

# Discursive Superstructures of Ideology by Brazilian Political TikTokers

Marcos Echevarria-Eirea\*<sup>1</sup> and João Cardoso L. Camargos<sup>†2</sup>

<sup>1</sup>*European University Institute*

<sup>2</sup>*The University of North Carolina at Chapel Hill*

---

\*[marcos.echevarriaeirea@eui.eu](mailto:marcos.echevarriaeirea@eui.eu)

†[joao\\_c\\_camargos@unc.edu](mailto:joao_c_camargos@unc.edu)

## Abstract

The seminal work of Converse (1964) alludes to the importance of political “pundits” for the formation and dissemination of ideological belief systems. Despite their theorized importance to the nature of ideology, there is still little empirical evidence corroborating their impact on “discursive superstructures” that cohere sets of plausibly independent policy stances. This is particularly the case within the—arguably already established—political arena of social media, where such processes are likely to occur organically among political elites and citizens. To the best of our knowledge, only Green (2024) studies these platforms as the “agora” of rhetorical structure formation, but he focuses solely on X, the US, and does not fully delve into political influencers outside of traditional experts as potential pundits. To fill these gaps, we employ TikTok data in Brazil, with which we test the presence and coherence of political discursive superstructures formed by Political Social Media Influencers. We apply BERTopic, an embeddings-based machine learning model that captures the meaning of individual words and their contextual use in sentences, to categorize the content of posts into topics. Understanding discourse structures as networks, we employ pivot scaling to position content pertaining to different topics in a low-dimensional space, and Mixed Graphical Models (MGMs) to form networks describing the presence and coherence of a discursive superstructure. Preliminary results point to the existence of a discursive superstructure of politics by Brazilian political influencers.

**Keywords:** discursive superstructures; belief systems; Political Social Media Influencers; Tik-Tok

## Introduction

In Converse's (1964) foundational work, ideology is conceived as a structured system of beliefs, where attitudes and ideas are constrained together in an apparently coherent manner. Individuals, however, do not inherently identify the constraints that bind these systems together, nor are prevailing relationships among issues obvious *a priori*. These connections and boundaries are socially constructed before they can be learned, which occurs through the development of "discursive superstructures" in political systems; citizens' engagement with these structures determines their capacity to map their material interests and psychological dispositions onto existing political cleavages (Jost et al. 2009).

Despite their importance, the scrutiny of discursive superstructure formation remains limited. This is in part due to the difficulty of pinning down the 'minuscule proportion' of the population (Converse 1964, p. 8) that is responsible for cohering collections of interests into full-fledged ideologies (Noel 2012), as well as a suitable medium in which they do so. To study them, we leverage the rising relevance of social media as a political sphere (Highfield 2016) and, in particular, the impact of online influencers. We reconceptualize Political Social Media Influencers (PSMIs) as a new group of pundits capable of engaging in the "creative synthesis" (Converse 1964) of discursive superstructures and belief systems. We also move beyond the US as the locus of discursive superstructures (and social media) research (**other**; Green 2024) by concentrating on Brazil as a case study. This is an interesting context for comparison with the US because Latin America is a region where ideology is typically considered non-structured, party systems are unstable (Mainwaring & Scully 1995), and political cues are missing (Zechmeister & Corral 2013). **Emphasize further the move beyond traditional pundits and why it might be important (e.g., newspaper article/TV talk show vs TikTok post).**

With these goals in mind, we design a novel approach for sampling PSMIs. The method consists of capturing the most influential Brazilian political "TikTokers" by measuring the quantity and quality of their output within a known "online political space", which we observe with specific hashtag markers. With this approach, we identify 108 PSMIs and collect a total of 5,533 of their posts during the election period. We then categorize PSMI content into topics with BERTopic (Grootendorst 2022), an embeddings-based machine learning model that

captures the meaning of individual words and their contextual use in sentences. By understanding discursive superstructures as networks of political themes that “speak to each other”, we capture their presence with the help of Mixed Graphical Models (MGMs) (Haslbeck & Waldorp 2020). Finally, we assign ideological positions with pivot scaling (Green 2024) to build networks of positional variance that describe the consistency of relationships across discursive superstructures, thus testing their coherence.

Preliminary results indicate that Brazilian political TikTokers are contributing to a discursive superstructure of beliefs. **Implications for non-expert pundits and discursive superstructure formation.** Furthermore, these findings align with recent evidence that ideologies in Latin America are less diffused than often assumed, even if they rest on different dimensions and connections than in the Western world (de la Cerda 2024; Martínez-Gallardo et al. 2023).

## **Theory**

### **Belief Systems and Discursive Superstructures of Ideology**

Converse (1964) conceives ideology as a system of beliefs, a configuration of ideas and attitudes in which some form of constraint or functional interdependence binds the (policy) elements together. Such a definition entails that citizens can organize their political beliefs in a structured manner, where certain beliefs lead to others, and are hence “coherent”. As a result, we would expect, in the US today, for instance, that a conservative identification or ideology typically entails a more stringent stance on immigration.

Here, however, we are not interested in the partisan or identity valence that different systems might take, but rather how, and if, these systems come to form coherently. Building on Green (2024), we seek to understand why these policy packages are apparently logical to voters. If issue stances do not *need* to deductively follow each other and yet they form ideological relationships, how do constrained belief systems hold together? In other words, how do the bounds of belief systems come to be?

Green (2024) turns to political coalitions as the antecedent of ideologies that prescribe constrained sets of beliefs. Policy preference necessitates cooperation to advance a common agenda in a majority-grounded democratic environment. In turn, to build and maintain coali-

tions, there must be mechanisms that permit the signaling of this continued cooperation. For political parties, for example, these tend to be formal, as they periodically meet to establish and vote for a policy platform. During these processes, it is coalition members who establish and assign the ideas and values that prescribe their diverse policy demands (Noel 2012).

The resulting collection of ideas, principles, and norms that coalitions generate to cohere their disparate policy stances creates a “discursive superstructure” of ideology, which, in combination with a “motivational substructure”, helps citizens organize their belief systems in politically meaningful ways (Green 2024). This understanding of discursive superstructures mirrors the notion of a network, where different ideas (nodes) should speak to each other (edges) to form a coherent structure of beliefs. Similarly, other authors have considered ideology as interrelated networks of beliefs, opinions, and values (Boutyline & Vaisey 2017; Jost et al. 2009). Therefore, we may employ network models to ascertain their existence and coherence.

#### **Theoretically define presence + coherence.**

Testing for the presence of these structures is especially compelling in Latin America. The patterns of ideological attachment in this continent are particularly relevant, where political parties and party systems are weaker, in contrast to the US and Western Europe. Latin American voters often exhibit higher electoral volatility, which reflects their institutional fragility and fluidity of political representation (Mainwaring & Scully 1995), making ideological consistency more elusive. This also occurs in our case study, Brazil (Moreira & Camargos 2025). Evidently, however, it does not hold in the same magnitude for every country in the region (Mainwaring & Pérez-Liñán 2014), and new evidence suggests that Latin America’s parties are as structured as those in Europe, albeit only *vis-à-vis* an economic dimension (Martínez-Gallardo et al. 2023). Moreover, emerging evidence suggests that ideological beliefs in Latin America may be shaped less by formal party systems and more by alternative identity-driven movements, such as charismatic leadership and ideological coalitions (de la Cerda 2024). Brazil is thus an interesting, arguably hard, scenario to evaluate the concept of discursive superstructures, especially on social media, where research has concentrated on the US (Green 2024).

## Political Social Media Influencers

Having covered the importance of discursive superstructures to belief systems and their formation mechanisms, we now consider who—and how—performs such creation processes. Like Green (2024), we focus on “political pundits”, as opposed to parties and politicians, whose communication strategy suffers from the professional implications of politics, such as appealing to pivotal voters, which limits their scope of discussion.

Political pundits are ‘informal political elites who disseminate political analyses and opinions to audiences they have cultivated for that purpose’ (Wan and Green 2023, p. A2). These are *informal* in that they do not directly pursue office, and *elites* because audiences seek out their perspectives on political issues. Pundits face incentives to be ideological first and partisan second, even when clearly aligned with a party (Green 2024). Although they do not need to “win” as many voters as possible, they do strive to cultivate and maintain their audiences. To do so, they must develop reputations for persuasively defending and advancing their worldviews to the public, and thus must discuss a wide range of issues that might not necessarily reference public policy directly, while remaining within ideological group norms (Green 2024). It is through these “proposals” that they render plausibly independent issue stances ideologically coherent, making them ideological entrepreneurs in the “creative synthesis” of discursive superstructures and, ultimately, belief systems (Converse 1964).

However, we go beyond Green’s (2024) classification of pundits by delving deeper into the context of social media, turning our attention to Political Social Media Influencers (PSMIs). These are not necessarily “traditional pundits” in the sense of opinion leaders of some standing, expertise, or recognition, but rather users with a consistent political output on social media; in other words, potentially anyone. Whereas Green (2024) studies a known set of pundits with a certain level of public recognition, we consider any possible commentator, including those with no prior or acknowledged understanding or experience of politics. We thus gather a much more “amateur” population of opinion entrepreneurs. These are nevertheless an important addition to the classic conceptualization of “pundit”, emergent from the advent of social media and its increasing politization (Sehl & Schützeneder 2023). Indeed, influencers have been shown to affect the information and opinion-forming behaviors of young people (Andı 2021; Schmuck

et al. 2022) and the perceived social duty to keep informed (Wunderlich et al. 2022). Similar to traditional pundits, PSMIs seek to draw other people’s attention to their own ideas and opinions (Kümpel et al. 2015), and inform their audiences on different themes and issues (Sehl & Schützeneder 2023). They discuss explicit issues that refer to public policies, but also cover lifestyle-based topics that may be of collective concern and impact society (Suuronen et al. 2022), helping assimilate topic and belief relationships. **Is this akin to the TV revolution?**

The literature on PSMIs does not provide an entirely concrete, self-contained definition of this type of influencers (Bause 2021; Riedl et al. 2023; Sehl & Schützeneder 2023). We follow the general understanding of PSMIs as content creators on social media whose content is clearly centered around politics. They can be distinguished from other influencers “in their willingness to associate their online influence with political and social causes” (Goodwin et al. 2023). Their followership is normally platform-built (**sourceriedle**), unlike celebrities who are generally famous prior to and regardless of joining a platform (e.g., actors, athletes). However, this is not always the case. For example, journalists and commentators are normally known for their work outside of social media, but may employ these platforms to voice their political opinions and enhance their readership, thus becoming PSMIs.

Based on our theoretical framework, and given that this remains somewhat generic, we specify two additional requisites. (1) PSMIs must not be part of or running for an elective institution or office-seeking group (i.e., parties). As such, we do not include politicians, for example, as these hold or contest power (Green 2024). We do, however, include ex-politicians who turn influencers, as long as they devote their content to political discussion and do not (officially) seek (re-)election. (2) PSMIs must be an individual or a small group of identifiable people. **This criterion ensures that they remain political “elites”, purposely sought after and followed by an audience. You are recognizable, not for your expertise, but for your content.** Organizations, including newspapers and news channels, that are not attributable to one individual or a small group of individuals are excluded. Likewise, meme accounts that re-post and edit political content—both humorous and serious—can display complex narratives and ideas (e.g., Munger and Li 2025), but do not conform to the definition of a pundit if they are anonymous/non-identifiable. Instead, we do include identifiable users who use humor and

memes to deliver their own original political content.<sup>1</sup>

### *TikTok*

Political pundits communicate across numerous and differing venues. We focus on social media because we aim to incorporate a broader political discussion from which to infer a discursive superstructure—while not all pundits can access or host talk programs and shows, or write books and articles, social media platforms present an incredibly accessible avenue of communication with the wider public.

Although research is beginning to emerge, TikTok as a political app remains relatively understudied. The platform reached an estimated 2 billion users in 2024 (Ceci 2025), with an estimated 100 million TikTok users in Brazil.<sup>2</sup> For comparison, X (formerly Twitter) has 16 million active users in the country (Statista 2025). Politics is also prevalent on TikTok (Diaconescu 2024; Herrman 2020). For instance, it became the locus of youth political discussion during the last two US elections (Maheshwari & Kircher 2024; Medina Serrano et al. 2020) and the Israeli-Palestinian conflict (Cervi & Marín-Lladó 2022). In Brazil, political candidates use TikTok to reach younger voters (Lima et al. 2023), for which they must adapt to the platform’s “affordances” (Chagas & Stefano 2023), that is, idiosyncrasies in the way users interact with and perform actions linked to the specific features of a platform (Bucher and Helmond 2018). A quick search on TikTok for “#eleicoes2022”, for example, returns over 400,000 posts, showcasing the amplitude of political discussion around the most recent election. For comparison, Facebook and Instagram return 890,000 and 1 million posts respectively (X does not display the total number) but also host more users (approximately 112 and 141 million respectively) (Dixon 2025a, 2025b).<sup>3</sup>

**A note on the impact of the algorithm: More popular topics are likely to emerge at the expense of others in our sample. The implication is that the algorithm, and the consideration**

---

<sup>1</sup>Like Riedl et al. (2023), we do not distinguish between PSMIs that operate with financial gains in mind and those who do not monetize their views.

<sup>2</sup>Several sources place the total number of users around this quantity (<https://worldpopulationreview.com/country-rankings/tiktok-users-by-country>; <https://www.shopify.com/blog/tiktok-statistics>; <https://www.castmagic.io/post/tiktok-usage-by-country>)

<sup>3</sup>For more information on TikTok, see [Appendix A](#).

of PSMIs of the algorithm, is part of the discursive superstructure formation mechanism. +

## Sources

TikTok’s own affordances, user base, and political discussions make it an adequate avenue on which to examine discursive superstructures and belief systems, permitting an interesting theoretical test across platforms and political contexts.

## Methodology

### Data and Sampling

Sampling on social media is always an arduous task. To complicate matters, we lack a list of PSMIs on TikTok from which to sample, let alone those specific to Brazil. Hence, there is no obvious starting point. Previous studies have overcome this issue in various ways. Green (2024) employs a snowball technique through X Lists, which, although not entirely comprehensive, provide an initial point of departure by signaling commonality among the accounts listed. With a set of “seed” known pundits, the author can identify other users who appear in the same lists as likely pundits. This is a feature of X not available on TikTok, unfortunately. Guinaudeau et al. (2022), who seek to gather US accounts that produce political content on TikTok, also undertake a snowball approach. They kick off with a set of general political hashtags in the US (e.g., #politics, #democrats) to identify users who focus on politics. From the videos produced by this initial sample, they further collect users who engage with them as potential additional political accounts. Using their content, they train a neural network classifier to distinguish between political and non-political posts, which is ultimately used to categorize users as either political or not based on the content they upload. Although powerful, this avenue involves the expensive effort of manually labeling a significant number of accounts and their posts, forcibly including non-political users who are not of interest here.

Instead, we draw up our own approach. Like Guinaudeau et al. (2022), we start with a set of hashtags and keywords with which to query TikTok’s Research API.<sup>4</sup> However, we limit these to the names of the main political candidates—in the case of Brazil, Lula and Bol-

---

<sup>4</sup>All TikTok data for this project is collected through TikTok’s Research API (<https://developers.tiktok.com/products/research-api/>) via Python- and R-based calls.

sonaro—in their different forms (e.g., “inacioluladasilva”, “luladasilva”, “lula”) and our calls to two months before the election (approximately one month per round, from September 2 to October 30, 2022). We also exclude videos uploaded outside of Brazil. Hashtags serve as categorization cues for users to ‘locate, self-organize, and collectively contribute to the information streams’ (Zulli and Zulli 2022, p. 3), resulting in the convergence of content around a topic or event, especially if contentious and highly publicized (Blevins et al. 2019). We thus postulate that this produces a set of videos in which most, if not all, Brazilian PSMIs will be represented, given the centrality of the presidency in organizing political competition in the country (Melo & Câmara 2012). **Additional data set collection + explain.**

This initial step captures over 720,000 videos and 320,000 unique users. In order to separate our potential PSMIs from the rest, we devise a score based on the quantity and quality of the videos produced by each user. We presume that influencers will consistently (measure of quantity) post with higher degrees of engagement and reach (measure of quality), given their focus on political content and existing audiences. We thus calculate:

$$S_i = E_i \cdot \log(1 + V_i) \cdot N_i \quad (1)$$

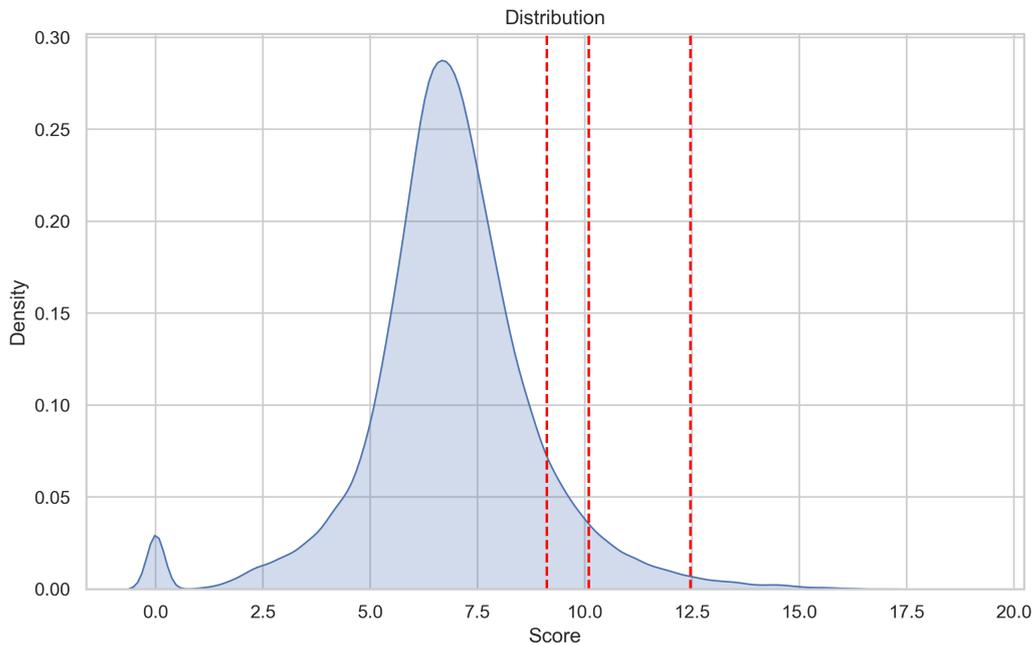
Here, the score ( $S$ ) for user  $i$  equals the median engagement ( $E$ )—made up of the sum of likes, comments, and shares—, multiplied by the log of the median view count ( $V$ ), multiplied by the number of posts ( $N$ ) by user  $i$ . We take the median to avoid the impact of outliers and log views to reduce the weight of viral videos and larger followings, given that content creators might have niche audiences.<sup>5</sup> Furthermore, to avoid one-off posts that become viral by celebrities or influencers who do not focus on politics, we set an arbitrary filter of at least three videos in the sample. Finally, we concentrate on the highest-scoring accounts. We are interested in the top percentiles because we seek PSMIs with some degree of influence as “pundits” of political discourse. [Figure 1](#) displays the distribution of accounts based on these sampling metrics.

**90% + why we focus on this group and not everyone: influential and consistent PSMIs are**

---

<sup>5</sup>We test whether different versions of our sampling score produce similar samples, obtaining very high sample consistency across a number of alterations and tests; see [Appendix B](#).

Figure 1: Distribution of Accounts Based on Sampling Score [Move to appendix](#)



*Note:* Vertical lines represent the 90, 95, and 99% cut-offs respectively. Sample score is logged for ease of visualization. Natural log + 1 handles values of the sample score that equal zero to avoid infinite instances. Only users with at least three videos are represented here.

most interesting and to cut down on the potential number of PSMIs, as not all in the sample are PSMIs.

At the 90% cut-off, this process produces an initial set of 7,836 users. From this group, we manually confirm users that fall under our definition of PSMI,<sup>6</sup> thus removing politicians, (anonymous) meme accounts, news and other organizations, and remaining celebrities and non-political influencers, yielding a final sample of 108 PSMIs. We collect their content for analysis during **X** for a total of **X** posts. [Justify period.](#)<sup>7</sup>

An additional possibility to increment the number of posts is to expand the collection window beyond the election for a comparison of discursive superstructures and belief systems between election and non-election periods.

---

<sup>6</sup>See our labeling codebook in [Appendix C](#).

<sup>7</sup>For general descriptives on the data see [Appendix D](#).

## Categorizing Content

To determine the presence and shape of a discursive superstructure, we must first categorize the content of PSMIs into coherent themes. To this end, we utilize word embeddings, a technique in Natural Language Processing (NLP) that represents words as vectors in a high-dimensional space. Its underpinning assumption is known as the distributional hypothesis, which posits that words appearing in similar contexts tend to have similar meanings. In particular, we apply BERTopic, a model that builds on transformer-based embeddings to generate themes that are semantically meaningful and context aware (Grootendorst 2022), calculating the semantic similarity among words (Sy et al. 2024).<sup>8</sup>

Fitting this model involves four main steps (for a detailed description, see Grootendorst [2022]). First, we embed documents to create representations in a vector space that can be compared semantically, assuming that documents containing similar topics will be semantically similar. Here, we use the widely popular “all-MiniLMM-L6-v2”,<sup>9</sup> a multilingual sentence transformer that includes Portuguese. Our documents are not entire posts, but sentences. This is because BERTopic only generates one topic per document, but posts on TikTok may discuss various topics in connection. By embedding at the sentence level, we allow for this possibility.<sup>10</sup> The textual information that we gather from each post includes the caption, hashtag/s, and subtitles if available.<sup>11</sup> Regarding preprocessing, we only remove capital letters, links, and leading and trailing spaces, as BERTopic does not demand highly preprocessed texts (Grootendorst 2022). **Transcribe videos?**

Second, we reduce the dimensionality of embeddings to avoid the distortion of spatial locality in high-dimensional spaces. For this purpose, we use UMAP, shown to adequately preserve the local and global features in this type of task (McInnes et al. 2018). These reduced em-

---

<sup>8</sup>Interestingly, BERTopic performs particularly well on social media data in comparison to other models (Grootendorst 2022).

<sup>9</sup>Hugging Face: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2>.

<sup>10</sup>We use the spaCy Python library (<https://spacy.io/>) to separate sentences in posts, in particular, the largest, best-performing model for the Portuguese language: [https://spacy.io/models/ptpt\\_core\\_news\\_lg](https://spacy.io/models/ptpt_core_news_lg).

<sup>11</sup>By focusing on the textual content of posts, we do not consider other aspects of language and communication, such as pitch, emphasis, or body expressions. Nevertheless, when measuring the *content* of a video for topic classification purposes, non-textual data arguably add little to the literal message conveyed.

beddings are, third, clustered with the HDBSCAN algorithm to identify semantically coherent groups (topics), a density-based algorithm that finds clusters of varying shapes and sizes, handles noise efficiently, and combines well with UMAP reductions (Allaoui et al. 2020).

Finally, to represent topics, we calculate a class-based Topic-Frequency-Inverse-Document-Frequency (c-TF-IDF) score, capturing the most representative words in each cluster based on their frequency and rarity in the corpus. Importantly, by iteratively merging the c-TF-IDF representations of the least common topic with its most similar one, we can reduce the number of topics to any arbitrary number.

Here, we merge clusters into 100 potential topics. Following previous studies (e.g., Barberá et al. 2019), this is to grant enough topic spread while preserving cohesiveness and keeping the (manual) labeling process manageable. In addition, we run the model multiple times with different topic numbers ( $K$ ), using 10-fold cross-validation and computing the coherence and diversity of resulting topics—two goodness-of-fit indicators of BERTopic (Grootendorst 2022). We find that  $K = 100$  fits the data best.

Although not all topics have entirely straightforward interpretations, we find that most can be easily labeled, albeit not all are political (e.g., “Books”, “TikTok’s algorithm”). We thus categorize topics as either “Political”, “Other” (clearly not about politics), and potentially political (e.g., “News”). This latter category, “Political/Other”, **explanation**. We also find some topics to be quite similar in word representation and content, which we merge to avoid potential source bias.<sup>12</sup>

## **Inferring a Discursive Superstructure**

### *Structure*

Once the content is collected and categorized, we seek to infer the structure of discourse displayed by our PSMI sample. As explained above, we think of discursive superstructures as networks of connections among different topics. Therefore, we fit a Mixed Graphical Model (MGM) (Haslbeck & Waldorp 2020) on our data to estimate pairwise associations between topics (binary variables), controlling for the presence of other topics. This approach infers

---

<sup>12</sup>For all topic information, see [Appendix E](#).

network ties between topics using a graphical LASSO (Friedman et al. 2008), a penalized maximum likelihood estimator of the inverse covariance matrix (precision matrix). The LASSO penalty shrinks weak associations to exactly zero, removing edges between conditionally independent topics and producing a sparser, more interpretable network.<sup>13</sup> This enables us to understand topic co-occurrence patterns and describe the presence and structure of a discursive superstructure.

Furthermore, we extract common centrality measures of interest: Strength centrality, which captures the influence of a topic on its immediately connected neighbors; closeness centrality, which measures the average distance from a topic to all others in the network, indicating how quickly its influence can spread through the network; and betweenness centrality, which identifies topics that act as bridges to others, showcasing their importance in unifying distinct themes in the network. Finally, to incorporate the uncertainty of the imputation and generate inference from the network, we run a nonparametric bootstrap to generate confidence intervals on these measures (1,000 resamples).

### *Coherence*

Crucially, we are interested in the coherence of the discursive superstructure, i.e., whether differences in how PSIMs talk about topic “A” correspond to the differences in how PSIMs talk about topic “B” and vice versa. For example, this may take the form of influencers who praise Bolsonaro discussing immigration differently from those who praise Lula. To test this, we follow Green (2024) in quantifying the variation in how PSIMs discuss diverse political themes, and, importantly, how such variation in the discussion of one theme is related to variation in the discussion of another, or as he puts it, ‘a rhetorical “what goes with what”’.

To ascertain these relationships, we must first pin down the position of sentences (our documents in this context) on different topics, for which we utilize pivot scaling (Hobbs & Green 2025), a form of principal component analysis conducted on a standardized and truncated word co-occurrence matrix. It outputs orthogonal dimensions that explain decreasing degrees of variance in that matrix. The result can be understood as reflecting meaningful variation in what

---

<sup>13</sup>Building a network without considering the relationship between variables in a highly-interactive context like this one would produce an overly-saturated network with little substantive interest.

PSMIs talk about when discussing each of the given topics.<sup>14</sup>

With these positions, we may analyze the extent to which differing topics systematically covary. We first average sentence scores on the first ten dimensions of each topic by each PSMI, assigning them a score of zero if they did not partake in a given topic. This produces a matrix with rows representing PSMIs, and columns representing their average sentence score on each dimension of each topic. Once again, we apply an MGM to this data. The final network has nodes as dimensions of topics, and edges as a non-zero relationship between two dimensions. We are thus able to quantify associations between the way political themes tend to be discussed by PSMIs. Finally, we aggregate the edges for each topic for ease of interpretation and calculate the centrality measures of interest.<sup>15</sup>

## Results

### The Presence of a Discursive Superstructure

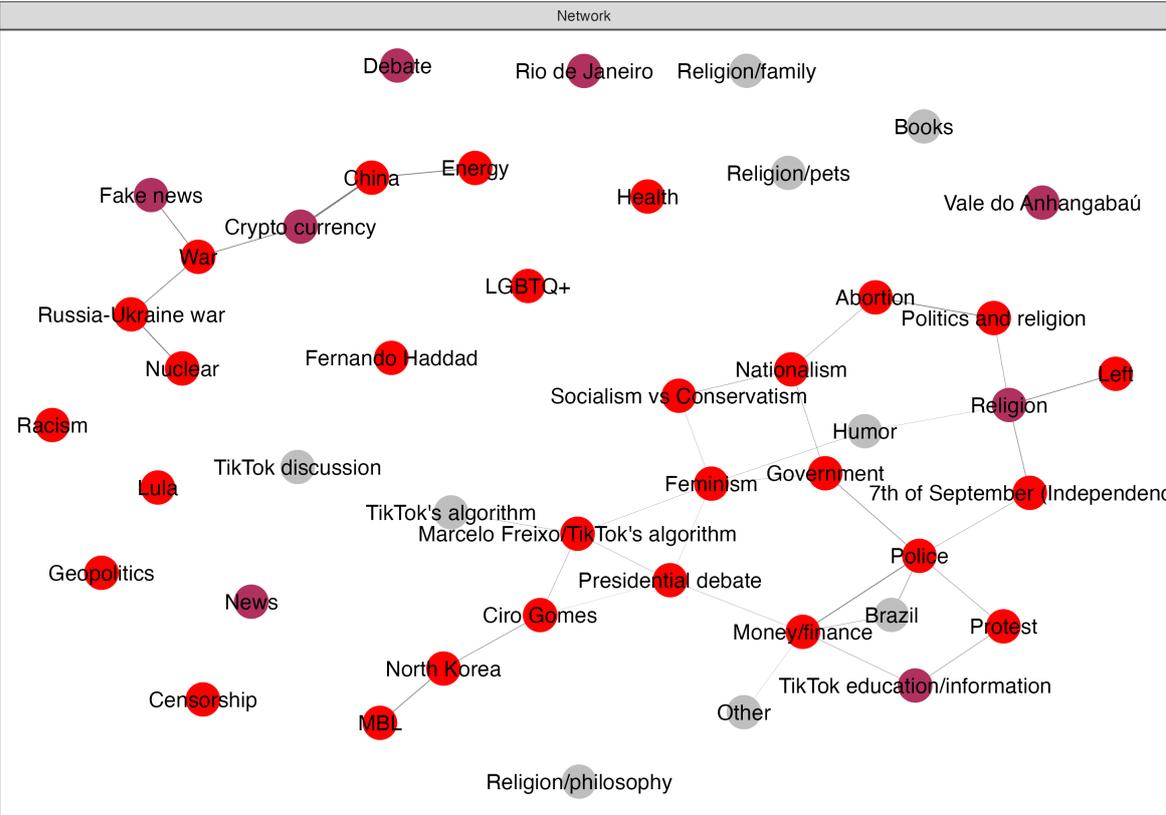
The initial MGM result is displayed by [Figure 2](#). We find preliminary evidence of a discursive superstructure, as there is a significant degree of connection among a large number of topics. **Counterfactual explain: what would we see were there no superstructure?** These also seem to make intuitive sense; for example, the bigger cluster is composed of links between “Abortion” and “Politics and religion”, which is in turn connected to “Religion”; “Socialism vs Conservatism” is directly connected with “Nationalism” and “Feminism”, and indirectly connected to “Government”, “Police”, and “Protest”, as well as other interesting themes. On the other cluster, we observe sensible connections among “War”, “Russia-Ukraine war”, “China”, “Nuclear”, and “Energy”, as well as “Fake news”. Surprisingly, the topic “Lula” stands by itself. This is likely due to its low variance, as most PSMIs discuss it. Similarly, we remove the topic “Bolsonaro” due to zero variance in our sample. The low variance stems from the time frame of video collection—the election—when all PSMIs are going to discuss the main candidates. We, therefore, expand the collection of posts to a six-year period between 2019 and 2025.

---

<sup>14</sup>We are currently unable to apply pivot scaling to all topics due to a low document (sentence) frequency in some of them; an issue that could be resolved by increasing our sample size of PSMIs and/or their posts.

<sup>15</sup>For entire details on this procedure, see [Appendix F](#).

Figure 2: Network of Topics



Type of topic • Other • Political • Political/Other

Note: Topic “Bolsonaro” is removed due to zero variance, as all PSIMs mention it; topic “Nicolas Ferreira” is also removed due to its mention frequency being too low.

Figure on network measures + confidence intervals + discussion.

The Cohesiveness of a Discursive Superstructure

Conclusion

## References

- Allaoui, M., Kherfi, M. L., & Cheriet, A. (2020). Considerably improving clustering algorithms using umap dimensionality reduction technique: A comparative study. *International conference on image and signal processing*, 317–325.
- Andi, S. (2021). How and why do consumers access news on social media. *Reuters institute digital news report*, 23.
- Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., & Tucker, J. A. (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), 883–901.
- Bause, H. (2021). Politische social-media-influencer als meinungsführer? *Publizistik*, 66(2), 295–316.
- Blevins, J. L., Lee, J. J., McCabe, E. E., & Edgerton, E. (2019). Tweeting for social justice in #Ferguson: Affective discourse in Twitter hashtags. *New Media & Society*, 21(7), 1636–1653.
- Boutyline, A., & Vaisey, S. (2017). Belief network analysis: A relational approach to understanding the structure of attitudes. *American journal of sociology*, 122(5), 1371–1447.
- Bucher, T., & Helmond, A. (2018). The Affordances of Social Media Platforms. In J. Burgess, T. Poell, & A. Marwick (Eds.), *The SAGE Handbook of Social Media* (pp. 233–254). SAGE Publications Ltd.
- Ceci, L. (2025, May 17). *Number of TikTok users worldwide from 2018 to 2029*. Statista. Retrieved March 13, 2025, from <https://www.statista.com/forecasts/1142687/tiktok-users-worldwide>
- Cervi, L., & Marín-Lladó, C. (2022). Freepalestine on TikTok: From performative activism to (meaningful) playful activism. *Journal of International and Intercultural Communication*, 15(4), 414–434.

Chagas, V., & Stefano, L. d. M. (2023). Estratégias de uso do tiktok por políticos brasileiros. *Revista de Sociologia e Política*, 31, e022.

Converse, P. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and its discontents* (pp. 206–61). New York: Free Press of Glencoe.

de la Cerda, N. (2024). Beyond partisanship: Political identities in comparative perspective.

Diaconescu, A. (2024, November 27). *Victory of Romania's 'TikTok Messiah' is an alarm bell for Europe*. Euronews. Retrieved November 28, 2024, from [https://www.euronews.com/2024/11/27/romania-tiktok-messiah-candidate-is-a-stark-premiere-for-europe?utm\\_source=newsletter&utm\\_campaign=today\\_newsletter&utm\\_medium=referral&insEmail=1&insNltCmpId=209&insNltSldt=10080&insPnName=euronewsfr&isIns=1&isInsNltCmp=1](https://www.euronews.com/2024/11/27/romania-tiktok-messiah-candidate-is-a-stark-premiere-for-europe?utm_source=newsletter&utm_campaign=today_newsletter&utm_medium=referral&insEmail=1&insNltCmpId=209&insNltSldt=10080&insPnName=euronewsfr&isIns=1&isInsNltCmp=1)

Dixon, S. J. (2025a). *Leading countries based on Facebook audience size as of February 2025*. Statista. <https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/>

Dixon, S. J. (2025b). *Leading countries based on Instagram audience size as of February 2025*. Statista. <https://www.statista.com/statistics/578364/countries-with-most-instagram-users/#:~:text=As%20of%20February%202025%2C%20India,UAE%2C%20Brazil%2C%20and%20Bahrain.>

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441.

Goodwin, A., Joseff, K., Riedl, M. J., Lukito, J., & Woolley, S. (2023). Political relational influencers: The mobilization of social media influencers in the political arena. *International Journal of Communication*, 17, 21.

Green, J. (2024). The rhetorical “what goes with what”: Political pundits and the discursive superstructure of ideology in us politics. *Public Opinion Quarterly*, 88(4), 1166–1190.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Guinaudeau, B., Munger, K., & Votta, F. (2022). Fifteen seconds of fame: Tiktok and the supply side of social video. *Computational Communication Research*, 4(2), 463–485.

Haslbeck, J. M., & Waldorp, L. J. (2020). Mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, 93, 1–46.

Herrman, J. (2020, June 28). *TikTok Is Shaping Politics. But How?* The New York Times. Retrieved November 28, 2024, from <https://www.nytimes.com/2020/06/28/style/tiktok-teen-politics-gen-z.html>

Highfield, T. (2016). *Social media and everyday politics*. Polity Press.

Hobbs, W., & Green, J. (2025). Categorizing topics versus inferring attitudes: A theory and method for analyzing open-ended survey responses. *Political Analysis*, 33(3), 231–251.

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual review of psychology*, 60(1), 307–337.

Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News sharing in social media: A review of current research on news sharing users, content, and networks. *Social media+ society*, 1(2), 2056305115610141.

Lima, J., Santana, M., Correa, A., & Brito, K. (2023). The use and impact of tiktok in the 2022 brazilian presidential election. *Proceedings of the 24th annual international conference on digital government research*, 144–152.

Maheshwari, S., & Kircher, M. M. (2024, October 21). *Trump vs. Harris: The 2024 Election Has Taken Over TikTok*. The New York Times. <https://www.nytimes.com/interactive/2024/10/21/business/media/2024-election-tiktok-trump-harris.html>

- Mainwaring, S., & Pérez-Liñán, A. (2014). *Democracies and dictatorships in latin america: Emergence, survival, and fall*. Cambridge University Press.
- Mainwaring, S., & Scully, T. (1995). *Building democratic institutions: Party systems in latin america* (Vol. 277). Stanford University Press Stanford.
- Martínez-Gallardo, C., Cerda, N. d. l., Hartlyn, J., Hooghe, L., Marks, G., & Bakker, R. (2023). Revisiting party system structuration in latin america and europe: Economic and socio-cultural dimensions. *Party Politics*, 29(4), 780–792.
- McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- Medina Serrano, J. C., Papakyriakopoulos, O., & Hegelich, S. (2020). Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. *12th ACM Conference on Web Science*, 257–266.
- Melo, C. R., & Câmara, R. (2012). Estrutura da competição pela presidência e consolidação do sistema partidário no brasil. *Dados*, 55, 71–117.
- Moreira, T., & Camargos, J. C. L. (2025). “O Campeão Voltou”? A Natureza Substantiva da Ideologia de Massa no Brasil. *Opinião Pública*, 31, 1–37.
- Munger, K., & Li, V. (2025). Thirst traps and quick cuts: The effects of tiktok “edits” on evaluations of politicians. *Social Media+ Society*, 11(2), 20563051251329990.
- Noel, H. (2012). The coalition merchants: The ideological roots of the civil rights realignment. *The Journal of Politics*, 74(1), 156–173.
- Riedl, M. J., Lukito, J., & Woolley, S. C. (2023). Political influencers on social media: An introduction. *Social Media+ Society*, 9(2), 20563051231177938.

Schmuck, D., Hirsch, M., Stevic, A., & Matthes, J. (2022). Politics—simply explained? how influencers affect youth’s perceived simplification of politics, political cynicism, and political interest. *The International Journal of Press/Politics*, 27(3), 738–762.

Sehl, A., & Schützeneder, J. (2023). Political knowledge to go: An analysis of selected political influencers and their formats in the context of the 2021 german federal election. *Social Media+ Society*, 9(2), 20563051231177916.

Statista. (2025). *Leading countries based on number of x (formerly Twitter) users as of February 2025*. Statista. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>

Suuronen, A., Reinikainen, H., Borchers, N. S., & Strandberg, K. (2022). When social media influencers go political: An exploratory analysis on the emergence of political topics among finnish influencers. *Javnost-The Public*, 29(3), 301–317.

Sy, C., Maceda, L., Flores, N., & Abisado, M. (2024). Unsupervised machine learning approaches in nlp: A comparative study of topic modeling with bertopic and lda. *Int J Intell Syst Appl Eng*, 12(21s), 3276–3283.

Wan, A., & Green, J. (2023). *Political Pundits and the Maintenance of Ideological Coalitions*. <https://osf.io/xfy8r>

Wunderlich, L., Hölig, S., & Hasebrink, U. (2022). Does journalism still matter? the role of journalistic and non-journalistic sources in young peoples’ news related practices. *The International Journal of Press/Politics*, 27(3), 569–588.

Zechmeister, E. J., & Corral, M. (2013). Individual and contextual constraints on ideological labels in latin america. *Comparative Political Studies*, 46(6), 675–701.

Zulli, D., & Zulli, D. J. (2022). Extending the Internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform. *New Media & Society*, 24(8), 1872–1890.

# SUPPLEMENTARY MATERIALS

## Discursive Superstructures of Ideology by Brazilian Political TikTokers

### Contents

<b>A</b>	<b>TikTok’s Affordances</b>	<b>A2</b>
<b>B</b>	<b>Sampling Score Robustness Checks</b>	<b>A4</b>
<b>C</b>	<b>PSMI Labeling Codebook</b>	<b>A6</b>
	C.1 Criteria for PSMI Classification . . . . .	A6
	C.2 Labels and Examples . . . . .	A7
	C.3 Procedure for Manual Labeling of Accounts . . . . .	A9
<b>D</b>	<b>Data Description</b>	<b>A10</b>
	D.1 PSMI vs non-PSMI Comparison . . . . .	A10
	D.2 Post Distribution of PSMIs . . . . .	A11
	D.3 Engagement and Reach (PSMIs): Total and Distribution . . . . .	A12
<b>E</b>	<b>BERTopic</b>	<b>A14</b>
	E.1 Topics and Labels . . . . .	A14
	E.2 Selection of Topic Number . . . . .	A14
<b>F</b>	<b>Pivot Scaling</b>	<b>A15</b>
	F.1 Pivot Scaling . . . . .	A15
	F.2 Confirming the Validity of Positions . . . . .	A15
	F.3 Topic Covariance Network . . . . .	A15
	<b>References</b>	<b>A16</b>

## A TikTok's Affordances

Each social media platform has its own “affordances”. Affordances shape the conditions determining which actions are not only available but also appealing and acceptable, molding each platform’s unique culture of interactions and thereby influencing user behavior in areas like content production and engagement (Umansky and Pipal 2023).

TikTok is available as an app for mobile devices or through a web browser. Its content takes the form of individual posts in short-video format, which can be as simple as text superimposed onto a background or more complex, including images, sounds, or special filters and effects (Medina Serrano et al. 2020). These can be annotated with text, captions, and hashtags. Users can concatenate posts—akin to retweets on X—known as ‘duets’, where another user’s video appears alongside one’s own video. Interactions and diffusion among users can also take the form of likes, reposts, or comments, which together with views, propagate a post through the platform’s recommendation algorithm (TikTok 2020). It is through these affordances that users can voice, amplify, discuss, and interact with opinions and ideas on the platform.

Similarly to X and Instagram, TikTok classifies videos using hashtags, with the most popular labelled as ‘trending’ (Medina Serrano et al. 2020). It also shares Instagram’s emphasis on visual communication (Hase et al. 2023) but remains unique in several key aspects. Music, for instance, is a central component of videos; users can play background songs or create original audio clips that can later be used by others, supporting the delivery of content (Medina Serrano et al. 2020).

Most importantly, however, is that the curation of content, i.e., the posts a user comes across, is predominantly centered around interactions instead of networks (Cervi et al. 2021), which is how content is traditionally filtered on X, Facebook, or YouTube. Users mostly consume content by viewing an algorithmically generated feed of videos on the ‘For You’ page, which is the default location when entering TikTok. This feed evolves based on the user’s watching, sharing, liking, and commenting patterns, rather than on her social connections (i.e., following network) (TikTok 2020). Such particularity incentivizes users to engage with *content*, not creators or friends, and encourages imitation and replication of successful posts (Zulli and Zulli 2022). It also means that ordinary users have a greater probability of going viral on

TikTok, where having a large base of followers is not necessary (Guinaudeau et al. [2022](#)), in turn enlarging the number of people who can contribute to narratives and discussions (Medina Serrano et al. [2020](#))—a process shown to be driven by a small number of accounts in other platforms (e.g., Stieglitz and Dang-Xuan [2012](#)).

## B Sampling Score Robustness Checks

Table 1: Correlations Across Sampling Score Variations **redo with extended sample of videos**

Score	Correlation method	Correlation with original
Without logging views	Spearman’s Rank	0.97*
Without logging views	Kendall’s Tau	0.87*
Sum instead of multiply	Spearman’s Rank	0.96*
Sum instead of multiply	Kendall’s Tau	0.84*
Without number of posts	Spearman’s Rank	0.91*
Without number of posts	Kendall’s Tau	0.77*
Mean instead of median	Spearman’s Rank	0.90*
Mean instead of median	Kendall’s Tau	0.76*

*Note:* The Spearman’s Rank Correlation measures the monotonic relationship between two variables (i.e., do they move in the same order, not necessarily linearly). It converts the values of both variables into ranks (ties get averaged) and computes the Pearson correlation between those ranks. The closer the score is to 1 (scale is 0 – 1), the more it means that as one goes up, the other almost always goes up in the same order. Kendall’s Tau-b is another nonparametric test, more robust to ties. The difference with Spearman’s Rank is that it measures the proportion of concordant vs. discordant pairs of observations, rather than direct rank correlations. The interpretation is similar, but the scale ranges from –1 to 1. For both measures: statistical significance at  $< 0.05$  (i.e., there is a statistically significant monotonic association between the two scores), signaled with \*. Estimates shown to 2 d.p.

**Justify choice (vs median engagement if one has the information why not use it? More precise score.)** Also explain scale: all measured on a numeric (ratio) scale, but they differ substantially in range and variability. Median engagement ( $E_i$ ): typically spans a few orders of magnitude (from a few interactions to perhaps thousands).

Median views ( $V_i$ ): usually exhibits the largest variation, often spanning several orders of magnitude more than engagement (from hundreds to millions). The log transformation of  $(1 + V_i)$  explicitly reduces this skew and makes it more comparable in scale to engagement.

Number of posts ( $N_i$ ): tends to vary the least—most users post only a few times, while a few post very frequently.

Thus, while all three components are on the same numeric scale, their distributions differ markedly. The log transformation on views and the use of medians (rather than means) help to stabilize the scale across components and reduce the influence of extreme values. However, the components are not standardized or normalized to the same range before multiplication; instead, their raw (but transformed) magnitudes jointly determine the score.

explanation of why not standardized: It captures absolute magnitude, meaning that a user

with much higher engagement or reach will naturally dominate the ranking — which is exactly what you want if you're identifying “influencers.”

If your goal is ranking or filtering high-performing accounts, then you should not normalize or standardize before multiplying. It maintains the real magnitude of engagement and reach, which is precisely what distinguishes influencers from ordinary users.

You've already taken two important steps to mitigate scale distortions:

Using medians (reduces outlier impact),

Applying log to views (compresses heavy-tailed variation).

## C PSMI Labeling Codebook

### C.1 Criteria for PSMI Classification

Table 2: Criteria for PSMI Classification

Must	Must not	May
Have content clearly centered around politics	Post infrequently about politics (e.g., during peak salience)	Have non-politics content, as long as content remains mostly about politics
Be an identifiable individual	Be an organization or an unidentifiable user (e.g., news organization, anonymous meme account)	Be a small group of identifiable people
	Be part of or running for an elective institution or office-seeking group	Be an ex-politician not seeking (re-)election officially
		Have a platform-based or platform-enhanced followership

## C.2 Labels and Examples

Table 3: Labels and Examples

Label	Description	Examples
Celebrity	Users followed for other reasons than their political content. They either do not cover politics at all or only during high periods of salience. These include influencers with platform-built fame who focus their content outside of politics.	<a href="#">danieldasilvva</a> <a href="#">essemenino</a>
Ex-politician (PSMI)	User who ran for an elective institutional position in the past (outside of our data collection period, i.e., 2022 election) but no longer does, and who dedicates their account to political content, although not for official (re-)election purposes. They fall into our sample of PSMIs, but we add the distinctive Ex-politician label.	<a href="#">manueladavila</a>
Journalist (PSMI)	Journalists who use TikTok to voice their opinions fall within our definition of PSMIs. We give them the additional label of “journalists” to differentiate their platform-enhanced followership from influencers only known from social media.	<a href="#">aaronтура</a> <a href="#">realpfigueiredo</a>
Memes	These are anonymous/non-identifiable accounts that might post political content, by either uploading unedited or edited existing political videos. They might add a message with captions and can convey complex ideas, including support or criticism towards politicians and parties. Some are entirely about politics, others only post political content occasionally.	<a href="#">scfvrtа</a> <a href="#">antigopt</a>
Organization	These are non-identifiable accounts that might post political content (e.g., news channels). However, it is not possible to attribute the content to one individual or a small number of individuals. They might post neutral or partisan political content.	<a href="#">metropolesoficial</a> <a href="#">institutolula</a>
Politician	User who ran for or held an elective institutional position during and around the data collection period (2022 general election) or still does.	<a href="#">aaronтура</a> <a href="#">realpfigueiredo</a>

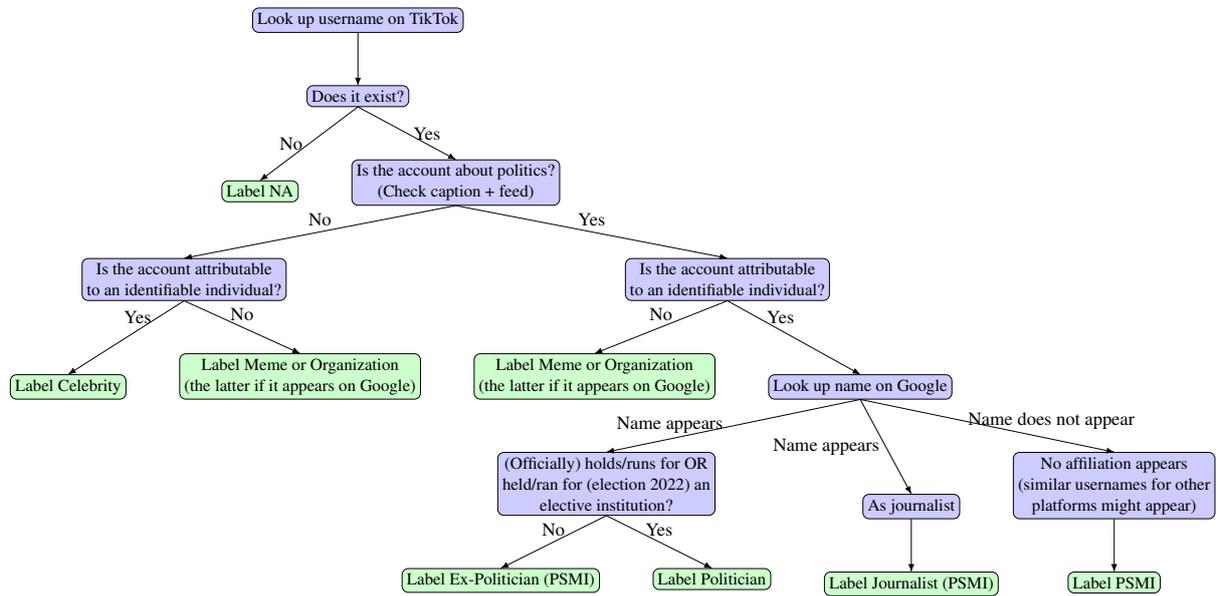
Table 3: Labels and Examples

Label	Description	Examples
PSMI	Account that is clearly centered around politics, identifiable, and by a user who does not run for or hold an elective institutional position.	<a href="#">eupresley</a> <a href="#">lucas_ale1</a>
NA	Accounts that no longer exist; these have been deleted since posting about the election in 2022.	

*Note:* Accounts may belong to more than one category, labeled accordingly. As long as an account is labeled as PSMI, we include it in our sample.

### C.3 Procedure for Manual Labeling of Accounts

Figure 3: Procedure for Manual Labeling of Accounts

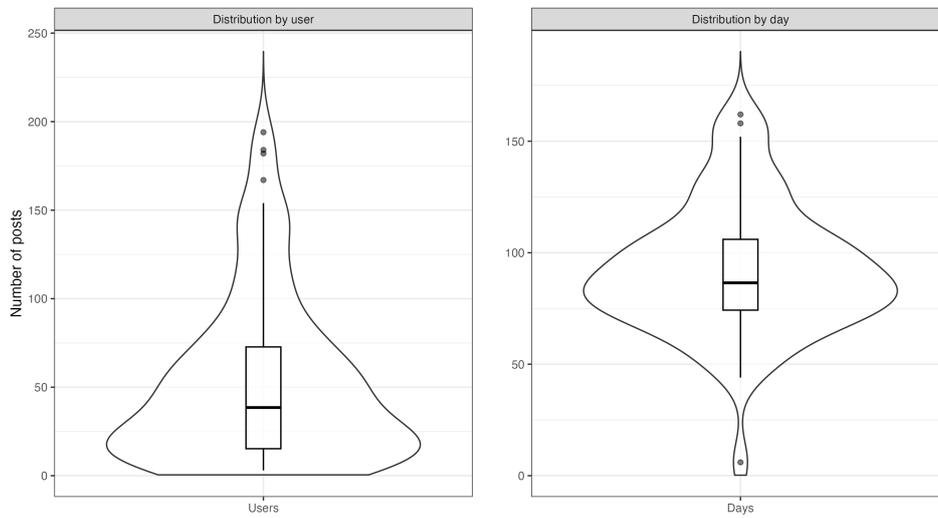


## **D Data Description**

### **D.1 PSMI vs non-PSMI Comparison**

## D.2 Post Distribution of PSMIs

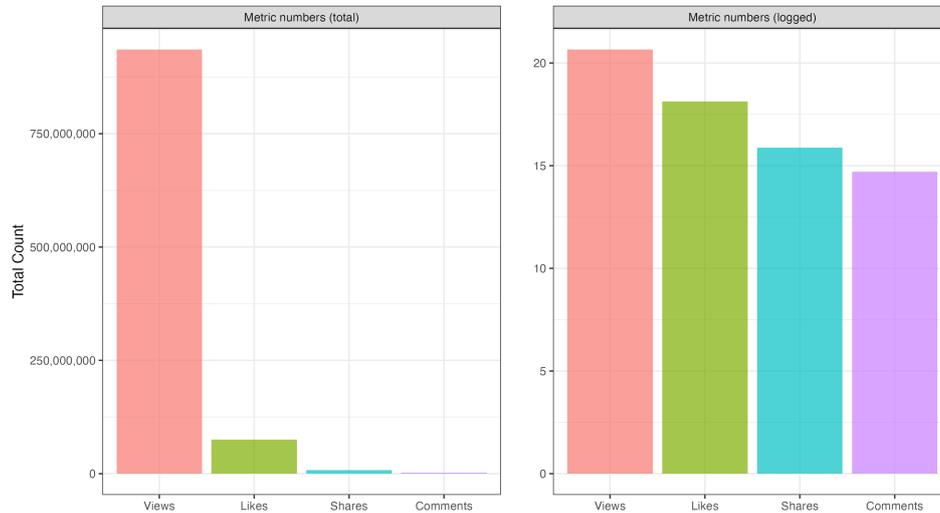
Figure 4: Violin Plot of Posts per Account and per Day



*Note:* The figure displays a rotated kernel density plot on each side to show the distribution of the data, including the median, quartiles, and density. The sample represented is the videos produced by Brazilian PSMIs during the two-month election period.

### D.3 Engagement and Reach (PSMIs): Total and Distribution

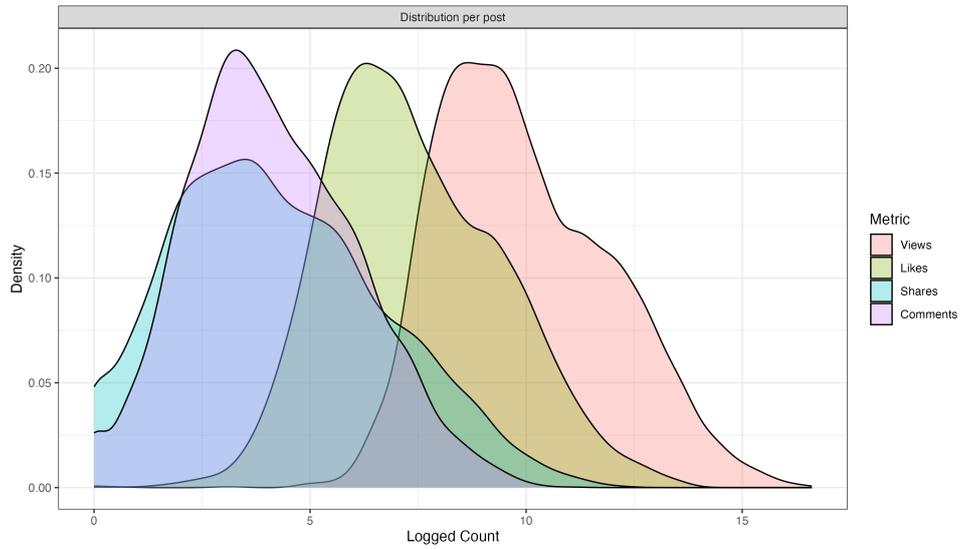
Figure 5: Total Engagement and Reach



*Note:* The left panel displays absolute counts, while the right panel displays logged values for ease of comparison across metrics. The sample represented is the videos produced by Brazilian PSMIs during the two-month election period.

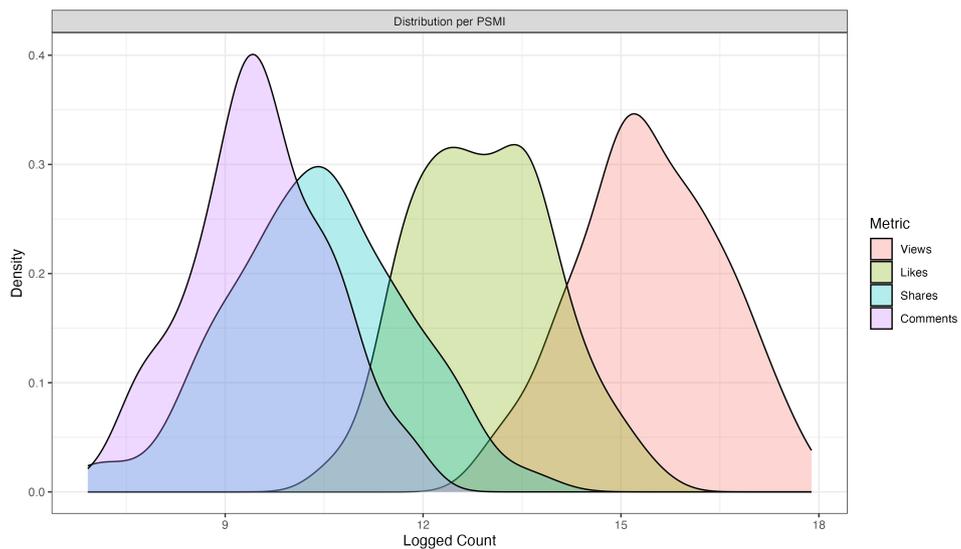
Users mostly interact with PSMIs by viewing their videos. TikTok considers a video as ‘viewed’ as soon as it starts (TikTok 2025), regardless of whether the user watches it entirely, making it difficult to determine the extent to which views are a meaningful interaction. It is, nevertheless, the most common form of engagement by far; an unsurprising finding, given the way it is counted and its low-effort implications.

Figure 6: Engagement and Reach Distribution per Post



*Note:* Natural log + 1 handles values that equal zero to avoid infinite instances. The sample represented is the videos produced by Brazilian PSIMs during the two-month election period.

Figure 7: Engagement and Reach Distribution per PSIM



*Note:* Natural log + 1 handles values that equal zero to avoid infinite instances. The sample represented is the videos produced by Brazilian PSIMs during the two-month election period.

# **E BERTopic**

## **E.1 Topics and Labels**

## **E.2 Selection of Topic Number**

## **F Pivot Scaling**

### **F.1 Pivot Scaling**

### **F.2 Confirming the Validity of Positions**

### **F.3 Topic Covariance Network**

## References

- Cervi, L., Tejedor, S., & Marín Lladó, C. (2021). TikTok and the new language of political communication. *Cultura, Lenguaje y Representación*, 26, 267–287.
- Guinaudeau, B., Munger, K., & Votta, F. (2022). Fifteen seconds of fame: Tiktok and the supply side of social video. *Computational Communication Research*, 4(2), 463–485.
- Hase, V., Boczek, K., & Scharrow, M. (2023). Adapting to Affordances and Audiences? A Cross-Platform, Multi-Modal Analysis of the Platformization of News on Facebook, Instagram, TikTok, and Twitter. *Digital Journalism*, 11(8), 1499–1520.
- Medina Serrano, J. C., Papakyriakopoulos, O., & Hegelich, S. (2020). Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. *12th ACM Conference on Web Science*, 257–266.
- Stieglitz, S., & Dang-Xuan, L. (2012). Political Communication and Influence through Microblogging—An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior. *2012 45th Hawaii International Conference on System Sciences*, 3500–3509.
- TikTok. (2020, June 28). *How TikTok recommends videos #ForYou*. TikTok. Retrieved March 12, 2025, from <https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you>
- TikTok. (2025, February). *Video play metrics*. TikTok Business Help Center. Retrieved March 22, 2025, from <https://ads.tiktok.com/help/article/video-play?lang=en>
- Umansky, N., & Pipal, C. (2023, September 21). *Dances, Duets, and Debates: Analysing political communication and viewer engagement on TikTok*. <https://doi.org/10.31219/osf.io/g8dbu>
- Zulli, D., & Zulli, D. J. (2022). Extending the Internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform. *New Media & Society*, 24(8), 1872–1890.