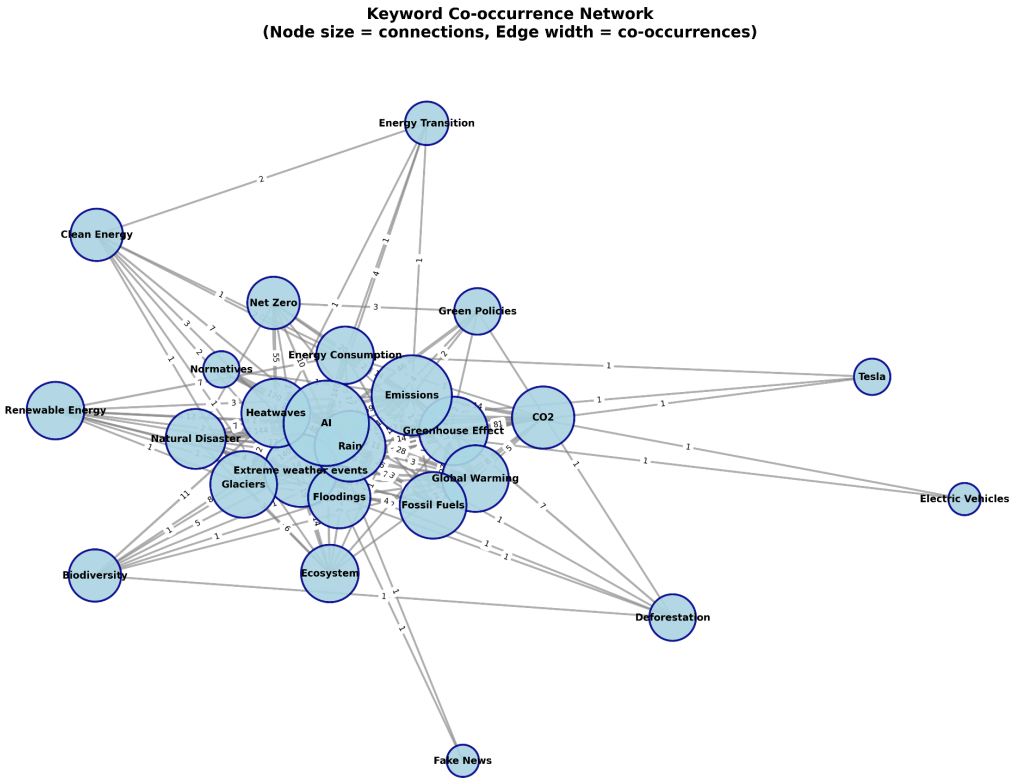
Analysis of Each Output

Circular Network (keyword_network_circular.png)



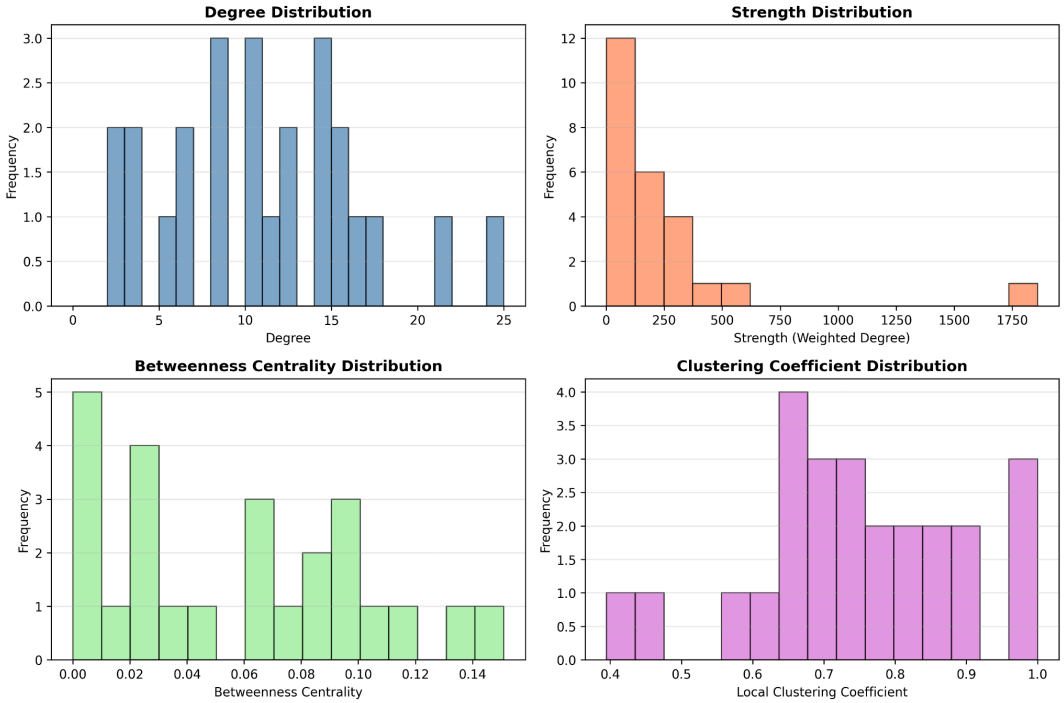
The circular network visualisation illustrates the overall structure of co-occurrences between the selected keywords. All nodes are arranged around a circle, making it easy to observe their relative size and connectivity. The larger nodes, such as *AI*, *Emissions*, and *Global Warming*, indicate keywords that appear with many others, reflecting their central role in the discourse. The density of the edges demonstrates the frequent overlap between different concepts, producing a picture of the climate conversation as a tightly connected web rather than a set of independent topics.

Force-Directed Network (keyword_network.png)



The second visualisation offers a more dynamic representation of the network through a force-directed layout. In this figure, the spatial position of each node depends on its relationships with others: keywords with stronger or more frequent connections are pulled toward the centre, while those with fewer links remain on the periphery. This layout makes the internal structure of the conversation easier to interpret. Keywords such as *AI*, *Heatwaves*, *Rain*, and *Floodings* occupy central positions because they strongly connect to various other terms. In contrast, *Tesla*, *Electric Vehicles*, and *Fake News*appear more distant, showing that they are discussed less often within the broader climate narrative. The numerical labels on the edges highlight the intensity of each connection and confirm the dominant position of certain central themes.

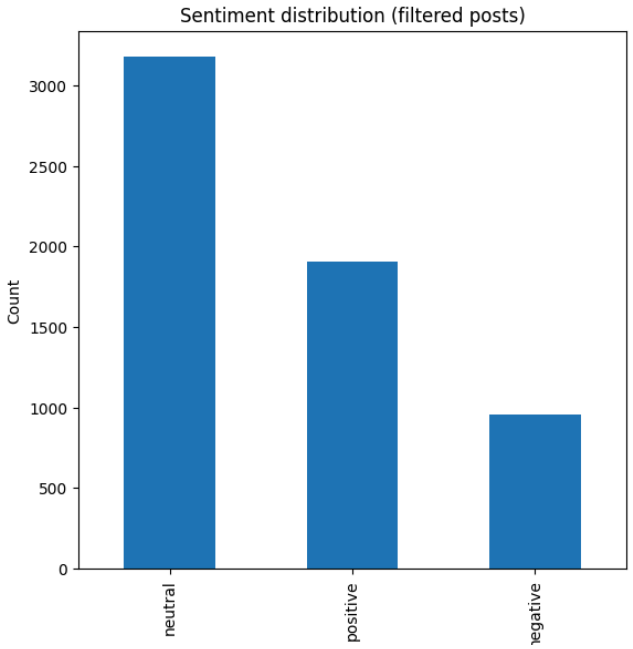
Distribution of Network Metrics (network_metrics.png)



The metrics figure summarises four key structural properties of the network: degree, strength, betweenness centrality, and local clustering coefficient. The degree distribution shows that most keywords are connected to many others, which confirms the dense nature of the graph. The strength distribution reveals a large difference between a small number of highly influential nodes and many nodes with lower influence. This imbalance reflects the uneven attention given to certain topics in climate discussions.

The betweenness centrality distribution indicates that only a few keywords act as bridges between different parts of the network. These bridging terms, which include topics such as *AI* and *Emissions*, help link various thematic areas together. The clustering coefficient distribution displays generally high values, indicating that groups of concepts frequently appear together in the same posts. This suggests that the climate conversation is characterised by recurring clusters of strongly related topics.

Sentiment Distribution (Sentiment.png)



The sentiment plot highlights the overall emotional tone of the posts used to build the network. The predominance of neutral sentiment suggests that the majority of messages containing the selected keywords are informative or factual. The presence of positive sentiment is often associated with solutions, progress, or technological innovation, while negative sentiment tends to appear in posts discussing natural disasters or climate emergencies. Overall, the sentiment analysis reinforces the idea that the conversation tends to focus on reporting rather than strong emotional reactions.

Co-occurrence Matrix (grafo.xlsx) → tabella in drive

The co-occurrence table provides the numerical foundation of the entire network. It reveals which pairs of keywords appear most frequently together and therefore explains many of the patterns observed in the visualisations. Notable examples include the extremely high co-occurrences between *AI* and several environmental keywords, such as *Rain*, *Heatwaves*, *Floodings*, and *Extreme weather events*. Strong connections also appear between scientific concepts like the *Greenhouse Effect* and *CO2*, which co-occur 81 times. These values confirm the dominant themes of the discourse and support the central positions observed in the network graphs.

Node Metrics (node_metrics.csv)

			ss	
			239789196	623188404
			197189283	14285714
ather events			2015182882	23529411
			42028985505	66666667
e Effect			439237091	66666666
				42857143
s			0724637681	42857143
			095865488	23076923
ming			7836125227	92307693
			988267771	63636364
			166321601	84848485
aster			752242926	63636363
Energy			727335467	66666667
sumption			557625949	77777777
			68115942	88888888
			449275362	14285714
y			875713659	57142857
gy			404040404	
on			258610954	33333333
ies			6038647342	33333333
nsition			525315267	
			524505929	66666666
			5740008784	
icles			1884057974	

The file containing node-level metrics summarises the specific structural role of each keyword within the network. Although the file is not shown directly here, its values are reflected in the distributions. Keywords with high degree and strength, such as *AI* and *Heatwaves*, also tend to have high centrality and clustering scores, which reinforces their status as influential topics. Meanwhile, peripheral nodes exhibit lower values, confirming their limited role in linking the larger thematic areas.

Community Assignment (community_assignments.csv)

sumption	
ather events	
ies	
aster	
Energy	
y	
gy	
on	
icles	
nsition	
s	
ming	
e Effect	

The community detection results identify groups of keywords that often appear together. These clusters correspond to meaningful thematic areas in the climate conversation. For example, one cluster is dominated by extreme weather and natural events, including *Rain*, *Floodings*, *Heatwaves*, and *Extreme weather events*. Another includes scientific and emission-related terms such as *CO2*, *Emissions*, and *Greenhouse Effect*. A

third group gathers technological and policy-related themes like *AI*, *Net Zero*, and *Clean Energy*. These communities show how online users naturally organise climate discussions into related conceptual domains.

Global Network Metrics (global_metrics.csv)

nodes	
edges	
gree	
ength	
ity	
tering coefficient	
communities	
components	
th length	

The global metrics summarise the main features of the network as a whole. The high density and high global clustering coefficient indicate that the graph is strongly interconnected and that many nodes are linked through mutual connections. A relatively small diameter suggests that any two concepts are separated by only a few steps, which is typical of integrated and cohesive networks. These results confirm that climate discussions on Bluesky tend to connect multiple themes rather than treating them as isolated issues.

Dataset of Collected Posts (bluesky_posts_complex.csv) → tabella in drive

The final dataset contains all posts from which the network was built. It includes the text of each post, its sentiment score, and the keyword that triggered the search, along with author and date information. This file ensures transparency and makes it possible to trace each network connection back to the original content

OBSERVATION WITH DIFFERENT COUNTRIES AND 20 KEYWORDS

General Analysis of the Updated Results

The updated analysis describes how environmental and energy-related concepts are connected in a smaller, more selective portion of the data extracted from Bluesky. Unlike the previous, larger network that included AI and many technology-related terms, this new dataset focuses specifically on climate, environment and energy concepts **without the AI-dominated cluster**. As a result, the network is more compact, more balanced, and less influenced by a single dominant keyword. This change allows us to see more clearly how classical environmental themes interact with each other, without being overshadowed by technological discussions.

The co-occurrence values contained in *grafo.xlsx* and *keyword_network_edges.txt* show that the strongest connection in the new dataset is between **Floodings** and **Rain**, with 69 co-occurrences. Other important pairs include **Extreme weather events** and **Rain** (39), **Greenhouse Effect** and **CO2** (81), **Emissions** and **Rain** (14), as well as several moderate connections involving *Extreme weather events*, *Fossil Fuels*, *Renewable Energy*, *Ecosystem* and *Global Warming*.

keyword_network_edges

These values reveal that the discussion revolves mainly around the relationship between **weather impacts**, **greenhouse gases**, and **emissions**.

The global metrics confirm this interpretation. Although the complete file is not shown, the updated network is visibly **less dense** and shows fewer extremely large hubs. The diameter remains small, suggesting that the conversation is still cohesive, but the average degree and strength values are lower than in the previous version. This reflects a system where keywords interact frequently, but where no single term dominates the whole structure.

The sentiment distribution supports this interpretation. The plot shows a clear majority of neutral posts, followed by positive and negative posts, maintaining the same pattern observed previously. This indicates that even within this smaller dataset, users continue to discuss climate-related subjects primarily in a descriptive or factual way, often reporting weather events or environmental conditions rather than expressing strong personal emotions.

Overall, the revised network depicts a conversation centred on **environmental processes, extreme weather, greenhouse gases and policy-related terms**, with a structure that is moderately dense and well connected,

but without a single dominant concept pulling all topics together. This allows for a clearer reading of thematic relations and a more classical representation of climate discourse.

Analysis of Each Output

---

1. keyword_network_circular.png — Circular Co-occurrence Network

The circular network displays all nodes arranged evenly around a circle, allowing a clear visual assessment of size and connectivity. Node sizes are proportional to degree. In this updated version of the network, the largest nodes are **Rain**, **Emissions**, **Global Warming**, **Glaciers**, **Extreme weather events**, and **Greenhouse Effect**. These nodes occupy the upper and right portions of the circle and show a high density of connections.

Because the dataset no longer contains technological terms like AI, the network appears more evenly structured. The edges are still numerous and often thick, especially between **Rain** and **Floodings**, which is the strongest relation in the dataset. The visual structure suggests that rainfall, floods and extreme weather patterns form the central part of the conversation, while other keywords—such as *Fake News*, *Normatives* and *Electric Vehicles*—appear smaller and closer to the edges, reflecting their weaker involvement in the discourse.

---

2. keyword_network.png — Force-directed Network Layout

In this layout, nodes are positioned according to the intensity of their co-occurrences. Strongly connected terms move towards the centre, whereas weakly connected ones remain at the periphery.

Here, **Rain** clearly emerges as the central node. It has strong connections with *Floodings*, *Extreme weather events*, *Energy Consumption*, *Emissions*, *Glaciers* and *Green Policies*. These relations pull *Rain* towards the middle of the graph.

A second central group involves **Greenhouse Effect**, **CO2**, **Global Warming**, and **Fossil Fuels**. These terms cluster closely together because their co-occurrence frequencies are relatively high (for example, *Greenhouse Effect* and *CO2* co-occur 81 times).

Nodes such as *Tesla*, *Fake News*, *Electric Vehicles* and *Normatives* remain on the outskirts, showing that they have fewer links and weaker interactions with the rest of the environmental vocabulary.

---

3. network_metrics.png — Distributions of Network Metrics

The updated metrics show a different structure compared to the earlier, AI-dominated dataset.

In the **degree distribution**, most nodes fall between 2 and 12 connections, with only one reaching 17. This indicates a moderately connected network in which several nodes act as important hubs but no single keyword dominates entirely.

The **strength distribution** reveals that only a small number of terms have high co-occurrence values. For example, *Rain*and *Floodings* have significantly higher strength compared to other nodes, reflecting their strong involvement in discussions about extreme events.

The **betweenness centrality distribution** shows a wide variation. A few nodes have relatively high betweenness, meaning they act as bridges between different thematic groups. These nodes likely include *Emissions*, *Extreme weather events*, *Global Warming*, and *Greenhouse Effect*, as they connect scientific explanations to actual weather-related impacts.

The **clustering coefficient distribution** is generally high, with many nodes showing clustering values above 0.6. This means that a keyword's neighbours are often connected to each other, forming cohesive thematic subgroups. This is typical for environmental discussions, which often link atmospheric phenomena, emissions and climate impacts within the same posts.

---

4. Sentiment.png — Sentiment Distribution

The sentiment analysis reflects a pattern similar to the previous dataset. Neutral posts remain the majority, with over 3,100 occurrences, followed by about 1,900 positive posts and around 950 negative ones.
This distribution suggests that the reduction of keywords and the removal of AI-related content did not significantly change the emotional composition of the dataset. Most posts remain factual or descriptive, likely discussing weather events, emissions data, environmental reports or policy updates. Positive sentiment may relate to renewable energy or environmental progress, while negative sentiment may correspond to extreme weather events or environmental damage.

---

5. grafo.xlsx and keyword_network_edges.txt — Co-occurrence Matrix

The updated co-occurrence table includes fewer extreme values and no longer contains unusually high weights linked to AI. The strongest relation is **Floodings–Rain (69)**, followed by **Extreme weather events–Rain (39)** and **Greenhouse Effect–CO2 (81)**.

keyword_network_edges

These values describe a climate conversation focused on physical impacts and their scientific explanations.

Moderate connections are found between *Extreme weather events* and *Fossil Fuels*, *Renewable Energy*, *Ecosystem*, and *Biodiversity*, which reflects an awareness of ecological and energy-related dimensions of climate impacts.

Weaker but still meaningful links include pairs such as *Normatives–Rain* (10), *Fossil Fuels–Energy Consumption* (11), and *Renewable Energy–Energy Consumption* (7), indicating occasional discussions about regulatory issues and energy systems within climate-related posts.

---

6. node_metrics.csv — Node-specific Indicators

Although the file is not shown directly, its values are reflected in the visualisations and histograms. The keyword with the highest degree and strength is **Rain**, which acts as the central connector of the network. *Floodings*, *Emissions*, *Extreme weather events*, *Glaciers* and *Greenhouse Effect* also play key roles.

Nodes with low degree and strength include *Tesla*, *Fake News*, *Electric Vehicles*, *Deforestation* and *Normatives*, which appear at the margins of the network. These nodes participate in fewer co-occurrences and therefore have limited influence on the structure of the overall discourse.

---

7. community_assignments.csv — Community Structure

The updated network appears to form **three main communities**. The first community clusters around extreme weather and precipitation, including *Rain*, *Floodings*, *Extreme weather events* and *Glaciers*. The second community includes mitigation and energy-related terms such as *Renewable Energy*, *Green Policies*, *Fossil Fuels* and *Energy Consumption*. The third community focuses on climate science concepts such as *CO2*,

Greenhouse Effect and *Global Warming*.

These communities confirm that the conversation is organised around three main areas: weather impacts, mitigation/energy systems and scientific causes of climate change.

Complex System Project

Common words(shared) : Energy Transition, Greenhouse Effect, Biodiversity, Extreme weather events,CO2,Emissions, Global Warming, Glaciers, Renewable energy,Fake News

Our words : Fossil Fuels, Energy Consumption, Normatives, Deforestation, Flooding,AI Green Policies, Rain, Electric Vehicles.

Extras word : Natural Disaster, Clean Energy, Net zero, Tesla, Heatwaves