Community-based Stance Detection

Emanuele Brugnoli[1,2,3,*], Donald Ruggiero Lo Sardo[1,2,3]

[1]Sony Computer Science Laboratories Rome, Joint Initiative CREF-SONY, Piazza del Viminale 1, 00184, Rome, Italy.
[2]Centro Studi e Ricerche Enrico Fermi (CREF), Piazza del Viminale 1, 00184 Rome, Italy.
[3]Dipartimento di Fisica - Sapienza Università di Roma, P.le A. Moro 2, 00185 Rome, Italy.

Abstract

Stance detection is a critical task in understanding the alignment or opposition of statements within social discourse. In this study, we present a novel stance detection model that labels claim-perspective pairs as either aligned or opposed. The primary innovation of our work lies in our training technique, which leverages social network data from X (formerly Twitter). Our dataset comprises tweets from opinion leaders, political entities and news outlets, along with their followers' interactions through retweets and quotes. By reconstructing politically aligned communities based on retweet interactions, treated as endorsements, we check these communities against common knowledge representations of the political landscape. Our training dataset consists of tweet/quote pairs where the tweet comes from a political entity and the quote either originates from a follower who exclusively retweets that political entity (treated as aligned) or from a user who exclusively retweets a political entity from an opposing ideological community (treated as opposed). This curated subset is used to train an Italian language model based on the RoBERTa architecture, achieving an accuracy of approximately 85%. We then apply our model to label all tweet/quote pairs in the dataset, analyzing its out-of-sample predictions. This work not only demonstrates the efficacy of our stance detection model but also highlights the utility of social network structures in training robust NLP models. Our approach offers a scalable and accurate method for understanding political discourse and the alignment of social media statements.

Keywords

Stance Detection, Polarisation, Social Networks

1. Introduction

Stance detection is a critical task within the domain of natural language processing (NLP). It involves identifying the position or attitude expressed in a piece of text towards a specific topic, claim, or entity[1, 2]. Traditionally, stances are classified into three primary categories: *favor*, *against*, and *neutral*. This classification enables a detailed description of textual data, facilitating a deeper insight into public opinion and discourse dynamics.

In recent years, the proliferation of digital communication platforms such as social media, forums, and online news outlets has resulted in an unprecedented volume of user-generated content. This surge underscores the necessity for automated systems capable of efficiently analyzing and interpreting these vast text corpora. Stance detection addresses this need by providing tools that can systematically assess opinions and reactions embedded within texts, thus offering valuable applications across various fields including social media analysis [3, 4], search engines [5], and linguistics [6].

According to the last report of World Economic Forum [7], the increase in societal polarization features among the top three risks for democratic societies. While a macroscopic increase of polarization has been observed, an understanding of the microscopic pathways though which it develops is still an open field of research. Through stance detection it would be possible to reconstruct these pathways down to the individual text-comment pairs.

Stance detection, has been explored across various fields with differing definitions and applications. Du Bois introduces the concept of the stance triangle, where stance-taking involves evaluating objects, positioning subjects, and aligning with others in dialogic interactions, emphasizing the sociocognitive aspects and intersubjectivity in discourse [6]. Sayah and Hashemi focus on academic writing, analyzing stance and engagement features like hedges, self-mention, and appeals to shared knowledge to understand communicative styles and interpersonal strategies [8]. Küçük and Can define stance detection as the classification of an author's position towards a target (*favor*, *against*, or *neutral*), highlighting its importance in sentiment analysis, misinformation detection, and argument mining [9]. These diverse approaches underscore the multifaceted nature of stance detection and its applications in enhancing the understanding of social discourse, academic rhetoric, and online content analysis. For a review of the recent developments of the field we refer to Alturayeif et al. [2] and AlDayel et al. [3].

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy
*Corresponding author.
emanuele.brugnoli@sony.com (E. Brugnoli);
donaldruggiero.losardo@sony.com (D. R. Lo Sardo)
0000-0002-5342-3184 (E. Brugnoli); 0000-0003-3102-6505 (D. R. Lo Sardo)

In this work, we propose a novel approach to training stance detection models by leveraging the interactions within highly polarized communities. Our method utilizes tweet/quote pairs from the Italian political debate to construct a robust training set. We operate under the assumption that users who predominantly retweet a particular political profile are likely in agreement with the statements made by that profile. We restricted our analysis to retweet since this form of communication primarily aligns with the endorsement hypothesis [10]. Namely, being a simple re-posting of a tweet, retweeting is commonly thought to express agreement with the claim of the tweet [11]. Further, though retweets might be used with other purposes such as those described by Marsili [12], the repeated nature of the interaction we observe in our networks reduces the probability that the activity falls outside of the endorsement behavior.

Conversely, while quoting a tweet works similarly to retweeting, the function allows users to add their own comments above the tweet. This makes this form of communication controversial regarding the endorsement hypothesis, as agreement or disagreement with the tweet depends on the stance of the added comment. On the other hand, the information social media users see, consume, and share through their news feed heavily depends on the political leaning of their early connections [13, 14]. In other words, while algorithms are highly influential in determining what people see and shaping their on-platform experiences [15], there is significant ideological segregation in political news exposure [16]. It is therefore reasonable to expect that users who almost exclusively retweet a political entity (party, leader, or both) use quote tweets to express agreement with statements posted by that entity and disagreement with statements posted by political entities ideologically distant from their preferred one. Additionally, the quote interaction perfectly encapsulates the stance triangle described by Du Bois [6].

In order to correctly assess political opposition we construct a retweet network and use the Louvain community detection algorithm [17] to characterize leaders and, through label propagation, the followers that align with their views.

Through these community labels we construct a dataset of claim-perspective couples by annotating tweet-quote pairs from profiles that clearly express political alignment as *favor* and annotating tweet-quote pairs in which the profiles come from different communities as *against*. Finally, we use a pretrained BERT model for Italian language and fine-tune it to the classification task.

This methodology aims to enhance the accuracy of stance detection models by incorporating real-world patterns of agreement and disagreement observed in polarized online environments. Further, it enables an unsupervised training paradigm that can be scaled to very large datasets.

In the following sections, we will outline the data gathering approach used for the dataset. Subsequently, we will describe the community detection methods employed to identify leaders and users within the Italian political discourse. We will then discuss the model architecture and its training process. In the results section, we will evaluate the model's performance and present our findings. Finally, the conclusion will address potential future developments, the implications of our work, and its limitations.

2. Results

In this study, we focus on a comprehensive set of Italian opinion leaders active on Twitter/X, including the official profiles of major news media outlets as well as prominent politicians and political parties. The profiles of news media outlets are further classified according to assessments provided by NewsGuard, which categorize them as either questionable or reliable sources. This classification is crucial for evaluating the quality of the information these outlets disseminate, particularly regarding their reputation for spreading misinformation. For the selected leaders, we collected all tweets produced from January 2018 to December 2022. The general public (followers) is identified based on their RTs to the content produced by these leaders. See Materials and Methods for details on the data collection process. Using this node configuration, we construct a bipartite network with two layers: leaders and followers, where the links represent the number of RTs by the latter of tweets made by the former. If a group of followers retweets tweets from two different leaders, it indicates that these leaders are likely communicating similar messages or viewpoints. To analyze these relationships more deeply, we perform a monopartite projection onto the leader layer. This projection, detailed in Materials and Methods, simplifies the network by concentrating solely on the leaders and the connections between them that are inferred from their shared followers. Panel (A) of Figure 1 shows the RT network of leaders aggregated in terms of communities identified through an optimized version of the Louvain algorithm [17]. The *a posteriori* analysis of the political leaders in each group reveals that the clustering algorithm effectively identified communities that align with the political affiliations of the leaders in each cluster [18, 19]. Specifically, the Left-leaning community includes political entities such as +Europa, Azione, Enrico Letta, and Nicola Fratoianni; the Right-leaning community features leaders from FdI, FI, and Lega; and the Five Star Movement (M5S) community includes key figures like Giuseppe Conte and Luigi Di Maio. An interesting observation from the network configuration is the clustering of questionable news sources. These profiles consistently group within the same com-

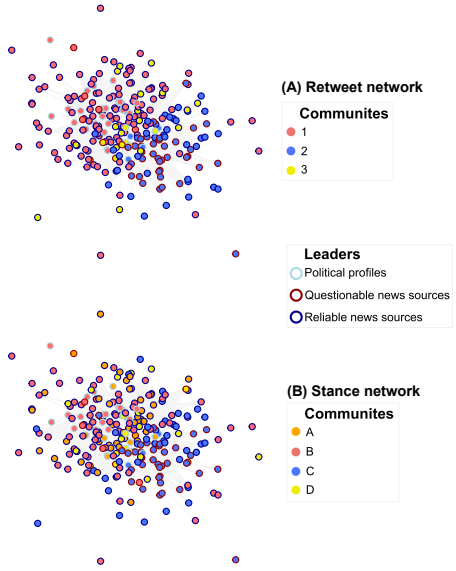

Figure 1: Projection of the follower-leader bipartite network onto the layer of leaders. In both **(A)** and **(B)**, the edges represent connections between leaders based on follower activity. **(A)** The edge weights are derived from the number of shared followers who retweeted content from both leaders. **(B)** The edge weights are based on the positive difference between favoring and against quote tweets made by shared followers on the content produced by the two leaders. In these visualizations, the node positions remain constant, providing a consistent framework for comparison. Node colors refer to communities as a result of running an optimized version of the Louvain algorithm. Nodes frame colors refer to the different types of leaders: political entities (azure), questionable news sources (dark red), and reliable news sources (dark blue).

munity, suggesting a potential alignment or affinity with specific political leanings or ideologies.

Leveraging the political bias of followers in our Twitter network, we build a very large dataset of tweet-quote pairs, each annotated with the corresponding stance (*favor* or *against*), as better described in Materials and Methods. Since this method assigns the stance to each pair in an unsupervised manner, to ensure that our approach is performing correctly, we randomly selected 500 pairs (250 favor and 250 against) and manually annotated their stance. We then compared the results of the automatic annotation with the manual annotation. The results, shown in Appendix - Table 3, indicate a high level of accuracy in favor and against classifications, with a small number of neutral cases. The dataset serves as training set for fine-tuning UmBERTo [20], an Italian language model based on the RoBERTa architecture [21], to assign stance labels to claim-perspective pairs. The fine-tuning process is performed using 5-fold cross-validation. The optimal performance for each fold is assessed by measuring the accuracy, i.e., the ratio of correctly predicted instances (both true *favor* and true *against*) to the total number of instances. The best-trained models from each fold demonstrate nearly identical performance, as shown by the average accuracy and F1-scores reported in the following table. The best model from fold 3 is identified

	Overall		Favor		Against	
	$\overline{\text{Acc}}$	(SD)	$\overline{\text{F1}}$	(SD)	$\overline{\text{F1}}$	(SD)
Training	0.863	(10^{-5})	0.863	(10^{-5})	0.864	(10^{-5})
Test	0.846	(10^{-6})	0.846	(10^{-6})	0.846	(10^{-5})

Table 1
Average performance of the best models from each fold on the training set and the test set. The table reports the mean and standard deviation (SD) for each metric considered: Accuracy for the overall model, and F1-score for each individual class.

as the highest performing and is therefore used in the following analyses. The corresponding confusion matrices for both the training and test sets are provided in Appendix - Table 5.

Given the imbalance in the label distribution of the claim-perspective dataset, we use 41,347 pairs – each annotated as favor and previously removed to create a balanced training set – as an additional test set to evaluate the model's performance. The model achieves an accuracy of 83.6% when predicting the stance of these pairs.

The model is then applied to classify all the collected tweet-quote pairs based on their stance. Thus, following the same procedure used to construct the RT network of leaders, we develop the stance network and analyze its community structure. In this case, the weight of a link in the bipartite follower-leader network represents the positive difference between the number of favoring and against quotes from a follower on the leader's tweets. Panel (B) of Figure 1 shows the stance network of leaders aggregated in terms of communities identified through the Louvain algorithm. The node positions in this representation are the same as those in the RT network, providing a consistent framework for comparison. More formally, to evaluate the differences in clustering assignments between nodes present in both the retweet network and the stance network, we perform a clustering comparison. Namely, we use the contingency table [22] associated with both the representations to compute community overlap. Figure 2 shows the comparison results broken down by source type: political entities and news outlets. While clusters C and D of the stance network primarily align with clusters 2 and 3 of the RT network, respectively, clusters A and B of the stance network mainly represent a refinement of cluster 1 from the RT network. This suggests that even in the stance network, the emerging communities align with the political affiliations of the leaders within each cluster.

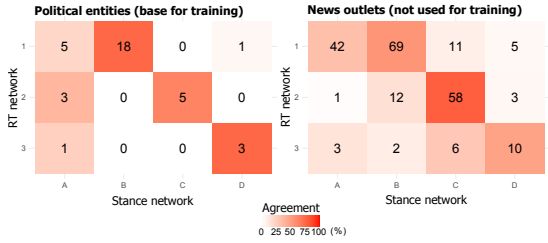

Figure 2: Contingency table associated with retweet network and stance network. Data is broken down by source type: political entities and news outlets.

Although the tweet-quote pairs used to train the model include only tweets from political entities, the result is significant. The training set does not include pairs where the quote comes from a follower who exclusively retweets political entities from the same ideological community as the tweet's author. This demonstrates the model's ability to reconstruct communities through precise classification of textual pairs.

The contingency table for news outlets, while displaying less pronounced patterns overall, still demonstrate clear coherence in classification between the retweet network and the stance network. This is particularly remarkable considering that these profiles were not included in the model's training set. The recovery of the retweet network's community structure within the stance network suggests that the model successfully generalizes across profiles with differing linguistic constraints, with only a minimal loss in accuracy, while still allowing for the reconstruction of group affiliations.

3. Discussion

Stance detection remains a vital yet challenging area in natural language processing (NLP), traditionally limited by the constraints of supervised learning. The availability of large language corpora, where interaction networks can be reconstructed, offers a novel approach that incorporates the social and dynamic aspects of stance, as outlined by Du Bois in his work on the stance triangle [6].

Our model addresses a more complex task compared to other state-of-the-art models. While existing models typically classify a user's stance on specific topics, our model classifies claim-perspective pairs into *favor* and *against* categories. This requires a deeper analysis of the relational stance between multiple interacting users and their statements.

Despite this increased complexity, our model achieved results comparable to those of existing state-of-the-art models [23, 24]. This success supports the hypothesis that in-group/out-group determinants, well-documented in opinion dynamics, significantly explain the variation in behaviors [25].

Moreover, our model's ability to reconstruct communities based on the accurate classification of textual pairs (as shown in Figure 2) underscores its potential for community reconstruction in scenarios where the interaction network is not provided.

Importantly, this approach also opens avenues for studying network dynamics based on the probability of agreement between account pairs. This has significant implications for understanding and potentially mitigating coordinated attacks, such as disinformation campaigns and political propaganda. By identifying patterns of agreement and disagreement, we can better detect and analyze the strategies behind these coordinated efforts, enhancing our ability to safeguard democratic processes and public discourse.

4. Materials and Methods

Data Collection. Our dataset comprises approximatively 15 million tweets collected by monitoring the activity of 583 profiles that reflect Italian online social dialogue (e.g., *La Repubblica*, *Il Corriere della Sera*, *Il Giornale*). Profiles were selected based on the list of news sites monitored by NewsGuard, a news rating agency dedicated to assigning reliability scores. According to NewsGuard, this list covers approximately 95% of online engagement with news, providing near-comprehensive coverage of news-related dialogue [26].

Additionally, we included Italian political entities in the list of profiles. This inclusion encompasses all major political parties and their leaders (e.g., *Giorgia Meloni* and *Fratelli d'Italia*, *Elly Schlein* and *PD*, *Giuseppe Conte* and *M5S*). For a complete list of the monitored political profiles see Appendix - Table 4.

For each monitored profile, we collected all tweets from January 2018 to December 2022 using the Twitter/X API before the limitations introduced by the new management[1]. We also gathered all retweets (RTs) and quotes (QTs) of this content within the same time frame, limited to those tweets that gained at least 20 RTs or 10 QTs. The following table provides a detailed breakdown of the data matching these criteria.

Category	Profiles	Tweets	RTs	QTs
News	329	279,793	16,365,178	3,587,830
Politics	38	101,017	15,385,363	2,388,621
TOTAL	367	380,810	31,750,541	5,976,451

Table 2
Breakdown of the dataset.

[1]https://twitter.com/XDevelopers/status/1621026986784337922

Community Detection. In order to reconstruct the discourse communities from the twitter activity we built a retweet network. In the context of the data collection strategy previously described, most RTs are from a non-monitored user (a *follower*) to one of the users monitored (a *leader*), excluding a few RTs from one leader to another (45, 299). We can therefore consider this network as a bipartite network, i.e. a network where all links are from one node type to another, with 367 leaders and 934, 394 followers, connected through links with a weight w_{xi} equal to the number of RTs from the follower x to the leader i.

To identify communities among leaders we assume that leaders with the same readership are more likely to be in the same political community. We therefore constructed a monopartite network by projecting on the leader layer, i.e. we construct a network from the set of all length two paths assigning weights that are the product of the path's links.

We used the Bipartite Weighted Configuration Model (BiWCM) to statistically validate our bipartite projection [27]. BiWCM accounts for weighted interactions and preserves the strength of nodes in both layers, ensuring that our observed co-occurrences are not due to random chance but represent genuine structural patterns in the data. In order to find political communities in the network, we applied the Louvain algorithm 1000 times and selected the solution that minimized modularity, i.e., the strength of division of the network into clusters, with higher values indicating a structure where more edges lie within communities than would be expected by chance [28].

The same procedure was followed to construct the stance network and study its community structure. In this case, the weight of a link in the bipartite follower-leader network indicates the fraction of favoring quotes from the follower to the leader's tweets.

Claim-Perspective Pairs Selection. To construct a dataset of claim-perspective text pairs annotated with the corresponding stance (*favor* if the perspective supports the claim, *against* otherwise), we first identified users who clearly expressed an (almost) absolute preference for a single political entity through their retweet activity. Specifically, for each follower, we calculated the distribution of their RTs across the political entities defined in Table 4. Then, we filtered those who allocated at least 80% of their RTs to a single political entity. Some users, although meeting the previous requirement, may not have had a sufficient level of retweet activity during the analyzed period to be considered inclined towards a particular political entity. For example, a user who has only given one retweet to the set of political profiles would appear totally inclined towards a particular entity. To reduce the uncertainty arising from the indiscriminate inclusion of all profiles satisfying the high retweet activity requirement for a single political entity, we calculated for each follower x the total number of retweets of content produced by the set of political entities $\mathscr{P}$ defined in Table 4 and excluded the bottom 80% of the resulting distribution (i.e., we imposed $|\mathrm{RT}_x(\mathscr{P})| > 7$). For the remaining users, we then assigned the label *favor* to those quotes of tweets from their preferred political entity and the label *against* to those quotes of tweets from entities belonging to other political communities, as determined by the community detection analysis. This procedure resulted in the creation of a dataset containing 243, 277 unique claim-perspective (tweet-quote) pairs, each annotated with the corresponding stance. Since the label distribution of the dataset was unbalanced towards *favor* (specifically, 142, 312 *favor* and 100, 965 *against*), we randomly removed 41, 347 *favor* pairs to obtain a balanced training set for the stance model. The removed pairs were later used as additional test set to evaluate the model's accuracy.

Stance model. We initialized our model starting from UmBERTo [20], an Italian language model based on the RoBERTa architecture [21]. Specifically, we relied on the cased version trained using SentencePiece tokenizer and Whole Word Masking on a large corpus, encompassing around 70 GB of text. This makes it highly effective for various natural language processing tasks in Italian, as it leverages a vast and diverse dataset to understand the nuances of the language [29, 30]. The pretrained model was then fine-tuned on the constructed dataset of tweet-quote pairs to create a tool capable of inferring the stance of claim-perspective text pairs: *favor* if the perspective agrees with the claim, and *against* otherwise. To input the text pairs into the pretrained model, we utilized UmBERTo's special tokens. Specifically, we concatenated the tweet and quote as

```
<s> + tweet + </s></s> + quote + </s>,
```

where `<s>`, `</s></s>`, and `</s>` represent the start, separation, and end tokens, respectively. Since we set `max_seq_length = 256`, which limits the total number of tokens that can be processed by the model, in cases where the concatenated strings exceeded this limit, the longer text between the tweet and the quote was truncated. This ensures that the input remains within the model's processing capacity while preserving as much information as possible from both texts. Conversely, shorter concatenated strings were padded using the special token `<pad>` until they reached the 256-token limit. Tweets and quotes were preprocessed before being concatenated by removing URLs, mentions, non-UTF-8 characters, line breaks, and tabs.

The pretrained UmBERTo model was imported into Python from the HugginFace Transformers library [31] as a model for sequence classification. The fine-tuning procedure enabled the model to output the probability dis-

tribution over the stance labels by minimizing the cross-entropy loss between the predicted labels and the true labels, effectively learning to classify the stance of claim-perspective pairs. We chose to perform 5-fold cross-validation to ensure the reliability of the results [32]. Namely, the data was first partitioned into 5 equally (or nearly equally) sized segments or folds. Subsequently 5 iterations of training and testing are performed such that within each iteration a different fold of the data is held-out for testing while the remaining 4 folds are used for learning. Thus, for each training-test split, we fine-tuned the UmBERTo model for 4 epochs using a batch size of 64 (for both training and testing) and an improved version of the Adam optimizer [33] with a learning rate of $5e-5$ and a weight decay of 0.01 for regularization. The chosen hyperparameters are among those recommended in the literature[34, 21].

5. Conclusion

This study introduces a novel stance detection model that significantly advances the understanding of alignment and opposition in social discourse. By leveraging social network data from X (formerly Twitter), we developed a robust training technique that utilizes interactions within politically aligned communities. Our approach involved curating a dataset of tweet/quote pairs, where the quotes are derived from users' interactions with leaders and politicians. This dataset facilitated the training of a BERT model, which achieved a state of the art accuracy of approximately 85%.

Our findings underscore the efficacy of using social network structures to train NLP models, demonstrating that retweet interactions can serve as reliable indicators of political alignment. This methodology not only enhances the scalability of stance detection but also offers a nuanced understanding of political discourse on social media platforms. By reconstructing and validating politically aligned communities through expert knowledge, our model provides a robust framework for analyzing the alignment of social media statements.

The implications of this work extend beyond stance detection, offering potential applications in monitoring political sentiment, identifying misinformation, and understanding public opinion dynamics. Future research could explore the integration of additional social network features and exploring the capacity of the model to generalize to other domains, interaction types and understanding how stance propagates within networks.

Additionally, investigating the role of specific linguistic markers like adverbs across different languages and cultures can reveal universal and language-specific determinants of stance.

While our model shows promising results, it also relies heavily on the assumption that retweets are mainly a form of endorsement, and that quotes within one's own political community are all in agreement and that outside of one's political community they are all in disagreement. While the high level of polarization observed in these networks support the validity of these assumptions, it also restricts the applicability of the model to domains where polarization is evident and these assumptions are valid.

Acknowledgments

We extend our deepest gratitude to Vittorio Loreto, the director of the Sony Computer Science Laboratories (CSL) and Professor at La Sapienza University of Rome, for his invaluable support and sponsorship of this research. His guidance was pivotal for the successful completion of our study. We also thank the anonymous reviewers for their insightful suggestions, which have greatly contributed to enhancing the quality of this work.

References

[1] D. Küçük, F. Can, Stance detection: Concepts, approaches, resources, and outstanding issues, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 2673–2676.

[2] N. Alturayeif, H. Luqman, M. Ahmed, A systematic review of machine learning techniques for stance detection and its applications, Neural Computing and Applications 35 (2023) 5113–5144.

[3] A. Aldayel, W. Magdy, It is more than what you say!: Leveraging user online activity for improved stance detection, 2019. URL: https://2019.ic2s2.org/, 5th International Conference on Computational Social Science, IC2S2 2019 ; Conference date: 17-07-2019 Through 20-07-2019.

[4] A. Gupta, S. Mehta, Automatic stance detection for twitter data, in: 2022 1st International Conference on Informatics (ICI), IEEE, 2022, pp. 223–225.

[5] T. Draws, K. Natesan Ramamurthy, I. Baldini, A. Dhurandhar, I. Padhi, B. Timmermans, N. Tintarev, Explainable cross-topic stance detection for search results, in: Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, 2023, pp. 221–235.

[6] J. W. Du Bois, The stance triangle, Stancetaking in discourse: Subjectivity, evaluation, interaction 164 (2007) 139–182.

[7] World Economic Forum, Global Risks Report 2024, Technical Report, World Economic Forum, 2024. URL: https://www.weforum.org/publications/global-risks-report-2024/.

[8] L. Sayah, M. R. Hashemi, Exploring stance and engagement features in discourse analysis papers., Theory & Practice in Language Studies (TPLS) 4 (2014).

[9] D. Küçük, F. Can, Stance detection: A survey, ACM Computing Surveys (CSUR) 53 (2020) 1–37.

[10] C. Becatti, G. Caldarelli, R. Lambiotte, F. Saracco, Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections, Palgrave Communications 5 (2019). doi:10.1057/s41599-019-0300-3.

[11] D. Boyd, S. Golder, G. Lotan, Tweet, tweet, retweet: Conversational aspects of retweeting on twitter, in: 2010 43rd Hawaii International Conference on System Sciences, 2010, pp. 1–10. doi:10.1109/HICSS.2010.412.

[12] N. Marsili, Retweeting: Its linguistic and epistemic value, Synthese 198 (2021) 10457–10483.

[13] W. Chen, D. Pacheco, K.-C. Yang, F. Menczer, Neutral bots probe political bias on social media, Nature Communications 12 (2021). doi:10.1038/s41467-021-25738-6.

[14] B. Nyhan, J. Settle, E. Thorson, M. Wojcieszak, P. Barberá, A. Y. Chen, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, S. González-Bailón, A. M. Guess, E. Kennedy, Y. M. Kim, D. Lazer, N. Malhotra, D. Moehler, J. Pan, D. R. Thomas, R. Tromble, C. V. Rivera, A. Wilkins, B. Xiong, C. K. de Jonge, A. Franco, W. Mason, N. J. Stroud, J. A. Tucker, Like-minded sources on facebook are prevalent but not polarizing, Nature 620 (2023) 137–144. doi:10.1038/s41586-023-06297-w.

[15] P. Gravino, D. R. Lo Sardo, E. Brugnoli, Cross-platform impact of social media algorithmic adjustments on public discourse, ArXiv (2024). doi:10.48550/arXiv.2405.00008.

[16] S. González-Bailón, D. Lazer, P. Barberá, M. Zhang, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Freelon, M. Gentzkow, A. M. Guess, S. Iyengar, Y. M. Kim, N. Malhotra, D. Moehler, B. Nyhan, J. Pan, C. V. Rivera, J. Settle, E. Thorson, R. Tromble, A. Wilkins, M. Wojcieszak, C. Kiewiet de Jonge, A. Franco, W. Mason, N. Jomini Stroud, J. A. Tucker, Asymmetric ideological segregation in exposure to political news on facebook, Science 381 (2023) 392–398. doi:10.1126/science.ade7138.

[17] V. D. Blondel, J.-L. Guillame, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 10008 (2008). doi:10.1088/1742-5468/2008/10/P10008.

[18] E. Brugnoli, P. Gravino, D. R. Lo Sardo, V. Loreto, G. Prevedello, Fine-grained clustering of social media: How moral triggers drive preferences and consensus, in: A. P. Rocha, L. Steels, H. J. van den Herik (Eds.), Proceedings of the 16th International Conference on Agents and Artificial Intelligence, ICAART 2024, Volume 3, Rome, Italy, February 24-26, 2024, SCITEPRESS, 2024, pp. 1405–1412. doi:10.5220/0012595000003636.

[19] M. Pratelli, F. Saracco, M. Petrocchi, Entropy-based detection of twitter echo chambers, PNAS Nexus 3 (2024) pgae177. doi:10.1093/pnasnexus/pgae177.

[20] L. Parisi, S. Francia, P. Magnani, Umberto: an italian language model trained with whole word masking, https://github.com/musixmatchresearch/umberto, 2020.

[21] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv (2019). doi:10.48550/arXiv.1907.11692.

[22] S. S. Brier, Analysis of contingency tables under cluster sampling, Biometrika 67 (1980) 591–596.

[23] A. Rashed, M. Kutlu, K. Darwish, T. Elsayed, C. Bayrak, Embeddings-based clustering for target specific stances: The case of a polarized turkey, in: Proceedings of the International AAAI Conference on web and social media, volume 15, 2021, pp. 537–548.

[24] S. Shi, K. Qiao, J. Chen, S. Yang, J. Yang, B. Song, L. Wang, B. Yan, Mgtab: A multi-relational graph-based twitter account detection benchmark, arXiv preprint arXiv:2301.01123 (2023).

[25] S. Rathje, J. J. Van Bavel, S. Van Der Linden, Outgroup animosity drives engagement on social media, Proceedings of the National Academy of Sciences 118 (2021) e2024292118.

[26] NewsguardTech.com, Social impact report 2021, 2022. Available from https://www.newsguardtech.com/wp-content/uploads/2022/01/NewsGuard-Social-Impact-Report-1.21.22.pdf (accessed Nov 27, 2023).

[27] M. Bruno, D. Mazzilli, A. Patelli, T. Squartini, F. Saracco, Inferring comparative advantage via entropy maximization, Journal of Physics: Complexity 4 (2023) 045011. doi:10.1088/2632-072X/ad1411.

[28] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113. doi:10.1103/PhysRevE.69.026113.

[29] F. Bianchi, D. Nozza, D. Hovy, FEEL-IT: Emotion and sentiment classification for the Italian language, in: O. De Clercq, A. Balahur, J. Sedoc, V. Barriere, S. Tafreshi, S. Buechel, V. Hoste (Eds.), Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Online, 2021, pp. 76–83.

[30] F. Tamburini, How "bertology" changed the state-of-the-art also for italian nlp, in: Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it 2020, Online, 2020.
[31] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Huggingface's transformers: State-of-the-art natural language processing, arXiv (2019). doi:10.48550/arXiv.1910.03771.
[32] P. Refaeilzadeh, L. Tang, H. Liu, Cross-Validation, Springer US, Boston, MA, 2009, pp. 532–538. doi:10.1007/978-0-387-39940-9_565.
[33] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv (2017). doi:10.48550/arXiv.1711.05101.
[34] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv (2018). doi:10.48550/arXiv.1810.04805.

		Automatic		
		Favor	**Against**	Σ
Manual	**Favor**	221	7	228
	Against	16	209	225
	Neutral	13	34	37
	Σ	250	250	500

Table 3
Comparison between manual and automatic annotation for 500 randomly selected tweet-quote pairs. The F1 score for the Favor category is 0.86, and for the Against category, it is 0.86 as well. These results indicate a strong agreement between manual and automatic annotation methods, especially considering that the unsupervised stance classification method does not account for labels other than Favor and Against, while some contents were manually classified as Neutral.

Political entity	**Twitter profiles**
+Europa	*piu_europa, emmabonino*
Articolo Uno	*articolounodp, robersperanza*
Azione	*azione_it, carlocalenda*
Cambiamo!	*giovannitoti*
Coraggio Italia	*coraggio_italia, luigibrugnaro*
Democrazia e Autonomia	*movimentodema*
Europa Verde	*europaverde_it, angelobonelli1*
FdI	*giorgiameloni, fratelliditalia*
FI	*forza_italia, berlusconi*
ItalExit	*gparagone*
IV	*italiaviva, matteorenzi*
Lega	*legasalvini, matteosalvinimi*
M5S	*giuseppeconteit, mov5stelle, luigidimaio*
ManifestA	*manifesta_it*
NcI	*maurizio_lupi*
PD	*pdnetwork, enricoletta, sbonaccini, ellyesse*
Potere al Popolo	*potere_alpopolo*
Rifondazione comunista	*direzioneprc*
SI	*si_sinistra, nfratoianni*
Unione di Centro	*antoniodepoli*
Unione Popolare	*unione_popolare, demagistris*

Table 4
List of Twitter profiles related to the main political entities active in Italy during the five-year period 2018-2022.

		Predicted		
		Favor	**Against**	Σ
Actual	**Favor**	70,690	10,082	80,772
	Against	10,517	70,255	80,722
	Σ	81,207	80,337	161,544

(a) training set

		Predicted		
		Favor	**Against**	Σ
Actual	**Favor**	16,929	3,264	20,193
	Against	2,740	17,453	20,193
	Σ	19,669	20,717	40,386

(b) test set

Table 5
Confusion matrices for both the (a) training and (b) test sets.