Building a Trusted Network of Energy Experts on 2 Twitter, Through Graph Traversal Strategies and Active 3 Node Classification

Vincenzo De Leo[1]; Michelangelo Puliga[1]; Martina Erba[2]; Cesare Scalia[2]; Andrea Filetti[2]; Alessandro Chessa[1,3]

[1] Linkalab, Complex Systems Computational Laboratory, Viale Elmas, Cagliari, Italy
[2] Eni S.p.A, Piazzale Enrico Mattei, 1, 00144, Rome, Italy
[3] Data Lab, Luiss University - Viale Romania, 32, 00197, Rome, Italy
E-mail: vincenzo.deleo@linkalab.it

Abstract

In this study we analyze the Twitter (now X) Friendship Network, focusing on users relevant to the energy sector, spotlighting experts, professionals, and businesses connected as 'following'. By analyzing their connections, we identify clusters within this network, revealing how they are grouped based on their roles. We show how the natural formation of these clusters on social media platforms like Twitter, significantly impacts public discourse around energy and related critical issues such as climate change. We also highlight how the dynamical interplay of misinformation leads to the formation of polarized user groups that often result in disengagement from online discussions. These clusters define small groups with shared ways of communicating.

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

Unlike broader networks, information exchanges here are sparse, typically involving accounts set up for precise business aims. Moreover, we exploit a Machine Learning approach to detect key members in these specialized groups and to uncover how these groups stay connected. This approach let us gain insights into corporate communication on social media, offering a fresh view on professional networking. Our findings highlight how companies in the energy sector use Twitter to coordinate their activities, with key institutions playing a central role in keeping these networks organized.

Keywords: Twitter; Energy experts; Graph traversal; Node classification; Social media; Social networks; Machine learning; Network analysis

1. Introduction

In recent years the role of online social media has become of strategic importance in many business sectors, including marketing and advertising, retail, and e-commerce but also more in general for education, research, policymaking, and more. It serves as a common platform where companies, institutions, and individuals can freely express their opinions trying to impact the public discourse. However, the vast flow of information on these platforms, often coming from untrusted sources, is unranked and continuously refreshed, leading to a mix of accurate and misleading statements. Without a structured system to verify content, misinformation can easily spread, leaving news open to misinterpretation and potential misuse.

In this work, we propose a novel approach to understanding online discourse around energy and climate change, a topic of vital global importance. By analyzing Twitter networks with Machine Learning (ML) techniques, this study offers insights into how information and misinformation spread, how communities of experts influence public opinion, and how digital platforms can both hinder or help public understanding of complex issues.

ML, a subset of the broader field of artificial intelligence, involves algorithms and statistical models that allow computers to perform tasks without explicit instructions, instead relying on pattern learning and inference reasoning. It's instrumental in analyzing vast amounts of data to uncover hidden patterns or predict future events. In the context of social networks like Twitter, ML can classify nodes (individual accounts or entities) and their connections, revealing how these nodes cluster into communities or subgraphs (i.e. a portion of a graph that consists of a subset of nodes from the original graph along with all the edges that connect nodes inside that subset), based on shared interests or interactions.

Twitter fosters the exploration of the information produced in the media, through a connection mechanism that enables one user to follow others and see their posts without the need for a mutual following. This unique setup means you can have a wide audience without needing to follow everyone back, focusing on updates from chosen contacts. We tapped into this mechanism to study how energy sector professionals connect, using a method that identifies and follows these connections to map out a network of industry-related discussions, named Friendship Network. Similar to research on global health organization networks, this approach helps us to detect key individuals and groups in the energy debate, using technology to filter relevant participants for closer examination [1].

After mapping the connections between Twitter users interested in energy topics and refining the data with an ML classifier, we examined a detailed subnetwork, or subgraph, featuring thousands of nodes (users). This sub-graph was then analyzed for additional connections, particularly focusing on mentions among users. Essentially, we looked at two types of connections forming a complex, layered network, where the same users could interact in different ways. This method helped us identify distinct groups within the network, each with unique patterns of communication such as how often they tweet, mention, or retweet each other. The specific objective of this work is to explore the dynamics within Twitter communities discussing energy and climate change, employing machine learning to classify nodes and understanding how these nodes create specialized subgraphs or communities around these pivotal topics. The chosen theme related to energy and climate change was due to the critical and urgent nature of these issues. Climate change represents one of the most significant challenges of our time, impacting global ecosystems, economies and societies. The role of energy transition in mitigating climate change effects is equally crucial, necessitating broad public engagement and informed policymaking. Notably, our analysis reveals heterogeneous communication styles across these groups. For example, energy institutions play a pivotal role, typically acting as connectors among various smaller groups, while companies tend to interact less directly with each other. This approach provides a rich understanding of the discussion dynamics in social networks, spotlighting the intricate web of relationships among institutions, companies, experts, and journalists interested in energy topics. Each identified sub-group within the larger network has its unique way of engaging and sharing information, highlighting the diversity of dialogue in the digital realm of energy discussions.

2. Background

Researchers like Barreto and Alzanin et al. have shown that social media can magnify misinformation, leading to "echo chambers" where biased information circulates among like-minded users, as identified by Del Vicario et al. Despite access to original content, its meaning can quickly become distorted, affecting how easily it can be found and understood. This overflow of information, combined with natural cognitive biases, fosters misinformation and polarizes users, a phenomenon also noted by Bessi et al. in the context of platforms like Facebook and YouTube. Weng et al. explained how this leads to a competition for attention, with certain topics becoming viral as they are amplified by users with many connections, while other topics fade away unnoticed [2-7].

The intense flow of messages and alerts on social media can overwhelm users, significantly reducing their likelihood to engage in conversations. Research, like the study by Gunaratne et al. highlights this by showing that when users receive 30 or more notifications an hour, they often withdraw from discussions. This phenomenon, known as "information overload." has been linked to reduced communication quality, ineffective marketing efforts, and lower productivity and performance, potentially leading to significant economic impacts, with losses estimated at 650 billion USD in 2007 [8, 9]. The term "death by information overload" has been used to describe this overwhelming state [1].

In essence, the flood of information, alongside deliberate algorithmic shaping of social media into echo chambers, and the creation of polarized user bubbles, makes accessing high-quality content challenging, thus dilute the quality of public discourse[4]. In this environment, identifying and amplifying voices of authority and communities that share common goals and engage with quality content becomes crucial for cutting through the noise of these vast networks.

The methodological approach of using machine learning for node classification and graph analysis finds resonance in the literature. Similar techniques have been employed in studies aiming to understand the structure and dynamics of social networks, including the generation of realistic synthetic social networks [10, 11]. These approaches underline the value of advanced computational methods in dissecting the complexities of online social interactions. Moreover, the findings underscore the role of key influencers and organizations within the Twitter communities discussing these themes, which is a common thread in the literature on social network analysis and community detection [12]. Understanding these roles is crucial for strategizing more effective communication and engagement on environmental issues, as indicated by studies that explore the impact of influencers in disseminating scientific knowledge and countering misinformation [13]. In the digital world of social media, identifying relevant groups and conversations is a challenge intertwined with navigating the vast networks of users. Community detection, as outlined by Fortunato and Rani et al, involves finding groups within these networks, defined by specific criteria such as shared interests or viewpoints [14, 15]. For example, Williams noted that distinct communities form among users engaged in debates on topics like climate change, identifiable through their interactions and shared references [16]. The key lies in defining how connections are made, whether through mentions, shared hashtags or mutual followers, 88 to uncover these "communities of practice" (organized groups of people with a common interest in a specific technical or business domain that regularly collaborate to share information, improve their skills, and actively work on advancing their knowledge of the domain) [2]. Modern network analysis uses multilink or multilayer graphs to represent these intricate relationships, offering a detailed view of how various user groups communicate and connect. This approach allows for a nuanced understanding of social media landscapes, highlighting the importance of identifying not just any connections but the right ones that reveal the structure and dynamics of user communities. The study's results, focusing on the dynamics within Twitter discussions around energy and climate change, reveal intricate community structures and influencers driving the discourse. Comparing these findings with similar research uncovers a nuanced landscape of online conversation regarding climate change.

One significant aspect is the growing polarization around climate change on social media, as highlighted in recent analyses [17]. For instance, a study on Twitter conversations related to the Conference of the Parties (COP) climate summits found an increase in polarization, with user ideologies showing more distinct clusters by COP26 compared to earlier summits (Nature Climate Change). This polarization is characterized by the division into majority and minority actors, with the majority leaning towards pro-climate views. This phenomenon underlines the ideological divide and it's deepening over time, which parallels the findings of the study in the discussion, emphasizing the role of key influencers and the ideological segmentation within Twitter communities focused on climate change and energy.

Moreover, the examination of climate change sentiment on Twitter using a multi-dimensional approach, including stance, sentiment, and topics, underscores the complexity of public opinions on climate change [18]. This approach, contrasting with more traditional sentiment analysis, enables a more comprehensive understanding of how climate change is perceived across different global regions, aligning with the objectives of the discussed study to map out specialized communities and understand their discourse patterns.

Lastly, the influence of fossil fuel firms in reframing online climate and sustainability communication is another critical point of comparison [19]. The strategic communication by these firms, alongside NGOs and IGOs, shapes the public discourse on sustainability and climate change, revealing the power dynamics at play within online conversations. This mirrors the discussed study's insights into how different stakeholders, including companies and institutions, influence Twitter discussions on energy and climate change.

In our research, we're looking at social networks in a new light, moving beyond traditional Community Detection

(CD) that focuses on how tightly-knit groups of accounts are. Instead, we're diving into the attributes of individual accounts, categorizing them as we go through the web of connections. This process, known as node classification, allows us to sort accounts based on their specific characteristics.

Furthermore, we're incorporating the concept of multiplex networks into our analysis to understand better how the same accounts interact across different contexts—like how someone might be connected in one way through public mentions and in another through direct friendships [20]. Multiplex networks, which capture these layered interactions, have been instrumental in analysing complex systems in a variety of fields, from biology to societal structures, and even in the study of transportation and infrastructure [21-23].

The multiplex network approach we used in our study adds a layer of depth by tracking how individuals engage across different types of communication. This method expands traditional ways of measuring importance (centrality) in a network by considering both direct connections and the broader context of those connections across different layers of interaction. For instance, the concept of multi degree reflects how well-connected someone is within each layer of the network. By evaluating centrality across multiple layers, we can better understand an individual's role in spreading information through various channels.

When it comes to identifying communities within these multiplex networks, our goal is to uncover cohesive groups that remain consistent across different forms of interaction [24]. This process reveals the network's underlying structure by considering the full spectrum of connections. Additionally, using algorithms that evaluate how segmented (modular) these networks are, we gain insight into how information flows within them. Generally, networks with higher segmentation see slower information spread, highlighting the intricate balance between connectivity and the dynamics of information transmission.

3. Methods

The innovation introduced in this work regarding the creation of the skilled user network is related to the fact that the generation of the network of interactions is not due to any mentions, retweets, or replies present in the texts of tweets downloaded from Twitter through a filter based on a Bag of 139 Word (BOW) of terms associated with the topic in question, but it is related to a seed of verified users, known as players who play a top-level role in the issues related to the topic in question. Starting from the seed of verified users, then, through the Twitter API, the descriptions of the first-level Friends were downloaded; through manual classification work, the descriptions were classified in terms of whether they were relevant or not to the topics of the discussion in question. This classification was used to train a supervised ML model to further classify, this time automatically, the Friends descriptions of the first-level users (only those classified as actors participating in discussions related to the topics of the discussion at hand), correspond to the second level Friends of the users belonging to the network seed. The final network is therefore composed of three levels of users: the seed (in the order of tens), their first-level Friends manually classified (in the hundreds) and their second-level Friends classified automatically by the model ML (in the order of thousands).

Once the network described above was obtained, the timelines of the discussions on the Twitter social network of users of all three levels were then downloaded, to be able to analyse the discussion topics, study the discussion communities from the topological properties of the graph, and the study of the texts of the tweets, discover the volume peaks of the various comparisons between users in the time series and more.

More details about methods can be found in the supplementary material [3].

4. Data Preparation

To obtain a network of high-quality users on a specific topic of discussion, it is necessary to choose the list of users of the initial list, the seed of the network, shown in the supplementary material, with the best possible criteria available [4].

Given the context of the discussion, it was decided to extract the starting seed from a list of energy sector companies in Italy; the list was obtained through queries made with the Bloomberg terminal which returned a list of 32 companies for which we made queries to the Twitter API to obtain user descriptions defined on the social network. From the descriptions above, it has been possible to manually perform a first filter that allowed the users of the seed to be selected based on whether or not there were terms related to energy and the green economy in their user descriptions defined on Twitter; this preliminary operation led to a list of 14 users, shown in the supplementary material, belonging to the network seed [5].

Having available the list of Twitter accounts of users belonging to the seed of the network that is being built, query codes of the Twitter API have been activated to perform get_friends_list() operations to obtain the list of all Friends of the users of the seed, obtaining a list of 740 unique users (see supplementary material for details) [6]. To manually classify the 740 users identified as described above, the descriptions of the users who make up the seed, shown in the table above, were taken into consideration, to extrapolate from these the terms that were considered most relevant to the object of the research of this project; some of these terms are, for example, energy, gas, methane, fuel, etc. This led to obtaining a selection of 210 users that wrote posts in the English language and that are classified as "related to energy discussions" [7].

Having available the list of Twitter accounts of the first-level Friends selected and classified as positive, query codes of the Twitter API have been activated to carry out get_friends_list() operations to obtain the list of all the Friends of the First-level friends selected and classified as positive, obtaining a list of second-level Friends consisting of more than 60 unique users.

Once the datasets of the descriptions classified manually have been properly defined, this has been used to develop an ML model able to automatically classify the list of second-level Friends. To find the most accurate classifier for our purposes we performed a Grid Search analysis over the following ML algorithms:

By comparing all the results obtained for all the models with the variation of all the parameters within the ranges of allowed values, it turned out that the most performing model from the point of view of accuracy is the Decision Tree Classifier ,for which accuracy of 0.80 was obtained with the following values of the parameters: criterion='gini', splitter='best' min_samples_split=0.02, min_samples_leaf=0.04, max_features=0.55, random_state=581.

The model obtained with the techniques described in the previous paragraph was then applied to the second-level Friend dataset to obtain the automatic classification of the Friends based on the content of their descriptions set in their Twitter social network profile.

The classification results have been that the number of users classified as positive has been equal to 4620. From a manual sample check on the results, it has been found that the automatic classification led to statistically significant results relative to the quality of the selected users. The statistical analysis conducted to assess the quality of the selected users involved 189 drawing a random sample of 100 users from those chosen by the ML model. This analysis included manually reviewing the account descriptions and the texts of their most recent tweets (the last 5)

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

taken from their timelines. The outcome of this analysis revealed that 95% of these descriptions were perfectly relevant to the expected profiles, 4% of the descriptions were somewhat ambiguous and could not be confidently classified, and 1% of the descriptions did not match the expected profiles **(Table 1)**.

Consequently, the final dimensions of the dataset classified as positive, considering the seed together with the first level Friends manually classified as positive and the second level Friends automatically classified as positive with ML techniques, are equal to 4844 not unique users (14 from the seed, 210 from the first level Friends and 4620 from the second level Friends) More details about data preparation can be found in the supplementary material [8].

Table 1. Tested models descriptions.

Model Name	Description
K Neighbors	The K Neighbors model belongs to the category of instance-based learning, where Predictions are made based on the majority class of its k nearest neighbors in the feature space.
SGD	The SGD Classifier model is distinguished by its ability to efficiently handle large-scale datasets due to its stochastic gradient descent optimization, offering faster training times compared to traditional batch gradient descent algorithms.
Decision Tree	The Decision Tree model stands out for its ability to partition the feature space into a hierarchical structure of decision nodes, allowing for interpretable and easily visualized decision-making processes.
Gradient Boosting	Gradient Boosting is a machine learning model known for its iterative optimization approach, wherein subsequent models are trained to correct errors made by the previous ones, leading to enhanced predictive performance compared to individual models.
Random Forest	The Random Forest model stands out due to its ensemble approach, combining multiple decision trees to enhance prediction accuracy and robustness.
Multi-Layer Perceptron	The Multi-Layer Perceptron model stands out for its ability to handle complex non-linear relationships and patterns in data due to its deep architecture with multiple hidden layers.

5. Networks Construction

The datasets in the previous paragraph were used to build the interaction network between users. The criterion used to relate users to each other is the status of Friend, that is, the fact that a user appears in the friend list of another user. By applying this criterion, a network has been created with 4844 unique users and 13806 links. In addition to this Friendship Network, we downloaded the timelines of all these users and built another network based on the mentions and the retweets in the discussions of these users, obtaining a Mentions Network with 2094 nodes and 32824 links in the greatest cluster of connected users (starting nodes were the same of the Friendship Network, but using this approach can generate several small isolated clusters of users that we ignored in our study since they do not contribute to interesting results).
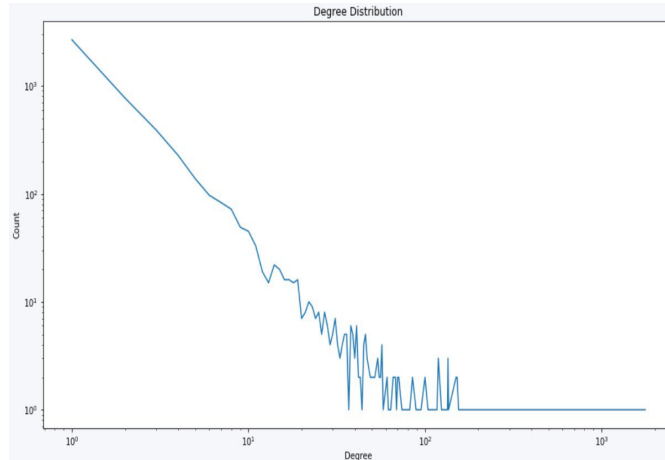
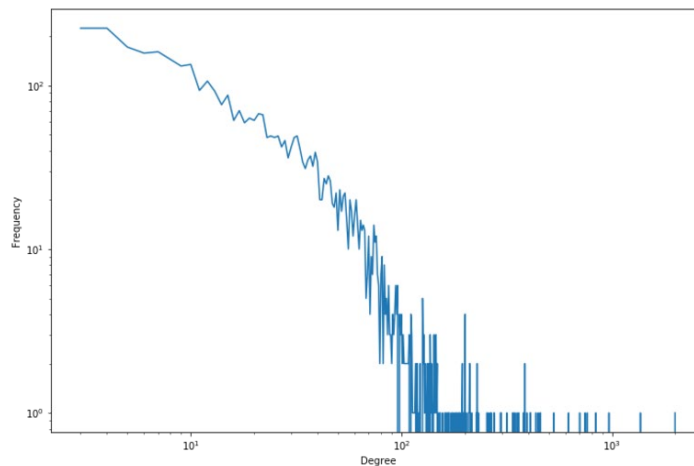**Figure 1.** Degree distribution of friendship network.

Figure 2. Degree distribution of mentions network.

This allowed us to compare the two networks in terms of centrality measures and communities and evaluate the differences and strengths of both solutions. For example, it can be seen that the degree distribution of the Friendship Network, shown in **Figure 1** is more related to a power-law trend than the one of the Mentions Network, shown in **Figure 2** with the consequent presence of hubs, as discussed in the Results section. More details about network construction can be found in the supplementary material [9].

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

5.1 Users Classification Based on Company, Institution, and Expert User Labels

To better understand the role of the hubs we created another classifier this time targeting 4 node categories (institutions, companies, experts, and others).

With manual training, we taught the ML systems to recognize these categories and apply the labels to all nodes. The data shows that the majority of the hubs belong to the "institution" category as demonstrated by the higher average degree shown in **Table 2**.

This data has been prepared using a Stratified K-Folds cross-validator to provide train/test indices to split data into train/test. The choice of this particular cross-validator has been mandatory because of the inequality between the classes in the dataset.

The classifier has been developed using a Naive Bayes classifier since it is particularly suited for imbalanced data sets.

Users have been classified using their description as the main feature of the classifier. After a community detection analysis, we have discovered that in the Friendship Network, there are dedicated communities for each of the selected categories (Institutions, Companies, etc.) confirming that they have well-separated roles and ways of communication.

Also, the Mentions Network is strongly influenced by the role of institutions (see supplementary material for details) [10]. This result confirms that, whatever the way we make the network, this category is the one that keeps strong the interactions between the other players and, therefore, the one that drives the discussion.

6. Results

The two networks, Friendship and Mentions, were analyzed individually with the classical complex network methods, and with the modern multiplex methodology. **Figure 3** shows a representation of the Friendship Network while **Figure 4** shows a representation of the Mentions Network; in both, the hubs are highlighted, and detected communities are distinguished based on the color of the nodes. The basic idea is that the acquaintances among the nodes can be expressed both by the friendship relationship and the mentions. While the friendship relation is a direct unweighted link, the mention is a direct and weighted relationship (counting the number of times node A mentions node B); the Mentions Network can therefore be used to purge the weak links and to find the users that engage in frequent conversations, performing in this way an active node classification [11].
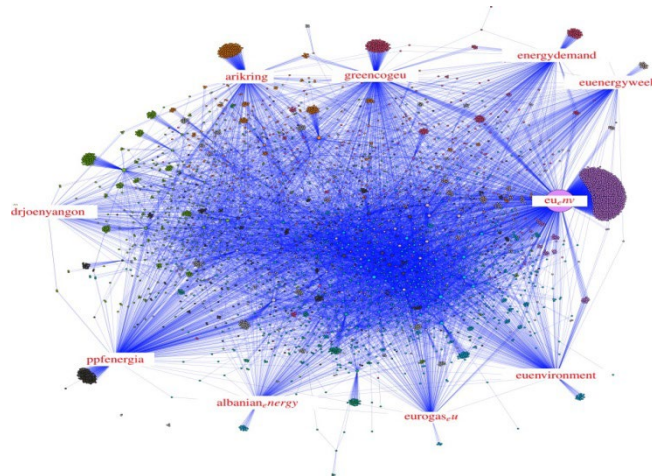
Figure 3. The Friendship Network.

In summary, we recovered 5k nodes with 14k links for the Friendship Network and 2k nodes with 32k links for the Mentions Network; interestingly, the overlap between the links of the two networks is not large, only 5% of the links is in common in the intersection of the two networks, indicating that being a follower is not equivalent, at least in the community of practice of the Energy, to engage in conversations and cite other users. An important tool to study the networks is represented by the degree distribution. Fitting the tail of the degree distribution with a power law reveals the presence or absence of a power law distribution of the degree.

$$y = x^{-\alpha}$$

As it is explained in, the fact shown in the Network Construction section (i.e. the different power law trends in the two networks) led to broad implications for the structure and dynamics of complex systems [25]. A strongly scale-free structure is empirically rare while for most networks, log-normal distributions fit the data as well or better than power laws, so this means that the Friendship Network is an ideal benchmark to characterize the energy discussions. This has profound implications for the structure of the networks that in the power law case are dominated by large hubs, nodes having a large number of friendship connections or citations. These hubs are the main originators of information in the network and those that ensure that a piece of news or a message is delivered to the final recipient. To better understand the role of the hubs we created another classifier this time targeting 4 node categories ("institutions", "companies", "experts", and "others"). With manual training, we taught the ML system to recognize

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

these categories and apply the labels to all nodes. The data in **Table 2** show that the majority of the hubs belong to the "institution" category and have a high average degree. The prevalence of institutions among the most connected nodes implies that this network is working mostly due to the role of the institutions that act as a hub connecting all companies and expert users. A list of the most connected nodes is presented in **Table 3.** Another possible analysis is the community detection of the single multiplex network composed of friendship and mention links, as shown in **Figure 5**. The procedure to find communities in the multilayer or multiplex network has been described in the supplementary material [12].
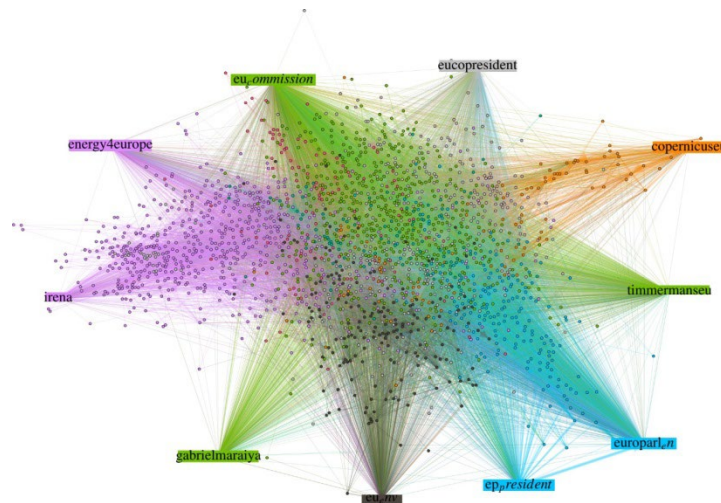

Figure 4. The Mentions and Retweets Network.

Table 2. The users classified by role with their average measures.

Role	Avg Followers	Avg Friends	Avg Statuses	Avg Degree
Company	13004.6	1090.8	4296.7	6.5
Expert User	10022.9	1503.2	8417.6	4.3
Institution	17148.8	1488.1	7200.6	12.0
Other	21210.4	1636.2	9600.0	5.6

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

The results are interesting and presented by summary statistics in **Table 4** while the density of links is in **Table 5.** The community (labeled as "6"), with 1062 nodes, has a high average degree, which is a signature of the presence of many institutions, (21% of the total nodes), the same community is also balanced Expert Users and Companies. The community represents the dense core of companies and institutions that are closer in business and social networks. The other communities appear to be less connected (lower average degree). Notice that there is no clear correlation in this table between the social network popularity measures and the composition of the community in terms of roles. Interestingly there is also a large community (labeled as "3") having the largest fraction of expert users; this community has indeed a low average degree. Experts, in other words, are less likely to connect and form clusters instead they compete for the same space in terms of user attention, likes, and so on.

These results align perfectly with others present in the literature on the topic of energy and climate change [26, 27]. Previous works like Cortés P.A. and Quiroga R. (2023) compared diverse platforms through which climate change discussions are propagated and highlighted the unique role of Twitter in facilitating real-time and widespread public engagement. Pearce W. et al. **(**2014) study complements the methodology of identifying active nodes within communities of practice, described in this work, suggesting that significant climate events or publications can act as catalysts for community formation and engagement on Twitter.

Table 3. The most important nodes by degree and their role.

Twitter Account	Degree	Reference
EU_ENV	1771	Institution
GreenCogEU	530	Institution
arikring	411	Expert User
EnergyDemand	292	Institution
EUEnvironment	288	Institution
euenergyweek	239	Institution
DrJoeNyangon	214	Expert User
Albanian_Energy	181	Company
Eurogas_Eu	168	Company
Climate_Rescue	152	Other
EU_ManagEnergy	152	Institution
Ed_Crooks	151	Expert User

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)
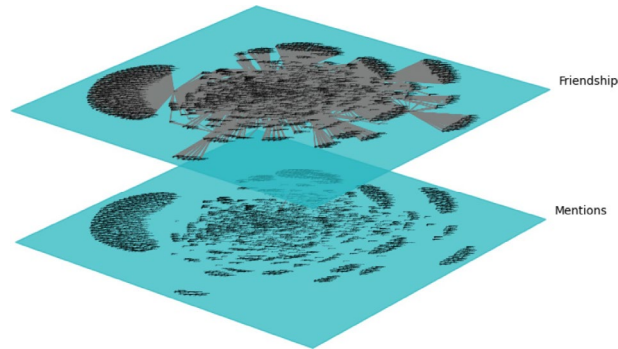

Figure 5. The Multiplex Network.

Table 4. Community detection on multiplex network, with the social impact of each community with at least 10 nodes each.

Community	Avg Foll.	Avg Friends	Avg Tweets	Avg Deg.	Expert %	Comp. %	Instit. %	Oth. %	Nodes
0	15991.0	1724.3	10211.0	6.6	28.9	25.2	12.2	33.7	5273
3	13652.8	932.0	3600.8	3.5	38.5	5.6	15.1	40.9	1475
6	15244.2	1165.7	4262.9	12.3	27.9	26.6	20.9	24.6	1062
16	17942.6	1454.8	7368.2	3.8	34.5	3.7	9.1	52.7	804
7	25605.0	849.7	3817.6	3.5	21.5	9.1	23.8	45.7	265
8	4666.2	1233.8	4171.8	2.4	11.4	6.3	13.9	68.4	158
30	1284.4	1810.0	1781.3	2.2	44.4	11.1	11.1	33.3	18
18	8014.8	764.5	4662.2	3.2	16.7	0.0	0.0	83.3	12

Table 5. The density of links between the classified nodes allows repetition as we have two equivalent layers of the multiplex network.

obj_start	obj_stop	Counts	Cross Density
Institution	Institution	2146	0.005441
Institution	Other	3762	0.003562
Other	Other	9328	0.003297
Institution	ExpertUser	2853	0.003243
Institution	Company	1559	0.002783

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

Other	Institution	2836	0.002685
Other	ExpertUser	6251	0.002653
Company	Institution	963	0.001719
Company	Company	1324	0.001664
Other	Company	2027	0.001351
ExpertUser	ExpertUser	2638	0.001344
Company	Other	1865	0.001243
ExpertUser	Other	2823	0.001198
ExpertUser	Institution	991	0.001126
Company	ExpertUser	1242	0.000994
ExpertUser	Company	816	0.000653

The results provided by Dellmuth L. and Shyrokykh K. (2023) extend the potential of Twitter data, discussed in our research, to inform not only community dynamics but also policy development and governance strategies in the context of climate change. Fownes J. R. et al. (2018) already underlined years ago the platform's role in shaping public opinion and the importance of understanding the network dynamics that facilitate these discussions, as we mentioned in this document [28-30] . Erokhin, D. and Komendantova, N (2023), and Kolic et al. (2022) both provided insights into the structure and content of climate change discussions on Twitter: their analyses, which included the examination of controversial discussions and the application of unsupervised methods to quantify conversation structures, present methodologies, and findings that enrich the understanding of how Twitter serves as a battleground for climate change narratives [31, 32]. Torricelli et al. (2023) illustrated how external events influence online discourse, which aligns with the idea of active node classification by showing that specific events can activate or intensify discussions within identified communities of practice. These studies, alongside our approach of active node classification, underscored the complexity of Twitter as a space for both consensus-building and controversy around climate change. Moreover, these results extend previous literature on active node classification and multiplex network analysis [33-36]. For example, compared with Laishram, R. et al. (2019), our study shares a focus on multiplex networks but stands out for its application of machine learning to classify active nodes, enriching the understanding of community formation and interaction patterns; similar to Himelboim, I. et al. (2017), it delves into the structural analysis of Twitter networks but it distinguishes itself by identifying active participants within communities, offering insights into the roles individuals play in disseminating information and shaping discourse; while Abitbol, J.L. et al. (2021) and Orhan, Y.E. et al. (2023) explore Twitter interactions from socioeconomic and political perspectives, respectively, this study complements these approaches by highlighting the practical applications of identifying communities of practice, especially in professional and thematic areas; the focus on active node classification in identifying communities of practice integrates with Li, J. and Chang, X. (2023), emphasizing the proactive identification of influential nodes within communities that could be pivotal in misinformation campaigns or fact-checking initiatives; both studies utilize Twitter data but apply different frameworks and analyses to achieve their respective goals, showcasing the diverse applications of social media

analysis in understanding both community behaviors and strategies for information verification. Hanteer O. and Rossi L. (2019) share a thematic approach with our work, yet our emphasis on active node classification within a multiplex framework offers a more dynamic view of community interactions and information flow, focusing on user engagement and network activity [37,38].

The application of multiplex network analysis in this paper is a sophisticated approach that considers the complexity of social interactions on platforms like Twitter. Studies like Kivelä et al. (2014) have emphasized the importance of multiplex networks in understanding the multifaceted relationships and interactions in complex systems, suggesting that single-layer network analysis might overlook crucial dynamics. The community detection approach, especially within multiplex networks, has been extensively studied. This work introduces an innovative method for analyzing Twitter data to identify and analyze communities based on their interactions and topics of discussion, that integrates and extends the standard approach described in Zhai X. et al. (2018). This approach leverages ML for classifying nodes and employs graph traversal within a multiplex network framework, emphasizing the importance of different types of interactions (such as mentions and friendships) in understanding community dynamics. Much like the work of Fortunato (2010), this paper's methodology highlights the significance of identifying sub-communities or groups within larger networks based on their interactions and connections; the ability to detect these communities allows for a deeper understanding of how information flows and how cohesive groups form around specific topics[39-42]. The emphasis on institutions as key nodes within the network aligns with findings in the literature on the role of influential users or opinion leaders in shaping discourse on social media. For example, studies like Cha et al.(2010) have explored the concept of influence on platforms like Twitter, identifying that users with the most followers are not necessarily the most influential in terms of retweets or mentions, which is a nuanced view also reflected in this paper's analysis. Finally, the notion that the structure of social networks affects the spread of information and user engagement can be related to the findings of Gonçalves et al. (2011), who discussed the saturation of information on social media and its impact on user attention and content dissemination [43].

7. Conclusions and Future Research

The work described in this manuscript is significant for several reasons: i) it offers a sophisticated method to analyze social networks, specifically Twitter, by identifying communities of practice; this contributes to a deeper understanding of how individuals and groups interact within the platform, especially around specific topics or interests; ii) by focusing on active node classification and employing graph traversal strategies within a multiplex framework, the research provides insights into how information flows and evolves within and across different communities on Twitter; this is crucial for studying dissemination patterns, including how knowledge, beliefs, and misinformation spread; iii) the methodology helps pinpoint active nodes or key influencers within communities, which can be invaluable for various stakeholders, including marketers, policymakers, and social scientists, to understand who drives conversations and how they might influence the broader network; iv) understanding communities of practice on social media can inform more effective communication strategies for organizations

looking to engage with specific audiences, influence public opinion, or market their products. It can also aid policymakers and activists in crafting messages that resonate with their intended audiences.

Graph traversal methods benefit from node active classification made by machine learning techniques. Guessing the role of a node in advance is a way to save computing time, avoiding costly explorations on the node neighbors, and, at the same time, it is a method to improve the effectiveness of subgraph reconstruction. With this in mind, this work explores how a small seed of nodes belonging to the world of energy generates a large graph of energy experts, institutions, and other companies that form a Friendship Network on Twitter. We then complemented the analysis by looking for the multi-layer network of the same users when they mention each other in tweets over the space of several years: the Mentions Network is the second layer inter-connected to the friendship one. Community detection, and node popularity measures (such as centrality, degree, assortativity, and Twitter statistics) together with specialized ML classifiers allow for further reconstruction of communities of practice by checking if a node is an institution, a company, or an energy expert. This fine-grain ML classification is useful to understand the functioning of the overall multi-layer network and to reveal how such a network is maintained by the institution nodes that create the majority of the re-shared contents and are actively mentioned by companies and experts. In this network, the peculiar behavior of the institutions is the glue that maintains the network functioning, and the information flows from them to the companies and the experts. This feature indicates that the institutions are using Twitter to communicate their decisions, new regulations, suggestions, and news, experts are commenting on the novelties, and companies reacting to the announcements with new business strategies. Finally, companies are using Twitter for marketing campaigns, events, and general promotion; they do not mention each other and tend to not participate in discussions with the general public, a communication style that is followed by experts. The approach revealed, with numerical indicators how the network is shaped by the role of each node, and how the energy domain is sensitive to regulations and news from official institutions.

From a methodological point of view, we introduced a hybrid approach of machine learning and network techniques that uses the best of both worlds. In fact from one side, we were able to understand concepts such as node popularity via centrality measures on all layers of the multi-layer network for instance; from the other thanks to the classification of nodes according to their role we were able to recognize the communication style. This, in turn, empowers researchers to discover the communities of practice and study the communication strategies of each group of users. Finally, the multi-layer network theoretical framework can be considered the best and most appropriate methodology for community detection, with a strategy that considers also the temporary links formed by node mentioning. It is worth remembering that, despite the friendship relation, not all nodes engage in discussions. This finding, at first, is rather surprising as one imagines a social network as a place of discussion, while our analysis showed that many nodes (such as companies) are, instead, using Twitter for its informative abilities. As stated previously they do not engage in discussions, they consume information from other nodes and produce news about their campaigns, and business achievements. Future research will try to make this technique more universal, and more "ready to use", as at this moment it requires a multi-step phase of classification, graph traversal, and filtering.

Ideally, an intelligent system will start classifying users with minimal human intervention, just looking at their contents and profiles. We foresee the utility of this solution, but also the challenge of a methodology of automatic node classification, that is mostly "unsupervised" in the detection of the groups.

The practical application of the study lies in its ability to analyze and map out the dynamics within Twitter discussions related to energy and climate change, using ML techniques. By identifying key influencers and how communities form around discussions of energy and climate change, stakeholders can develop more effective communication strategies. This could involve tailoring messages to specific audiences or engaging directly with influential users to amplify important messages about climate action and energy transition. Understanding the public discourse on social media can provide policymakers with insights into public opinion and concerns. This can help in crafting policies that are more aligned with the public's understanding and expectations, thereby increasing the effectiveness and public acceptance of these policies. The study's methodology can be used to identify misinformation spread within online discussions on climate change and energy. By mapping out the nodes and subgraphs where misinformation proliferates, efforts can be directed to counteract these narratives with factual information, thus improving public knowledge of these critical issues. For researchers and advocacy groups, understanding the landscape of online discussions on energy and climate change can aid in identifying gaps in public knowledge, emerging trends, and areas of significant debate. This can guide future research directions and advocacy campaigns to address these gaps and leverage areas of public interest. Companies, especially those in the energy sector, can use insights from the study to understand public sentiment towards different energy sources, sustainability practices, and climate change policies. This can inform corporate social responsibility strategies, communication plans, and product development to align with public concerns and values. Overall, the importance of this study extends across various domains, offering a valuable tool for engaging with and influencing the discourse on energy and climate change in the digital age. By leveraging social media analytics, stakeholders across the board can better navigate the complex landscape of public opinion and action on these pressing global issues.

8. Author Contributions Statement

M.P., A.C. and V.D.L. performed the analysis, M.E. C.S. and A.F. provided support for elaboration, discussion, and theoretical framework in the business processes.

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors.

Informed Consent

This article does not contain any studies with human participants performed by any of the authors.

Figure Permissions

This article does not contain any images that is subject to copyright or ownership rights.

Competing Interests
The authors declare no competing interests.

Data Availability
The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Lopreite, Milena, et al. "A social network analysis of the organizations focusing on tuberculosis, malaria and pneumonia." *Soc Sci Med* 278 (2021): 113940.

[2] Barreto, Ana Margarida. "The word-of-mouth phenomenon in the social media era." *Int J Mark Res* 56.5 (2014): 631-654.

[3] Alzanin, Samah M., & Aqil M. Azmi. "Detecting rumors in social media: A survey." *Procedia Compute Sci* 142 (2018): 294-300.

[4] Del Vicario, Michela, et al. "The spreading of misinformation online." *Proc Natl Acad Sci* 113.3 (2016): 554-559.

[5] Eismann, Kathrin. "Diffusion and persistence of false rumors in social media networks: implications of searchability on rumor self-correction on Twitter." *J Bus Econ* 91.9 (2021): 1299-1329.

[6] Bessi, Alessandro, et al. "Users polarization on Facebook and Youtube." *PloS One* 11.8 (2016): e0159641.

[7] Weng, Lilian, et al. "Competition among memes in a world with limited attention." *Sci Rep* 2.1 (2012): 335.

[8] Gunaratne, Chathika, et al. "The effects of information overload on online conversation dynamics." *Comput Math Organ Theory* 26 (2020): 255-276.

[9] Roetzel, Peter Gordon. "Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development." *Bus Res* 12.2 (2019): 479-522.

[10] Shirzadian, Pouyan, et al. "A time evolving online social network generation algorithm." *Sci Rep* 13.1 (2023): 2395.

[11] Davies, Alex, & Nirav Ajmeri. "Realistic Synthetic Social Networks with Graph Neural Networks." *arXiv preprint arXiv:2212.07843* (2022).

[12] Zhu, Yanqiao, et al. "A survey on deep graph generation: Methods and applications." *Learn Graphs Conf,* PMLR, (2022).

[13] Liu, Zijing, & Mauricio Barahona. "Graph-based data clustering via multiscale community detection." *Appl Netw Sci* 5.1 (2020): 3.

[14] Fortunato, Santo. "Community detection in graphs." *Phys Rep* 486.3-5 (2010): 75-174.

[15] Rani, Seema, & Monica Mehrotra. "Community detection in social networks: literature review." *J Inf Knowl Manag* 18.02 (2019): 1950019.

[16] Williams, Hywel TP, et al. "Network analysis reveals open forums and echo chambers in social media discussions of climate change." *Glob Environ Change* 32 (2015): 126-138.

[17] Falkenberg, Max, et al. "Growing polarization around climate change on social media." *Nat Clim Change* 12.12 (2022): 1114-1121.

[18] Effrosynidis, Dimitrios, et al. "Correction: Exploring climate change on Twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters." *PLoS One* 18.10 (2023): e0293151.

[19] Debnath, Ramit, et al. "Do fossil fuel firms reframe online climate and sustainability communication? A data-driven analysis." *Npj Clim Action* 2.1 (2023): 47.

[20] Horvat, E.-A., & Zweig, K. A. (2014). Multiplex networks. In R. Alhajj, & J. Rokne (Eds.), *Encycl Soc Netw Anal Min* (pp. 1019-1023).

[21] Kinsley, Amy C., et al. "Multilayer and multiplex networks: An introduction to their use in veterinary epidemiology." *Front Vet Sci* 7 (2020): 596.

[22] Xiong, Xiaojin, et al. "Coevolution of relationship and interaction in cooperative dynamical multiplex networks." *Chaos: Interdiscip J Nonlinear Sci* 34.2 (2024).

[23] Shanmukhappa, Tanuja, et al. "Multi-layer public transport network analysis." *2018 IEEE Int Symp Circuits Syst (ISCAS)*. IEEE, 2018.

[24] Magnani, Matteo, et al. "Community Detection in Multiplex Networks." 2021, arXiv:1910.07646.

[25] Broido, Anna D., & Aaron Clauset. "Scale-free networks are rare." *Nat commun* 10.1 (2019): 1017.

[26] Cortés, Pablo A., & Riva Quiroga. "How academic research and news media cover climate change: a case study from Chile." *Front Commun (Lausanne)* 8 (2023).

[27] Pearce, Warren, et al. "Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report." *PloS One* 9.4 (2014): e94785.

[28] Dellmuth, Lisa, and Karina Shyrokykh. "Climate change on Twitter: Implications for climate governance research." *Wiley Interdiscip Rev: Clim Change* 14.6 (2023): e848.

[29] Fownes, Jennifer R. et al. "Twitter and climate change." *Sociol Compass* 12 (2018).

[30] Erokhin, D., & N. Komendantova. "Climate Discussion on Twitter." *Climate Mis/Disinformation in a Post-Factual Era*, 2023, 447.

[31] Kolic, Blas, et al. "Quantifying the structure of controversial discussions with unsupervised methods: a look into the Twitter climate change conversation." *arXiv preprint arXiv:2206.14501* (2022).

[32] Torricelli, Maddalena, et al. "How does extreme weather impact the climate change discourse? Insights from the Twitter discussion on hurricanes." *Plos Clim* 2.11 (2023): e0000277.

[33] Laishram, Ricky, et al. "Crawling the community structure of multiplex networks." *Proc AAAI conf artif intell* . 33. 01. (2019).

[34] Himelboim, Itai, et al. "Classifying Twitter topic-networks using social network analysis." *Soc media soc* 3.1 (2017): 2056305117691545.

[35] Abitbol, Jacob Levy, & Alfredo J. Morales. "Socioeconomic patterns of twitter user activity." *Entropy* 23.6 (2021): 780.

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

[36] Orhan, Yunus Emre, et al. "Building Political Hashtag Communities: A Multiplex Network Analysis of US Senators on Twitter during the 2022 Midterm Elections." *Computation* 11.12 (2023): 238.

[37] Li, Jiexun, & Xiaohui Chang. "Combating misinformation by sharing the truth: a study on the spread of fact-checks on social media." *Inf Syst Front* 25.4 (2023): 1479-1493.

[38] Hanteer, Obaida, & Luca Rossi. "An innovative way to model twitter topic-driven interactions using multiplex networks." *Front Big Data* 2 (2019): 9.

[39] Kivelä, Mikko, et al. "Multilayer networks." *J Complex Netw* 2.3 (2014): 203-271.

[40] Zhai, X., et al. "Null Model and Community Structure in Multiplex Networks." *Sci Rep*, vol. 8, 2018, p. 3245, doi:10.1038/s41598-018-21286-0.

[41] Fortunato, Santo. "Community detection in graphs." *Phys Rep* 486.3-5 (2010): 75-174.

[42] Cha, Meeyoung, et al. "Measuring user influence in twitter: The million follower fallacy." *Proc Int AAAI Conf Web Soc Media* Vol. 4. No. 1. 2010.

[43] Goncalves, Bruno, et al. "Validation of Dunbar's number in Twitter conversations." *arXiv preprint arXiv:1105.5170* (2011).