Max Raphael **Mynter**

PHYSICS || SOCIOLOGY || TECHNOLOGY & MANAGEMENT

[redacted] | [redacted] | in/maxmynter

January 6, 2022

Research Proposal: DPhil Social Data Science [about 1570 words and two figures]

The taxonomy of political opinion dynamics in online forums by means of natural language processing and unsupervised machine learning methods

The ongoing digitization of our social interactions has shifted some political discourse into the online realm. However, this exchange of ideas differs from the real-world interaction. The statements are recorded, asynchronously accessible, and the parties are anonymous to each other. Therefore, some argue that the internet provides a platform for egalitarian discussion. Nonetheless, by joining specific communities and following only selected content creators, some people end up in echo chambers where their political beliefs are not challenged but endlessly reinforced. The storming of the U.S. capitol, riots in the wake of the George Floyd murder, and protests against the COVID-countermeasures are only some offline world events that were - at least to some extent - fueled by these online discourses and the resulting polarization. These circumstances beg the question: Is the internet a threat to democracy and our civil society?

This question, however, cannot be investigated without a proper understanding of online political discourse. With my doctoral research, I want to contribute to its taxonomy by developing and employing methods for large-scale analysis of digital traces of internet users, specifically their expressed political opinion in posts in online forums and the resulting polarization of communities.
I want to work on the following questions: What are the unique characteristics of online political discourse and polarization. How do both relate to real-world events? How are the opinion dynamics in the period before events such as elections? Does the online discussion influence their outcome, and how does this outcome influence subsequent discourse?

In doing so, the main contribution of my doctoral research will be methodological. I want to develop methods that utilize the unstructured data of online discussions to provide quantitative insights for this research. Therefore, I have to identify and operationalize necessary variables. How can political opinion be measured, how the polarization of a community, and which of these metrics are significant?

There has been prior work within the field of online political discourse that provided both theoretical insights and methodological development of natural language processing (NLP). For example, the authors [1] have employed NLP methods alongside social network analysis on Twitter data to quantify public response to actions of prominent political leaders in Indonesia.

Findings regarding polarization suggest that changes in individual political polarization occur with exposure to new information [2]. This information can be introduced by highly influential users in online communities, which are often partisan. While frequency and vocality of expression differ between influencers of different political slants [3], there are competing findings on whether exposure to opposing views triggers stronger polarization of users [4] or has a moderating effect [5]. Nonetheless, analysis of Twitter data showed that online polarization has generally increased over the past years [6]. This insight is consistent with recent findings [7] that suggest a significant increase in polarization on the Reddit online community around the 2016 U.S. presidential election. Different authors [8] conducted a similar study in in 2019, where they

developed an NLP framework to uncover linguistic dimensions of political polarization in social media. Other (OII affiliated) work introduced a framework for finding linguistic features that are maximally informative about the edge topology of a social network and use it to detect polarized concepts in online discussion forums [9].

During exploratory research for this proposal, I have applied the π^{LO} measure of [8] to the reddit politosphere dataset of [9] that comprises posts on over 600 political subreddits from the period of 2008 to 2019. Specifically, I analyze posts in the `news`, `worldnews`, and `politics` communities to infer polarization with π^{LO}, called the leave-one-out estimator. This quantity is defined as

$$\pi^{LO} = \frac{1}{2}\Big(\frac{1}{|L|}\sum_{i\in L}\hat{\mathbf{q}}_i \cdot \hat{\mathbf{p}}_{-i} + \frac{1}{|R|}\sum_{i\in R}\hat{\mathbf{q}}_i \cdot (1-\hat{\mathbf{p}}_{-i})\Big). \tag{1}$$

Here, $\hat{\mathbf{q}}_i = \frac{\mathbf{c}_i}{m_i}$ with $\mathbf{c}_i$ is the vector of empirical token frequencies per Redditor. A token is a unique one or two word expression. The variable m_i is the total token count. The sums run over each Redditor i. The vector $\hat{\mathbf{p}}_{-\mathbf{i}}$ is computed as $\hat{\mathbf{p}}_{-\mathbf{i}} = \hat{\mathbf{q}}^{L\,i} \oslash (\hat{\mathbf{q}}^{L\,i} + \hat{\mathbf{q}}^{R\,i})$, with $\oslash$ denoting the element-wise division.

This quantity captures differences in average token usage by partisanship with $\hat{\mathbf{p}}_{-\mathbf{i}} = 0.5$ indicating equal use. The dot product then weighs this token usage difference with the token frequency by user i. It has an upper bound of 1 if solely other users of their same political stance use the token(s) used by Redditor i. The lower bound, 0, is attained if user i uses only tokens used by Redditors from the opposition within the political spectrum. Each sum is normalized by the number of users with each political affiliation. The quantity π^{LO} is then the average of the partisan polarization.

To calculate π^{LO}, one needs a partisanship assignment for every user. For the exploratory analysis of this proposal, I inferred political alignment by taking the number-of-posts majority class of each user in forums (subreddits) that I hand-labeled with their political affiliation. Users with less than five posts in indicator subreddits are excluded.

As displayed in the figure 1 we find significant polarization different from random noise. For reference, the U.S. congress shows polarization of $\pi^{LO} = 0.53$. However, the polarization stays constant over the analyzed time. A t-test shows no difference in polarization with $p = 0.89$. This result is consistent with the results from [7] who only found a polarization increase before the period analyzed in this work.

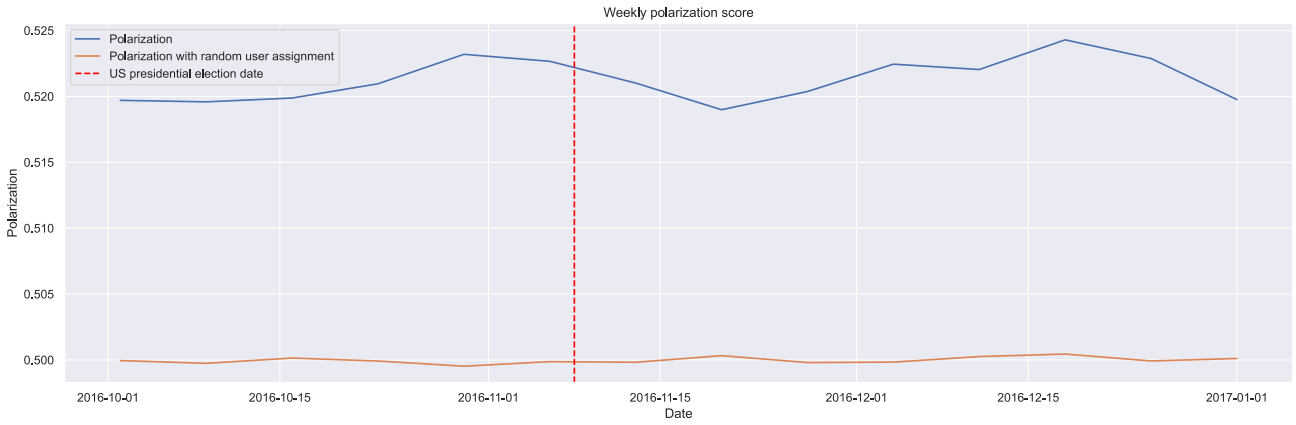

Figure 1: Weekly polarization in the analyzed Reddit communities (blue) and polarization with random partisanship assignment (orange).

Nonetheless, we find differences in the polarizations depending on political leaning. as evident from figure 2. Left-leaning individuals polarizations show a more extensive spread than the polarization for right-leaning individuals.

Yet, methodological issues remain — first, the selection and labeling of subreddits. I assigned static labels, but the opinions of individuals and communities can shift over time. Additionally, I excluded specific subreddits to which we cannot

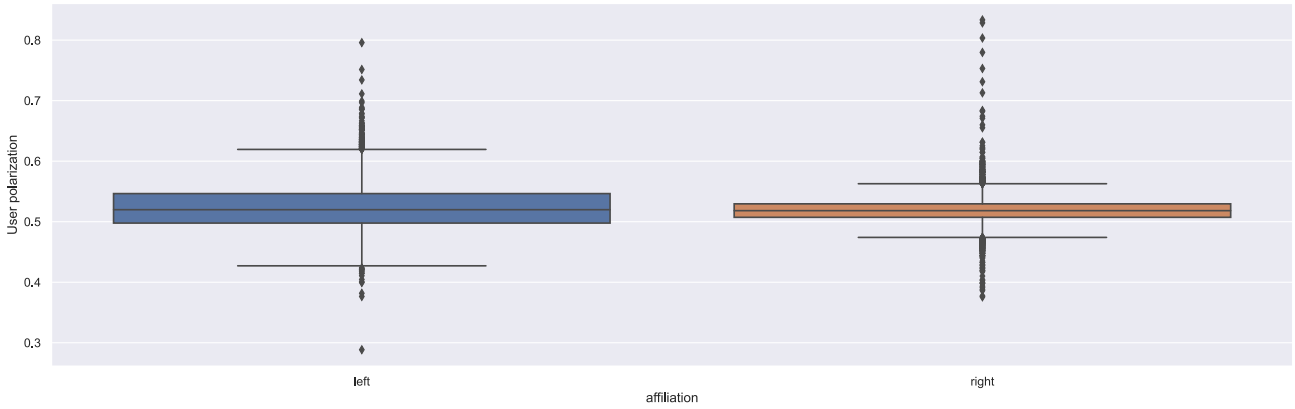

Figure 2: User polarization distribution over the whole time interval.

assign a left/right/centrist label. This exclusion can have complex effects on the results. Next, rigorous standards for partisanship assignment result in a substantial reduction of the final corpus size. Many posts in the analyzed sample are from users to whom no political affiliation could be assigned or with only a limited number of posts in the subreddits we use for partisanship assignments. Therefore, in the context of this analysis, their partisanship is unclear, and information-carrying data might have been excluded.
Second, the classification approach simplifies political opinions to a spectrum with binary poles, which has been criticized as it is insufficient to capture political opinions fully [10].
Third, the Reddit online forum is US-dominated, but users from the whole world can participate in the discussion. Additionally, the forum is prone to bots employed by different actors to propagate certain political opinions. These bots might skew the results of our analysis.
Fourth, the polarization measure (equation 1) and the composite dot products only use unigram and bigram featurization and no context understanding. This procedure probably foregoes some predictive power that could be extracted from the data.

To prepare methodological improvements, I evaluated the log-odds ratio with uninformative prior to infer idiosyncratic language as described in [11]. With this analysis, I find strong evidence for partisan-specific vocabulary. Left-leaning individuals use words such as *Capitalism*, *Class*, or *Socialism*, whereas right-leaning individuals' most idiosyncratic words are *Hillary* and *Trump*.

Therefore, I am confident that I can address the above shortcomings of this exploratory analysis with my graduate research. The findings on idiosyncrasy suggest that developing an unsupervised algorithm that can then infer user partisanship on a large scale is possible. This approach would allow labeling of far more data to evaluate polarization and, in the process, enable the extension of π^{LO} to more than two political poles. One could employ the same methodology to detect and analyze the influence of bots in online discussions. Additionally, a novel, extended measure that incorporates meanings of words and not only token counts will contribute to more detailed results.
On top, the analysis should be extended beyond Reddit data. One could collect and evaluate data from other sources like Twitter or 4chan. An unsupervised method to infer political slant is necessary for the latter since posts are anonymous. Additionally, I will extend the analysis beyond the single election event discussed in this proposal.

This preliminary analysis and prior findings in the literature present the taxonomy of online discourse as a worthwhile endeavor. My research will uncover insights about opinion dynamics in the online political discourse. Additionally, novel NLP methods incorporating context for partisanship assignment and direct quantification might be helpful beyond this primary application. These results will aid future research in the computational social science domain and equip lawmakers and NGOs with the knowledge to address the democracy threatening side effects of political discourse on the internet.

I believe that I am well prepared to conduct and grow with the proposed research during a DPhil program. My previous

education in social sciences enabled me to work with social science concepts. In contrast, my physics degrees equipped me with the necessary analytical skills to understand, extend, and develop the necessary algorithms. Furthermore, my education in technology management enables me to properly manage these tasks. Additionally, I have visited UC Berkeley for the last fall semester and attended lectures such as NLP to specifically prepare me to work in this interdisciplinary domain. Lastly, with my work at Allianz Global Investors and numerous academic research projects such as my final thesis successfully completed, I feel confident to provide novel insights and methods to the field.

References

[1] A. Budi and W. A. Pamungkas, "Partisanship in crisis: Public response to covid-19 pandemic in indonesia," *Jurnal Ilmu Sosial dan Ilmu Politik*, vol. 24, no. 1, pp. 15–32, 2020.

[2] H. A. Prasetya and T. Murata, "A model of opinion and propagation structure polarization in social media," *Computational Social Networks*, vol. 7, no. 1, pp. 1–35, 2020.

[3] J. Jiang, X. Ren, and E. Ferrara, "Social media polarization and echo chambers: A case study of covid-19," *arXiv preprint arXiv:2103.10979*, 2021.

[4] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9216–9221, 2018.

[5] P. Barberá, "How social media reduces mass political polarization. evidence from germany, spain, and the us," *Job Market Paper, New York University*, vol. 46, 2014.

[6] V. R. K. Garimella and I. Weber, "A long-term analysis of polarization on twitter," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, 2017.

[7] I. Waller and A. Anderson, "Quantifying social organization and political polarization in online platforms," *Nature*, pp. 1–5, 2021.

[8] D. Demszky, N. Garg, R. Voigt, J. Zou, M. Gentzkow, J. Shapiro, and D. Jurafsky, "Analyzing polarization in social media: Method and application to tweets on 21 mass shootings," *arXiv preprint arXiv:1904.01596*, 2019.

[9] V. Hofmann, J. B. Pierrehumbert, and H. Schütze, "Modeling ideological agenda setting and framing in polarized online groups with graph neural networks and structured sparsity," 2021.

[10] S. Bagui, C. Wilber, and K. Ren, "Analysis of political sentiment from twitter data," *Natural Language Processing Research*, vol. 1, no. 1-2, pp. 23–33, 2020.

[11] B. L. Monroe, M. P. Colaresi, and K. M. Quinn, "Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict," *Political Analysis*, vol. 16, no. 4, p. 372–403, 2017.