

Tell Me Who You Are and I Will Predict Your Vulnerability to Political Persuasion Techniques

Alessia Antelmi¹, Lucio La Cava² and Arianna Pera³

¹ Università degli Studi di Torino, alessia.antelmi@unito.it,

² Università della Calabria, lucio.lacava@dimes.unical.it,

³ IT University of Copenhagen, arpe@itu.dk

1 Introduction

In recent years, social media have played a crucial role in casting political discourse. Their pervasive influence has reshaped the landscape of political communication and engagement, and they have emerged as integral tools for politicians to disseminate their ideas, connect with constituents, and rally support in ways that were previously unimaginable. This digital evolution has introduced a new dimension to political communication: strategic messaging aimed at persuading and mobilizing the public. Leveraging the algorithmic capabilities of these platforms, politicians strategically tailor their messages to resonate with specific audiences, amplifying their impact in a targeted manner (for instance, with the aid of political micro-targeting [13]) and crafting their messages to influence public sentiments via digital political persuasion.

Scholars are increasingly turning their attention to this evolving landscape. While the majority of research has delved into the realm of computational propaganda, which encompasses techniques involving automated accounts and message amplification [4], a promising direction is focusing on political persuasion within digital contexts. This approach seeks to understand the dynamics of persuasive messaging employed by politicians on social media, shedding light on the strategies, tactics, and especially effects of these communication efforts. Rather than solely concentrating on detecting bots or automated activities, this line of inquiry explores how politicians strategically shape public perceptions and elicit desired responses. Despite the clear interest of various stakeholders in such a phenomenon, especially after the scandals of the 2016 US Elections and the Brexit [6], existing literature in the persuasion context mainly focuses on offline surveys [11] or controlled experiments [12], without considering an organic observation of social media. An additional difficulty in such a landscape is given by the role the individual characteristics of target users play in the delivery of persuasive messages. For this reason, the personality of each social media user should be taken into account when evaluating the effects of political persuasion on the online audience.

To shed some light on this connection, we propose a framework as part of an ongoing experimental work for the investigation of the relationship between persuasion techniques used in social media and the personality traits of the audience of such messages in political contexts. Our proposed computational framework includes both Natural Language Processing (NLP) and Network Science methods with the aim of raising the general public awareness of political instruments and encouraging the development of critical reasoning given the urge to fight fake news dissemination.

2 Proposed Framework and Experimental Setup

To evaluate the association between persuasion techniques used by politicians and the characteristics of their audience on social media, we first defined the composition of such audience. Specifically, we focused on persuasion messages disseminated by politicians through their tweets. We then identified users influenced by these specific persuasion strategies as those who engage in retweeting such persuasive tweets, as retweeting content has been recognized to be a sign of agreement [2]. In such a setting, we present an under-development framework with four main components. We emphasize that our proposed framework is conceived to be highly modular, making it possible to replace any of the components with more suitable alternatives as needed.

1. **Persuasion techniques detection tool.** The first component is a Machine Learning (ML)-based algorithm aimed at extracting persuasive messaging techniques used by politicians on social media. In a political setting, persuasion techniques carry similarities with propaganda as, in both cases, communication strategies are used to influence public opinion. Thus, we propose to consider propaganda techniques detection methods as a reference for our tool. In such a context, scholars have taken into account different types of techniques: among them, Da San Martino et al. [5] proposed a list of 18 strategies and a BERT-based model for their detection. Our aim is to extend this model to a multilingual version.
2. **Personality traits extraction tool.** The second component of our framework is a set of measures to extract personality traits given user-generated textual content and the social network in which the user is embedded. In particular, we plan to perform our analysis by considering the following measures:
 - Prediction scores of users' classification into their *Myers-Briggs Type Index* (MBTI) personality type based on text samples from social media. A multilingual model such as *Twisty* [10] will be considered as a starting point.
 - Emotion prediction scores of user-generated content. BERT-based tuned models trained on annotated corpora, such as the one presented by Bianchi et al. [1], represent references for such an analysis. Further feasible approaches include semantic networks shaping affective patterns of concepts derived from text (e.g., *forma mentis* networks [9]).
 - Sentiment prediction scores of user-generated content. Such a task requires language-specific models for sentiment analysis, such as the BERT-based one presented by Bianchi et al. [1].
 - Social dimensions of user-generated content. Choi et al. [3] proposed an LSTM-based model taking into account word embeddings to extract language pragmatics from text. We aim to extend such an algorithm to the multilingual case in order to apply it in our framework.
 - Network measures as proxies for the user's role in the social network. For instance, centrality measures, such as page rank and betweenness centrality, and their standard deviations can be used to evaluate the relevance and influence of a user.
3. **Association measure between persuasion techniques and personality.** The third component of the proposed framework involves quantifying the link between persuasion techniques and the personality of affected users. Given user-specific vectors

of personality trait scores and binary indicators of the user’s endorsement of persuasion techniques, we plan to apply Canonical Correlation Analysis (CCA) [7] since canonical loadings serve as quantitative indicators of the extent to which personality traits contribute to each canonical variable. As such, CCA facilitates our exploration of the specific personality traits that exhibit stronger associations with particular persuasion strategies.

4. **Prediction of user vulnerabilities to political persuasion techniques.** The last component of the framework encapsulates and builds upon the insights derived from the previous stages of the pipeline. Specifically, we intend to leverage unsupervised learning techniques, ranging from K-means to more complex approaches, to identify unique user clusters. This approach will enable us to pinpoint and predict user vulnerabilities to persuasion techniques based on shared characteristics and susceptibility patterns.

To conduct our examinations, we are exploiting the dataset introduced by Pierri et al. [8], containing Italian-language political discussions that occurred throughout the 2022 Italian elections, and also including a mapping between political representatives and their parties. Elections are perfect times for analyzing political persuasion strategies, as heightened stakes and the competitive environment often accentuate the diverse array of methods employed to sway public opinion.

3 Conclusion

In this extended abstract, we propose a comprehensive under-development framework designed to explore the intricate interplay between political persuasion techniques prevalent on social media platforms and the personality traits of users influenced by such messaging dynamics (e.g., within the context of the 2022 Italian General Elections). By leveraging the capabilities of an ML model to categorize persuasion strategies embedded in textual content, employing NLP tools to extract nuanced linguistic dimensions, utilizing network metrics to determine users’ roles within the social network, and harnessing statistical measures of association, we aim to derive insights into the correlations between distinct personality traits and specific persuasion strategies.

Our proposed framework, which is currently being validated through experiments and will be released as a project associated with multiple research works, holds the potential to raise consciousness among the broader public, with a particular emphasis on the social media user demographic. By shedding light on the mechanisms of political persuasion, our effort seeks to foster the critical thinking skills of individuals in the face of misinformation while promoting an environment conducive to well-informed discourse.

Ethics Statement

The proposed analyses use aggregated data and adhere to Twitter’s terms of data usage, with user identification limited to numerical IDs and election candidates’ usernames.

While the study acknowledges potential harm in associating personality traits with political behavior, it emphasizes the reporting of aggregated results as hypotheses rather than comprehensive descriptions, condemning any intentions of degradation or persuasion based on political orientation or personality traits while acknowledging concerns about potential misuse.

References

1. Bianchi, F., Nozza, D., Hovy, D., et al.: FEEL-IT: Emotion and sentiment classification for the Italian language. In: Proceedings of the eleventh workshop on computational approaches to subjectivity, sentiment and social media analysis. Association for Computational Linguistics (2021)
2. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: 2010 43rd Hawaii international conference on system sciences. pp. 1–10. IEEE (2010)
3. Choi, M., Aiello, L.M., Varga, K.Z., Quercia, D.: Ten social dimensions of conversations and relationships. In: Proceedings of The Web Conference 2020. pp. 1514–1525 (2020)
4. Da San Martino, G., Cresci, S., Barrón-Cedeño, A., Yu, S., Di Pietro, R., Nakov, P.: A survey on computational propaganda detection. In: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. pp. 4826–4832 (2021)
5. Da San Martino, G., Yu, S., Barrón-Cedeno, A., Petrov, R., Nakov, P.: Fine-grained analysis of propaganda in news article. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). pp. 5636–5646 (2019)
6. Grassegger, H., Krogerus, M.: The data that turned the world upside down. *Vice Motherboard* 28 (2017)
7. Hotelling, H.: Relations between two sets of variates. In: Breakthroughs in statistics: methodology and distribution, pp. 162–190. Springer (1992)
8. Pierri, F., Liu, G., Ceri, S.: ITA-ELECTION-2022: A multi-platform dataset of social media conversations around the 2022 Italian general election. arXiv preprint arXiv:2301.05119 (2023)
9. Stella, M., De Nigris, S., Aloric, A., Siew, C.S.: Forma mentis networks quantify crucial differences in STEM perception between students and experts. *PLoS one* 14(10), e0222870 (2019)
10. Verhoeven, B., Daelemans, W., Plank, B.: Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. In: Proceedings of the Tenth international conference on language resources and evaluation (LREC’16). pp. 1632–1637 (2016)
11. Weeks, B.E., Ardèvol-Abreu, A., Gil de Zúñiga, H.: Online influence? Social media use, opinion leadership, and political persuasion. *International journal of public opinion research* 29(2), 214–239 (2017)
12. Zarouali, B., Dobber, T., De Pauw, G., de Vreese, C.: Using a personality-profiling algorithm to investigate political microtargeting: assessing the persuasion effects of personality-tailored ads on social media. *Communication Research* 49(8), 1066–1091 (2022)
13. Zuiderveen Borgesius, F., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., de Vreese, C.H.: Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review* 14(1), 82–96 (2018)