Social Media vs. Surveys: A New Scalable Approach to Understanding Legislators' Discourse[*]

Valentina González-Rostani[†] José Incio[‡] Guillermo Lezama[§¶]

This paper explores how legislators use social media, specifically investigating whether their posts reflect the concerns expressed by their legislative party peers in an anonymous survey. Utilizing data from Twitter, we compare legislators' social media posts with their responses in PELA, a survey of parliamentary elites in Latin America. We propose a novel, and scalable method for analyzing political communications, employing OpenAI for topic identification in statements and BERTopic analysis to identify clusters of political communication. This approach enables a thorough and detailed examination of these topics over time and across political parties. Applying our method to statements made by members of the Chilean Congress, we observe a general alignment between the preferences stated in surveys by elites and the prominence of these issues on Twitter. This result validates Twitter as a tool for predicting politicians' preferences. Our methodological approach enhances our understanding of political communication and strategy, offering valuable tools for analyzing political rhetoric over time.

[*]Please do not cite or circulate without authors' permission.

[†]University of Pittsburgh, Department of Political Science. Email: mag384@pitt.edu.
[‡]Pontificia Universidad Católica del Perú, Department of Sociology. Email: jincio@pucp.pe.
[§]University of Pittsburgh, Department of Economics. Email: guillermo.lezama@pitt.edu.
[¶]Authors in alphabetical order.

INTRODUCTION

Social media has revolutionized how elected representatives communicate with their constituencies, fellow legislators, and the media. This paradigm shift from traditional communication methods to digital platforms presents a unique intersection of transparency and strategy within political discourse.[1] While these platforms facilitate a more direct and immediate connection with the public, they also introduce a nuanced battlefield for politicians. This digital arena compels them to continually balance between voicing their genuine policy preferences and molding their rhetoric to align with what is electorally advantageous. As highlighted by Alesina and Cukierman (1990), this trade-off encapsulates one of the main dilemmas politicians face: the choice between promoting policies that resonate with their core values and those that maximize their reelection prospects.

The complexity of this scenario has increased in the digital age. Blum, Cormack, and Shoub (2023) point out that the wide range of communication channels available to politicians, especially social media channels, greatly expands their ability to influence public perception and gain support. However, a critical question arises: Do these digital interactions reflect politicians' genuine policy preferences and priorities, or are they strategic tactics aimed at winning votes and retaining office? This research note explores this dilemma, comparing the characteristics of political communication in the digital era with the preferences expressed in an anonymous survey of parliamentary elites.

Surveys focusing on the political elite, such as the Parliamentary Elites in Latin America Observatory (PELA hereafter, PELA-USAL, 2018), offer valuable insights into the beliefs and attitudes prevalent among political leaders. Their responses to PELA are often used as a baseline for legislators' opinions and attitudes. However, these surveys often fail to shed light on the specific content politicians choose to communicate to their constituents. In addition, since the sample is taken once per legislature, it does not allow us to observe changes in legislators' opinions or reactions to events between surveys.

Fortunately, the extensive social media data collection allows us to observe their public views in real time, providing a new avenue for exploring these communication strategies and broader aspects of elite political behavior and public policy formulation. Politicians widely use social networks. For example, every U.S Congress member maintains a Twitter account (Golbeck

[1] Refer to Siegel et al. (2022); Blum, Cormack, and Shoub (2023); Cormack (2016)

et al., 2018), with similarly high usage rates observed in Europe (Scherpereel, Wohlgemuth, and Lievens, 2018) and Latin America (Munger et al., 2019).[2] Therefore, understanding these political communications is crucial as even ostensibly cheap talk political speeches can have profound implications (Farrell, 1995). They can shape public attention, influence policy actions (Jones and Baumgartner, 2004), and affect how the electorate perceives their representatives' performance in the legislature (Grimmer, 2013).

This study makes two key contributions to understanding political communication. First, we analyze Twitter data and contrast it with elites' survey preferences to gain insights into legislators' daily strategic communications. This approach provides a unique perspective on how and when politicians diverge in their beliefs and attitudes from what they communicate on social media. The variation (or lack of) between public and private legislators' statements can give us evidence for different dynamics, such as the relationships between politicians and their parties.

Our second major contribution lies in the layout of a scalable methodological approach to identifying issues in political communications and the disparity between these and politicians' private stances. The first step is to classify legislators' tweets into political topics. We used the chat completion feature of OpenAI API to classify statements through 'system instructions,' following the issues asked in the PELA survey, as common problems. Then, we compare the issue salience in Twitter with the importance assigned in PELA. Our second step is to employ an unsupervised BERT-based topic analysis (BERTopic, Grootendorst, 2022) to discover emerging clusters within political discourse. This step allows us to unveil other salient topics that could inform projects such as PELA to include in their future questionnaires. To test our approach, we map the daily communication of Chilean members of Congress from March 2014 (when they assumed office) to December 2014,[3] and we provide descriptive plots of this data with insights into how elite strategic communication evolves over time and differs from stated preferences.

Our findings demonstrate the efficacy of utilizing Twitter as a proxy for legislators' privately stated preferences. First, we show that a substantial proportion (24%) of public statements on Twitter align with the topics in PELA's questionnaire. This observation suggests that legislators do indeed employ Twitter as a platform to convey their opinions on critical issues. Even more

[2]Twitter serves as a critical platform for politicians, enabling them to engage in policy discussions, disseminate updates about their activities (Golbeck, Grimes, and Rogers, 2010; Golbeck et al., 2018), communicate with the electorate (Hemphill, Otterbacher, and Shapiro, 2013), and enhance their visibility in various media outlets (Graham et al., 2013).

[3]We selected this year because it coincides with the year of the PELA data collection.

interesting is that, for the Chilean case, the degree of importance (rankings) that legislators expressed in the anonymous survey largely coincides with the importance they assign to it in their public statements. Second, the analysis using unsupervised clustering of tweets demonstrates additional clusters of political communication, such as international affairs. Finally, we also demonstrate that a significant proportion of tweets do not align with any of these common issue labels or other political content, which suggests the need for cautious use of tweets without prior filtering. For instance, many statements refer to greetings and interactions with other accounts but are not directly tied to a policy topic.

Our results and our methodological proof of concept make a compelling case for the importance of exploring political communications on social media. These platforms provide data publicly available to everyone and with great time granularity. To our knowledge, no prior study has juxtaposed legislators' declared preferences with their actual communication strategies in such a comprehensive manner. Moreover, our methodology not only maps the strategic communication tendencies of legislators, but also offers a scalable approach for analyzing legislative bodies at large. It provides an insightful lens through which the strategic nature of political communication can be understood. Our study also shows GPT-3.5's effectiveness in identifying policy issues in political discourse, underscoring the efficiency and multi-language applicability of modern Large Language Models compared to traditional human coding methods, aligning with increasing academic focus in this field (Gilardi, Alizadeh, and Kubli, 2023; Wu et al., 2023).

POLICY-SEEKING OR VOTE-SEEKING?

While surveys highlight the policy preferences of parliamentary elites, social media analysis uncovers a range of the political communication strategies they employ. These strategies may fall into three categories: first, policy-seeking, where politicians champion their preferred policies; second, communication of party policy preferences, tailored to core supporters; and third, deviation from their actual policy preferences, possibly neglecting personal priorities in favor of issues relevant to the electorate, which signals a move toward vote-seeking behavior. If politicians are policy-seeking, alignment between their private and public policy priorities is expected. However, if they are office-seeking, a divergence between their views and social media statements is likely.

Building on the notion of politicians as policy-seeking, we should expect their communications to align with their stated preferences and priorities. This perspective is supported by scholars like Jacobs and Shapiro (2000) and Shapiro and Jacobs (2000), who argue that many politicians prioritize achieving specific policy outcomes over securing reelection. These policy-driven politicians exploit lapses in public political engagement to steer the agenda towards their preferences (Shapiro and Jacobs, 2000) and strive to align public opinion with their views by providing explanations for any inconsistencies (Grose, Malhotra, and Parks Van Houweling, 2015).

Another plausible reason for the alignment between stated policy positions and political communication could be the legislators' focus on catering to their core party supporters (e.g., Wright, 1989). For instance, these supporters play an important role in nomination processes and primaries (Bawn et al., 2012; Fenno, 1978; Gerber and Morton, 1998) as well as in general elections (Holbrook and McClurg, 2005). Moreover, as Egan (2013) and Kastellec et al. (2015) point out, the policy preferences of these highly engaged party members are more likely to be in sync with those of policy-oriented politicians.

Though there are incentives to maintain alignment of preferences and actions, legislators do not always act to support their own views. Deviations could result from adapting to a particular context to avoid losing their electoral support. For instance, Mayhew (2004) emphasizes Congress members' constant pursuit of reelection, which, applied to our analysis, could lead them to utilize platforms like social media for direct communication with constituents. This need to remain electorally relevant may result in politicians adapting their communication to suit changing public sentiments (Stimson, MacKuen, and Erikson, 1995). Barberá et al. (2019) show that legislators tend to align their discourse with public concerns, often following rather than leading on public issues. Canes-Wrone and Shotts (2004) echo these conclusions, highlighting how shifts in public opinion directly influence the preferences and behaviors of political figures.

Moreover, the need for strategic positioning can lead to selective communication. Politicians might choose to avoid discussing certain topics that are important to them but potentially unfavorable in the public eye. Milita et al. (2017) suggest that in such cases, remaining silent on certain issues can be more beneficial than creating ambiguity (which would increase saliency). Additionally, Cormack (2013) and Gonzalez-Rostani (2023) demonstrate that politicians can either portray themselves ideologically or focus on specific issues, such as pro-redistributive policies, in their campaign communications, depending on the coverage of swing districts

and the targeting of certain audiences. This nuanced approach to communication reflects a calculated effort to balance personal policy preferences with the demands of vote-seeking and maintaining a favorable public image.

Empirical Strategy and Data

Our case focuses on the members of the Chilean Congress elected during the 2013 election. This group includes all members of the House of Representatives and the Senate. They assumed their legislative roles in March 2014. We utilize two primary sources of data for our analysis. First, to gauge legislators' preferences, we rely on the PELA-USAL (2018), which provides a comprehensive anonymous survey of legislators' characteristics and policy positions gathered after each legislative election.[4] Surveys are a conventional method for exploring voters' policy preferences or ideological views, and the PELA survey applies this methodology to legislators.

We analyze legislators' responses to a question about the relevance assigned to 12 critical issues, including inflation, education, health, and corruption, among others. Legislators rated the importance of these issues on a scale from 1 to 10.[5] For example, a rating of 10 for education and 8 for health indicates a higher significance of education for that legislator. We focus on these questions to examine the alignment between a politician's stated concern for an issue and its prominence in their communications with constituents. This data allows us to examine whether a legislator's stated concern for an issue correlates with its prominence in their public messaging.

Due to PELA's anonymity, we could not directly match legislators' responses to their Twitter postings. Consequently, we compared responses aggregated at the party level. We quantify the importance each party assigns to each issue by ranking the relevance of each topic to each legislator and then averaging these rankings by party. For instance, if education is the top issue for legislator A and the fourth for legislator B, their party's average ranking for education would be 2.5.

Second, to investigate elite public communication, we implement a scalable method to analyze politicians' statements via their Twitter activity. We initially compiled a list of elected officials and their Twitter handles, finding that 92% of legislators had an account at the time

[4]The 54th Congress is the most recent data available in the PELA survey and was conducted in 2014. For more details, refer to Appendix B.

[5]See the issues and the party rankings on Table B.3

we collected the data. Utilizing these handles, we downloaded legislators' past tweets through the Twitter API,[6] generating a comprehensive dataset of 122,245 tweets spanning from March 2014 to December 2014.[7] We rely on Twitter statements, as it is a reliable indicator of the significance legislators assign to different political issues, and to represent their ideology on par with legislative speeches and roll-call votes (Barberá et al., 2019).

Our approach includes an innovative strategy designed to categorize the content of these tweets, aligning them with the common issues assessed in the PELA survey. We employed OpenAI technology to identify them. We began by developing a prompt to recognize the issues of interest.[8] We then rigorously validated this prompt by randomly sampling tweets from our dataset, ensuring accurate identification and avoiding both the omission of relevant tweets and the mislabeling of others. Human coders were utilized for this validation phase. Once we generated a reliable prompt, we applied it across all statements, resulting in a categorical classification for each tweet.

This method represents a significant advancement over previous research methodologies, which have primarily relied on hand-coded content analysis (e.g., Casiraghi, 2021; Hawkins et al., 2019; Jenne, Hawkins, and Silva, 2021; Jagers and Walgrave, 2007) or automated text analysis using dictionaries (e.g. Pauwels, 2011; Elçi, 2019; Gründl, 2020; Gennaro, Lecce, and Morelli, 2021). For instance, while insightful, hand-coding is limited by its time-consuming nature and resource demands. Furthermore, it is susceptible to human biases, potentially impacting the reliability of measurements. Similarly, using dictionaries requires extensive knowledge of the case under study and carries the risk of researcher arbitrariness in word selection and of being highly context-specific. The approach of using predefined seed words or dictionaries is particularly problematic, as politicians often reference topics in various ways, with context playing a crucial role in understanding the issues being discussed.

Moreover, recent work has shown that ChatGPT can match or even outperform hand coding. Gilardi, Alizadeh, and Kubli (2023) demonstrates that ChatGPT classifies documents better than crowd-workers (especially M-Turk) in tasks involving annotation and topical classification of documents (Gilardi, Alizadeh, and Kubli, 2023; Kocoń et al., 2023). Furthermore, ChatGPT-3 has proven to be very efficient in identifying latent topics, such as hate speech.(Ji et al.,

[6] These tweets were retrieved in February 2023.

[7] In this analysis, we focused exclusively on legislators from the seven political parties represented in PELA (141 legislators). Out of these, 129 politicians possessed Twitter accounts. Among them, 97 actively tweeted during the analyzed period.

[8] See appendix C.1 for more details.

2023) and populism (Bellodi et al., 2023). The classification accuracy of this procedure can be visually corroborated by the word clouds representing each sub-group of tweets, as depicted in Figure D.1 in the appendix.

To determine the relevance of each topic to legislators, we categorize every Twitter statement as either pertaining to one of the predefined issues (as identified by PELA) or as unrelated. This method mirrors our approach to the PELA data, where we assess the salience of each topic in relation to the legislators' total body of tweets. We calculate the proportion of tweets dedicated to each topic during the entire period by each legislator. Subsequently, we rank the topics for each legislator based on these proportions. This ranking serves as an indicator of the importance legislators place on each topic.

We then employ BERTopic analysis in conjunction with OpenAI to classify tweets into topic clusters without imposing pre-defined structures. This approach enables us to automatically identify clusters that are not included in the PELA options but still demonstrate significant saliency. Our analysis focuses on uncovering these topics, as they may be more indicative of the interconnections within social media discourse. Additionally, this method informs the designers of elite surveys about potentially important issues, such as international affairs related to domestic policies, which are currently not covered in the survey but could be considered for future inclusion.

These procedures provide us with the necessary data to construct a tracker for monitoring issue salience, such as the daily ratio of legislators' tweets pertinent to topics like corruption, and to gauge the significance of each issue according to party affiliation. In what follows, we first present plots of the comparisons between stated survey preferences and political communication; then, we illustrate the potential of our methods by showing the dynamics over time on a specific issue and the content of some of the top issues discussed.

DISCUSSION

Figure 1 displays the distribution of legislators' rankings (based on frequencies) for the top 3 topics, as determined by their own parties' responses in PELA. We observe that the topics ranked highly in PELA are also frequently mentioned in the top 3 positions on Twitter. Figure 2 shows the distribution of legislators' Twitter rankings for the three least-relevant topics, according to their own parties. This figure reveals that topics considered least relevant in PELA also receive less attention on Twitter by the legislators of that party. This alignment between survey

preferences of elites and public Twitter statements suggests a convergence between the policy preferences expressed privately in surveys and those shared publicly. This finding carries two implications. First, it indicates a general alignment between private and public stances, supporting the notion that politicians may be more policy-driven or less driven by strategic behavior in their public communications. Second, it validates using Twitter as a reliable tool for assessing policy preferences, offering the added benefit of providing more time-sensitive, granular data.[9]

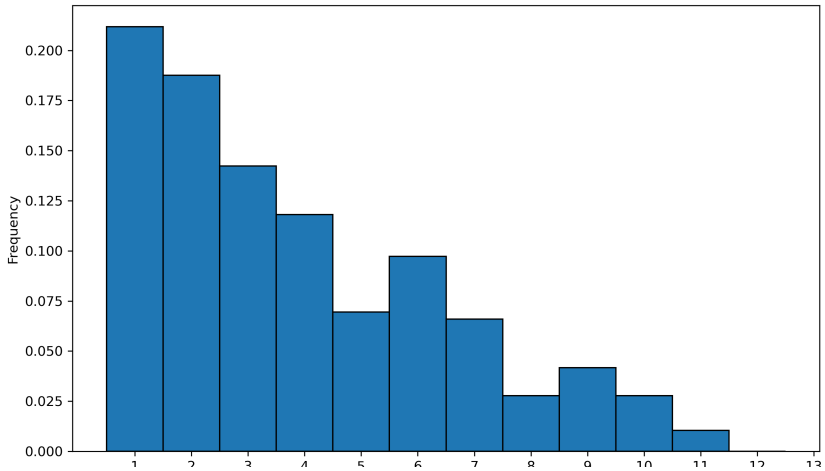

Figure 1: Histogram of Twitter Ranking Values for the Most Relevant Issues on PELA (Relative Terms).
Note: This figure displays topics that are ranked in the top 3 (priority 1 to 3) by parties in PELA (2014). The x-axis shows the corresponding rankings on Twitter along with their frequencies. The data comprise all tweets produced by members of Congress, collected from Twitter from Marc 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI.

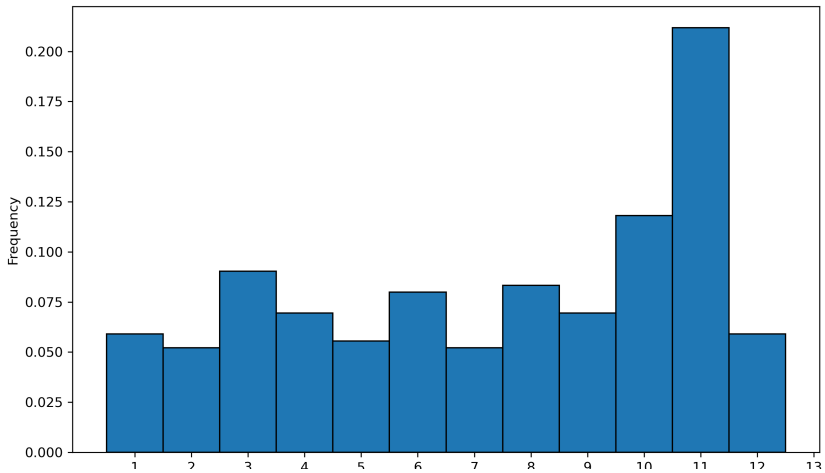

Figure 2: Histogram of Twitter Ranking Values for the Least Relevant issues on PELA (Relative Terms).
Note: This figure displays topics that are ranked in the top 3 (priority 1 to 3) by parties in PELA (2014). The x-axis shows the corresponding rankings on Twitter along with their frequencies. The data comprise all tweets produced by members of Congress, collected from Twitter from March 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI.

[9]This paper adds to others that have also found similar results regarding the usefulness of Twitter's data in studying different phenomena, such as identifying policy agendas and issues of importance to the political elite (Hemphill, Russell, and Schöpke-Gonzalez, 2021) or predicting legislator ideology (Barberá et al., 2019).

We then compare the issues' ranking given by the legislators (grouped by parties) in PELA versus the issue's saliency in Twitter.[10] Figure 3 shows for each topic the average distance between legislators' Twitter rankings (grouped by parties) and their respective parties' rankings in PELA. Positive values indicate higher rankings on Twitter than in PELA, implying less salience of the topic on Twitter than its measure of party concern in PELA. The dotted lines between -2 and 2 aim to highlight when rankings are similar, as a difference of 2 in a ranking is likely not meaningful.

While there is not much divergence, this plot still reveals variations across specific topics and parties. Notably, crime and drug trafficking appear as concerns for legislators in survey responses but are less salient in their daily political communications. Interestingly, this applies to parties in PELA regardless of their focus on crime, with PS averaging 5.07 and RN at 2.35. This evidence suggests that while legislators recognize the importance of some issues, they strategically prefer to state them in an anonymous survey rather than publicly mention them.

Conversely, the topic of conflict between powers (e.g., executive and legislative) seems less relevant in survey responses but is frequently addressed on Twitter, indicating higher saliency. However, this could be explained by our classifier algorithm, which classified tweets into this topic, often referring to other state powers as part of legislative activity and not necessarily about conflicts.

Other significant divergences, nearly two ranking levels apart, include inflation, where the right-wing parties RN and UDI highlight more on Twitter despite not considering it a primary issue (8.75 and 7.91 in PELA). RN's Twitter emphasis on education also differs from its lower PELA ranking of 5.5, compared to other parties' ratings below 3.75. This reflects the extensive Twitter dialogue on education in 2014, indicating RN's alignment with prevailing trends.

In sum, while there are broad similarities between public salience and survey responses, notable differences in particular topics and among specific parties are evident. Our approach can also open up new veins of analysis to analyze the divergences and similarities between issues and parties in more detail. As we have seen, there are issues where all parties choose to recognize their importance or talk about them in private and not in public or vice versa. On the other hand, thanks to the granularity of the data in social networks, we can pinpoint the strategic behaviors of legislators and parties in particular situations.

[10] Refer to Table D.4 for the saliency of each issue.

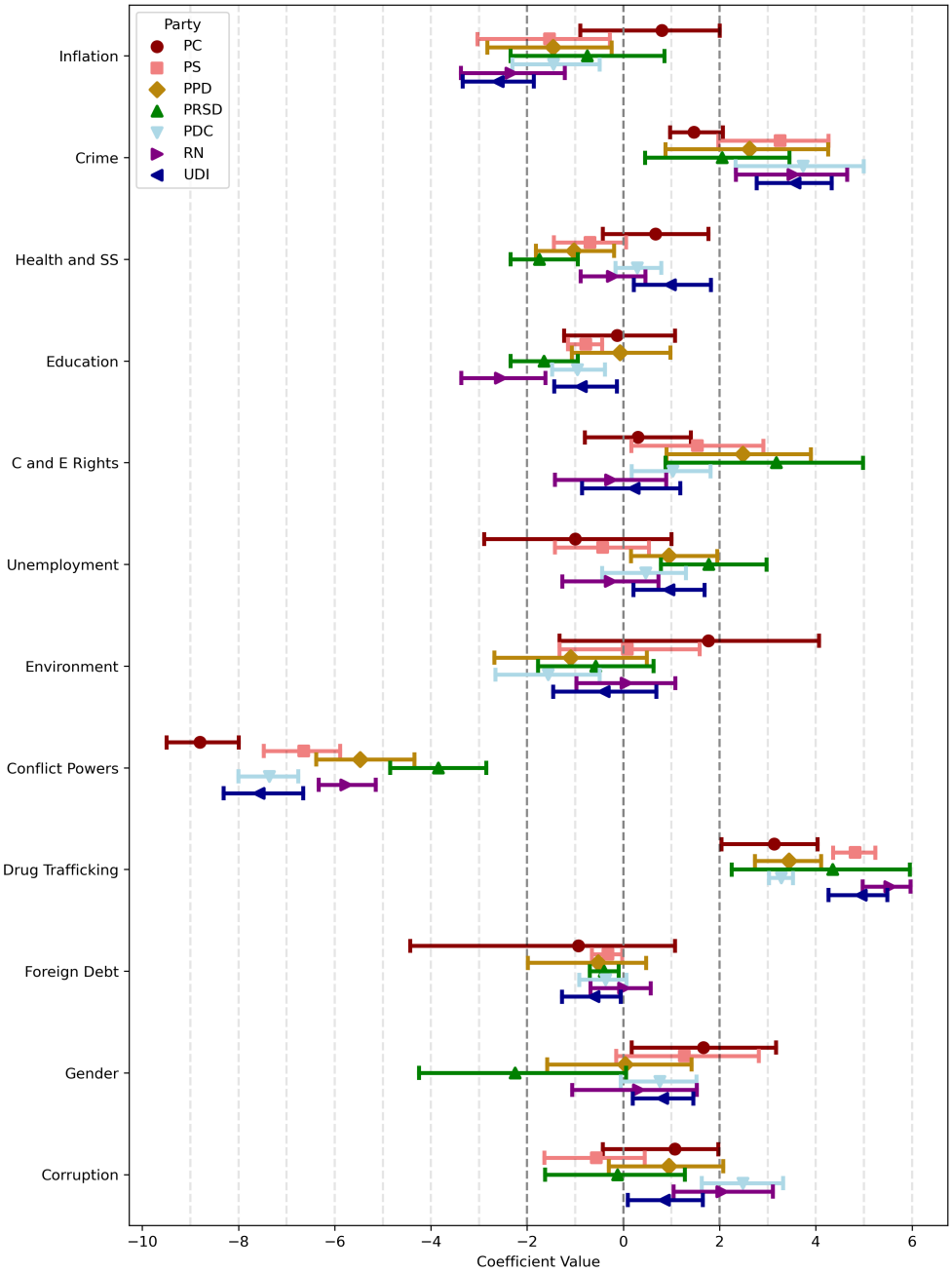

Figure 3: Coefficient Plot with Difference between Twitter and PELA Rankings

Note: This figure plots the distance between Twitter rankings and PELA (2014) rankings on the x-axis. It calculates each legislator's distance from their respective party's average ranking in PELA. A value of 0 indicates no difference between the rankings; positive numbers signify greater relevance on PELA than on Twitter; negative numbers indicate lesser relevance on PELA than on Twitter. The lines represent the 95% confidence intervals obtained through bootstrapping with 1000 simulations. The y-axis lists the issues included in the PELA survey, with different colors representing the various parties. The order of parties is based on ideology, ranging from the leftist (PC) to the rightist (UDI) See Table A.1. The data comprise all tweets produced by members of Congress, collected from Twitter from March 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI.

Finally, we employ BERTopic analysis to identify clusters of tweets by topic and use OpenAI to label each topic. Table F.6 shows the results of this analysis. Twitter statements predominantly focus on social media engagement (31.6%),[11] government issues and reforms (29.6%), and

[11] Here are included social media interactions such as greetings to their followers, promotion of events, statements regarding national holidays, etc.

the sharing of national news (10.6%). Our analysis identifies several issue-related clusters that are not covered by PELA, such as 'Venezuela and Human Rights,' 'Tax Reforms,' 'Natural Resources,' 'Labor Rights and Domestic Workers,' 'Public Transport,' and international affairs such as the Middle East conflict. These insights indicate the growing importance of international affairs to politicians and other domestic areas of interest, such as taxes and labor rights, providing crucial information for survey designers.

TRACING THE EVOLUTION OF ISSUE SALIENCE ACROSS TIME

To demonstrate the application and potential of our analysis, Figure 4 illustrates how our approach facilitates the evaluation of issue salience over time. We detail the daily proportion of tweets on education (Panel a) and gender (Panel b) from members of Congress. The plot indicates that education is more salient than gender, with noticeable peaks in salience at specific times.

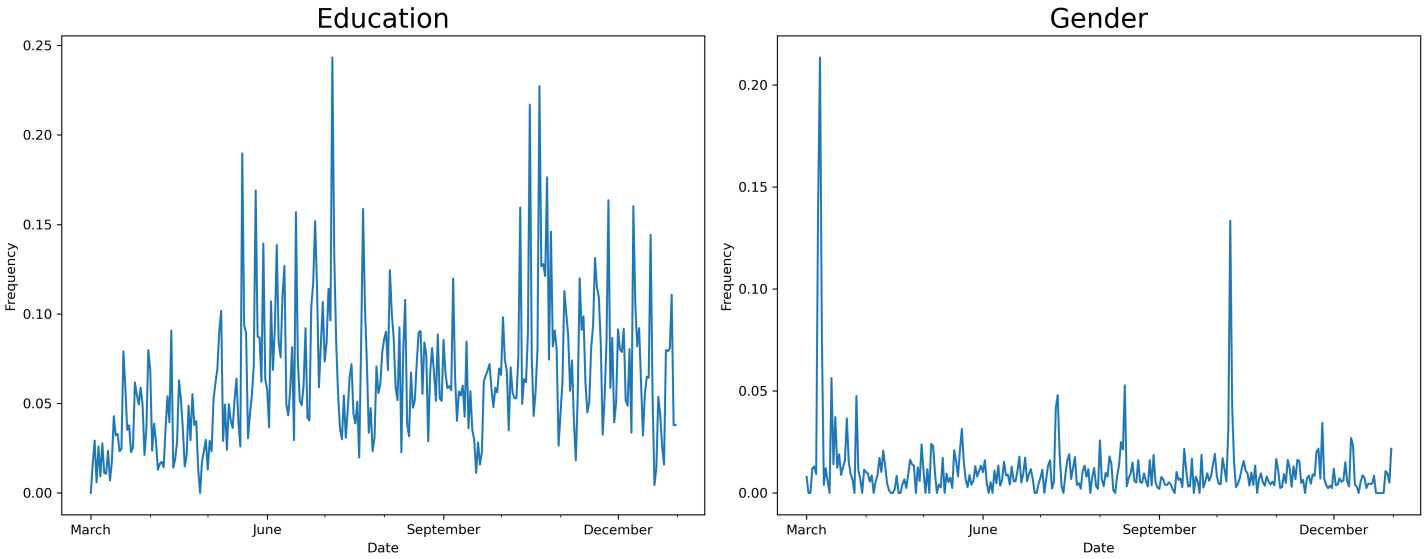

Figure 4: Salience of Education, and Gender in Members of Congress' tweets.
Note: This figure illustrates the shifting salience of education (Panel a) and gender issues (Panel b) over time in legislators' tweets. The data comprise all tweets produced by Members of Congress, collected from Twitter from March 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI. Saliency is calculated as the ratio of the number of tweets on a given day about topic i to the total tweets of that day. The x-axis represents time, while the y-axis represents frequency. See the evaluation of all topics in Appendix (Figure E.2).

Regarding education, a peak representing 25% of all daily tweets is observed at the beginning of August. This aligns with the Bachelet government's unveiling of an educational reform roadmap. A subsequent peak occurred on October 21, corresponding with the legislative approval of the Ley de Inclusión Escolar (School Inclusion Law; Law No. 20.845). This law fundamentally changes student admissions, terminates shared financing in state-funded schools,

and prohibits profit-making by these institutions. Similarly, tweets related gender issues rose to 20% on March 8th, International Women's Day, with another significant increase on October 10th, following the Chamber of Deputies' establishment of the Ministry of Women and Gender Equity. Consequently, our analysis effectively identifies temporal shifts in legislators' focus. These shifts particularly underscore issues such as gender, which, despite appearing less prominent in the PELA (2014) analysis, still exhibit sporadic peaks in salience.

CONCLUSION

This study demonstrates that analyzing legislators' daily political communications through social media provides valuable, real-time, and scalable insights into the microdynamics of political elites. By contrasting these communications with elites' stated preferences in PELA, we gain an understanding of which topics and parties are more likely to align or diverge in private and public contexts. This analysis also allows for a more detailed analysis of the dynamics of the public agenda. For example, it is possible to model the factors determining whether an issue prevails or ceases to be relevant over time. It is also possible to assess which issues are most likely to be disruptive in the short or medium term, which could inform projects such as PELA to include such issues in their field planning. Our proof of concept, mapping the daily discussions of the 54th Chilean Congress members in 2014, reveals significant trends in legislators' strategic behavior through simple descriptive analyses and the potential of our methodology.

Our methodology can be adapted to different legislative contexts and can address key social science questions about elite behavior. This includes examining the events that lead to greater divergence and exploring issues related to issue ownership within parties. This approach facilitates a comprehensive analysis of various legislator groups with high temporal precision and enables comparison with their declared preferences.

This paper pioneers the analysis of political speeches using OpenAI, which offers highly accurate classifications in multiple languages. This method is considerably less resource-intensive than hand-coding, and it is less arbitrary, and more updated than dictionary-based approaches. We expect that this tool will provide fellow academics, policymakers, and the public with a deeper understanding of how political elites tackle key policy issues. Future research could explore the temporal dynamics leading to significant deviations from stated preferences. It could also analyze which legislators are more prone to divergence, considering additional factors beyond party affiliations, like age and gender.

REFERENCES

Alesina, Alberto, and Alex Cukierman. 1990. "The Politics of Ambiguity." *The Quarterly Journal of Economics* 105 (4).

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker. 2019. "Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data." *American Political Science Review* 113 (4).

Bawn, Kathleen, Martin Cohen, David Karol, Seth Masket, Hans Noel, and John Zaller. 2012. "A theory of political parties: Groups, policy demands and nominations in American politics." *Perspectives on Politics* 10 (3).

Bellodi, Luca, Massimo Morelli, Antonio Nicolò, and Paolo Roberti. 2023. "The Shift to Commitment Politics and Populism: Theory and Evidence.".

Blum, Rachel, Lindsey Cormack, and Kelsey Shoub. 2023. "Conditional Congressional communication: how elite speech varies across medium." *Political Science Research and Methods* 11 (2).

Canes-Wrone, Brandice, and Kenneth W. Shotts. 2004. "The Conditional Nature of Presidential Responsiveness to Public Opinion." *American Journal of Political Science* 48 (4): 690–706.

Casiraghi, Matteo CM. 2021. "'You're a populist! No, you are a populist!': The rhetorical analysis of a popular insult in the United Kingdom, 1970–2018." *The British Journal of Politics and International Relations* 23 (4).

Cormack, Lindsey. 2013. "Sins of Omission: Legislator (Mis)Representation in Constituent Communications.".

Cormack, Lindsey. 2016. "Extremity in Congress: Communications versus Votes." *Legislative Studies Quarterly* 41 (3): 575–603.

Egan, Patrick J. 2013. *Partisan priorities: How issue ownership drives and distorts American politics*. Cambridge University Press.

Elçi, Ezgi. 2019. "The Rise of Populism in Turkey: A Content Analysis." *Southeast European and Black Sea Studies* 19 (3).

Farrell, Joseph. 1995. "Talk is Cheap." *The American Economic Review* 85 (2).

Fenno, Richard F. 1978. "Home style: House members in their districts." *(No Title)* .

Gennaro, Gloria, Giampaolo Lecce, and Massimo Morelli. 2021. Mobilization and the Strategy of Populism Theory and Evidence from the United States. Technical report CEPR Discussion Papers.

Gerber, Elisabeth R., and Rebecca B. Morton. 1998. "Primary election systems and representation." *The Journal of Law, Economics, and Organization* 14 (2).

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT outperforms crowd workers for text-annotation tasks." *Proceedings of the National Academy of Sciences* 120 (30): e2305016120. Publisher: Proceedings of the National Academy of Sciences.

Golbeck, Jennifer, Brooke Auxier, Abigail Bickford, Lautaro Cabrera, Meaghan Conte McHugh, Stephani Moore, Jacquelyn Hart, Justin Resti, Anthony Rogers, and Jenna Zimmerman. 2018. "Congressional twitter use revisited on the platform's 10-year anniversary." *Journal of the Association for Information Science and Technology* 69 (8).

Golbeck, Jennifer, Justin M. Grimes, and Anthony Rogers. 2010. "Twitter use by the U.S. Congress." *Journal of the American Society for Information Science and Technology* 61 (8).

Gonzalez-Rostani, Valentina. 2023. "Elections, Right-wing Populism, and Political-Economic Polarization: The Role of Institutions and Political Outsiders.".

Graham, Todd, Marcel Broersma, Karin Hazelhoff, and Guido van 't Haar. 2013. "Between Broadcasting Political Messages and Interacting with Voters." *Information, Communication & Society* 16 (5).

Grimmer, Justin. 2013. *Representational style in Congress: What legislators say and why it matters*. Cambridge University Press.

Grootendorst, Maarten. 2022. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint arXiv:2203.05794* .

Grose, Christian R., Neil Malhotra, and Robert Parks Van Houweling. 2015. "Explaining Explanations: How Legislators Explain their Policy Positions and How Citizens React." *American Journal of Political Science* 59 (3).

Gründl, Johann. 2020. "Populist ideas on social media: A dictionary-based measurement of populist communication." *New Media & Society* .

Hawkins, Kirk A., Ryan E. Carlin, Levente Littvay, and Cristóbal Rovira Kaltwasser. 2019. *The ideational approach to populism: Concept, theory, and analysis*. Routledge.

Hemphill, Libby, Annelise Russell, and Angela M. Schöpke-Gonzalez. 2021. "What Drives U.S. Congressional Members' Policy Attention on Twitter?" *Policy & Internet* 13 (2): 233–256. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/poi3.245.

Hemphill, Libby, Jahna Otterbacher, and Matthew Shapiro. 2013. What's congress doing on twitter? In *Proceedings of the 2013 conference on Computer supported cooperative work*. San Antonio Texas USA: ACM pp. 877–886.

Holbrook, Thomas M., and Scott D. McClurg. 2005. "The Mobilization of Core Supporters: Campaigns, Turnout, and Electoral Composition in United States Presidential Elections." *American Journal of Political Science* 49 (4): 689–703.

Jacobs, Lawrence R., and Robert Y. Shapiro. 2000. *Politicians don't pander: Political manipulation and the loss of democratic responsiveness*. University of Chicago Press.

Jagers, Jan, and Stefaan Walgrave. 2007. "Populism as political communication style. An empirical study of political parties' discourse in Belgium.".

Jenne, Erin K., Kirk A. Hawkins, and Bruno Castanho Silva. 2021. "Mapping Populism and Nationalism in Leader Rhetoric Across North America and Europe." *Studies in Comparative International Development* 56 (2): 170–196.

Ji, Yunjie, Yan Gong, Yiping Peng, Chao Ni, Peiyan Sun, Dongyu Pan, Baochang Ma, and Xiangang Li. 2023. "Exploring ChatGPT's Ability to Rank Content: A Preliminary Study on Consistency with Human Preferences.". arXiv:2303.07610 [cs].

Jones, Bryan D., and Frank R. Baumgartner. 2004. "Representation and Agenda Setting." *Policy Studies Journal* 32 (1): 1–24.

Kastellec, Jonathan P., Jeffrey R. Lax, Michael Malecki, and Justin H. Phillips. 2015. "Polarizing the Electoral Connection: Partisan Representation in Supreme Court Confirmation Politics." *The Journal of Politics* 77 (3): 787–804.

Kocoń, Jan, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, Anna Kocoń, Bartłomiej Koptyra, Wiktoria Mieleszczenko-Kowszewicz, Piotr Miłkowski, Marcin Oleksy, Maciej Piasecki, Łukasz Radliński, Konrad Wojtasik, Stanisław Woźniak, and Przemysław Kazienko. 2023. "ChatGPT: Jack of all trades, master of none." *Information Fusion* 99: 101861.

Martínez-Gallardo, Cecilia, Nicolás de la Cerda, Jonathan Hartlyn, Liesbet Hooghe, Gary Marks, and Ryan Bakker. 2022. "Revisiting party system structuration in Latin America and Europe: Economic and socio-cultural dimensions." *Party Politics* .

Mayhew, David R. 2004. *Congress: The electoral connection*. Yale university press.

Milita, Kerri, Elizabeth N. Simas, John Barry Ryan, and Yanna Krupnikov. 2017. "The effects of ambiguous rhetoric in congressional elections." *Electoral Studies* 46: 48–63.

Munger, Kevin, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2019. "Elites Tweet to Get Feet Off the Streets: Measuring Regime Social Media Strategies During Protest." *Political Science Research and Methods* 7 (4): 815–834.

Pauwels, Teun. 2011. "Measuring Populism: A Quantitative Text Analysis of Party Literature in Belgium." *Journal of Elections, Public Opinion and Parties* 21 (1): 97–119.

PELA-USAL. 2018. "Base de Datos de Elites Latinoamericanas : Parliamentary Elites in Latin America Observatory (PELA-USAL)." *Universidad de Salamanca* .

Scherpereel, John A., Jerry Wohlgemuth, and Audrey Lievens. 2018. "Does Institutional Setting Affect Legislators' Use of Twitter?" *Policy & Internet* 10 (1): 43–60.

Shapiro, R., and L. Jacobs. 2000. "Who leads and who follows." *Decisionmaking in a glass house: Mass media, public opinion, and American and European foreign policy in the 21st century* 223.

Siegel, Alexandra, David Laitin, Duncan Lawrence, Jeremy Weinstein, and Jens Hainmueller. 2022. "Tracking Legislators' Expressed Policy Agendas in Real Time.".

Stimson, James A., Michael B. MacKuen, and Robert S. Erikson. 1995. "Dynamic representation." *American political science review* 89 (3).

Wright, Gerald C. 1989. “Policy voting in the US Senate: Who is Represented?” *Legislative Studies Quarterly* .

Wu, Patrick Y, Jonathan Nagler, Joshua A Tucker, and Solomon Messing. 2023. “Large Language Models Can Be Used to Estimate the Latent Positions of Politicians.”.

ONLINE APPENDIX

Contents

A Descriptive Data: Chilean Parties **18**

B PELA **18**
B.1 Data Description 18
B.2 Ranking 19

C Empirical strategy **20**
C.1 Prompt 20

D OpenAI Topics **21**
D.1 Ranking Calculation using Twitter 23

E Mapping Issues Over Time **24**

F BERTopic **24**

A DESCRIPTIVE DATA: CHILEAN PARTIES

	L-R Ideology.
Partido Comunista de Chile (PC)	1.22
Partido Socialista de Chile (PS)	3.11
Partido por la Democracia (PPD)	3.61
Partido Radical Socialdemócrata (PRSD)	3.88
Partido Demócrata Cristiano (PDC)	5.00
Renovación Nacional (RN)	7.11
Unión Demócrata Independiente (UDI)	8.94

Table A.1: Parties Chile
Note: The ideological position goes from 1 (extreme left) to 10 (extreme right). Source: Author's own elaboration based on data from Chapel Hill Expert Survey: Latin America (CHES:LA, Martínez-Gallardo et al., 2022)

B PELA

B.1 Data Description

A few databases help us understand the positions of legislators and parties in Latin America. The most notable effort is the Latin American Parliamentary Elites Project (PELA), which has collected legislators' opinions in 18 countries in the region since 1994 (PELA-USAL, 2018).

PELA is an invaluable resource that has contributed to the production of knowledge about legislators, their positions, and their relationships with parties. It has been the source of a vast academic literature [12]. PELA conducts one survey per legislature per country, deploying academics, including professors and graduate students, to conduct in-person interviews.

[12] For more reference: https://oir.org.espela/en/publications/

Party	Number of legislators	Mean ideology
UDI	17	7.65
RN	10	7.50
PDC	11	4.73
PRSD	4	4.25
PPD	9	3.75
Otros	7	3.57
PS	7	2.29
PC	3	1.00

Table B.2: PELA's coverage
The number of legislators represents the legislators by party interviewed during the PELA field. The Mean Ideology column represents the average by a party of the legislators' self-ratings on the left-right continuum, where one is left and ten is right. Source: PELA-USAL (2018)

The PELA survey was administered to a total of 68 legislators who were part of the 54th Chilean Congress. The following table details the number of legislators by party.

B.2 Ranking

For the ranking construction, the scores given by each legislator to the problems included in the PELA questionnaire were averaged by party.

Topic	Other	PC	PDC	PPD	PRSD	PS	RN	UDI
Inflation	9.07	8.50	8.50	8.33	9.25	9.29	8.75	7.91
Crime	5.71	8.33	3.36	4.17	5.75	5.07	2.35	2.35
Health and SS	1.86	2.83	2.05	3.61	3.75	3.07	2.95	2.71
Education	2.50	2.83	3.36	2.61	3.75	2.57	5.50	3.00
C and E Rights	6.57	5.50	6.55	4.56	6.13	5.71	8.30	8.56
Unemployment	6.36	7.50	6.23	6.89	5.13	6.43	6.65	5.35
Enviroment	6.71	5.33	6.09	7.39	4.88	6.00	7.20	7.50
Conflict Powers	10.36	10.50	10.45	8.89	6.25	10.14	7.90	10.74
Corruption	6.43	3.83	6.09	5.89	8.63	7.64	4.80	5.59
Drug Trafficking	6.50	7.17	7.95	7.22	5.75	6.14	5.50	5.24
Foreign Debt	11.57	10.83	11.23	10.78	10.50	11.57	10.75	11.18
Gender	4.36	4.83	5.82	6.83	8.25	4.36	7.35	7.65

Table B.3: Distribution of topics by parties
Note: Source: PELA & The score by topic represents the average ranking legislators gave to each topic.

C EMPIRICAL STRATEGY

Our Python function configures ChatGPT3-5 with a few instructions (detailed in the next section) and connects with OpenIA API. The function processes all tweets and executes the delivered instructions. Approximately 23% of the tweets are classified into pre-set topics (based on the topics proposed by PELA), and the remaining 77% remain unclassified.

We applied BERTopic to perform a latent topic analysis in the second step. Then, we connect to the OpenIA API to generate readable descriptions for the latent topics based on the most relevant words of these topics.

C.1 Prompt

The prompt we use to configure the language model parameters is detailed below. It is a complete list of instructions to avoid ambiguities during classification.

"Please classify the content of tweets from Chilean congressmen. Assign a number from the following list of topics based on the central theme or issue of the tweet. If the tweet's content does not align with any of these topics, assign a 0. The topics, along with some indicative keywords, are:

1. Inflación (Inflation) - Keywords: precios, costo de vida, aumento, economía, alza de precios, incremento, subida, devaluacion, alza de canasta basic
2. Inseguridad Ciudadana y delincuencia (Public insecurity and crime) - Keywords: seguridad, delitos, policía, crimen, desconfianza, delincuencia, vulnerabilidad, temor
3. Salud/Seguridad Social (Health/Social Security) - Keywords: hospitales, médicos, pensiones, salud pública, jubilaciones, covid, virus, vacunas, bienestar, paciente, essalud
4. Educación (Education) - Keywords: escuelas, universidades, estudiantes, reforma educativa, alfabetizacion, profesor, enseñanza, aprendizaje
5. Derechos de los grupos étnicos y culturales (Rights of ethnic and cultural groups) - Keywords: indígenas, cultura, lengua, diversidad, minorias, discriminacion, pluricultural, identidad
6. Desempleo y Subempleo (Unemployment and Underemployment) - Keywords: trabajo, empleo, economía laboral, oportunidades, tasa de desempleo, crisis laboral, subempleo, informal
7. Medio Ambiente (Environment) - Keywords: naturaleza, contaminación, protección ambiental, cambio climático, arboles, aves, incendios, sostenibilidad, biodiversidad, conservation
8. Conflictos entre los poderes del Estado (Conflicts between the powers of the State) - Keywords: congreso, gobierno, ley, constitución, separacion de poder, institucional, autonomia, legislativo, ejecutivo, judicial
9. Narcotráfico (Drug trafficking) - Keywords: drogas, narcóticos, fronteras, policía, narcotrafico, pandillas, delincuencia organizada, cocaína
10. Deuda Externa (External Debt) - Keywords: préstamos, FMI, crédito, finanzas, deuda externa, prestamo, banco mundial, financiamiento
11. Las desigualdades entre hombre y mujeres (Gender Inequality) - Keywords: género, igualdad, derechos de la mujer, brecha salarial, empoderamiento, feminismo, roles, equidad

12. Corrupción (Corruption) - Keywords: bribes, graft, dishonesty, embezzlement, kickbacks, nepotism, fraud, scandal, corrupt.

Analyze the tweets provided below, and for each, indicate only the topic number(s) it pertains to (NEVER A TEXT), based on the central theme of the tweet in relation to the top cs and keywords listed. If the tweet is unrelated to these topics, or if you're unable to determine the topic due to lack of context or clarity, assign a 0. Ensure to provide a classification only for tweets that have a clear and definite relation to the topics.

Remember, the classification should be based on concrete policy or political issues referenced in the tweet, not on general expressions or sentiments. Do it from a Chilean Perspective. Provide only the number(s) of the relevant topic(s), nothing else.")

D OpenAI Topics

OpenAI API allows access to language models and image generation through different Python packages.[13] ChatGPT's large language model has over 175 billion parameters that have been trained using a large amount of text from the Internet and other sources. The model has been trained using Reinforcement Learning from Human Feedback (RLFH).

The constructed function connected to the OpenAI API, and based on the prompt noted in the previous section, classified the tweets into one of the topics queried by PELA. The following table shows the frequency of tweets for each topic. The "Other" category includes all tweets that were not classified into any of the PELA topics.

Topic	Frequency (%)
Education	6.82%
Health and SS	4.36%
Conflict Powers	4.18%
Environment	1.68%
Corruption	1.22%
Crime	1.22%
Inflation	1.20%
Unemployment	1.16%
Gender	1.12%
C and E Rights	0.86%
Drug Trafficking	0.13%
Foreign Debt	0.08%
Other	76.77%

Table D.4: Frequency of Issues Over the Total Number of Tweets in the Period.
Note: This table illustrates the share of every issue that appeared in PELA 2014. The data comprise all tweets produced by Members of Congress, collected from Twitter from March 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI. Source: Tweets

To validate our procedure of classifying tweets into topics using OpenAI, we plotted the frequencies of the most important words for each topic using word clouds. As can be seen visually, the most frequent words in the corpus of each subgroup of tweets have a relationship with the topic to which they were classified.

[13]Here are some examples: https://platform.openai.com/overview

Figure D.1: Most Frequent Words by Issue

Note: These figures plot the most frequent words for each of the topics in the analyzed period. The data comprise all tweets produced by Members of Congress, collected from Twitter from March 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI.

D.1 Ranking Calculation using Twitter

	PC	**PDC**	**PPD**	**PRSD**	**PS**	**RN**	**UDI**
Inflation	9.30	7.05	6.88	8.50	7.75	6.41	5.30
Crime	9.80	7.10	6.79	7.80	8.33	5.88	5.92
Health and SS	3.50	2.33	2.58	2.00	2.38	2.72	3.68
Education	2.70	2.40	2.54	2.10	1.79	2.94	2.12
C and E Rights	5.80	7.57	7.04	9.30	7.25	8.03	8.78
Unemployment	6.50	6.69	7.83	6.90	6.00	6.38	6.30
Environment	7.10	4.52	6.29	4.30	6.08	7.25	7.10
Conflict Powers	1.70	3.10	3.42	2.40	3.50	2.12	3.16
Drug Trafficking	10.30	11.24	10.67	10.10	10.96	11.03	10.18
Foreign Debt	9.90	10.86	10.25	10.10	11.25	10.75	10.56
Gender	6.50	6.57	6.88	6.00	5.62	7.66	8.46
Corruption	4.90	8.57	6.83	8.50	7.08	6.84	6.44

Table D.5: Average Ranking by Party
Note: We average the legislators ranking for each topic on Twitter by party. Source: Tweets

E Mapping Issues Over Time

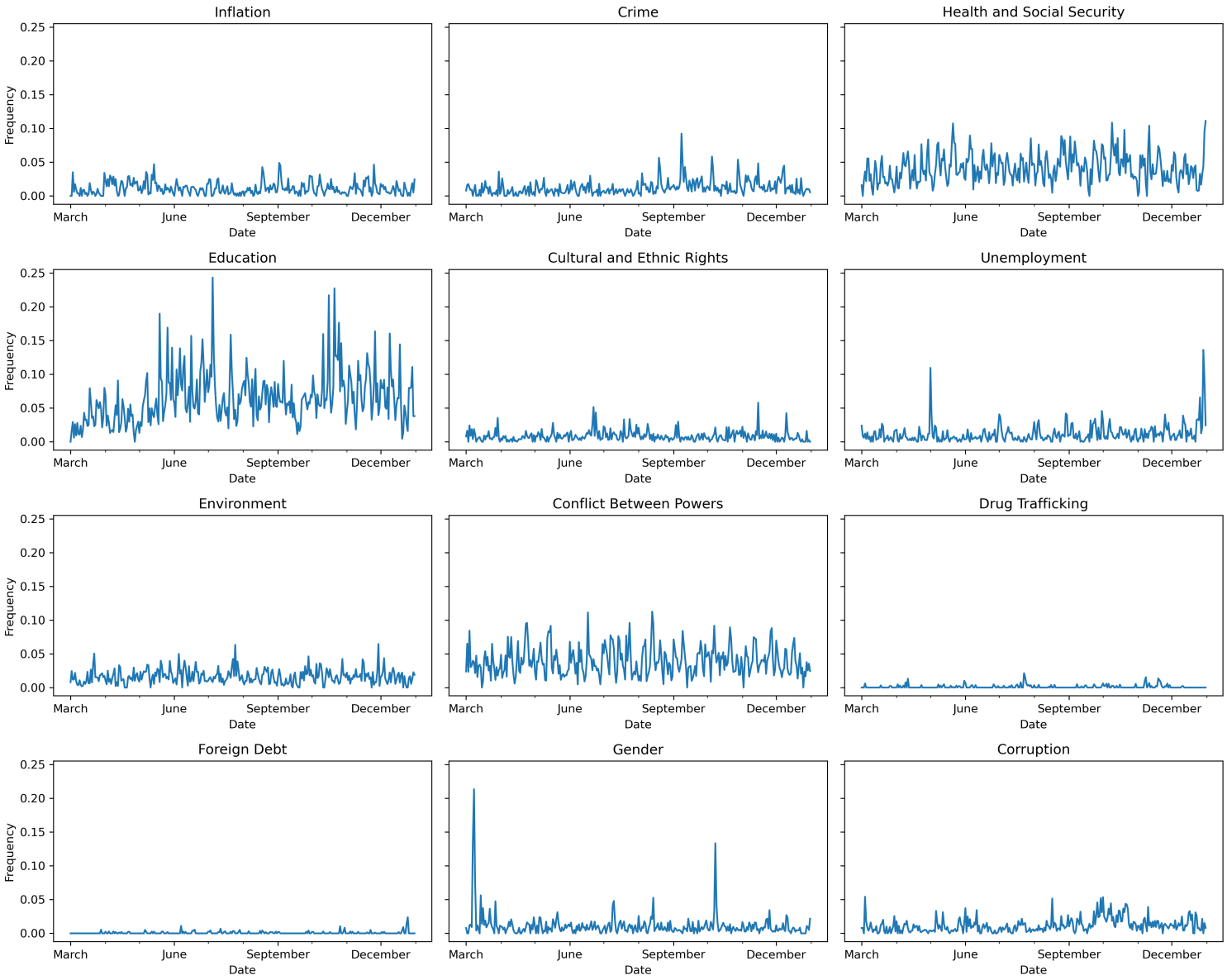

Figure E.2: Salience of PELA issues in Members of Congress' Tweets.
Note: This figure illustrates the shifting salience of every issue that appeared in PELA 2014 over time in legislators' tweets. The data comprise all tweets produced by Members of Congress, collected from Twitter from March 2014 to December 2014. Tweets have been classified as relevant to these issues using OpenAI. Saliency is calculated as the ratio of the number of tweets on a given day about topic i to the total tweets of that day. The x-axis represents time, while the y-axis represents frequency.

F BERTopic

We employed a clustering technique that leverages HuggingFace transformers and TF-IDF, as demonstrated (BERTopic, Grootendorst, 2022), to identify the underlying semantic structure and latent themes within the discourse of immigration. BERTopic, built upon the Bidirectional Encoder Representations from Transformers (BERT) architecture, offers an advanced approach to extracting and categorizing latent topics from textual data. Unlike conventional methods, BERTopic captures contextual relationships between words, resulting in more coherent and

interpretable topics. We perform clustering using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). The language is set to 'multilingual,' and the number of topics we determined was 40. This means that after training the topic model, the number of topics was reduced to 40 using a c-TF-IDF calculation.

Once clusters are established, we employ the OpenAI API to retrieve topic representations from the documents within each cluster and identify keywords for each cluster.

Table F.6: Distribution of Topics by Share

Topic	Share
Social Media Engagement	31.6%
Reform and Policy Discourses	29.6%
National News Distribution	10.6%
Venezuela and Human Rights	6.9%
Tax Reforms and its implications	6.6%
Natural resources (e.g., water)	6.2%
Congress and sport initiatives	3.1%
Free Software and Technology	2.9%
Public Health Infrastructure	2.5%
Gender Equity and Women's Ministry	2.4%
Enhancing Quality of Life	2.3%
Labor Reform and Domestic Workers	2.3%
Disaster Response and Aid	2.1%
Radio Broadcasts and Interviews	2.1%
Media Engagement and Interviews	2.0%
Middle East Conflict	1.8%
Public Transport Discussions	1.8%
Pet Ownership Responsibility	1.4%
Community Support and Assistance	1.4%
Food Industry and Health	1.2%
Cannabis Legalization Debate	1.1%
Broadcasting Pre-Announcements	1.1%
Energy Policy Discussions	1.1%
Civic Engagement and Public Safety	1.0%
Remote Areas and Environmental Protection	1.0%
Civil Union and Equality Rights	0.9%
Therapeutic Abortion Debate	0.6%
Disability Rights and Inclusion	0.6%
Organ Donation Awareness	0.4%
Electoral System Analysis	0.4%
Breast Cancer Awareness	0.4%
Expressions of Gratitude	0.4%
Cultural and Social Reflections	0.4%
Cultural Programs and Tributes	0.4%
Educational Disruptions	0.2%
Taxation and Fiscal Policy	0.2%
Public Health Preparedness	0.2%
Protests and Indigenous Rights	0.2%
Health and Public Service Announcements	0.2%

Note: This table showcases the most relevant topics identified using BERT OpenAI topic analysis. The data comprise all tweets produced by Members of Congress, collected from Twitter from March 2014 to December 2014. 'Topics' refers to the automatic labels generated by OpenAI, based on representative terms and documents; 'Share' denotes the proportion of tweets in each cluster relative to the total number of tweets.