Agreement and disagreement on climate crisis: insights from Twitter during the Conferences of the Parties

Liliana Martirano[1], Lucio La Cava[1], Francesco Gullo[2], and Andrea Tagarelli[1]

[1] University of Calabria, 87036 Rende (CS), Italy
[2] UniCredit, Rome, Italy
{liliana.martirano, lucio.lacava, tagarelli}@dimes.unical.it, fgullo@acm.org

1 Introduction and problem definition

Social media platforms are recognized as essential spaces where people express their opinions and engage in discussions. Both public figures and ordinary individuals create and share content, resulting in a democratization of information that embraces a wide range of perspectives. In such a scenario, understanding users' *agreement* and *disagreement* can lead to a deeper awareness of emerging topics and critical issues; identifying the appropriate model of user behavior in online debates still remains an open challenge. Users often engage only with like-minded individuals, prioritizing information that confirms their existing beliefs, but they need to understand the reasons for disagreement to develop more persuasive arguments and deal with counterarguments effectively.

For more than 10 years, researchers have been concerned with the classification of positions of agreement and disagreement on a local (*sentence-level*) and global (*user-level*) scale in online debates [6], [10]. More recently there has been a growing research interest for developing effective methods for reconstructing signed networks [9], predicting sign in sparse social networks [7], enhancing structural balance of signed networks [5], [1]. Signed networks help to capture complex relationships between entities, such as support and opposition, conflict and cooperation, trust and distrust. Formally, a signed (undirected) network is defined as $G = (V, E^+, E^-)$, where V is the set of vertices and E^+ (resp. E^-) is the set of positive (resp. negative) edges.

Our goal is to investigate on key elements for modeling a signed network of users discussing the climate crisis on Twitter, summarizing their agreement or disagreement w.r.t. multiple sub-topics. We analyze their interactions (retweets and replies) and their texts' similarity, spotting the main climate-related topics. Our study is conducted on posts published on Twitter before, during and after COPs (Conferences of the Parties), also known as UNFCCCs (United Nations Climate Change conferences), being the foremost global forums for multilateral discussion of climate change matters.

2 Preliminary results

We leveraged the most representative dataset involving climate social debate so far [2]. It consists of a large corpus of tweets spanning 7 years, from COP20 (2014) to COP26 (2021), gathered through the corresponding set of hashtags "cop2x", with $x \in \{0, ..., 6\}$, for a time windows of six months before and after the conference date. We hydrated

Table 1. Top user pairs in COP20 by number of retweets, in descending order of retweet count, and corresponding users statistics.

# retweets(u,v) + # retweets(v,u)	cos_sim(u,v)	# tweets(u)	# retweets(u,·) + # retweets(·, u)	# tweets(v)	# retweets(v,·) + # retweets(·,v)
611	0.4951	715	711	744	1960*
471	0.4973	47	2708	760	17248**
335	0.5601	364	364	744	1960*
271	0.6684	221	496	760	17248**
189	0.5047	423	1274	760	17248**
184	0.3813	179	690	80	1553
170	0.5858	43	892	760	17248**
151	0.4970	1140	884	168	4843
147	0.4706	43	178	760	17248**
146	0.6960	276	278	301	867

the corresponding Tweet IDs through the Hydrator tool,[3] following the official Twitter guidelines. We selected only English-language posts. In the present work, we focus on the first and least numerous COP20 (244833 posts and 61495 users) for illustrative purposes, but the following applies to each of the 7 conferences.

We are interested in investigating whether retweets and replies are good proxies for users agreement and disagreement. To this purpose, we first construct a **user interaction network** with edges weighted by type and number of interactions occurred between the user pairs, ignoring texts and relying solely on retweets and replies. We additionally construct a fully-connected **user similarity network** based on the cosine similarity of users' tweets, which provides additional information on user pairs since the signed networks inferred exclusively on the interactions are known to be sparse [7]. We group tweets by user and concatenate them after a cleaning procedure involving the normalization of hashtags to terms and the removal of links and emoji. The resulting texts are encoded using Sentence-BERT [8], a neural language model framework based on Transformer models, resulting in a sentence embedding for each user based on the content of the tweets. We compared the embeddings by cosine-similarity to find sentences with a similar meaning, and built a network of the users with edges weighted according to the similarity score of each user pair.

We first investigated the user pairs with the highest number of retweets. As shown in Table 1, we spotted that 9 out of 10 user pairs have cosine similarity scores greater than or approximately equal to 0.5. The fact that the score is not directly proportional to the number of retweets of the pair depends on the users' participation in other conversations, so they share a large piece of text but the remaining is different based on their interactions with other users. On the other hand, we spotted the user pairs with highest cosine similarity scores; we notice that user pairs with cosine similarity tending to 1 have a high number of retweets (typically, ordinary users retweeting authoritative users). We can therefore assert that the retweet interaction corresponds to a form of agreement in our data [3], even at the user-level.

We then looked at the posts with the highest number of replies. We note that although the texts contain clear opposition, some user pairs have high similarity values. This is partly due to the use of similar terms in the debate, but mostly due to the involve-

[3] https://github.com/docnow/hydrator

Table 2. Subset of replies in disagreement to the original post (first line). The cosine similarity score is wrt the author of the post.

text	cos_sim
Protest image at the Nazca Lines, Peru, today. Full picture	1.0000
HUGE HUGE mistake. Protest image at the Nazca Lines, Peru, today	0.4267
Did you make any damage to Nazca Lines?	0.6954
[...] and stop the bullshit, u just can't protest without puting in danger anything?	0.4720
You had no right to do that in a protected zone. It is disrespectful to our country	0.4420
This is an act of vandalism on a historical landmark not a protest!	0.3937
Walking without autorization in a restricted area!!! WTF	0.5140

Table 3. Main topics discussed in COP20

Main Topics
$T_{\langle 20,0\rangle}$: Impact of human activities and efforts to address climate change
$T_{\langle 20,1\rangle}$: International climate negotiations, agreements, and commitments
$T_{\langle 20,2\rangle}$: Urgent need for action to protect the Earth and save the planet
$T_{\langle 20,3\rangle}$: Need for global cooperation and sustainable solutions
$T_{\langle 20,4\rangle}$: Impact of deforestation and reforestation on ecosystems and landscapes

ment of the users in other discussions: the pairs shown in table 2 are involved together in replies of at least one other thread, and some have common retweets. Employing the opposite approach, we investigated the negative scores with higher absolute value and spotted that they do not necessarily correspond to more interactions between the user pair; therefore, we cannot assume that replies are a proxy for disagreement at user-level.

Generally speaking, we spotted that users contributing to discussions involving different sub-topics tend to have low cosine similarity in absolute value, whereas if they discuss the same sub-topic, even if not directly with each other, they tend to have high absolute scores. In the negative case it is less pronounced since negative values never deviate too far from 0, but they gain more meaning for individual arguments. Having therefore figured out that the specific subject of debate cannot be disregarded, we employ BERTopic [4] to extract relevant and meaningful sub-topics. We report the top-5 (by volume of tweets) discussed topics for COP20 in Table 3.

3 Proposed network model

In the light of the above discussed experimental results, we propose a COP network model of user agreements and disagreements, which also integrates information from user-provided textual contents.

We are given, for each COP $c \in [20..26]$, a corpus $\mathscr{T}_c = \{t_{\langle c,1\rangle}, ..., t_{\langle c,n\rangle}\}$ containing all tweets written during c, including six months before and after the conference dates. For each COP c, we infer a **multilayer signed network** $G_c = \langle T_c, V_c, V_{\langle c,T\rangle}, E^+{}_{\langle c,T\rangle}, E^-{}_{\langle c,T\rangle}\rangle$ such that: T_c is the set of topics $T_{\langle c,i\rangle}$ (layers) of COP c, V_c contains all users who interacted during COP c with at least one tweet in $\mathscr{T}_c$ $V_{\langle c,T\rangle}$ contains all users who interacted during COP c with at least one tweet in $\mathscr{T}_c$ about topic T_i, and $E^+{}_{\langle c,T\rangle} = \{(u,v)\,|\,u,v \in V_{\langle c,T\rangle} \wedge s_{\langle c,T_i\rangle}(u,v) = +1\}$ and $E^-{}_{\langle c,T\rangle} = \{(u,v)\,|\,u,v \in V_{\langle c,T\rangle} \wedge s_{\langle c,T_i\rangle}(u,v) = -1\}$, where $s_{\langle c,T_i\rangle}(u,v) \in \{+1,-1\}$ represent the *agreement* (resp. *disagreement*) between u and v based on topic T_i in COP c. Each user appears in one or more layers, i.e., $V_c = \bigcup_{T_i} V_{\langle c,T_i\rangle}$.

We devise a novel strategy for deriving signs for pairs of users regardless of the specific topic. First, for each layer, if the user pairs have a suitable percentage of retweets, we assign a positive sign, i.e., $s_{\langle c,T_i \rangle}(u,v) = +1$. Otherwise, if they have commented on at least one same thread, their sign reflects the difference between a predefined threshold and their similarity score, computed by combining the similarity of texts with their sentiment scores. For the latter, we focus on *aspect-based sentiment analysis* [11], as tweets, despite being short texts, can refer to multiple aspects of the discourse with possibly different polarities. The same approach can be extended to predict the signs of pairs without direct interactions. The values of individual layers can be combined to obtain a single sign for each user pair summarizing the positions held in all the sub-topics.

Summary. We analyzed Twitter users' agreement/disagreement about the debate on climate-related issues around COPs, based on users' retweets and replies as well as tweet contents. We proposed a novel network model to infer a multilayer signed network for each COP, with layers corresponding to topics. We are currently validating the proposed model and investigating the aggregation function that combines the positions of user pairs with respect to specific sub-topics into a unique agreement or disagreement relationship.

References

1. Dinh, L., Rezapour, R., Jiang, L., Diesner, J.: Enhancing structural balance theory and measurement to analyze signed digraphs of real-world social networks. Frontiers in Human Dynamics 4, 1028393 (2023)
2. Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, N., Larosa, F., Sas, M., Mekacher, A., Pearce, W., Zollo, F., Quattrociocchi, W., Baronchelli, A.: Growing polarization around climate change on social media. Nature Climate Change 12(12), 1114–1121 (2022)
3. Gaumont, N., Panahi, M., Chavalarias, D.: Reconstruction of the socio-semantic dynamics of political activist twitter networks—method and application to the 2017 french presidential election. PloS one 13(9), e0201879 (2018)
4. Grootendorst, M.: Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv:2203.05794 (2022)
5. Kirkley, A., Cantwell, G.T., Newman, M.E.: Balance in signed networks. Physical Review E 99(1), 012320 (2019)
6. Murakami, A., Raymond, R.: Support or oppose? classifying positions in online debates from reply activities and opinion expressions. In: Coling 2010: Posters. pp. 869–875 (2010)
7. Nasrazadani, M., Fatemi, A., Nematbakhsh, M.: Sign prediction in sparse social networks using clustering and collaborative filtering. J. Supercomput. 78(1), 596–615 (2022)
8. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (11 2019), https://arxiv.org/abs/1908.10084
9. Xiang, B.B., Ma, C., Chen, H.S., Zhang, H.F.: Reconstructing signed networks via ising dynamics. Chaos: An Interdisciplinary Journal of Nonlinear Science 28(12) (2018)
10. Yin, J., Narang, N., Thomas, P., Paris, C.: Unifying local and global agreement and disagreement classification in online debates. In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis. pp. 61–69 (2012)
11. Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W.: A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Trans. on Knowledge and Data Engineering (2022)