Understanding Opinions Towards Climate Change on Social Media

Yashaswi Pupneja[1,2] **Joseph Zou**[3] **Sacha Levy**[4] **Shenyang Huang**[1,3]
[1]Mila [2]Université de Montréal [3]McGill University [4]Yale University
`yashaswi.pupneja@umontreal.ca` `yuesong.zou@mail.mcgill.ca`
`sacha.levy@yale.edu` `shenyang.huang@mail.mcgill.ca`

Abstract

Social media platforms such as Twitter (now known as X) have revolutionized how the public engage with important societal and political topics. Recently, climate change discussions on social media became a catalyst for political polarization and the spreading of misinformation. In this work, we aim to understand how real world events influence the opinions of individuals towards climate change related topics on social media. To this end, we extracted and analyzed a dataset of 13.6 millions tweets sent by 3.6 million users from 2006 to 2019. Then, we construct a temporal graph from the user-user mentions network and utilize the Louvain community detection algorithm to analyze the changes in community structure around Conference of the Parties on Climate Change (COP) events. Next, we also apply tools from the Natural Language Processing literature to perform sentiment analysis and topic modeling on the tweets. Our work acts as a first step towards understanding the evolution of pro-climate change communities around COP events. Answering these questions helps us understand how to raise people's awareness towards climate change thus hopefully calling on more individuals to join the collaborative effort in slowing down climate change.

1 Introduction

Climate change is a widely debated topic, and a significant number of individuals still denies the effect of climate change. Hence, we seek to understand how real-world events would influence opinions regarding climate change on social media platforms. According to a survey by the Pew Research Center, 15% of adults in the US do not believe that the Earth is warming, and 36% believe that it is due to natural causes rather than human activity. The spread of misinformation and propaganda campaigns on social media has contributed to skepticism around climate change [14].

To address these challenges, we aim to analyze the growth or contraction of communities of climate change supporters and non-supporters on social media platforms. This analysis will provide insights into how real-world events such as the Conference of the Parties (COP) events shape public opinion on climate change. By analyzing the patterns of social media interactions and the spread of climate change discourse on social media platforms, effective communication strategies can be developed to promote climate change awareness and counter misinformation campaigns. The stakeholders of this study include policymakers, environmental organizations, and the general public.

2 Background and Related Work

Climate Change on Social Media. Social media have created new arenas for public debates and revolutionized the discussion of prominent issues such as global climate change [20]. Various research on this topic have been conducted. [11] looked into user profile, tweet attitude, discussion

topics, and climate phenomenons. [2] built a random forest (RF) model to predict people's attitude towards climate change and environment protection based on psychological and demographic factors. Furthermore, [25] modeled the relationship between perceived climate change efficacy and news and user activities on social media by using multilevel regression.

Temporal Graph Learning. Social networks can be modeled as a temporal graph where users are nodes and interactions are edges while the nodes and edges in the graph would naturally change over time [23, 29, 28]. The dynamic community detection task which aims to identify groups of individuals who are closely connected to each other on the temporal graph [21, 7, 4, 22]. Qin et al. [21] and Boudebza et al. [4] both proposed novel methods to detect *stable communities*, groups of nodes forming a coherent community throughout a period of time. In this work, we first apply the well-know Louvain community detection algorithm [3] to identify the growth or contraction of user communities of climate change topics on twitter.

Misinformation detection. Misinformation related to climate change is a significant problem that has been widely researched. Several studies have examined the relationship between social media usage, belief in conspiracy theories, and misinformation, highlighting the gravity of the issue [12]. The adverse impact of such misinformation has been well-documented, underscoring the importance of inoculating people against misinformation related to climate change [18]. Research on debunking strategies has produced mixed results, highlighting the continued influence of misinformation and the need for effective measures [17, 5]. The identification of psychological factors that drive misinformation and hinder debunking [9] indicates the complexity of the issue and the need for targeted interventions to tackle it.

Political polarization detection. Political polarization [8] is one of the crucial socio-political issues in twenty-first century. It intensifies political debates and may even threaten civil society. [8] demonstrated that political polarization happens naturally when the observable indicators of policy outcomes are not monotonic. Hence due to its commonness and harmfulness, it is important to monitor political polarization and its impact especially in public discussions about climate. [16] demonstrated that public opinion toward climate politics is closely consistent with people's positions on parties and the US general public has become quite politically polarized in recent years. Further studies [6, 19] showed similar alignment between polarized public opinions toward climate politics and partisan sorting. Considering the effect of events, [13] highlighted a large increase in polarization in 2021 during the 26th United Nations Conference of the Parties on Climate Change (COP26) and attributed it to growing right-wing activity. Moreover, [26] investigated hostile communication between polarized competing groups.

3   Dataset

In this section, we first discuss how we constructed the twitter dataset which we used for this project in Section 3.1. Then we extract the user-user mentions temporal network from the full dataset to examine the user interactions on twitter 3.2.

3.1   Collecting Twitter Dataset

To understand how people engage with climate change related topics on social media, we first collect dataset of tweets from Twitter, a popular social network for discussing trending topics. In particular, we focus our collection process on climate change related tweets, posts made by users on Twitter. We leveraged a recently released Kaggle dataset [11, 10] containing climate change related tweets and their attributes for our study. However, the dataset lacks the raw tweet text and other information. Therefore, we use the tweet ID from the dataset to recollect the tweets using the Twitter API for Academic Research. Our dataset contains 13.6 millions tweets sent by 3.6 million users from 2006 to 2019. Each tweet contains various information such as the sender, any mentions of other users, timestamp, any hashtags and the raw text of the tweet. We have also collected dates for relevant Conference of the Parties (COP) Events in the duration of the dataset (as seen in Appendix C).

3.2   Constructing Temporal Graph

From the collected meta-data of tweets, we construct the user-user mentions network to be a weekly temporal graph. This network naturally shows how users interact with each other over time. However,

the edges might not necessarily indicate that two users shares the same opinion. For example, it is possible for a climate change supporter to mention another user who doesn't believe in climate change. In this work, we discretize the mentions network to be a weekly temporal graph. More specifically, we represent the temporal graph $\mathbf{G}$ as a series of graph snapshots, $\mathbf{G} = \{\mathbf{G}_t\}_{t=1}^{T}$ where $\mathbf{G}_t = \{\mathbf{V}_t, \mathbf{A}_t\}$ is the graph snapshot at time t and $\mathbf{V}_t, \mathbf{A}_t$ represent the set of nodes and the adjacency matrix of $\mathbf{G}_t$ respectively.

While processing the tweet dataset, we find that there exist a large number of users which only send climate related tweets sparsely. Therefore, the constructed temporal graph is disconnected and sparse thus forming isolated components rather communities of users. To increase the density of the network while focusing on users that are more active on climate change topics, we remove all users that have less than 100 edges (across all steps) from the network (both outgoing and incoming edges). In this way, we only preserve a dense core set of users.

4 Methodology
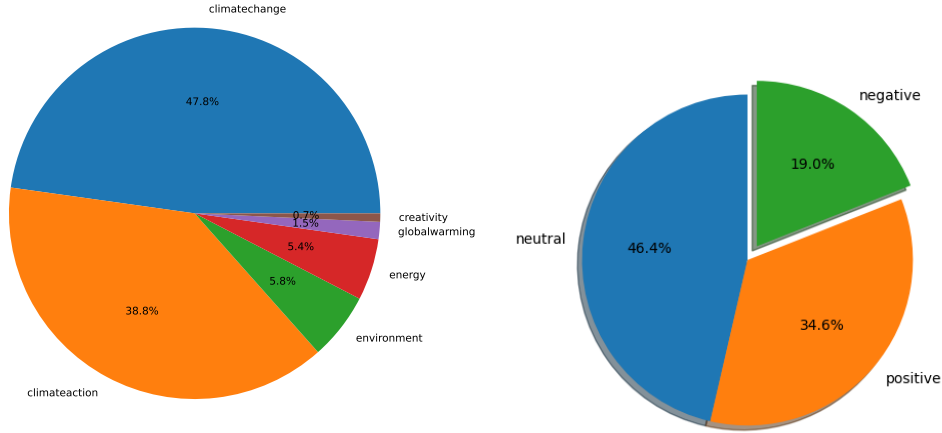
(a) Distribution of top Hashtags in tweets. (b) Distribution of *positive*, *negative* and *neutral* tweets.

Figure 1: Analysis of a). hashtags and b). sentiment in extracted tweets.

Hashtags Analysis. The top [YP: 8] most used hashtags from tweets are shown in Figure 1a. Some interesting hashtags include creativity, spirituality, economy, education and "its time to change". In particular, the hashtags themselves might already contain views or opinion towards climate change topics such as "its time to change".

Sentiment Analysis. Sentiment analysis is a technique used to determine the emotional tone behind a piece of text. We analyze the overall sentiment of tweets to understand the emotions associated with different topics related to climate change such as climate policy, environmental disasters or scientific research. Sentiment analysis also helps identify patterns in the emotional tone of different groups, such as climate change supporters or deniers. Here, we used the TextBlob library to detect the sentiments of the tweets which determined whether the texts were *positive*, *negative* or *neutral*. The results are summarized in the pie chart in Figure 1b. The majority of tweets are neutral while there remains 19.0% of tweets that have a negative sentiment. This shows that most users share a neutral sentiment towards climate change and it calls for more actions to engage the public in more positive involvement in climate change actions. We also acknowledge that negative sentiments such as fear and discontent are also considered as negative but can show awareness towards climate events.

Temporal Graph Analysis

Figure 2 shows the evolution of the number of edges and graph density over time. We also marked COP events by grey vertical lines. COP events are often accompanied by local peaks in the number of edges. This is likely because more discussion on climate change related topics happen around COP events annually. For graph density, we observe large peaks around COP 13, 14 and 17. The
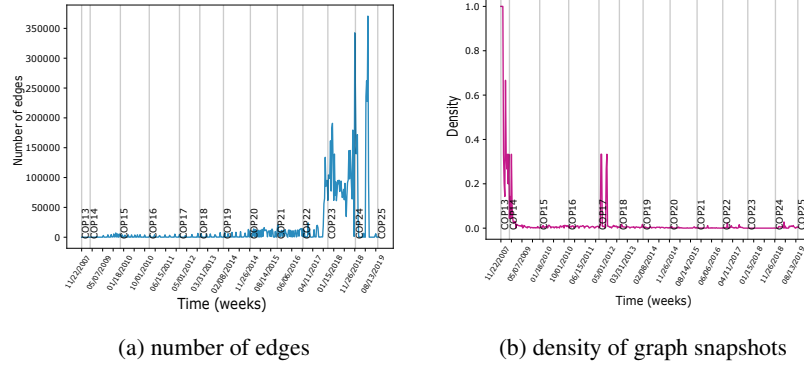
(a) number of edges

(b) density of graph snapshots

Figure 2: The evolution of a). the number of edges and b). graph density in the temporal user-user mentions network constructed from tweets.
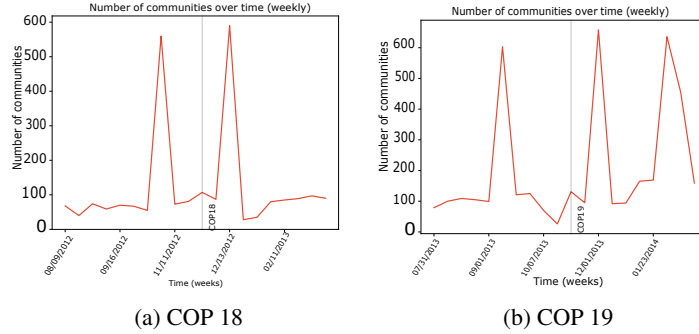

(a) COP 18

(b) COP 19

Figure 3: The number of communities detected around a). COP 18 and b). COP 19

peaks around COP 13 and 14 can be a result of noise in data collection in early years of twitter. The peaks around COP 17 signals increased discussion between active twitter users around this time. This coincides with the fact that COP 17 has one of the most significant achievements as all countries agree to the reduction of emissions, including the US and emerging countries such as Brazil, China, India and South Africa [1]. Details on the time of COP events are in Appendix C.

Community Detection on Temporal Graph. In this work, we aim to examine changes to the community structure of the user-user mentions network around the time of COP events. To this end, we utilize the fast and scalable Louvain algorithm [3] to detect communities on individual snapshots of the temporal graph. The Louvain algorithm optimizes the *modularity Q* of the graph, defined as follows in a directed graph,

$$Q = \frac{1}{2m} \sum_{ij} [\mathbf{A}_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \tag{1}$$

where $\mathbf{A}$ is the weighed adjacency matrix of the graph, k_i and k_j are the sum of the weights of edges of nodes i, j respectively, m is the sum of all of the edge weights in the graph, c_i and c_j are the communities of nodes i, j and lastly, δ is the kronecker delta function. The complexity of the Louvain algorithm is $O(n \cdot \log n)$ where n is the number of nodes. More details can be found in Appendix B.

Overall, The number of communities seem to increase over time with local peaks periodically. We also examine the community structure of graph snapshots surround COP events (10 weeks before and 10 weeks after) with the Louvain algorithm. Figure 3 shows the change in number of communities around COP 18 and COP 19. For COP 18 and 19, there is a decrease in the number of communities. This shows that larger communities are formed likely due to increased user discussions. However, for COP 20 and COP 21, there is an increase of communities. Therefore, it is difficult to clearly draw conclusions based on the number of communities around COP events. This is also expected as the Louvain algorithm is computed based the graph structure alone and doesn't utilize the tweet text and other edge features. More community detection results are reported in Appendix D.
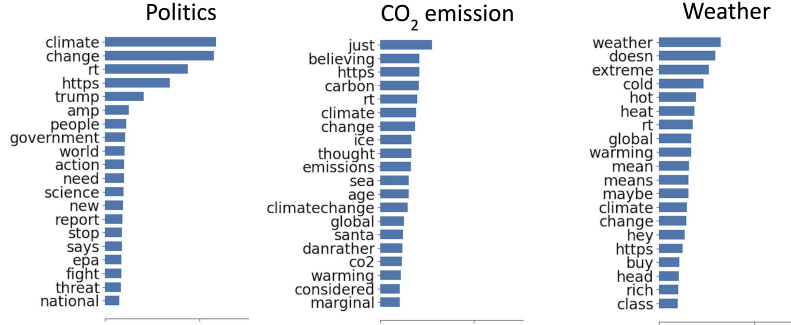
4

Figure 4: Significant topics extracted from the corpus.

Topic Modeling. Topic models are a class of unsupervised machine learning that allows us to extract topics from a large corpus of text data. They are powerful techniques that are commonly used for identifying patterns and themes in large text datasets. A typical topic model works by identifying groups of words that tend to co-occur in the same documents or tweets. These groups of words are known as "topics" and can provide insights into the underlying themes and issues that are present in the data. In our analysis of opinions on Twitter, we used topic modeling to identify the main topics of conversation related to climate change. By using the mini-batch non-negative matrix factorization (NMF) algorithm [15], we were able to efficiently extract topics from a large-scale dataset and gain insights into the key themes and issues that are salient to the public.

We first tokenized the tweets data and then trained the mini-batch NMF model on the preprocessed corpus by setting the number of topics as 10. We utilized the scikit-learn toolkit to run mini-batch NMF model. The training algorithm was run for iterations until convergence, i.e., a set of stable topics was obtained. After the matrix factorization had been done, we looked at the topic-word score matrix W and described each topic by the words with the highest relevance score. Figure 4 shows three obtained representative topics with their top 20 words about *politics*, *CO_2 emission*, and *weather*. The *politics* topic is related to politicians (e.g., Trump), organizations (e.g., EPA) and actions to address and mitigate climate change. The *CO_2 emission* topic includes global warming, Arctic ice melting, and sea level rising as consequences of CO_2 emission. The *weather* topic shows people's concern about extreme and abnormal weather caused by global warming.

5   Pathway to Impact

This project aims to provide valuable insights into opinions about climate change on Twitter, potentially impacting many stakeholders. For *policy makers*, our approach provides an understanding of public opinion and sentiment as guidance to decision-making. For *Researchers in social sciences*, our project contributes to the study of the patterns and dynamics of climate change-related online communication. For *journalists*, our study provides insights about when and how to report on climate change topics. For *the public*, this project can inform them about potential negative sentiment and opinions towards climate change events on social media thus help more individuals raise awareness of climate change topics. Future work can extend our work and apply more recent techniques such as Large Language Models [24] and Graph Neural Networks [27].

6   Acknowledgement

References

[1] Achievements of the conference of the parties. `https://www.activesustainability.com/climate-change/achievements-of-the-conference-of-the-parties/?_adin=02021864894`. Accessed: 2023-09-27.

[2] L. F. Beiser-McGrath and R. A. Huber. Assessing the relative importance of psychological and demographic factors for predicting climate and environmental attitudes. *Climatic change*, 149:335–347, 2018.

[3] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[4] S. Boudebza, R. Cazabet, O. Nouali, and F. Azouaou. Detecting stable communities in link streams at multiple temporal scales. In *Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part I*, pages 353–367. Springer, 2020.

[5] M.-p. S. Chan, C. R. Jones, K. H. Jamieson, and D. Albarracín. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Association for psychological Sciences*, 28(11):1531–1546, 2017.

[6] T. H. Y. Chen, A. Salloum, A. Gronow, T. Ylä-Anttila, and M. Kivelä. Polarization of climate politics results from partisan sorting: Evidence from finnish twittersphere. *Global Environmental Change*, 71:102348, 2021.

[7] R. Clark, G. Punzo, and M. Macdonald. Network communities of dynamical influence. *Scientific reports*, 9(1):1–13, 2019.

[8] A. K. Dixit and J. W. Weibull. Political polarization. *Proceedings of the National Academy of Sciences*, 104(18):7351–7356, 2007.

[9] U. K. Ecker, S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga, and M. A. Amazeen. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29, 2022.

[10] D. Effrosynidis, A. I. Karasakalidis, G. Sylaios, and A. Arampatzis. The climate change twitter dataset. *Expert Systems with Applications*, 204:117541, 2022.

[11] D. Effrosynidis, G. Sylaios, and A. Arampatzis. Exploring climate change on twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters. *Plos one*, 17(9):e0274213, 2022.

[12] A. Enders, J. Uscinski, and M. Seelig. The relationship between social media use and beliefs in conspiracy theories and misinformation. In *Polit Behav*, 2021.

[13] M. Falkenberg, A. Galeazzi, M. Torricelli, N. Di Marco, F. Larosa, M. Sas, A. Mekacher, W. Pearce, F. Zollo, W. Quattrociocchi, et al. Growing polarization around climate change on social media. *Nature Climate Change*, pages 1–8, 2022.

[14] C. Funk and M. Hefferon. Us public views on climate and energy. *Pew Research Center*, 25, 2019.

[15] C. Févotte and J. Idier. Algorithms for Nonnegative Matrix Factorization with the β-Divergence. *Neural Computation*, 23(9):2421–2456, 09 2011.

[16] L. C. Hamilton and K. Saito. A four-party view of us environmental concern. *Environmental Politics*, 24(2):212–227, 2015.

[17] S. Lewandowsky, U. K. H. Ecker, C. M. Seifert, N. Schwarz, and J. Cook. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3):106–131, 2012.

[18] S. v. d. Linden, A. Leiserowitz, S. Rosenthal, and E. Maibach. Inoculating the public against misinformation about climate change. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2017.

[19] A. M. McCright, R. E. Dunlap, and S. T. Marquart-Pyatt. Political ideology and views about climate change in the european union. *Environmental Politics*, 25(2):338–358, 2016.

[20] W. Pearce, S. Niederer, S. M. Özkula, and N. Sánchez Querubín. The social media life of climate change: Platforms, publics, and future imaginaries. *Wiley interdisciplinary reviews: Climate change*, 10(2):e569, 2019.

[21] H. Qin, R.-H. Li, G. Wang, X. Huang, Y. Yuan, and J. X. Yu. Mining stable communities in temporal networks by density-based clustering. *IEEE Transactions on Big Data*, 8(3):671–684, 2020.

[22] G. Rossetti and R. Cazabet. Community discovery in dynamic networks: a survey. *ACM computing surveys (CSUR)*, 51(2):1–37, 2018.

[23] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2020.

[24] R. Tang, Y.-N. Chuang, and X. Hu. The science of detecting llm-generated texts. 2023.

[25] L. Tuitjer and P. Dirksmeier. Social media and perceived climate change efficacy: A european comparison. *Digital Geography and Society*, 2:100018, 2021.

[26] A. Tyagi, J. Uyheng, and K. M. Carley. Affective polarization in online climate change discourse on twitter. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 443–447, 2020.

[27] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

[28] J. You, T. Du, and J. Leskovec. Roland: graph learning framework for dynamic graphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2358–2366, 2022.

[29] H. Zhou, D. Zheng, I. Nisa, V. Ioannidis, X. Song, and G. Karypis. Tgl: a general framework for temporal gnn training on billion-scale graphs. *Proceedings of the VLDB Endowment*, 15(8):1572–1580, 2022.

Table 1: Conference of the Parties (COP) Events present in the Twitter dataset.

Name	Date	Location
COP13	03 Dec 2007 - 17 Dec 2007	"Bali, Indonesia"
COP14	01 Dec 2008 - 12 Dec 2008	"Poznan, Poland"
COP15	07 Dec 2009 - 18 Dec 2009	"Copenhagen, Denmark"
COP16	28 Nov 2010 - 10 Dec 2010	"Cancun, Mexico"
COP17	28 Nov 2011 - 09 Dec 2011	"Durban, South Africa"
COP18	26 Nov 2012 - 07 Dec 2012	"Doha, Qatar"
COP19	11 Nov 2013 - 23 Nov 2013	"Warsaw, Poland"
COP20	01 Dec 2014 - 12 Dec 2014	"Lima, Peru"
COP21	30 Nov 2015 - 12 Dec 2015	"Paris, France"
COP22	07 Nov 2016 - 18 Nov 2016	"Marrakech, Morocco"
COP23	06 Nov 2017 - 17 Nov 2017	"Bonn, Germany"
COP24	03 Dec 2018 - 14 Dec 2018	"Katowice, Poland"
COP25	02 Dec 2019 - 13 Dec 2019	"Madrid, Spain"

A General Considerations

We encountered several general considerations that we believe are worth noting.

Ethical Considerations: Respecting the privacy and confidentiality of Twitter users is crucial, and we have taken steps to de-identify individual users in our analysis and focus on trends of the population instead of investigating individuals. However, we acknowledge that social media data analysis can raise ethical concerns, and urge researchers and practitioners to be mindful of potential harms and take appropriate measures.

Data Quality Considerations: Slang, buzzwords, and sarcasm can be challenging to interpret in social media data analysis. To address this, we employed natural language processing and sentiment analysis techniques, but acknowledge that our analysis may still be subject to noise and bias.

Limitations of Twitter Data: While Twitter can provide valuable insights into public opinions, it is important to recognize its limitations. Twitter users may not be representative of the broader population, and the topics and issues discussed on Twitter may not be indicative of those that are most relevant to the public.

Opportunities for Future Research: Our analysis suggests several avenues for future research on climate-change-related opinion on social media, such as exploring demographic differences in how climate change is perceived and discussed on Twitter, or tracking shifts in public opinion over time.

B Louvain Community Detection Algorithm

The Louvain algorithm has two phases which are interactively applied. First, all nodes in the graph is assigned its own unique community. Then, the algorithm decides if moving a node from its own community to its neighbors' community improves the modularity score Q. The node is then placed into its neighbor's community which sees the most increase in modularity. In the second phase, the algorithm groups all nodes in the same community and builds a new network where nodes are the communities from before. After this creation, the first phase then starts again. The algorithm terminates when no modularity increase can be observed. The complexity of the Louvain algorithm is $O(n \cdot \log n)$ where n is the number of nodes. Note that by design, the Louvain algorithm constructs hierarchical communities as it iteractively groups communities into larger ones.

C COP Events Details

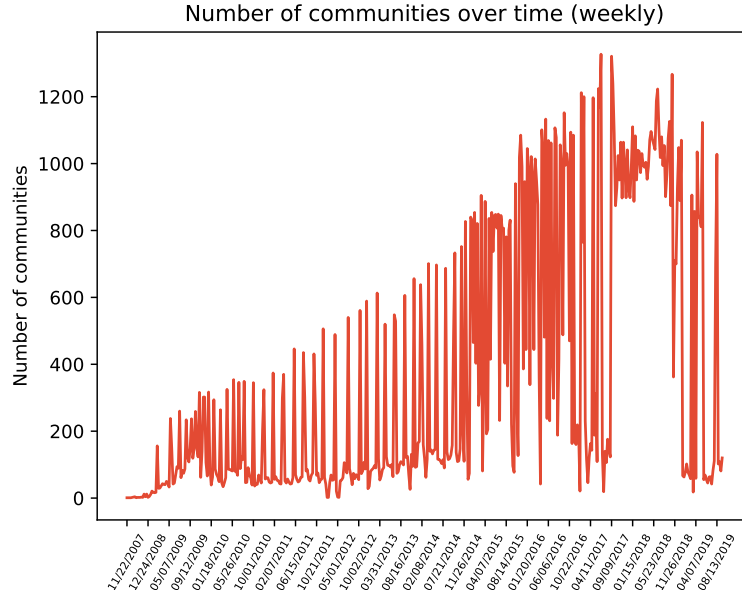We report the COP events present in the Twitter dataset in Table 1.

Figure 5: The evolution of number of clusters in each snapshot of the temporal graph.
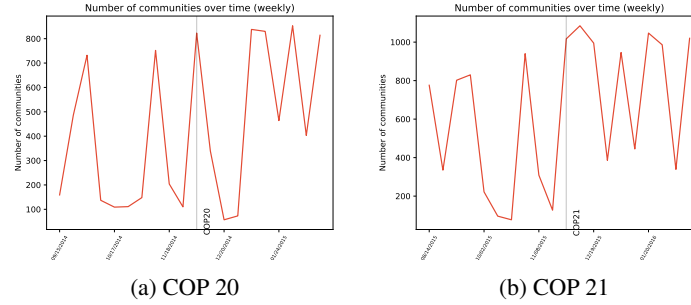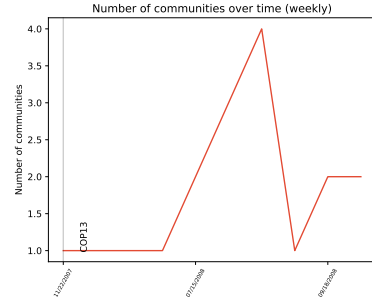

(a) COP 20

(b) COP 21

Figure 6: The number of communities detected by the Louvain algorithm around a). COP 20 and b). COP 21 event.

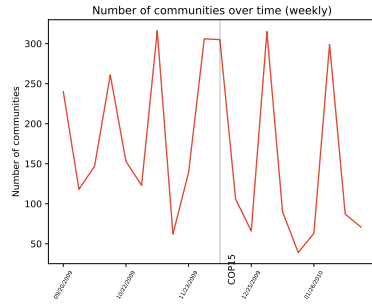# D   Additional Community Detection Results

Figure 5 shows the change in number of communities detected by the Louvain algorithm in the temporal graph. Here we provide additional community mining results for all the COP events present in our dataset including COP 13 to COP 24.
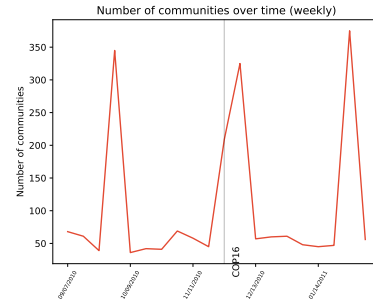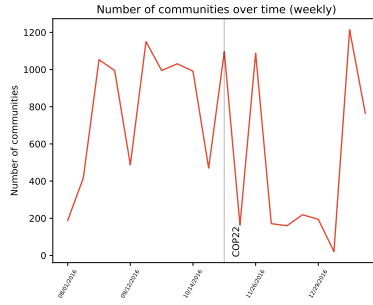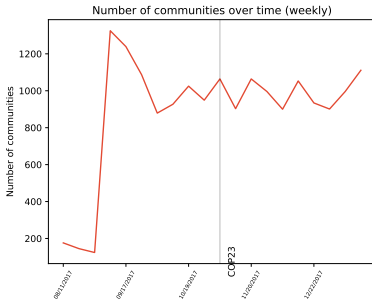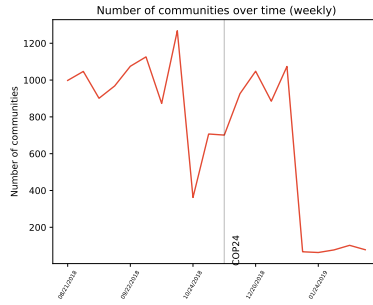
(a) COP 13
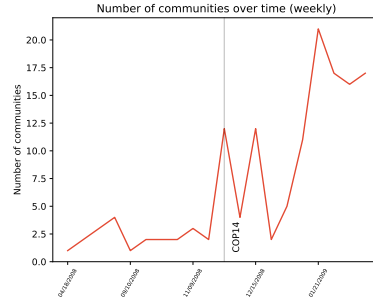
(b) COP 14

(c) COP 15

(d) COP 16
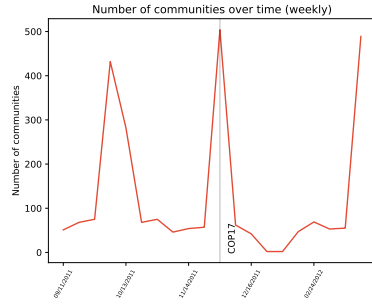
(e) COP 17

(f) COP 22

(g) COP 23

(h) COP 24

Figure 7: The number of communities detected by the Louvain algorithm around